Wiley Series on Mass Spectrometry
Dominic M. Desiderio and Nico M. M. Nibbering, Series Editors

MASS SPECTROMETRY IN STRUCTURAL BIOLOGY AND BIOPHYSICS

Architecture, Dynamics, and Interaction of Biomolecules

Second Edition

IGOR A. KALTASHOV
STEPHEN J. EYLES

WILEY

MASS SPECTROMETRY IN STRUCTURAL BIOLOGY AND BIOPHYSICS

Architecture, Dynamics, and Interaction of Biomolecules

Second Edition

IGOR A. KALTASHOV
STEPHEN J. EYLES
University of Massachusetts at Amherst

WILEY
A JOHN WILEY & SONS, INC., PUBLICATION

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

Library of Congress Cataloging-in-Publication Data:

Kaltashov, Igor A.
Mass spectrometry in structural biology and biophysics : architecture, dynamics, and interaction of biomolecules / Igor A Kaltashov, Stephen J Eyles. - 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-93779-2 (cloth)
1. Mass spectrometry. 2. Biophysics. 3. Biomolecules-Spectra. I. Eyles, Stephen J. II. Title.
QP519.9.M3K35 2012
610.28'4–dc23

2011040425

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE TO THE SECOND EDITION

The first edition of *Mass Spectrometry in Biophysics* was published over six years ago, and this field has experienced a truly transformative change during this period. Investigation of architecture and behavior of biopolymers (mostly proteins) by mass spectrometry (MS) was performed in the early 2000s in only a handful of laboratories around the world. The results of these studies were frequently met with skepticism outside of the MS community. However, by the end of the decade, MS had become an indispensable tool in experimental biophysics, which is capable of providing unique information on the conformation and dynamics of biopolymers, as well as their interactions with physiological partners. Not only has MS continued to progress at an accelerated pace throughout these years, but the scope of its applications in biophysics and structural biology also expanded very dramatically. As a result of these developments, some of the segments of the first edition became somewhat outdated and no longer provided adequate coverage of several key state-of-the-art techniques.

Another exciting change that has occurred in recent years is that MS-based studies of protein behavior are no longer confined to the realm of academic science. Indeed, the explosive growth of the biopharmaceutical sector in the past decade brings to the fore the need to have capabilities to analyze behavior of protein therapeutics and places a premium on developing analytical techniques able to handle these extremely complex species. Mass spectrometry can certainly fit the bill, and the gradual acceptance of these new tools within the biopharmaceutical industry and regulatory agencies as reliable methods to study architecture and dynamics of biomolecules is of little surprise to anyone.

In preparing the second edition of this book, our aim was to bring the reader up to date with the field by providing an expanded and up-to-date coverage of MS-based experimental methodologies in biophysics and structural biology, as well as addressing the specific needs of the new and rapidly growing segment of practitioners of this technique in the biopharmaceutical industry. We have tried as much as possible to preserve the original organization of the book, which proved very efficient in presenting the material. Introductory Chapters 1 and 2 were minimally changed, while Chapter 3 was updated to reflect, *inter alia*, introduction and rapid proliferation of Orbitrap mass analyzers and ion mobility spectrometers, as well as wide acceptance of the so-called electron-based ion fragmentation techniques (e.g., electron capture and electron transfer dissociation). The most extensive changes were made to Chapters 4–7, which present experimental methodologies used to probe various aspects of protein architecture and behavior under a variety of conditions. Similarly, very extensive revision was made to Chapter 8 (Chapter 9 in the first edition), which reflects a continuing expansion of MS into the realm of oligonucleotides, polysaccharides, and synthetic polymers, as well as polymer–protein conjugates.

Former Chapters 8 and 10 from the first edition were removed from this book. Indeed, the synergism between MS and other biophysical techniques (the topic of the former Chapter 8) is now commonly accepted, and in fact has become a defining element in the experimental design; many examples of this are dispersed throughout the text of the second edition. Studies of structure and behavior of biopolymers in the gas phase (the topic of the former Chapter 10) have now transformed into a separate field, and its careful and detailed consideration is no longer possible in this book given obvious space limitations. The exception is made for several gas-phase methods that are either already used to study solution structure (e.g., gas-phase H/D exchange to probe oligonucleotide conformations) or show promise in that

regard (e.g., measurements of biopolymer ion mobility in the gas phase). The final chapter of this book (chapter 9, which was old chapter 11) once again strives to go beyond routine measurements and considers several fields that are currently out of the reach of the commonly accepted MS based techniques (membrane proteins, protein aggregates, very large biopolymer assemblies, etc.).

Taken together, the second edition is a systematic presentation of a modern mass spectrometry-based armamentarium that can be used to solve a variety of challenging problems in biophysics, structural biology, and biopharmaceuticals. One of our goals was not only to provide practical advice, but also to arm the reader with a solid coverage of all relevant fundamental issues, including extensive references to, and examples from, the original published work. In addition to that, the book contains a large number of examples and illustrations taken from the work carried out in our laboratory, some of which have never been published. We are indebted to the past and present group members who provided this material (names in parentheses indicate present employment if different from UMass-Amherst): Dr. Dmitry R. Gumerov (Mersana Pharmaceuticals), Dr. Andras Dobo (Sigma-Aldrich-Fluka Europe), Prof. Hui Xiao (Albert Einstein School of Medicine), Dr. Anirban Mohimen (Vertex Pharmaceuticals), Prof. Wendell Griffith (University of Toledo), Dr. Joshua K. Hoerner (Schering-Plough Research Institute), Dr. Mingxuan "Sunshine" Zhang (Biogen IDEC), Dr. Virginie Sjoelund (National Institutes of Health), Dr. Rachael Leverence (University of Wisconsin), Dr. Agya Frimpong, Dr. Rinat R. Abzalimov, Dr. Cedric E. Bobst, Mr. Guanbo Wang, and Mr. Shunhai Wang.

We are also grateful to Professors Michael L. Gross (Washington University at St. Louis), S. Walter Englander (University of Pennsylvania Medical School), George H. Lorimer (University of Maryland at College Park), Virgil L. Woods, Jr. (University of California at San Diego School of Medicine), Roman A. Zubarev (Karolinska Institute), Lars Konermann (University of Western Ontario), Joseph A. Loo (UCLA), John Engen (Northeastern University), and Richard W. Vachet (UMass-Amherst) for numerous very helpful discussions over the past several years that have had direct impact on this book. We would also like to acknowledge our collaborators from industry who helped us better understand the unique needs of the biopharmaceutical sector and how they can be addressed using mass spectrometry tools: Dr. Pavel Bondarenko (Amgen, Inc.), Drs. Steven Berkowitz and Damian Houde (Biogen IDEC), and Drs. Philip Savickas, John Thomas, Melanie Lin, and Paul Salinas (Shire Human Genetic Therapies). Finally, we would like to acknowledge the National Institutes of Health and the National Science Foundation for their generous support of our own research efforts at the interface of biophysics and MS, many examples of which are presented in this book.

IGOR A. KALTASHOV

STEPHEN J. EYLES

University of Massachusetts at Amherst

PREFACE TO THE FIRST EDITION

Strictly speaking, the term *biophysics* refers to the application of the theories and methods of physics to answer questions in the biological arena. This obviously now vast field began with studies of how electrical impulses are transmitted in biological systems and how the shapes of biomolecules enable them to perform complex biological functions. Over time, biophysicists have added a wide variety of methodologies to their experimental toolkit, one of the more recent additions being mass spectrometry (MS). Traditionally limited to the analysis of small molecules, recent technological advances have enabled the field of MS to expand into the biophysical laboratory, catalyzed by the 2002 Nobel prize winning work of John Fenn and Koichi Tanaka. Mass spectrometry is a rapidly developing field whose applications are constantly changing. This text represents only a snapshot of current techniques and methodologies.

This book aims to present a detailed and systematic coverage of the current state of biophysical MS with special emphasis on experimental techniques that are used to study protein higher order structure and dynamics. No longer an exotic novelty, various MS based methods are rapidly gaining acceptance in the biophysical community as powerful experimental tools to probe various aspects of biomolecular behavior both *in vitro* and *in vivo*. Although this field is now experiencing an explosive growth, there is no single text that focuses solely on applications of MS in molecular biophysics and provides a thorough summary of the plethora of MS experimental techniques and strategies that can be used to address a wide variety of problems related to biomolecular dynamics and higher order structure. This book aims to close that gap.

We intended to target two distinct audiences: mass spectrometrists who are working in various fields of life sciences (but are not necessarily experts in biophysics) and experimental biophysicists (who are less familiar with recent developments in MS technology, but would like to add it to their experimental arsenal). In order to make the book equally useful for both groups, the presentation of the MS based techniques in biophysics is preceded by a discussion of general biophysical concepts related to the structure and dynamics of biological macromolecules (Chapter 1). Although it is not meant to provide an exhaustive coverage of the entire field of molecular biophysics, the fundamental concepts are explained in some detail to enable anyone not directly involved with the field to understand the important aspects and terminology. Chapter 2 provides a brief overview of "traditional" biophysical techniques with special emphasis on those that are complementary to MS and that are mentioned elsewhere in the book. These introductory chapters are followed by an in-depth discussion of modern mass spectrometric hardware used in experimental studies of biomolecular structure and dynamics. The purpose of Chapter 3 is to provide readers who are less familiar with MS with concise background material on modern MS instrumentation and techniques that will be referred to in the later chapters (the book is structured in such a way that no prior familiarity with biological MS is required of the reader).

Chapters 4–7 deal with various aspects of protein higher order structure and dynamics as probed by various MS based methods. Chapter 4 focuses on "static structures", by considering various approaches to evaluate higher order structure of proteins at various levels of spatial resolution when crystallographic and nuclear magnetic resonance (NMR) data are either unavailable or insufficient. The major emphasis is on methods that are used to probe biomolecular topology and solvent accessibility (i.e., chemical cross-linking and

selective chemical modification). In addition, the use of hydrogen–deuterium exchange for mapping protein–protein interfaces is briefly discussed. Chapter 5 presents a concise introduction to an array of techniques that are used to study structure and behavior of non-native protein states that become populated under denaturing conditions. The chapter begins with consideration of protein ion charge state distributions in electrospray ionization mass spectra as indicators of protein unfolding and concludes with a detailed discussion of hydrogen exchange, arguably one of the most widely used methods to probe the structure and dynamics of non-native protein states under equilibrium conditions. The kinetic aspects of protein folding and enzyme catalysis are considered in Chapter 6. Chapter 7 focuses on MS based methods that are used to extract quantitative information on protein–ligand interactions (i.e., indirect methods of assessment of binding energy). The remainder of this chapter is devoted to advanced uses of MS to characterize dynamics of multiprotein assemblies and its role in modulating protein function.

Complementarity of MS based techniques to other experimental tools is emphasized throughout the book and is also addressed specifically in Chapter 8. Two examples presented in this chapter are considered in sufficient detail to illustrate the power of synergy of multiple biophysical techniques, where some methods provide overlapping information to confirm the evidence, while others provide completely unique details. Chapter 9 presents a discussion of MS based methods to study the higher order structure and dynamics of biopolymers that are not proteins (oligonucleotides, polysaccharides, as well as polymers of nonbiotic origin). Chapter 10 provides a brief discussion of biomolecular properties in the gas phase, focusing primarily on the relevance of *in vacuo* measurements to biomolecular properties in solution.

This book concludes with a discussion of the current challenges facing biomolecular MS, as well as important new developments in the field that are not yet ready for routine use. Chapter 11 focuses on several areas where MS is currently making a debut. It begins with a discussion of novel uses of MS aimed at understanding "orderly" protein oligomerization processes, followed by consideration of "catastrophic" oligomerization (e.g., amyloidosis). This chapter also considers other challenging tasks facing modern MS, such as the detection and characterization of very large macromolecular assemblies (e.g., intact ribosomes and viral particles), as well as applications of various MS based techniques to study the behavior of a notoriously difficult class of biopolymers–membrane proteins. This chapter concludes with a general discussion of the relevance of *in vitro* studies and reductionist models to processes occurring *in vivo*.

Throughout the entire book, an effort has been made to present the material in a systematic fashion. Both the theoretical background and technical aspects of each technique are discussed in detail, followed by an outline of its advantages and limitations, so that the reader can get a clear sense of both current capabilities and potential future uses of various MS based experimental methodologies. Furthermore, this book was conceived as a combination of a textbook, a good reference source, and a practical guide. With that in mind, a large amount of material (practical information) has been included throughout. An effort has also been made to provide the reader with a large reference base to original research papers, so that the details of experimental work omitted in the book can easily be found. Because of space limitations and the vastness of the field, a significant volume of very interesting and important research could not be physically cited. It is hoped, however, that no important experimental techniques and methodologies have been overlooked. The authors will be grateful for any comments from the readers on the material presented in the book (Chapters 1, 3, 4, 5, 7, 10, and 11 were written mostly by I.K. and Chapters 6, 8, and 9 by S.E.; both authors contributed equally to Chapter 2). The comments can be e-mailed directly to the authors at kaltashov@chem.umass.edu and eyles@polysci.umass.edu.

We are grateful to Professors David L. Smith, Michael L. Gross, Max Deinzer, Lars Konermann, Joseph A. Loo, and Richard W. Vachet for helpful discussions over the past several years that have had direct impact on this book. We would also like to thank many other colleagues, collaborators, and friends for their support and encouragement during various stages of this challenging project. We are also indebted to many people who have made contributions to this book in the form of original graphics from research articles (the credits are given in the relevant parts of the text). We also thank the current and past members of our research group, who in many cases contributed original unpublished data for the illustrative material presented throughout. Finally, we would like to acknowledge the National Institutes of Health and the National Science Foundation for their generous support of our own research efforts at the interface of biophysics and mass spectrometry.

IGOR A. KALTASHOV

STEPHEN J. EYLES

University of Massachusetts at Amherst

CONTENTS

1

GENERAL OVERVIEW OF BASIC CONCEPTS IN MOLECULAR BIOPHYSICS

This introductory chapter provides a brief overview of the basic concepts and current questions facing biophysicists in terms of the structural characterization of proteins, protein folding, and protein–ligand interactions. Although this chapter is not meant to provide an exhaustive coverage of the entire field of molecular biophysics, the fundamental concepts are explained in some detail to enable anyone not directly involved with the field to understand the important aspects and terminology.

1.1. COVALENT STRUCTURE OF BIOPOLYMERS

Biopolymers are a class of polymeric materials that are manufactured in nature. Depending on the building blocks (or *repeat units* using polymer terminology), biopolymers are usually divided into three large classes. These are (1) polynucleotides (built of nucleotides); (2) peptides and proteins (built of amino acids); and (3) polysaccharides (built of various saccharide units). This chapter only considers general properties of biopolymers using peptides and proteins as examples; questions related to polynucleotides and polysaccharides will be discussed in some detail in Chapter 8.

All polypeptides are linear chains built of small organic molecules called *amino acids*. There are 20 amino acids that are commonly considered *canonical* or *natural* (Table 1.1). This assignment is based upon the fact that these 20 amino acids correspond to 61 (out of total 64) codons within the triplet genetic code with three remaining codons functioning as terminators of protein synthesis (1,2), although there are at least as many other amino acids that occur less frequently in living organisms (Table 1.2). Noncanonical amino acids are usually produced by chemical modification of a related canonical amino acid (e.g., oxidation of proline produces hydroxyproline), although at least two of them (selenocysteine and pyrrolysine) should be considered canonical based on the way they are utilized in protein synthesis *in vivo* by some organisms (3,4). Furthermore, new components can be added to the protein biosynthetic machinery of both prokaryotes and eukaryotes, which makes it possible to genetically encode unnatural amino acids *in vivo* (5,6). A peculiar structural feature of all canonical (with the exception of glycine) and most noncanonical amino acids is the presence of an asymmetric carbon atom (C_α), which should give rise to two different enantiomeric forms. Remarkably, all canonical amino acids are of the L-type. The D-forms of amino acids can also be synthesized *in vivo*, and are particularly abundant in fungi; however, these amino acids do not have access to the genetic code. The rise and persistence of homochirality in the living world throughout the entire evolution of life remains one of the greatest puzzles in biology; examples of homochirality at the molecular level also include almost exclusive occurrence of the D-forms of sugars in the nucleotides, while manifestations of homochirality at the macroscopic level range from specific helical patterns of snail shells to the chewing motions of cows (7,8).

Unlike most synthetic polymers and structural biopolymers (several examples of which will be presented in Chapter 8), peptides and proteins have a very specific sequence of monomer units. Therefore, even though polypeptides can be considered simply as highly functionalized linear polymers constituting a nylon-2 backbone, these functional groups, or *side chains*, are arranged in a highly specific order. All

Mass Spectrometry in Structural Biology and Biophysics: Architecture, Dynamics, and Interaction of Biomolecules, Second Edition.
Igor A. Kaltashov and Stephen J. Eyles.

TABLE 1.1. Chemical Structure and Masses of Natural (Canonical) Amino Acids

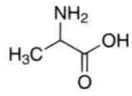
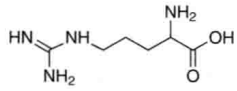
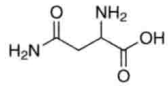
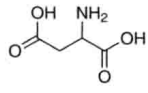
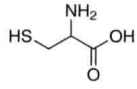
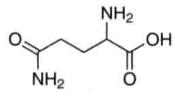
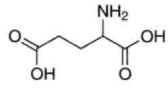
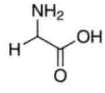
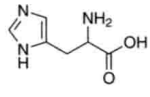
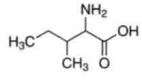
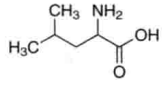
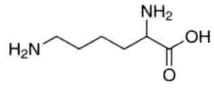
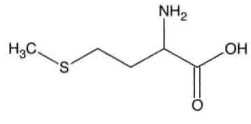
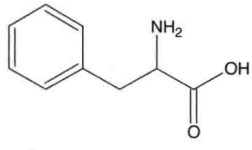
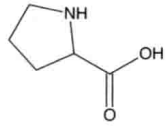
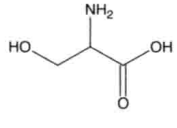
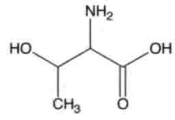
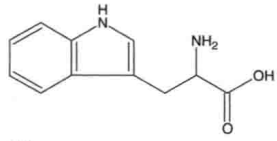
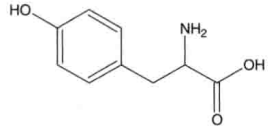
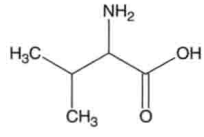
Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass[a] (Residue)	Average Mass (Residue)
Ala (A)	Alanine	C_3H_5NO		Nonpolar	71.037	71.079
Arg (R)	Arginine	$C_6H_{12}N_4O$		Basic	156.101	156.188
Asn (N)	Asparagine	$C_4H_6N_2O_2$		Polar	114.043	114.104
Asp (D)	Aspartic acid	$C_4H_5NO_3$		Acidic	115.027	115.089
Cys (C)	Cysteine	C_3H_5NOS		Polar/acidic	103.009	103.145
Gln (Q)	Glutamine	$C_5H_8N_2O_2$		Polar	128.059	128.131
Glu (E)	Glutamic acid	$C_5H_7NO_3$		Acidic	129.043	129.116
Gly (G)	Glycine	C_2H_3NO		Nonpolar	57.021	57.052
His (H)	Histidine	$C_6H_7N_3O$		Basic	137.059	137.141
Ile (I)	Isoleucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Leu (L)	Leucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Lys (K)	Lysine	$C_6H_{12}N_2O$		Basic	128.095	128.174
Met (M)	Methionine	C_5H_9NOS		Nonpolar/ amphipathic	131.040	131.199
Phe (F)	Phenylalanine	C_9H_9NO		Nonpolar	147.068	147.177
Pro (P)	Proline	C_5H_7NO		Nonpolar	97.053	97.117
Ser (S)	Serine	$C_3H_5NO_2$		Polar	87.032	87.078

TABLE 1.1. (*Continued*)

Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass[a] (Residue)	Average Mass (Residue)
Thr (T)	Threonine	$C_4H_7NO_2$		Polar/amphipathic	101.048	101.105
Trp (W)	Tryptophan	$C_{11}H_{10}N_2O$		Amphipathic	186.079	186.213
Tyr (Y)	Tyrosine	$C_9H_9NO_2$		Amphipathic	163.063	163.176
Val (V)	Valine	C_5H_9NO		Nonpolar	99.068	99.133

[a] See Chapter 3 for a definition of monoisotopic and average masses.

naturally occurring proteins consist of an exact sequence of amino acid residues linked by peptide bonds (Fig. 1.1*a*), which is usually referred to as the *primary structure*. Some amino acids can be modified after translation (termed *post-translational modification*), for instance, by phosphorylation, methylation, or glycosylation. Among these modifications, formation of the covalent bonds between two cysteine residues is particularly interesting, since such *disulfide bridges* can stabilize protein geometry, by bringing together residues that are distant in the primary structure into close proximity in three-dimensional (3D) space. The highly specific spatial organization of many (but not all) proteins under certain conditions is often referred to as *higher order structure* and is another point of distinction between them (as well as most biological macromolecules) and synthetic polymers. Although disulfide bridges are often important contributors to the stability of the higher order structure, correct protein folding does not necessarily require such covalent "stitches". In fact, cysteine is one of the least abundant amino acids, and many proteins lack it altogether. As it turns out, relatively weak noncovalent interactions between functional groups of the amino acid side chains and the polypeptide backbone are much more important for the highly specific arrangement of the protein in 3D space. Section 1.2 provides a brief overview of such interactions.

1.2. NONCOVALENT INTERACTIONS AND HIGHER ORDER STRUCTURE

Just like all chemical forces, all inter- and intramolecular interactions involving biological macromolecules (both covalent and noncovalent) are electrical in nature and can be described generally by the superposition of Coulombic potentials. In practice, however, the noncovalent interactions are subdivided into several categories, each being characterized by a set of unique features.

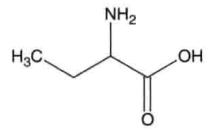
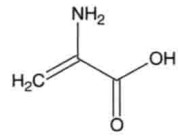
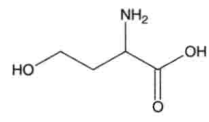
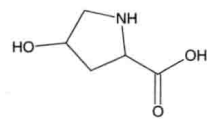
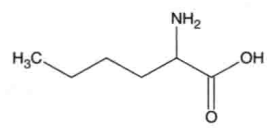
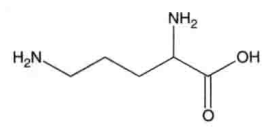
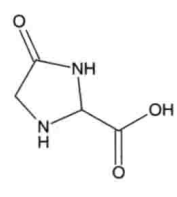
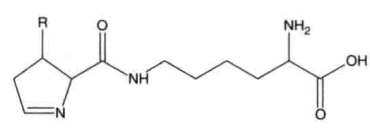
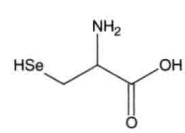
1.2.1. Electrostatic Interaction

The term *electrostatic interaction* broadly refers to a range of forces exerted among a set of stationary charges and/or dipoles. The interaction between two fixed charges q_1 and q_2 separated by a distance r is given by the Coulomb law:

$$E = \frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r} \qquad (1\text{-}2\text{-}1)$$

where ε_0 is the absolute permittivity of vacuum [8.85×10^{-12} C^2/N·m in Système International (SI)] and ε is the dielectric constant of the medium. Although the numerical values of the dielectric constants of most homogeneous media are readily available, the use of this concept at the microscopic level is not very straightforward (9,10). The dielectric constant is a measure of the screening of the electrostatic interaction due to the polarization of the medium, hence the difficulty in defining a single constant for a protein, where such screening depends on the exact location of the charges, their environment, and so on. Although in some cases the values of the "effective" dielectric constants for specific protein systems can be estimated based on experimental measurements of the electrostatic interactions, such an approach has been disfavored by many for a long time (11). This book will follow the example set by Daune (12) and will write all expressions with $\varepsilon = 1$.

TABLE 1.2. Chemical Structure and Masses of Some Less Frequently Occurring Natural (Noncanonical) Amino Acids

Symbol	Name	Molecular Formula (Residue)	Chemical Structure	Side-Chain Character	Monoisotopic Mass (Residue)	Average Mass (Residue)
Abu	2-Aminobutyric acid	C_4H_7NO		Nonpolar	85.053	85.106
Dha	Dehydroalanine	C_3H_3NO		Nonpolar	69.021	69.063
Hse	Homoserine	$C_4H_7NO_2$		Polar	101.048	101.105
Hyp	Hydroxyproline	$C_6H_{12}N_2O_2$		Polar	144.090	144.174
Nle	Norleucine	$C_6H_{11}NO$		Nonpolar	113.084	113.160
Orn	Ornithine	$C_5H_{10}N_2O$		Basic	114.079	114.147
Pyr	Pyroglutamic acid	$C_5H_5NO_2$		Moderately polar	111.032	111.100
Pyl	Pyrrolysine	$C_{11}H_{16}N_3O_2 + R$ (NH_2, OH, or CH_3)		Polar		
Sec	Selenocysteine	C_3H_5NOSe		Polar/acidic	144.960 (150.954[a])	150.039

[a] Most abundant.

Interaction between a charge q and a permanent dipole p separated by a distance r is given by

$$E = -\frac{qp \cdot \cos\theta}{4\pi\varepsilon_0 r^2} \qquad (1\text{-}2\text{-}2)$$

where θ is the angle between the direction of the dipole and the vector connecting it with the charge q. If the dipole is not fixed directionally, it will align itself to minimize the energy Eq. (1-2-2), that is, $\theta = 0$. However, if such energy is small compared to thermal energy, Brownian motion will result in