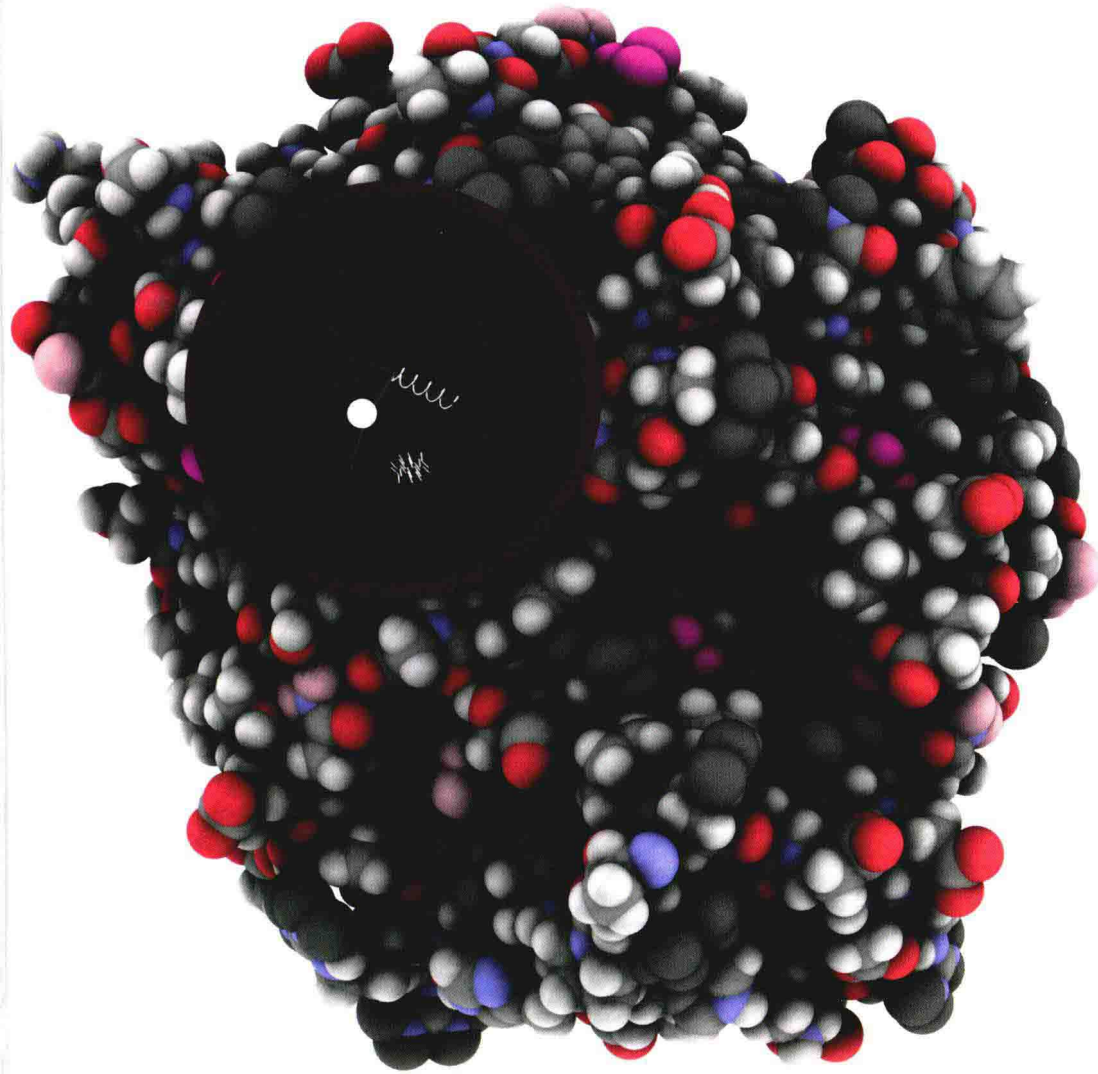
Concepts and Applications of Protein Structure

Steven Tiff



Concepts and Applications of Protein Structure

Edited by **Steven Tiff**

New York

Published by Callisto Reference,
106 Park Avenue, Suite 200,
New York, NY 10016, USA
www.callistoreference.com

Concepts and Applications of Protein Structure
Edited by Steven Tiff

© 2015 Callisto Reference

International Standard Book Number: 978-1-63239-122-3 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Copyright for all individual chapters remain with the respective authors as indicated. A wide variety of references are listed. Permission and sources are indicated; for detailed attributions, please refer to the permissions page. Reasonable efforts have been made to publish reliable data and information, but the authors, editors and publisher cannot assume any responsibility for the validity of all materials or the consequences of their use.

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy. Furthermore, the publisher ensures that the text paper and cover boards used have met acceptable environmental accreditation standards.

Trademark Notice: Registered trademark of products or corporate names are used only for explanation and identification without intent to infringe.

Printed in China.

Concepts and Applications of Protein Structure

Preface

Over the recent decade, advancements and applications have progressed exponentially. This has led to the increased interest in this field and projects are being conducted to enhance knowledge. The main objective of this book is to present some of the critical challenges and provide insights into possible solutions. This book will answer the varied questions that arise in the field and also provide an increased scope for furthering studies.

This book compiles the latest studies and advancements in understanding protein structure. Due to several advancements in the field of science and technology, experts today are able to explain various protein structures thoroughly. This book is a collection of several aspects and researches related to protein structures. Several experts have given their contribution in the compilation of data enclosed in this book. This book is a useful reference of knowledge for both students and experts dealing with protein structure.

I hope that this book, with its visionary approach, will be a valuable addition and will promote interest among readers. Each of the authors has provided their extraordinary competence in their specific fields by providing different perspectives as they come from diverse nations and regions. I thank them for their contributions.

Editor

此为试读,需要完整PDF请访问: www.ertongbook.com

Contents

Section 1

Introduction

1

An Evolutionary Biology Approach to Understanding Neurological Disorders

Azhari Aziz[1], Jasmina Ilievska[1,2],
Paul R. Fisher[2] and Naomi E. Bishop[1,3]
[1]*Cell Biology and Molecular Pathogenesis Laboratory,*
[2]*Molecular Cell Biology Laboratory,*
Department of Microbiology,
[3]*Olga Tennison Autism Research Centre,*
La Trobe University, Melbourne, Victoria,
Australia

1. Introduction

Many common human neurological disorders, including epilepsy, Alzheimer's disease, Parkinson's disease, autism spectrum disorders, and schizophrenia show complex heritability and genetics. While studies of single-gene diseases typically provide a more straightforward opportunity to understand the underlying molecular mechanisms of disease, complex diseases are more common and inherently more difficult to study. Nonetheless, researchers have begun to make dramatic inroads into the study of complex human diseases, including many neurological disorders, in the post-human genome sequence era. This is largely due to new technologies and resources that are promoting our understanding of protein structure and function, thereby facilitating the association of disease phenotypes with genetic loci. The online Mendelian inheritance in man (OMIM) database lists those genes implicated in human disease, and this highlights progress made in this field, where around 10% of human genes have a known disease-association (Amberger et al., 2009).

In the first few sections of this paper, we highlight the differences and similarities between simple and complex human genetic disorders, and key methods to study these disorders. We emphasize the key role comparative and evolutionary biology techniques play in increasing our understanding of the pathophysiology of complex human disorders, including in the assessment of the functional traits of gene products implicated in human disease. Several human neurological disorders are used to illustrate the power of this methodology. In the last sections of this paper, the significance and implications of comparative and evolutionary biology data are highlighted using schizophrenia, and autism as specific examples. The surprising recent links between neurological disorders and cancer are discussed in the final section. We conclude that exploration of the evolutionary history of human genes, and comparison of protein structure, helps us understand how and why human neurological disorders originated, influences the choice of appropriate animal

此为试读,需要完整PDF请访问: www.ertongbook.com

models for human disorders, and informs our interpretation of data from model organisms, including the evaluation of novel therapeutics. We conclude that comparative and evolutionary biology, including techniques facilitating the prediction of protein function, has a major role in facilitating further understanding of human neurological disorders and in the development of therapeutic interventions.

2. Classification of human genetic disorders

A genetic disorder is a disease caused by an abnormality, or abnormalities, in genetic material or genome architecture. Genetic disorders are traditionally subdivided into four types: (i) single-gene disorders (often referred to as Mendelian or monogenic diseases), (ii) mitochondrial genome disorders; (iii) chromosomal disorders (where there are gross changes in chromosome structure, such as loss, duplication and/or translocation) diseases, and (iv) multigenic or complex diseases.

Much progress has been made in the last few decades in identifying the molecular cause of many rare genetic diseases (Amberger et al., 2009). Most of these are highly-penetrant traits due to single-gene mutations, and therefore follow classical Mendelian inheritance patterns (Antonarakis & Beckmann, 2006). Good progress has also been made in understanding mitochondrial genome disorders, particularly in the 30 years since the publication of the reference sequence for human mitochondrial DNA (Anderson et al., 1981; Kumar, 2008; Tuppen et al., 2010). Furthermore, a wide variety of chromosomal disorders have been characterized, where defects can be visualised microscopically (Theisen A & Shaffer LG, 2010). These three classes of genetic disorders are individually rare, although chromosomal disorders are being reported slightly more frequently in recent decades, due to factors such as increased parental age (Jones, 2008; Fonseka & Griffin, 2011), and technological advances facilitating detection of smaller deletions and duplications (Berg et al., 2010; Shaffer et al., 2007; Slavotinek, 2008). Despite these broad categories frequently being used to classify human genetic disorders, it must be borne in mind that the phenotypic expression of genetic mutations varies, and this is discussed next, before we focus our attention on complex genetic disorders.

2.1 The complexity of single-gene disorders

It is now well-known, even for well-characterized Mendelian genetic disorders, that individuals with a specific mutation can display phenotypic differences. This includes variation in the age of onset, in the severity of disease symptoms, and/or in phenotypic characteristics. Indeed, phenotypic pleiotropy is the rule, rather than the exception, even for single-gene disorders (Nadeau, 2001). The variation in individual phenotype can be affected by environmental factors, allelic variation and/or 'modifier' genes. Modifier genes can affect transcription and levels of gene expression directly, or affect phenotype at the cellular, tissue, or organism-level (Nadeau, 2001). While increasing numbers of human modifier genes are being identified, most of the progress in understanding genetic modifiers is dependent on model organisms, such as mice, where gene targeting experiments can be carried out using inbred strains (Nadeau, 2001, 2003). One particularly relevant class of modifiers are referred to as protective alleles, as their presence prevents disease from occurring (Nadeau, 2001, 2003). These findings provide insights relevant to the development

of novel therapeutics, as therapeutics could be based on mimicking and/or possibly enhancing the effects of these protective alleles. This class of gene also means that the same genetic mutation can lead to a different disease phenotype depending on the genetic background (Lobo, 2008). Therefore, while genetic disorders are frequently classified as monogenic or complex (see below), the distinction between the two types is becoming increasingly blurred.

Other emerging aspects of monogenic disorders overlapping with those of complex disorders, are those due to the multi-functional nature of many genes. This multi-functionality can make it difficult to predict phenotype from genotype. This is illustrated by metabolic genes, where some gene products are referred to as 'moonlighting proteins' due to the multiple phenotypic effects of mutations (Jeffery, 2009; Sriram et al., 2005). Such multi-functionality also contributes to the observation that distinct phenotypes can be associated with different mutations in the same gene. For example, the *LMNA* gene encodes two proteins and is linked to five diseases (Vigouroux & Bonne, 2002), while mutation of the *ERCC2* gene may cause xeroderma pigmentosum (XP), Cockayne syndrome with XP, or trichothiodystrophy, three phenotypically different disorders (Lehmann et al., 2001). In other cases, different mutations in a single gene can cause different diseases via mechanistically different pathways. For example, the *FMR1* gene is considered 'a gene with three faces' (Oostra & Willemsen, 2009). Mutation of *FMR1* is best characterised as the cause of fragile X syndrome mental retardation, which is inherited in an X-linked dominant pattern, and is due to a lack of *FMR1* mRNA and protein expression. However, mutations leading to high levels of *FMR1* mRNA are linked to tremor/ataxia syndrome via mRNA 'toxicity', while the gene is also linked to premature ovarian insufficiency via a third uncharacterized molecular pathway, possibly affecting the production of *FMR1* mRNA isoforms (Oostra & Willemsen, 2009; Tassone et al., 2011).

Another factor complicating the phenotype of Mendelian disorders is the finding that heterozygotes for some recessively inherited Mendelian disorders, whom show no symptoms of the homozygotic phenotype, are at risk of an apparently unrelated disorder (Sidransky, 2006; Sriram et al., 2005). For example, patients who are heterozygous for the gene deficient in Gaucher disease are at an increased risk of neurodegenerative synucleinopathies, such as Parkinson disease (Sidransky, 2006). An additional complication arises in patients who show clinical symptoms consistent with a single-gene defect in a metabolic pathway, but do not have a complete deficiency in any one enzyme, but rather have multiple partial defects. This phenomenon is referred to as synergistic heterozygosity (Vockley et al., 2000).

Finally, while some genetic disorders are largely polygenic and complex in nature, a subset is inherited in a classical Mendelian manner (see Fig. 1). For example, with Alzheimer disease (ALZ) and Parkinson disease (PKD), a subset of the diseases (prefixed by the term 'familial') are inherited in a Mendelian manner. With PKD, around 5% of cases are due to mutations in one of several specific genes with either autosomal dominant or recessive inheritance patterns (Gasser, 2009; Lesage & Brice, 2009; Shulman et al., 2011), but PKD-associated genes with a more modest penetrance are now beginning to be identified (International Parkinson Disease Genomics Consortium, 2011; Liu et al., 2011; Shulman et al., 2011). With ALZ, around 0.1% of cases are inherited in an autosomal dominant manner, while one *APOE* allele, present in 2% of Caucasian populations, has recently been

reclassified from 'risk gene' status to being considered moderately penetrant with semi-dominant inheritance (Blennow et al. 2006; Genin et al., 2011). Nonetheless, ALZ in most patients is influenced by a combination of multiple genetic risk factors and protective alleles (Sherva & Farrer, 201; Waring & Rosenberg, 2008).

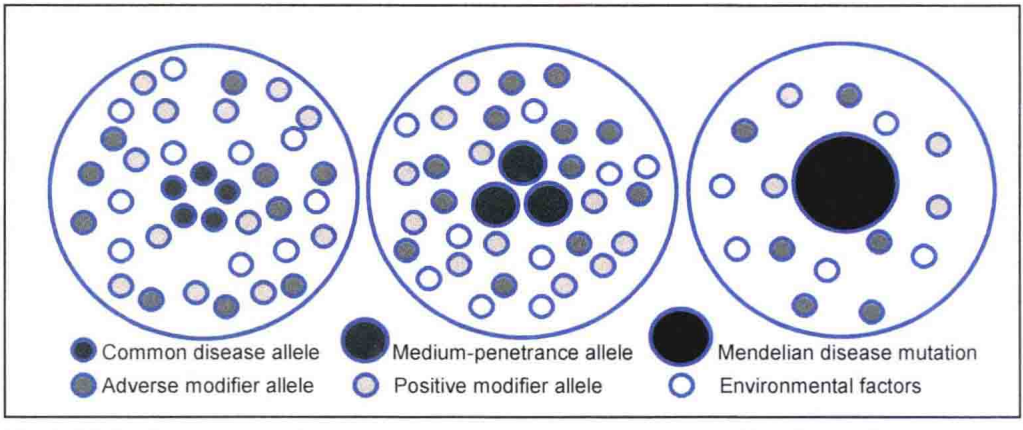


Fig. 1. Multiple genetic and evironmental variants in different combinations affect phenotypes. Common variants (left) or rare mutations (right), or alleles of medium penetrance (middle) can all cause human genetic disorders. Penetrance of a disease allele can be affected by so-called modifier alleles.

Therefore, Mendelian disorders have more in common with multi-factorial diseases than originally thought, and both are affected by genetic background and environmental conditions. Furthermore, the rare Mendelian forms of common complex disorders are providing key insights about the pathogenesis of many complex diseases by highlighting cellular pathways perturbed in the disease state (discussed further in Section 4.4) and this is leading to testable hypothese about disease etiology (Peltonen et al., 2006). Complex genetic disorders are discussed next, emphasizing the importance of evolutionary and comparative biology, while the relevance of these areas of research to multifunctional genes will be discussed further in Section 4.

2.2 Complex genetic disorders

While most Mendelian disorders are rare, there are over 7000 such disorders, and so they collectively affect hundreds of millions of people worldwide (Amberger et al., 2009). By contrast, most of the common disorders of children and adults are complex diseases, and a single highly-penetrant gene is not causative of the disease phenotype (see Fig. 1). Indeed, the causes of such disorders are usually heterogeneous, and a combination of effects from more than one gene, combined with non-genetic factors (environment), play a role in disease development (Davey Smith et al., 2005). Such disorders in children include mental retardation, autism spectrum disorders, attention deficit/hyperactivity disorder, and cancer. In adults, common complex disorders include schizophrenia, bipolar disorder, diabetes, coronary heart disease, hypertension, obesity, and cancer. The complex, multigenic, nature of these diseases has made them inherently more difficult to study. However, in the next

section of this paper, we will discuss the key methods used to determine the genetic underpinnings of common complex disorders. Understanding the etiology of these multifactorial diseases is essential for the development of effective means of treatment and/or prevention.

3. Studying complex human genetic disorders

Complex disorders often cluster in families without clearly demonstrating Mendelian inheritance patterns. This makes it difficult to determine the genetic versus non-genetic contribution to the disease phenotype, and to calculate the heritable component of the disorder. Below we will discuss methods used to establish the heritability of human complex disorders, generation of the genetic variation that underpins these disorders, and discuss how to establish which genes are responsible for complex human disease.

3.1 Heritability of complex human diseases

Heritability is usually defined as the proportion of total phenotypic variation that can be attributed to genetic variability (Lee et al., 2011; Visscher et al., 2008). While the interaction of environment on phenotype makes heritability difficult to measure accurately in some cases (Ober & Vercelli, 2011), methods of obtaining unbiased estimates of heritability from various types of pedigree data are well established for both continuous phenotypes and complex human disorders (Lee et al., 2011; Visscher et al., 2008). Furthermore, animal models are invaluable in dissecting aspects of genetic and environmental interactions that are more difficult to assess in human studies (Complex Trait Consortium, 2004) and is discussed further in Section 4.5.

For many human diseases, recent data suggest that the heritable component of many has previously been underestimated (Lee et al., 2011). This has been due to limitations of the methodology employed, as well as other factors, such as evidence demonstrating monozygotic twins are less genetically similar than once thought (Zwijnenburg et al., 2010). A further example is that of PKD, which was long considered a non-hereditary disorder (Shulman et al., 2011; Westerlund et al., 2010). Despite extensive efforts to find environmental risk factors for the disease, genetic variants now stand out as the major causative factor (Shulman et al., 2011; Westerlund et al., 2010; Wirdefeldt et al., 2011). This shift of focus away from environmental toxins, towards genetic contributions, is now leading to rapid progress in understanding PKD and in guiding the development of the next generation of therapeutics (Shulman et al., 2011).

While it is clear that genetics underpins the pathophysiology of complex human disorders, the genetic alleles contributing to the disease phenotype are not always inherited. *De novo* mutations are increasingly being implicated in human disease and, by definition, these mutations are not present in the biological parents of the affected individual. Nonetheless, depending on the severity of the phenotype, and on any effects on fitness, these novel mutations may be transmitted to subsequent generations. Indeed, the rise of techniques such as intracytoplasmic sperm injection (ICSI), can facilitate transmission of *de novo* mutations even if they lead to infertility (Jiang et al., 1999). There is a wide variety in *de novo* human germline mutations, and these can include duplications or deletions of various size, as well as alterations in the number of chromosomes (Arnheim & Calabrese, 2009). The frequency of

de novo mutations in germ-line cells increases with parental age, and can also be caused by environmental factors such as radiation exposure (Sasaki, 2006), genotoxic chemicals (Phillips & Arlt, 2009), or congenital viral infections (Ansari & Mason, 1977; Fortunato & Spector, 2003; Nusbacher et al., 1967; Vijaya-Lakshmi et al., 1999).

While *de novo* mutations typically occur in gametes, and are often defined as such, new mutations can also occur in the precursors of germ cells, leading to germline mosaicism (Arnheim & Calabrese, 2009), or occur post-fertilization, during embryonic/foetal development and somatic cells (Lupski, 2010). Indeed, new mutations can develop at any time, with cancer being the best-known example of a genetic disorder caused by somatic cell mutagenesis, while Proteus syndrome is a recently-identified disease linked to somatic mosaicism (Lindhurst et al., 2011). While mutations in the genome of somatic cells cannot be passed on to future generations, they may have detrimental effects. While *de novo* mutations are best-studied in cancer (see Section 6.2.3), a contributing role of somatic *de novo* mutations, such as those occurring during brain development, to neurological disorders (see Section 6.1.2) has not been explored.

Mutation frequencies also vary widely across the genome, and often concentrate at certain positions or 'hotspots' (see also Sections 3.2, 6.1.2 & 6.2.3), which have structural and functional features affecting mutagenesis (Ananda et al., 2011; Arnheim & Calabrese, 2009; Carvalho et al., 2010; Rogozin & Pavlov, 2003). For example, CpG context elevates the mutation rate by an order of magnitude (Schmidt et al., 2008), non-B DNA structures induced by palindromic AT-rich repeats facilitate recurrent translocations on chromosomes 11 and 22 at positions 11q23 and 22q11 (Kurahashi et al., 2006), while interspersed repetitive elements such as *Alu*, LINE, long-terminal repeats, and simple tandem repeats are frequently observed at breakpoints in the 9q34.3 subtelomere region (Yatsenko et al., 2009). However, at any given point, multiple mechanisms are acting, making prediction of mutational site and frequency difficult (Arnheim & Calabrese, 2009).

Despite *de novo* mutations most typically being deleterious, rather than neutral or advantageous, their very existence is evidence for ongoing adaptive evolution. Therefore, genetic disorders can be considered 'side-effects' or manifestations of the fundamental mechanisms that provide the genetic variation necessary for evolution to occur. The interaction between the evolutionary past of the human genome and human genetic disease is discussed next.

3.2 Human evolution and genetic disorders

Duplicated regions of DNA play a key role in the evolution of novel gene functions (Conant & Wolfe, 2008; Lynch & Conery, 2000; Ohno, 1970), but are also a source of genetic instability, leading to mutations implicated in both rare and common human genetic disorders (Marques-Bonet & Eichler, 2009). It is therefore relevant to explore the origins of human duplicated sequences. Current evidence indicates that many segmental duplications occurred the hominid lineage and, more specifically, in the common ancestor of African great apes (chimpanzee, gorilla, humans) after divergence from the Ponginae or Asian great ape (orangutan) lineage (Bailey & Eichler, 2006; Carvalho et al., 2010; Koszul & Fischer, 2009; Marques-Bonet & Eichler, 2009). A considerable portion of duplicated human sequences have also been found to correspond to expanded gene families, some of which show

signatures of positive selection (Marques-Bonet & Eichler, 2009). Duplications specific to the *Homo sapiens* lineage have also been detected, and include duplications in gene families implicated in neurotransmission, and these may play a role in higher-order brain function in humans (Han et al., 2009).

Different classes of repetitive DNA sequence have been identified (Bao & Eddy, 2002). Some, such as LINES (e.g. L1 family) or SINES (e.g. *Alu* family), are long and short retrotransposable elements, found interspersed throughout the genome. Others are concentrated in certain regions, such as centromeres and telomere-adjacent sequences. These latter regions are also sites of increased genomic instability, which are associated with disease-causing chromosomal breakpoints (Stankiewicz & Lupski, 2002). *Alu* elements have propagated to more than one million copies in primate genomes, and likewise contribute to human genomic diversity (Batzer & Deininger, 2002). Indeed, one in 50 individuals will carry a *de novo* L1 insertion, and one in 20 individuals a *de novo Alu* insertion (Collier & Largaespada, 2007). The active nature of many human retrotransposons is therefore linked to disease-causing somatic and germline mutations (Collier & Largaespada, 2007; Wallace et al., 1991; Oldridge et al., 1999; Claverie-Martin et al., 2003). Repeats may also contribute to DNA secondary structures that are more prone to breakage (Yatsenko et al., 2009). One novel aspect of our increased understanding of the role of repetitive DNA sequence in *de novo* mutations, and our ability to detect such sequences, is that this information can now be used to predict rearrangements that will contribute to genomic disorders (Carvalho et al., 2010; Ou et al., 2011; Sharp et al., 2006).

Therefore, while duplicated sequences in primate genomes predispose apes and humans to extensive genetic diversity and biological innovation, the downside is that many *de novo* genomic changes are mediated by recombination events between these duplications. This characteristic of hominids, and *Homo sapiens* in particular, makes humans particularly susceptible to genomic rearrangements. These rearrangements, in turn, then play a major role in human genetic disease pathogenesis (Inoue & Lupski, 2002; Marques-Bonet & Eichler, 2009). The evolutionary history of some specific genomic rearrangements is discussed next.

3.2.1 Evolutionary history of specific human disease mutations

Using an evolutionary perspective, we can use comparative genomic analyses to calculate the age of appearance of segmental duplications mediating specific disease-causing mutations. Such analyses have revealed that the segmental duplication flanking the Charcot-Marie-Tooth disease region on chromosome 17 (at position 17p12) has an origin in the hominoid ancestor after the divergence of chimpanzees and humans, those flanking the DiGeorge syndrome region on chromosome 22 (22q11.2) expanded after the divergence of hominoids from Old World monkeys, the duplications flanking the Angelman/Prader-Willi region on chromosome 15 (15q11-q13) began to expand before the divergence of the Old World monkeys, while the Smith-Magenis syndrome segmental duplications (17p11.2) date back to after the divergence of New World monkeys (Marques-Bonet & Eichler, 2009). These, and other similar data, have demonstrated that the predisposing genomic features contributing to many genomic disorders have emerged within the last 25 million years (Marques-Bonet & Eichler, 2009).

此为试读,需要完整PDF请访问: www.ertongbook.com