Preface

The thought of putting together the present book came up after we had just co-authored an extensive review on Arginine (Gupta and Uversky 2024a) which in turn led us to speculate that many of the properties of Arginine may in fact be due to its tendency to form clusters and repeat sequences (Gupta and Uversky 2024b). In the present book, we examine the biological importance of repeat sequences of other amino acids as well.

The first chapter presented by Yuzo Fujino and Yoshitaka Nagai discusses a general phenomenon of repeat expansion diseases associated with the expanded short tandem repeats (STRs). There are at least 50 such repeat expansion diseases, which originate from the STR expansions happening in coding or noncoding regions of the corresponding genes. In addition to the polyQ maladies, coding repeat expansion diseases include polyalaninine (polyA) and polyglycine (polyG) diseases, originating from the expansion of the CAG, GCN, and GGC repeat sequences, respectively. Noncoding repeat expansion diseases are caused by the expansion of trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide, or even dodecanucleotide sequences. The corresponding sequences can be translated via a noncanonical form of translation, repeat-associated non-AUG (RAN) translation, which is characterized by the AUG codon-independent and repeat sequence-dependent translation initiation, and which can happen from all reading frames.

This chapter also delves into how these expansions correlate with intrinsic disorder, which leads them also to play important roles in the dynamics of the reversible formation of the membrane less organelles in cells via liquid-liquid phase separation.

While the chapters in this book can be read in any order and independently, it will be helpful to read this first chapter as an overview of the fairly wide footprint of the repeat sequences in biology as well as medical sciences.

The second chapter by Raffaella Parlato and Patrick C.A. van der Wel is dedicated to the introduction of a class of protein aggregation diseases associated with the expansion of the repetitive polyglutamine (polyQ) segment (also known as polyQ or CAG-repeat expansion disorders or polyQ-related syndromes). This chapter is mostly focused on the polyQ expansion-driven Huntington's disease. Although nine neurodegenerative diseases are associated with expansion of polyQ region of specific proteins, the most well-known example is Huntington's disease, which arises as a consequence of the polyQ repeat expansion in the huntingtin (HTT) protein above a specific threshold (35 glutamines or 35 CAG codons in the *HTT* gene).

Biochemists sometimes tend to overlook that a large number of proteins have hydrophobic clusters on their surfaces. The third chapter by Fabio C.L. Almeida, Ramon Pinheiro-Aguiar and Ariana Azevedo Vasconcelos though deals with a special class of proteins, which do not have a hydrophobic core but have stable 3D structures secured by multiple disulfide (S–S) bonds and contain surface hydrophobic clusters important for their function. These proteins represent an interesting case of intrinsically disordered proteins that are capable of gaining stable structures as a result of the oxidative folding. Readers will note that the intrinsic disorder runs as a common thread in case of cellular functions of proteins with repeat sequences, which are discussed in various chapters.

The roles of various classes of amino acid clusters in protein-protein interactions is the theme of the fourth chapter. Here, Juan Martín Hernández Castillo, Albros Hermes Poot and Gabriel Del Río introduce Residue Cluster Classes (RCC), which are maximally connected clusters of amino acids found in protein structures. RCC serves as a useful means for the efficient classification of protein structure and prediction of protein function and protein-protein interactions.

The fifth chapter by Esteban Lanzarotti, Lucas Defelipe, Marcelo Marti and Adrian Turjanski specifically discusses aromatic clusters and also covers their importance in protein-drug complex formations. The importance of aromatic residues for structure and function of proteins is difficult to overestimate. It was shown that protein data bank (PDB) virtually does not contain "non-smelly" proteins, i.e., proteins that do not contain cysteine and aromatic residues (CFYWH-depleted), indicating that that unique 3D structures of monomeric globular proteins are crucially dependent on the presence of cysteine and aromatic residues. Overall, three major roles of aromatic clusters have been outlined, where intraprotein (IP) aromatic clusters are formed by interactions between residues of the same protein chain and thereby are only involved in protein folding; protein-drug (PD) aromatic clusters are present at the sites where drug-like ligands are bound; and protein-

protein (PP) aromatic clusters are found at the protein-protein interaction interfaces.

Cell-penetrating peptides represent another set of molecules, which have repeat amino acid sequences. These peptides (both of natural origin and artificially designed) can cross cell membranes without compromising their structural or functional integrity and thereby serve as promising means for the intracellular delivery of various compounds ranging from small molecules to peptides, proteins, and nucleic acids. The sixth chapter by Ngoc Hai Trieu Phong, Mohammed Kamal, Duy Binh Tran and Tohru Yamada covers the ever-increasing importance of such cell-penetrating peptides (many of which contain repeat amino acid motifs), particularly in the context of the design of novel drug-delivery systems with high efficiency and controllable specificity.

Elastin-like peptides gained attention initially as a fusion tag for convenient purification of recombinant proteins. These now constitute a hugely versatile class of molecules, which straddle a large number of areas in biotechnology and biomedical sciences. The seventh chapter by Takeru Nose and Keitaro Suyama is a good overview of these peptides and their importance in both biology and biotechnology.

All textbooks of biochemistry, enzymology, and protein science typically identify aromatic amino acids as chromophores responsible for the intrinsic fluorescence of proteins. The last and eighth chapter by Vishnu N. Vijayan, Karthika Kannan and Sushabhan Sadhukhan describe the fluorescence shown by some non-aromatic amino acids, when these occur as repeat sequences. Fluorescence-based tools and techniques have now become immensely powerful in the study of biological molecules and processes. The applications of the fluorescence shown by the non-aromatic amino acids which are described in this interesting chapter may turn out to be the proverbial tip of the ice berg.

We hope that the book will be found useful by diverse kinds of readers who are interested in the involvement of peptides and proteins in biology, biotechnology, drug delivery designs and sensors and many other areas.

We thank all the contributing authors for their efforts.

Editors
Munishwar Nath Gupta
Vladimir N. Uversky