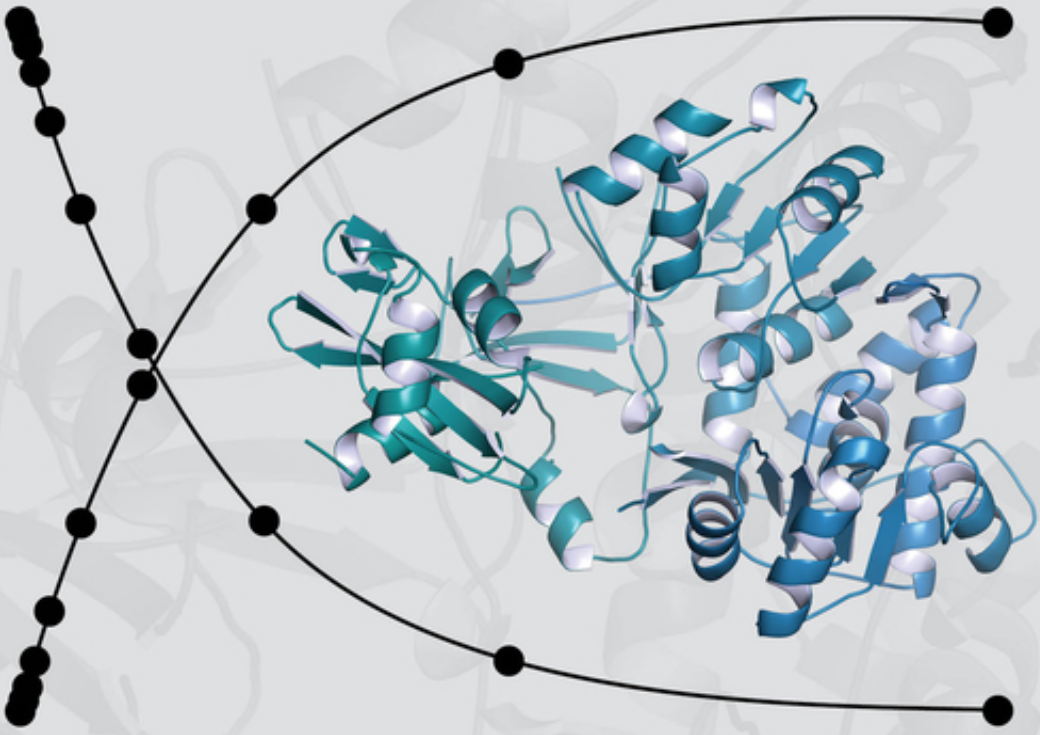
LABORATORY GUIDE TO ENZYMOLOGY

GEOFFREY A. HOLDGATE · ANTONIA TURBERVILLE · ALICE LANNE



WILEY

Laboratory Guide to Enzymology

Laboratory Guide to Enzymology

Geoffrey A. Holdgate, Antonia Turberville, and Alice Lanne

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Names: Holdgate, Geoffrey A., author. | Turberville, Antonia, author. | Lanne, Alice, author.
Title: Laboratory guide to enzymology / Geoffrey A. Holdgate, Antonia Turberville, Alice Lanne.
Description: Hoboken, New Jersey : Wiley, [2024] | Includes bibliographical references and index.
Identifiers: LCCN 2023046724 (print) | LCCN 2023046725 (ebook) | ISBN 9781394179794 (paperback) | ISBN 9781394179800 (adobe pdf) | ISBN 9781394179824 (epub)
Subjects: MESH: Enzymes–pharmacokinetics | Enzyme Assays–methods | Drug Discovery
Classification: LCC QP601 (print) | LCC QP601 (ebook) | NLM QU 135 | DDC 572/.7–dc23/eng/20231128
LC record available at https://lccn.loc.gov/2023046724
LC ebook record available at https://lccn.loc.gov/2023046725

Cover Design: Wiley
Cover Images: Courtesy of Geoffrey A. Holdgate; © StudioMolekuul/Shutterstock

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Contents

Preface

The application of enzymology is an essential approach to drug discovery. Many drug targets are enzymes and modulating their behavior can provide useful therapeutic intervention. As such, an understanding of the basic principles in the use of enzymes is important in identifying and characterizing molecules that may change the function of enzymes. Sometimes, however, the topic can appear difficult as it combines chemical and mathematical concepts that are often unfamiliar to the biologist. Fortunately, there are many useful textbooks that provide information in great detail, covering the many aspects of enzymology. This book is not meant to be a replacement for those textbooks, nor is it a repository for information; rather, it is a guide that helps readers navigate the field and gain a basic understanding of the principles and techniques used in the study of enzymes. In writing this book, we focused on providing the fundamental principles, concepts, protocols, and examples required to generate and analyze enzyme kinetic data. The book serves to provide an initial text that allows the reader to undertake and understand the experiments that are required in establishing assays and building screening approaches that are the bedrock of early drug discovery.

The book begins with an introduction to proteins and enzymes, illustrating the structural features that are key to protein function. Aspects of binding kinetics and thermodynamics are introduced and the importance of quality control when working with proteins is highlighted. The use of buffers in enzyme studies is described as the control of pH is a critical requirement when working to characterize enzymes and their inhibitors.

Steady-state assays and their design to understand enzyme function and to identify and characterize inhibitors and activators are discussed. The different types of mechanism of action for these modulators with respect to substrate concentration are introduced and data analysis methods are presented.

Finally, the different types of molecular interactions are presented and key examples of application in drug discovery are described. The appendices contain a range of key information that supplements the material in the main text.

We hope that this book will be a useful laboratory companion for life science students, academic and industrial researchers who are interested in learning about enzymology. It is our goal to provide a clear and concise introduction to the field, taking a stepwise journey through the equations and their derivations so that the mathematics is not overwhelming and the rationale for the models used is clear. In this way, we trust that the book will inspire readers to begin their journey into exploring enzymes, their mechanisms, and regulation.

Cheshire, 2024

Geoffrey A. Holdgate
Antonia Turberville
Alice Lanne

Acknowledgments

We are grateful to many people over many years for their contributions in turning the idea of a laboratory guide to enzymology into reality. This project certainly would not have been possible without the guidance and mentorship of Wal Ward. His experience and expertise guided enzyme kinetic studies at AstraZeneca and its predecessor companies for two decades. Without his teaching of the fundamental principles, his drive for quality, and wise advice, we would not have been able to continue to deliver leadership in enzymology to the company over the last 10 years. During that time, we have also been indebted to numerous other current and previous colleagues, including Rachel Grimley, Archie Argyrou, Bharath Srinivasan, Xiang Zhai, Astrid Kraal, Kara Herlihy, Hua Xu, Bryony Ackroyd, Sheffin Joseph, and Christopher Stubbs who have contributed ideas, discussion, challenge, and support. We especially thank Bharath Srinivasan and Gareth Davies for critical reading of the manuscript.

1

Introduction to Proteins and Enzymes

CHAPTER MENU

1.1 Protein Structure

Proteins are the central functional molecules of life, encoded by DNA, translated, and expressed to carry out the essential functions of the cell. The building blocks for proteins are amino acids: every amino acid contains a positively charged amine group (N-terminus), a negatively charged carboxyl group (C-terminus), a hydrogen atom, and an R group, all centered around a chiral carbon (alpha carbon, C_α) (Figure 1.1). The presence of a chiral carbon results in stereoisomerism; naturally occurring amino acids are *L*-isomers, and *D*-isomers can arise during chemical synthesis. There are 20 different R groups, which give rise to 20 different amino acids (Figure 1.2). Amino acids can be charged (negatively and positively), polar and non-polar. These different properties contribute to different bonding interactions and architecture of the protein (Section 1.1.4) [1, 2].

1.1.1 Primary Structure

Each protein is formed of a unique sequence of amino acids, which determines the properties of the protein. These are linked by covalent peptide bonds between the amino group of one residue and the carboxyl group of the next, forming long polypeptide chains of amino acids. The number and sequence of amino acids in a polypeptide chain is known as the primary (1°) structure of a protein and is determined by the DNA sequence of the gene. Mutations to the DNA sequence may lead to changes in the amino acids in the polypeptide chain, thus altering the primary structure of the protein [1, 2].

1.1.2 Secondary Structure

The secondary structure of proteins describes the layout of the protein backbone in three dimensions. This structure is formed from the individual peptide bonds between residues, which usually are planar and *trans* (with the exception of proline). There are common elements that often combine to contribute to the protein backbone describing its overall fold. Rotations around the

Laboratory Guide to Enzymology, First Edition. Geoffrey A. Holdgate, Antonia Turberville, and Alice Lanne.

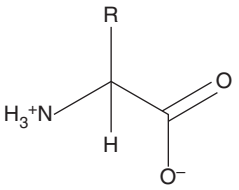


Figure 1.1 General amino acid structure.

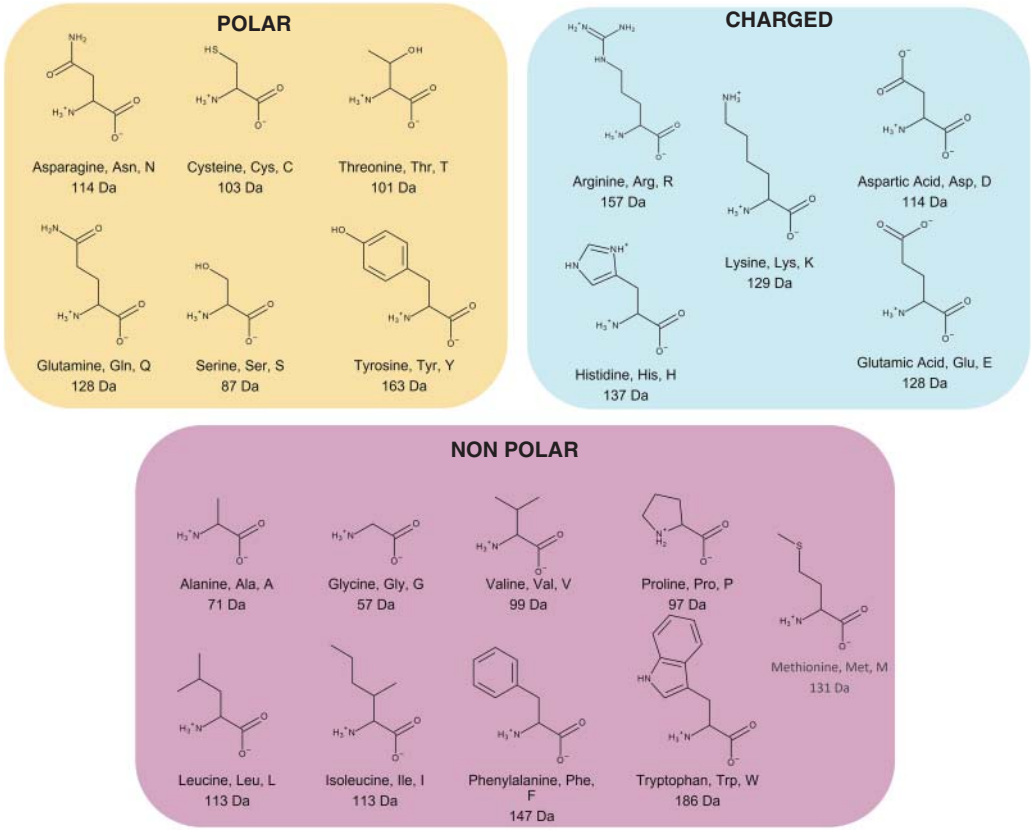


Figure 1.2 Chemical structure of amino acids.

peptide bond enable hydrogen bond formation between the carbonyl oxygen group and amide hydrogen atom of spatially adjacent amino acids, resulting in folding of the polypeptide chains into secondary (2°) structures. Hydrogen bonding can also occur between amino acid side chains. Common secondary structures include the alpha helix, the beta sheet, loops, and many protein structures contain a combination of all elements [1].

1.1.2.1 The Alpha Helix

One of the most important structural features is the alpha helix (Figure 1.3). This is a right-handed helical structure containing 3.6 amino acid residues in each turn. It is formed when each N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues before it in the polypeptide chain. This occurs as the C=O groups in the helix are parallel to the axis and are directionally aligned with the N-H groups to which the hydrogen bond is formed. The amino acid side chains are positioned away from the axis. Alpha helices can vary in length, although there are few

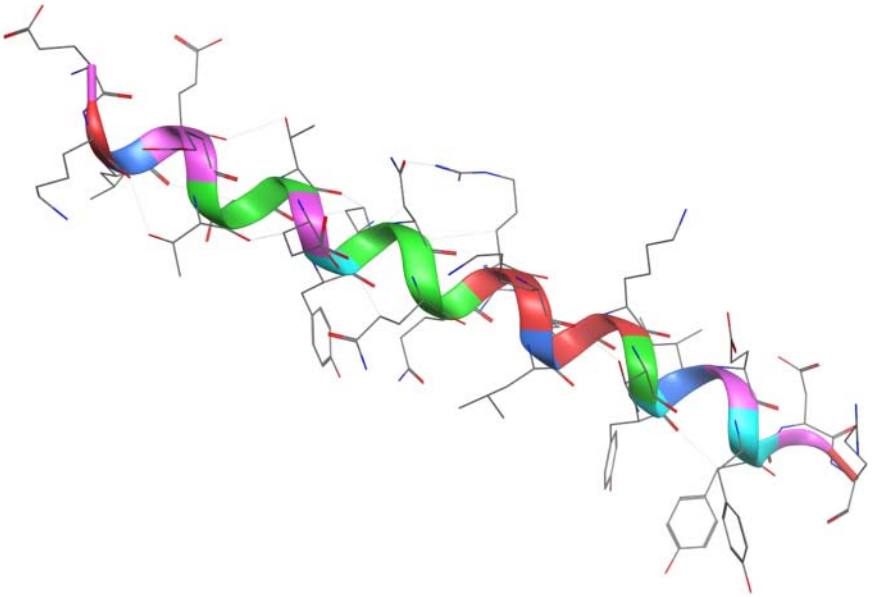


Figure 1.3 The alpha helix.
The structure of the alpha helix is shown: the backbone of the helix is represented in cartoon, and sticks show the amino acid side chains protruding from the backbone. The colors used are from the Clustal-X color scheme (Table 1.1).

examples of proteins where the helix length extends beyond 40 residues. Clearly, the first and last residue of an alpha helix cannot make hydrogen bonds to contribute to the helix, so these residues are often amino acids that can make hydrogen bonds with other parts of the protein or with the solvent. Some residues are more likely to form alpha helices than others, with alanine, leucine, arginine, methionine, and lysine having the highest propensity, although the tendency to form helices will depend on the identities of the neighboring residues. Conversely, residues such as aspartate, glycine, and proline tend not to form alpha helices. Proline cannot donate an amide hydrogen bond and also interferes sterically with the backbone of the preceding turn. However, proline may sometimes be positioned as the first residue in an alpha helix, providing structural rigidity to the helix. Often, alpha helices display an amphipathic nature, with hydrophobic residues located on one side and hydrophilic residues on the other. Another feature of alpha helices is that they tend to have a macrodipole, with the Nterminus being the positive pole. This arises as the individual microdipoles from the carbonyl groups of the peptide bonds in the helix align along the axis [1, 2].

1.1.2.2 The Beta Sheet

Another common structural motif in proteins is the beta sheet (Figure 1.4). When the backbone of a protein exists in an extended conformation (beta strand), it is possible for residues to make complementary hydrogen bonds with another beta strand. These interactions may occur when the chains are aligned in the same or opposite directions. When the chains are aligned in the same direction, the arrangement is termed a parallel beta sheet, and when the chains alternate in direction, it is termed an antiparallel beta sheet. Usually, an extensive hydrogen bond network is established where the N–H groups in the backbone of one strand establish hydrogen bonds with the C=O groups in the backbone of the adjacent strand. Often, beta sheets contain around 10 residues but can be much shorter (as low as 2 or 3 residues). Beta sheets often contain large aromatic residues

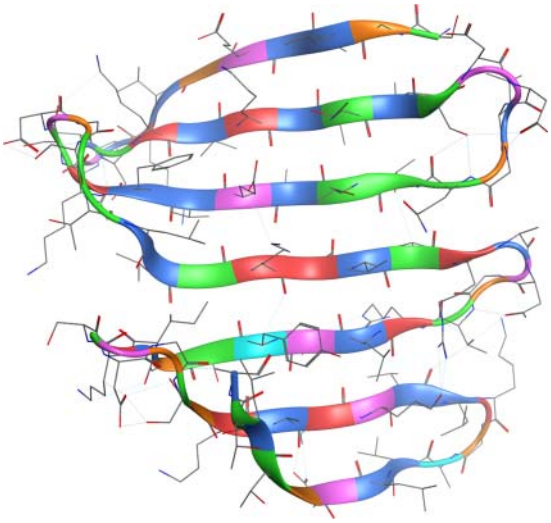


Figure 1.4 The beta sheet.
The structure of a beta sheet is shown: the backbone of the sheet is represented in cartoon, and sticks show the amino acid side chains protruding from the backbone.

(tyrosine, phenylalanine, and tryptophan) and branched amino acids (threonine, isoleucine, and valine) [1, 2].

1.1.2.3 Loops

There are segments of a protein that connect the alpha helix and beta sheet elements together, which in themselves do not have recognizable regular structural patterns. These secondary structural elements are termed loops (Figure 1.5). Loops are an important component of secondary structure, often containing as much as half of the total number of residues in a protein [3]. Loops often contribute significantly to the overall shape, dynamics, and physicochemical properties of the protein [4]. Loops are frequently located on the protein's surface in solvent-exposed

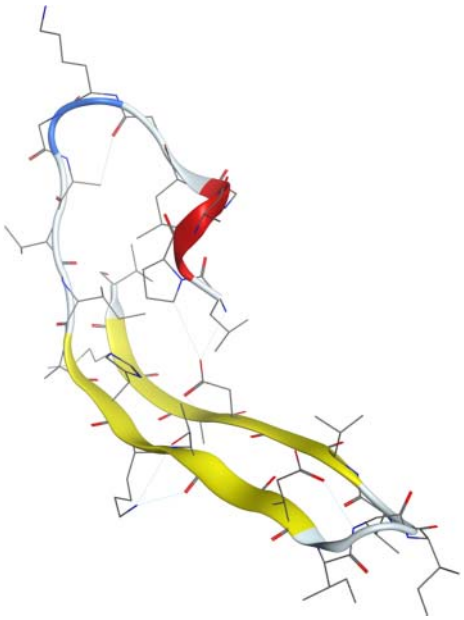


Figure 1.5 Loop region.
The structure of a loop region is shown: the backbone of the sheet is represented in cartoon, and sticks show the amino acid side chains protruding from the backbone.

Table 1.1 Clustal-X color scheme for coloring amino acids.

Clustal-X color scheme		
Category	**Color**	**Residue**
Hydrophobic	Blue	A, I, L, M, F, W, V
Positive charge	Red	K, R
Negative charge	Magenta	E, D
Polar	Green	N, Q, S, T
Cysteine	Pink	C
Glycine	Orange	G
Proline	Yellow	P
Aromatic	Cyan	H, Y
Unconserved	White	Any, gap

regions and are often involved in important interactions. Despite the lack of patterns, loops do not appear to be completely random structures, and they have been classified in various ways, including their geometrical shape [5]. However, even though their importance is recognized, loop structure remains difficult to predict.

The primary structure of a protein influences the secondary structure, with certain residues more likely to form one structure over the other; for example, proline residues are often called "helix breakers" as their cyclic nature induces a kink in the polypeptide chain and prevent alpha helix formation. Glycine residues, for example, also are frequently involved in tight turns as they are small and flexible [2].

The image for the loop structure has been colored by structure (in the program MOE2022; red: alpha helix, yellow: beta sheet, loop: white, turn: blue). The alpha helix and beta sheet above (Figures 1.3 and 1.4) have been colored using the Clustal-X color scheme (Table 1.1).

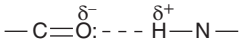
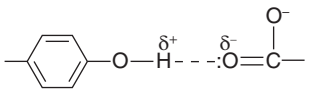
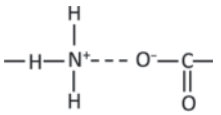
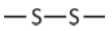
1.1.3 Tertiary Structure

The three-dimensional (3D) structure of a protein is defined by the position of all the atoms of the polypeptide chain arranged in 3D space. This is termed the tertiary (3°) structure, and it comprises the arrangement of the secondary structural elements, as described in Section 1.1.2, and involves numerous interactions between residues (Table 1.2). Proteins often contain regions that are conserved across or within families and which carry out similar functions in each. These are distinct folded units in a polypeptide chain that provide structural or functional features. For example, a protein could contain regulatory and catalytic domains. These structural elements may be domains, folds, and motifs. The combination of these elements in a single polypeptide chain may be sufficient to produce a fully functional protein without requirement for additional polypeptides [6].

1.1.3.1 Domains, Folds, and Motifs

Protein domains are areas of a protein that can fold stably and serve a specific purpose. Proteins may contain one or more domains (Figure 1.7). Each protein domain has a specific fold, which describes how the secondary structural elements in that domain are organized. Often, one fold

Table 1.2 Bonds and interactions in proteins.

Interactions	Description	Residues involved
Hydrogen bonds Interpeptide bond  Side chain bond 	A hydrogen bond is the electrostatic interaction between a hydrogen atom (covalently bound to a residue) and the electronegative atom (O, N) of another residue. In proteins, this occurs between the hydrogen atom of the N–H group of one residue and the oxygen atom of the O=C of another residue.	Polar
Ionic bonds 	An ionic bond is the electrostatic interaction between two groups of opposite charges.	Charged
Hydrophobic interactions 	Hydrophobic interactions describe the tendency of non-polar molecules to associate in aqueous solution, resulting largely from the favorable entropy produced by the breaking of the hydrogen bonds of the surrounding water.	Non-polar
Disulfide bonds 	Disulfide bonds are covalent interactions between two sulfur molecules; this type of interaction only occurs between cysteine side chains.	Cys

may be used by several different proteins to fulfill a range of activities. A structural motif is a small region of 3D structure that arises in a range of diverse proteins that may have a unique function. Some common protein motifs are shown in Figure 1.6.

An example of protein domains is shown in Figure 1.7.

Proteins are usually considered as layers of the secondary structural elements (alpha helices, α and beta sheets, β), and four types of proteins have been described based on the combinations of these elements. These are α/α (consisting of all α), β/β (consisting of all β), α/β (consisting of both α and β in varied regions), and α+β (where the α and β elements occur in different regions to each other).

1.1.4 Quaternary Structure

Whilst many proteins are made up of a single polypeptide chain, some proteins require multiple subunits to come together for functional activity. The coming together of multiple subunits forms the quaternary (4°) structure of a protein, and these can either be formed of multiple copies of the same (homopolymer) or different (heteropolymer) subunits. An example of a protein with a quaternary structure is hemoglobin, which is a heteropolymer formed of four subunits (two alpha and two beta subunits) [1, 2].

1.1.5 Protein Structure Prediction

The application of artificial intelligence has revolutionized protein structure prediction [7]. AlphaFold2 [8], DeepMind's machine-learning protein structure prediction program, released in

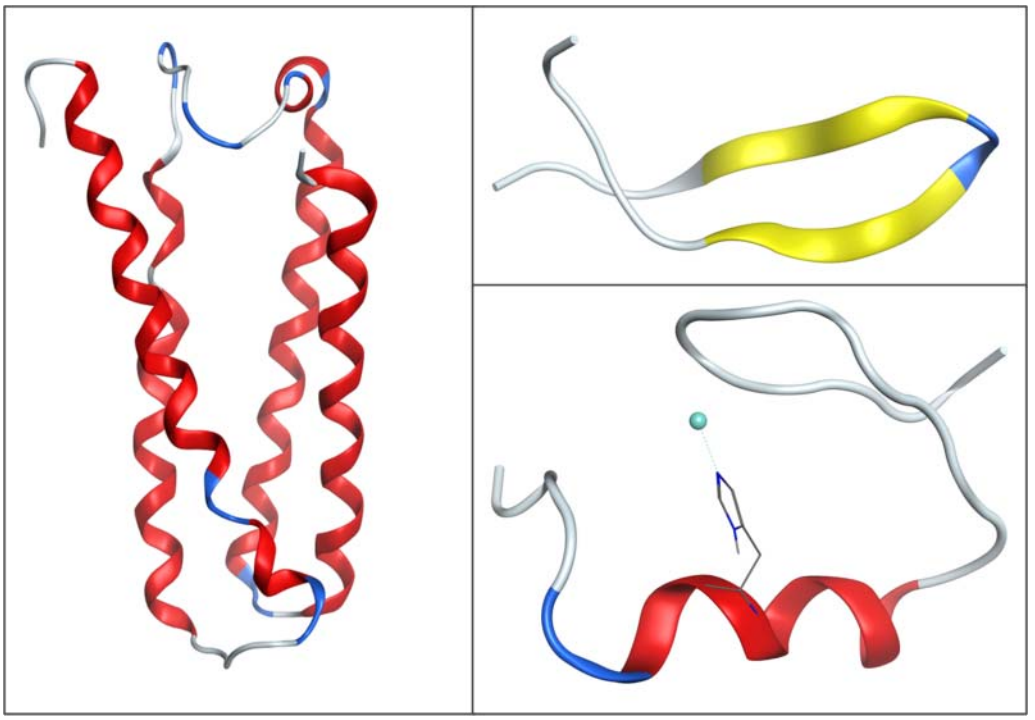
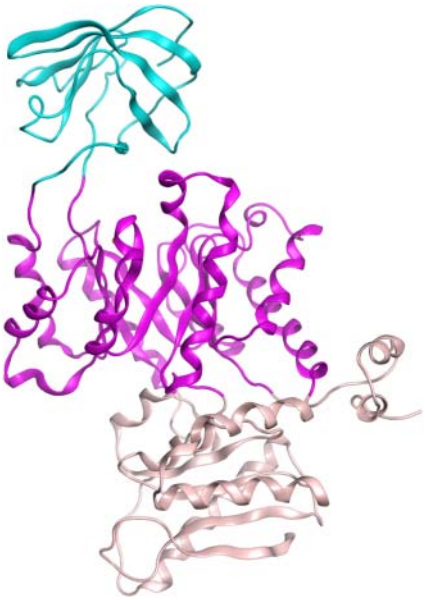


Figure 1.6 Common protein motifs.
Common protein motifs are shown in cartoon format: a four-helix bundle motif (left) is four α-helices packed together, typically in a lengthwise manner; a hairpin (top right) is a simple structure consisting of two antiparallel β-sheets joined by a loop; and a zinc finger motif (bottom right) is two beta strands with an alpha helix folded over to bind a zinc ion.

Figure 1.7 Multi-domain protein.
Pyruvate kinase contains an all-β-nucleotide-binding domain (in blue), an α/β-substrate-binding domain (in magenta), and an α/β-regulatory domain (in pink).



2021, enables the generation of high-confidence protein structures. This was the first indication that deep learning-based methods can now predict protein structures with an accuracy often comparable to that of experimental structures. AlphaFold2 generates 3D structures from input amino acid sequences by querying several databases of protein sequences and constructing a multiple sequence alignment (MSA). This enables the determination of the parts of the sequence that are more variable and allows the detection of correlations between them. The network also tries to identify proteins that may have a similar structure to the input, termed "templates", and constructs an initial "pair representation", which suggests which amino acids are likely to be in contact with each other. The MSA and the templates are passed through a transformer, which identifies the most useful information, and cyclically exchanges between the model and sequence alignment. Finally, the structure module incorporates the MSA representation and pair representation to construct a 3D model of the structure, which comprises a list of Cartesian coordinates representing the position of each atom of the protein. After generating a final structure, it is passed back through the process to refine the prediction. In 2022, AlphaFold2 was used to predict more than 200 million proteins from around 1 million different species. Now, another neural network, ESMFold [9] from Meta has been used to predict over 600 million proteins. Although this network is not currently as accurate as AlphaFold2, it is around 60 times faster at predicting structures for short sequences. Additionally, many of these structures are unlike anything in the current databases of experimentally determined protein structures or any of AlphaFold's predictions from known organisms. These breakthrough in artificial intelligence-based approaches demonstrate that structure prediction will become simpler, cheaper, and more accurate in future and will allow researchers to obtain structural information about proteins much more rapidly, allowing significant advances in the understanding of many areas of biology.

The AlphaFold source code can be found here.

https://github.com/deepmind/alphafold

The ESMFold source code can be found here.

https://github.com/facebookresearch/esm

1.2 Enzymes

Enzymes are proteins that carry out highly specific activities. Every enzyme contains an active site in which a specific molecule or reactant known as a substrate will bind, and the enzyme will convert the substrate into a product through a series of steps (Figure 1.4). Compounds that mimic the substrate are frequently used in drug discovery as a strategy to inhibit the targeted enzyme [10, 11].

The International Union of Pure and Applied Chemistry (IUPAC) Enzyme Commission (EC) developed a system to classify enzymes, to avoid confusion between different names for the same enzyme. This system is based on a numerical nomenclature, which describes the reaction catalyzed by a particular enzyme. Enzymes are split into six broad classes (Table 1.3), these are oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases [12]. Within these enzyme classes, there are further subdivisions of enzyme types. An example of a large subclass of enzymes are kinases, which phosphorylate their protein substrate (often another kinase) using adenosine triphosphate (ATP). This phosphorylation can act like a switch, activating or inactivating the protein substrate. Kinases are normally involved in complex signaling cascades in which there is a chain of kinases, each acting on the subsequent kinase in the cascade to bring about a biological effect.

Table 1.3 Enzyme classification.

Number	Class	Reaction type	Selected sub-classes
1	Oxidoreductases	Oxidation/reduction	1.1 acts on CH–OH group 1.2 acts on aldehyde group 1.3 acts on CH–CH group 1.4 acts on CH–NH_2 group
2	Transferases	Chemical group transfer reactions	2.1 transfers 1 carbon group 2.3 acyltransferases 2.4 glycosyltransferases 2.7 phosphotransferases
3	Hydrolases	Hydrolytic bond cleavage reactions	3.1 esterases 3.2 glycosidases 3.4 peptidases
4	Lyases	Non-hydrolytic bond cleavage or elimination reactions	4.1 C–C lyases 4.2 C–O lyases 4.3 C–N lyases
5	Isomerases	Rearrangement of atoms in molecules (isomerization)	5.1 racemases 5.3 intramolecular oxidoreductases 5.4 intramolecular transferases
6	Ligases	Bond synthesis to join two or more molecules together, coupled to hydrolysis (e.g. ATP)	6.1 C–O ligases 6.2 C–S ligases 6.3 C–N ligases

Notes: Enzymes in each class are subdivided using a second number, which more specifically defines the catalyzed reaction and classifies the sub-class. Third and fourth numbers classify each enzyme further into sub-sub-class and serial number to give each enzyme a unique identifier of the form EC 1.2.3.4, respectively.

1.2.1 Properties of Enzymes

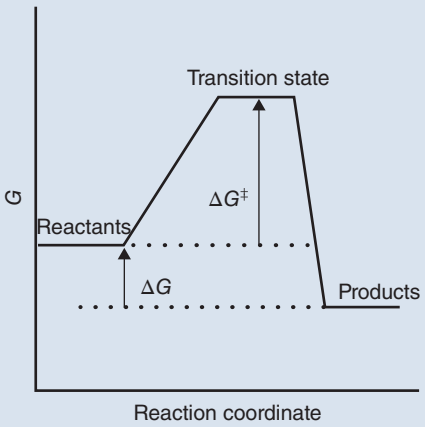
1.2.1.1 Catalysis

Enzymes are biological catalysts. Like all catalysts, they increase the rate of a reaction without perturbing the equilibrium position (bringing about the same rate enhancement in either direction) and remain unchanged after the reaction. This enables the essential reactions carried out within the cell to proceed rapidly enough for metabolism to be maintained. Enzyme catalysis is usually discussed in terms of a model known as transition state theory. Enzymes increase the rate of reaction by reducing the free energy of the transition state (Figure 1.3). The transition state is defined as the most unstable species on the reaction pathway, and so occurs at a peak in the free energy profile of a reaction. Compounds that are analogous to the transition state usually bind strongly and potently inhibit enzyme activity by outcompeting the substrate [12].

Key Concept: Transition State Theory and Free Energy Diagrams

All chemical reactions pass through an unstable intermediate termed the transition state, which is a transitory structure between those of the substrates and products. The lifetime of the transition state is purported to be around 10^{-13} seconds, similar to the time for a single bond vibration. Although direct observation of the structure of the transition state is not possible, it is central to understanding catalysis, because enzymes function by lowering the activation energy barrier by tightly binding to the transition state. This allows a greater proportion of the substrate to reach the energy needed to proceed to the product.

Often a free energy diagram is constructed as a simple 2D attempt to mimic the 3D potential energy surface, which describes the set of potential energy paths the components of the reaction follow. The diagram shows the initial energy of the substrates, the transition state, and the products. The difference between the substrates' ground state and the transition state is the activation energy.



The energy required to progress from the reactant ground state to the transition state is called the activation energy (E_A) or energy barrier and is the difference in free energy between the reactant or ground state and the transition state. The height of the activation energy barrier is related

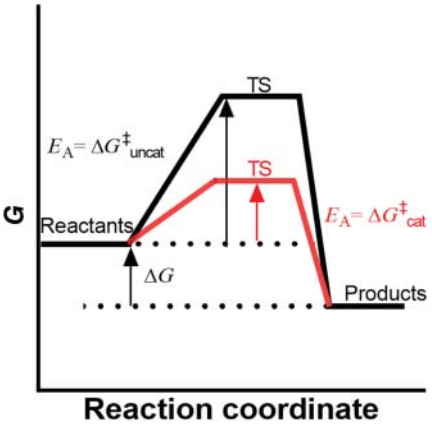


Figure 1.8 Energy profile diagram.
Reactants require energy (E_A) to reach their transition state and then form products. Products are formed through two routes: uncatalyzed (without enzyme, black) and catalyzed (with enzyme, red). The free energy of activation, $\Delta G^{\ddagger}$, is determined from the temperature dependence of the reaction rate, and the overall free energy change, ΔG, which is calculated from the reaction equilibrium constant.

to the rate of a chemical reaction. Enzymes work by reducing the size of this barrier and may increase the rates of reactions significantly (Figure 1.8).

Enzymes have unique structures that have evolved to enhance catalytic rates. The surface of the enzyme interacting with solvent water comprises mostly polar groups, whereas the interior of the protein is non-polar. Binding sites on the surface potentially allow for steep polarity gradients moving toward the interior of the protein, which provides the required environment for catalysis to occur. Additionally, hydrogen bonding, hydrophobic interactions, and charge–charge interactions involved in maintaining the enzyme's structure can be exploited to bind substrate molecules in a specific manner. In the absence of substrate, the binding site is filled with water molecules which must be displaced as the enzyme-substrate (ES) complex is formed. The energy of interaction must be great enough to expel this water and overcome the unfavorable entropy change associated with the formation of the complex. The reactivity is conferred by binding and introducing strain into the substrate so that it is biased towards the transition state. This ensures optimum binding of the transition state rather than the substrate itself, resulting in the exquisite alignment of the catalytic groups from the enzyme that will elicit the chemical step (Figure 1.9). This arrangement allows very large rate enhancements enabled by the stabilization of the transition state of the enzyme-catalyzed reaction relative to the uncatalyzed reaction. Enhancements in the rate of almost 10^{17}-fold have been observed in sweet potato β-amylase [13].

1.2.1.2 Specificity

Enzymes are incredibly specific, being able to discriminate between enantiomers (isomers of a molecule that are mirror images of each other), for example, the difference between L-amino acids (existing naturally) and D-amino acids and D-sugars (existing naturally) and L-sugars. This property also means only one stereoisomer of a drug may be capable of acting on the target; ibuprofen is an example. The active ingredient in ibuprofen is the *S*-isomer, which inhibits the cyclooxygenase enzyme (COX), but ibuprofen is a racemic mix of *R* and *S* isomers (Figure 1.10) [14, 15].

Enzymes are also able to catalyze the reactions they promote under surprisingly mild and economic conditions compared to other catalysts. Essentially, most enzymes seem to work best at temperatures close to body temperature, near neutral pH values, and at standard atmospheric pressures, qualities lacking in most industrial catalysts. For example, the Haber process is an

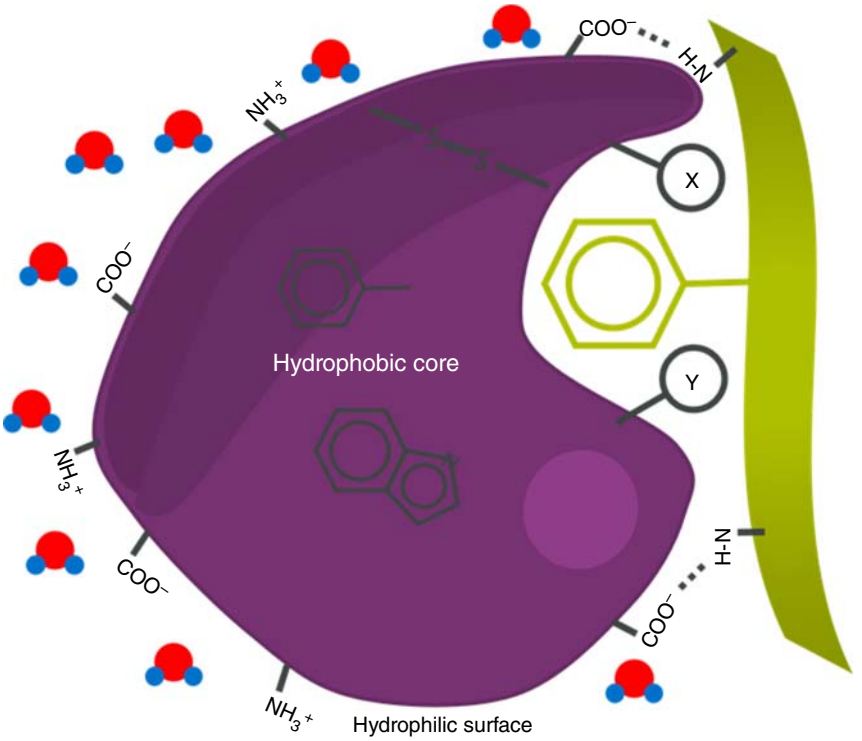


Figure 1.9 **General model of the interaction of a substrate with an enzyme**.
The hydrophobic core of the enzyme (purple) generally contains residues that have higher hydrophobicity. The hydrophilic surface contains residues, which interact with the solvent. Hydrogen bonding interactions are depicted between the enzyme and substrate (green), positioning the relevant substrate group close to the groups involved in catalysis (X and Y).

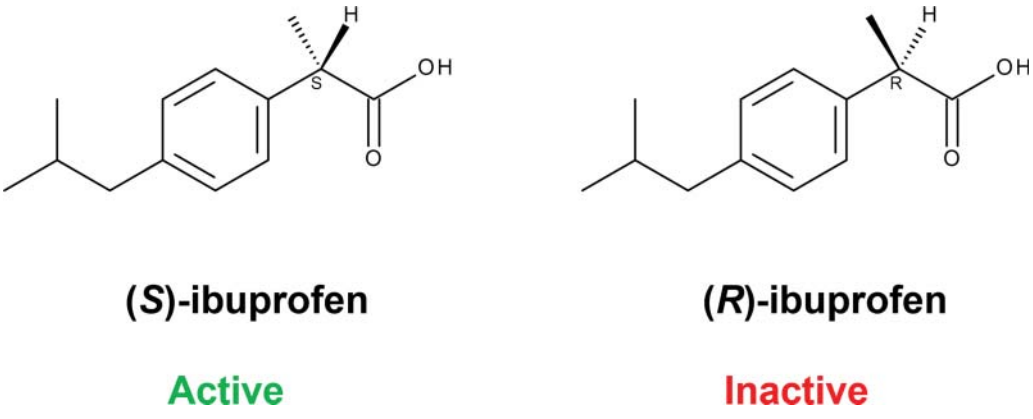


Figure 1.10 **Ibuprofen stereoisomers**.
Commercially bought ibuprofen is a mixture of stereoisomers; the *S*-isomer inhibits cyclooxygenase (COX), whereas the *R*-isomer is inactive against the enzyme.