

Human genomic diversity, viral genomics and proteomics, as exemplified by human papillomaviruses and H5N1 influenza viruses

Meena K. Sakharkar,^{1,2} Kishore R. Sakharkar^{1,3} and Vincent T.K. Chow^{4*}

¹Biomedical Engineering Research Centre, Nanyang Technological University, Singapore

²Advanced Design and Modeling Laboratory, Nanyang Technological University, Singapore

³OmicVista, Singapore

⁴Infectious Diseases Program and Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

*Correspondence to: Tel: +65 6516 3691; Fax: +65 6776 6872; E-mail: micctk@nus.edu.sg

Date received (in revised form): 28th January 2009

Abstract

The diversity of hosts, pathogens and host–pathogen relationships reflects the influence of selective pressures that fuel diversity through ongoing interactions with other rapidly evolving molecules in the environment. This paper discusses specific examples illustrating the phenomenon of diversity of hosts and pathogens, with special reference to human papillomaviruses and H5N1 influenza viruses. We also review the influence of diverse host–pathogen interactions that determine the pathophysiology of infections, and their responses to drugs or vaccines.

Keywords: genomes, H5N1 avian influenza virus, human papillomaviruses, major histocompatibility complex, single nucleotide polymorphisms

Introduction

The availability of complete genome sequences and the wealth of large-scale biological datasets provide an unprecedented opportunity to elucidate the genetic bases of human diseases and host–pathogen interactions. The diversity of hosts, pathogens and host–pathogen relationships reflects the influence of selective pressures that fuel diversity through ongoing interactions with other rapidly evolving molecules in the environment. Such influences add another source of genetic adaptability as cells adjust to new environments and out-manoeuvre pathogenic threats. The study of genetic variation in pathogens and hosts has practical significance for developing strategies to combat and control infectious diseases. Vaccines based on highly

polymorphic antigens may be confounded by allelic restriction of the host immune response. In addition, the study of the distribution of genomic polymorphisms among different hosts may provide information on responses to drugs or vaccines. Based on the concept that host damage is the most relevant outcome of the host–pathogen interaction, we need better to understand both host and pathogen polymorphisms in drug or vaccine design. A deeper knowledge of the diversity and nature of pathogens will provide valuable insights into genetic markers that may be useful for detection, identification and forensics. The ability to discriminate between virulent pathogens and their counterparts that are either less or not virulent is another major challenge. It is important to discover genetic

factors that mediate the virulence process in order to devise novel methods to prevent or treat disease. Similarly, understanding how the host responds to microbial invasion, and how the pathogen evades or manipulates the immune response to subvert the host, will contribute to the development of vaccines and other prophylactic strategies. Concomitantly, the unravelling of associations between hosts and pathogens will be highly relevant to the modelling of the population biology of multi-host pathogens and their impact on co-infections. This paper reviews and discusses specific examples concerning the issue of diversity of hosts and pathogens, and the influence of diverse host–pathogen interactions that determine the pathophysiology of infections, and their responses to drugs or vaccines.

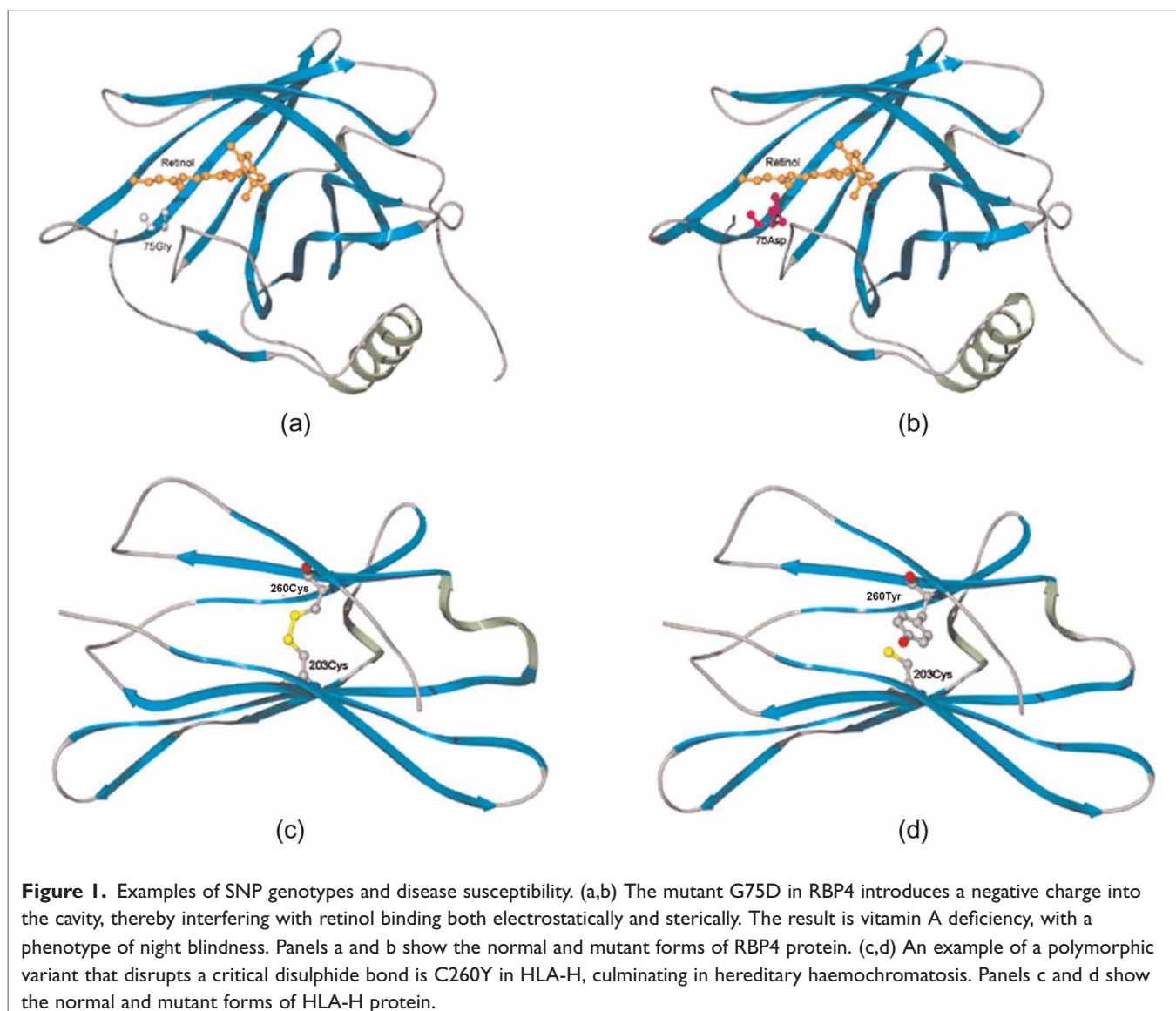
Human genome diversity and ‘SNPshots’

The genetic blueprint of an individual not only determines disease susceptibility, but also his/her response to drug treatment. Numerous genes are involved in drug response and toxicity, introducing a daunting level of complexity in the search for candidate genes. Thus, genetics—particularly gene polymorphisms—exert a significant impact on target discovery.

The HapMap constitutes a catalogue of common genetic variants that exist in humans. It describes what these variants are, where they occur in our DNA and how they are distributed among individuals within a specific population and among populations in different parts of the world. This project provides information that can link genetic variants to the risks for specific illnesses, which can lead to new methods for preventing, diagnosing and treating diseases.^{1,2} Differences in individual bases are by far the most common type of genetic variation. These genetic differences, known as single nucleotide polymorphisms (SNPs), represent DNA sequence variations that occur when a single nucleotide in the genome sequence is altered. SNPs are more common than other types of polymorphisms, and occur at a frequency of approximately one in 1,000 base pairs³ throughout the genome (including promoter regions and coding and intronic sequences). Some of these

differences may alter gene products in ways that confer susceptibility or resistance to diseases, or contribute to disease severity or progression. Although over 99 per cent of human DNA sequences are the same across the human population, DNA sequence variations impact on how humans respond to disease; to environmental stresses such as infections, toxins and chemicals; and to drugs and other therapies.⁴ Since genetic factors also affect response to drug therapy, SNPs can help to determine why individuals differ in their abilities to absorb or clear certain drugs, as well as to ascertain the mechanisms of adverse drug effects. Moreover, by affecting drug–target proteins such as G–protein–coupled receptors, enzymes, ion channels and proteins involved in detoxification pathways, non-synonymous coding SNPs (cSNPs) (namely substitutions resulting in alterations of encoded amino acids) significantly influence the diverse responses of efficacy and toxicity of therapeutic agents in the human population.⁵

In general, only about 5 per cent of the disease-causing non-synonymous mutations hitherto identified have direct effects on the catalytic or ligand-binding properties of enzymes and receptors.^{6–8} An interesting example is retinol-binding protein 4 (RBP4), the retinol-specific transport protein present in plasma. Elucidation of the crystal structures of different forms of RBPs have revealed their interactions with retinol, retinoids and transthyretin (TTR; one of the plasma carriers of thyroid hormones).⁹ The core of RBP is a beta-barrel whose cavity accommodates retinol. The retinol hydroxyl group is near the protein surface, in the region of the entrance loops surrounding the opening of the binding cavity, and participates in polar interactions. The G75D mutant introduces a negative charge into the cavity, thereby interfering with retinol binding both electrostatically and sterically. The result is vitamin A deficiency with a phenotype of night blindness.¹⁰ (Figure 1a, b). Furthermore, RBP4 is expressed and secreted by adipose tissue and is strongly associated with insulin resistance. A strong positive correlation also exists between RBP4 mRNA and adipose inflammation (monocyte chemoattractant protein-1 and CD68), and glucose transporter 4 mRNA.¹¹



Some non-synonymous cSNPs associated with human disorders disrupt important structural features of the affected protein. For example, a polymorphic variant that disrupts a critical disulphide bond is C260Y in HLA-H, culminating in hereditary haemochromatosis (Figure 1d).¹²

Human genome diversity and drug responses

The total number of SNPs reported in public SNP databases currently exceeds 9 million.¹³ Occasionally, an SNP may actually cause a disease and can therefore be exploited to search and isolate the disease-

causing gene. Since SNPs are genetic variations that occur at regular intervals and at high frequency throughout the human genome, they can be used as markers within the genome. If a particular marker is found to be common among individuals with a particular disease, it suggests that the gene involved is probably located near the marker.¹⁴ This renders SNPs of great value to biomedical research, and for developing pharmaceutical products or diagnostics.¹⁵

For virtually all medications, inter-patient variability in response to drug therapy is the rule rather than the exception. Tremendous progress has been achieved in understanding the molecular basis of drug action, and in elucidating genetic

determinants of disease pathogenesis and drug response.^{16,17} This inter-patient variability is potentially regulated by processes such as drug transport, drug metabolism, cellular signalling pathways (eg G-protein-coupled receptors) and response pathways (eg apoptosis, cell cycle control). Polymorphisms of drug-metabolising enzymes and/or pharmacological targets are frequently associated with adverse drug reactions or failure of efficacy.^{17–20} Drug responses may also be modulated by non-genetic factors, however, especially co-medications and co-morbidities.

Thus, SNPs potentially can be applied in the development of individualised medicine and can provide an important source of information for studying the relationship between genotypes and phenotypes of human diseases.²¹ A bottleneck occurs when linking information about the variation in human genes to the variation in drug responses (pharmacogenetics) and understanding how interacting systems of genes determine individual drug responses (pharmacogenomics).²² A fundamental challenge in analysing disease cSNPs is the relative scarcity of alleles that can be mapped to three-dimensional protein structures. In the future, it is envisioned that knowledge of an individual's SNP genotype may provide a basis for assessing susceptibility to a disease and the optimal choice of therapies.⁶

Viral genome diversity exemplified by papillomaviruses and influenza viruses

Viruses are exceptionally diverse in morphology, genetic organisation, replication strategies, virulence and many other characteristics. Viral genome sequences represent a treasure trove of essential information for understanding pathogenesis better, as well as developing novel diagnostics and antiviral therapies. The ability of various medically important viruses to develop high degrees of genetic diversity, and to acquire mutations to escape immune pressures, contributes to the difficulties in vaccine development.

Diversity of human papillomaviruses

There are more than 100 different types of human papillomavirus (HPV), the causative agent of benign papillomas or warts, and a cofactor in the development of carcinomas of the genital tract, skin, head and neck. HPVs are broadly divided into cutaneous and mucosal HPV types. Eight major proteins, designated as early (E) or late (L) gene products, are encoded by the HPV DNA genome. Proteins E1 and E2 are involved in viral replication, as well as the regulation of early transcription. E1 binds to the origin of viral replication (ORI) and exhibits ATPase as well as helicase activity, whereas E2 forms a complex with E1, facilitating its binding to the ORI.²³ Furthermore, E2 acts as a transcription factor that regulates early gene expression by binding to specific E2 recognition sites.²⁴ E4 plays important roles in promoting the differentiation-dependent productive phase of the viral life cycle.²⁵ The E5 protein supports HPV late functions²⁶ and disrupts major histocompatibility complex (MHC) class II maturation.²⁷ The E6 and E7 oncoproteins are mainly responsible for HPV-mediated malignant cell progression, leading ultimately to invasive carcinoma.²⁸ Finally, L1 and L2 are the major and minor capsid proteins, respectively, and HPV vaccines based on L1 are already in clinical use.²⁹

In 1999, we determined the complete nucleotide sequence of a novel genital HPV type from a female sex worker with a wart virus infection in Singapore — namely, HLT7474-S — which was designated as candidate HPV type 85 (HPV-85) by the Reference Center for Papillomaviruses, German Cancer Research Center, Heidelberg, Germany. Its genomic organisation and phylogenetic relationships were analysed.^{30,31} The DNA sequence of the L1 open reading frame (ORF) of HPV-85 shares similarities of 78.3 per cent, 78.1 per cent and 78.0 per cent with those of the most closely related known types (HPV types 39, 70 and 45, respectively), thus satisfying the criteria for a new HPV type, which is defined on the basis of a dissimilarity exceeding 10 per cent in the L1 gene.³² In addition, the E6 and E7 ORFs of HPV-85 exhibit highest percentage similarities of

79.7 per cent and 77.9 per cent to E6 of HPV-18 and E7 of HPV-59, respectively, thus reiterating the relatedness between HPV-85 and known genital HPVs belonging to group A7.³³

Phylogenetic trees^{34,35} based on the individual ORFs, putative proteins and long control regions (LCRs) reveal the relationships between HPV-85 and the high-risk HPVs from group A7. The E1, E2, E5 and L2 proteins of HPV-85 are more closely associated with those of HPV-70 and HPV-39. Greater similarities are observed for the E4, E6 and L1 proteins of HPV-85, however, compared with those of HPV-18 and HPV-45, and for the HPV-85 E7 protein and LCR compared with its HPV-59 counterparts. These data exemplify the diversity of HPV viruses, with particular reference to HPV-85 as a co-evolved member of the A7 group of genital HPVs.

HPVs and protein disorder

It is now recognised that many functional proteins or their long segments are devoid of stable secondary and/or tertiary structure, and exist instead as very dynamic ensembles of conformations. They are known by different names, including natively unfolded, intrinsically disordered, intrinsically unstructured, rheomorphic, pliable and different combinations thereof. Disordered proteins have high flexibility and are reported to be involved in regulation, signalling and control pathways in which interactions with multiple partners, as well as high-specificity and low-affinity interactions, are often required.³⁶

To elucidate whether intrinsic disorder plays a role in the oncogenic potential of different HPV types, we performed a detailed bioinformatics analysis concentrating on the E6 and E7 oncoproteins of high-risk and low-risk HPVs. Three high-risk (HPV-16, HPV-18 and HPV-85) and two low-risk (HPV-6 and HPV-11) HPV types were analysed in order to compare the extent of intrinsic protein disorder in these virus types. The amino acid sequences of the different HPV types were extracted from the Los Alamos National Laboratory (<ftp://ftp-t10.lanl.gov/pub/papilloma/SWISS-PROT-files>). Predictions of intrinsic disorder in HPV proteins were performed using a set of disorder predictors — that is,

POODLE-S³⁷ and POODLE-L.³⁸ We employed POODLE-S to analyse the E6, E7 and L2 proteins, since their sequences are only ~100 residues long, whereas POODLE-L was used for other proteins. The results are presented in Figure 2.

We conducted Tukey's multiple comparison test to compare the residue disorder values for each of the two HPV groups. The one-way analysis of variance (ANOVA) indicates that the E6 oncoproteins of oncogenic HPVs (HPV-16, HPV-18 and HPV-85) are significantly more disordered ($p < 0.001$) than those of non-oncogenic HPVs (HPV-6 and HPV-11). Thus, the results of this analysis are consistent with the conclusion that high-risk HPVs are characterised by an increased degree of intrinsic disorder of the E6 protein. The molecular basis of this disorder, in terms of protein sequence variation of virulent HPV types, is supported by experimental evidence for the transforming ability of E6 proteins of oncogenic HPVs.²⁸ Furthermore, the disorder trend is more significant for E6 than for E7, consistent with a previous report using the commercial software PONDR. The data also highlight the diversity of high-risk and low-risk HPVs at the protein structure level.³⁹ These intrinsic differences in E6 protein dynamics may be exploited and targeted pharmacologically — for example, using monoclonal antibodies or chemical inhibitors.

Certain non-oncoproteins of non-oncogenic HPVs exhibit greater 'disorder' — for example, L1 and L2 of HPV-6 and HPV-11, and E2 of HPV-6. This observation suggests that it is the disorder of critical viral protein(s) that defines the oncogenic capability and risk level of HPV. Our data illustrate an alternative computational approach to distinguishing between the non-oncogenic from the high-risk oncogenic HPV types. It will be interesting to determine whether the phenomenon of increased protein disorder of more virulent virus types can be generalised beyond the HPV family.

Genetic diversity of H5N1 influenza viruses

Since 2003, highly pathogenic avian influenza A H5N1 viruses have spread from Asia to Europe and

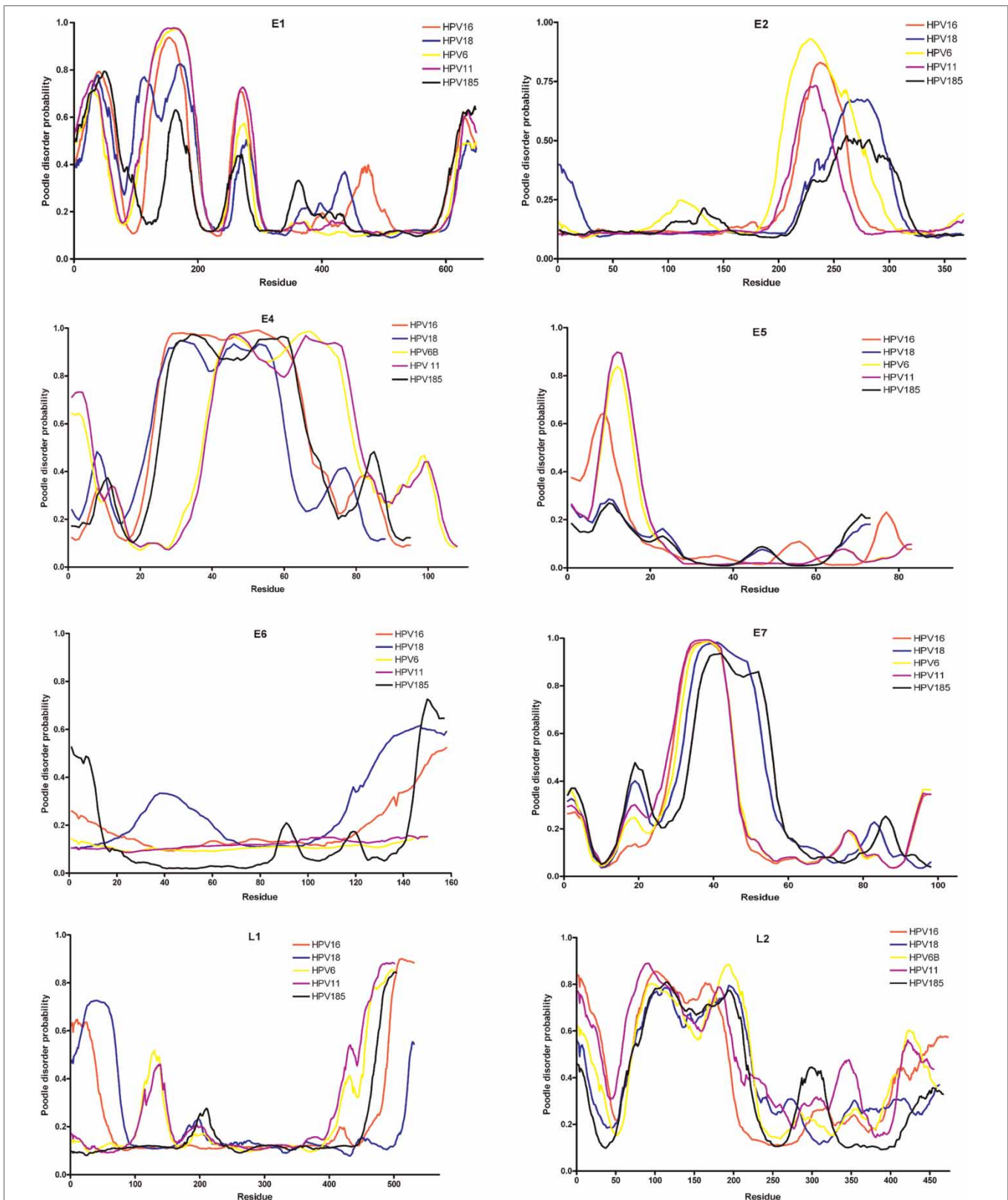


Figure 2. Protein disorder probability based on the results of Poodle prediction. The table illustrates Tukey's multiple test for comparing low-risk and high-risk HPVs. One-way ANOVA reveals the differences in disorder scores for the HPV proteins.

Africa, infecting wild birds, commercial poultry and humans with alarming fatality rates. Scrupulous surveillance and multidisciplinary interrogation of H5N1 evolution and ‘migration patterns’ are crucial for preventing further casualties in humans and poultry.⁴⁰ The escalating number of human cases of H5N1 virus infection has raised serious concerns about the potential emergence of an influenza pandemic attributed to a mutated H5N1 strain with efficient human-to-human transmissibility.⁴¹ The two major surface glycoproteins encoded by the segmented influenza virus RNA genome are haemagglutinin (HA) and neuraminidase (NA). HA is the major antigen for neutralising antibodies and is involved in the binding of virions to sialic acid-linked receptors on host cells. Infectivity of influenza virus depends on the cleavage of HA by specific host proteases, whereas NA mediates the release of progeny virions from the cell surface and prevents clumping of newly formed virus particles.^{42,43} HA is a ~550 amino acid polypeptide that forms homotrimers (spikes) on the exterior of the influenza virus particle. Nascent HA is directed to the cell membrane in an infected host cell and is anchored to the cell membrane by a short transmembrane region at the C-terminus. Its biological activation involves proteolytic cleavage of a specific region by host enzymes. The nascent HA is also subject to extensive post-translational glycosylation that serves as a mechanism for immune evasion. Introducing new mutations in these two proteins represents the major strategy used by H5N1 to expand its host range and to avoid recognition by the host immune system. Here, we illustrate this diversity of influenza viruses by comparing HA proteins from over 270 H5N1 strains. A total of 272 HA sequences were downloaded from the National Center for Biotechnology Information (NCBI) Influenza Resource database, and phylogenetic analyses were performed using the PHYLIP and Neighbourhood-Joining (NJ) method, with a bootstrap of 1,000. This phyloinformatics analysis revealed a distinct pattern of spatial clustering of the strains based on their geographical origin, rather than temporal clustering or according to host range (Figure 3). The dataset includes samples isolated up to 2006, and represents globally distributed locations

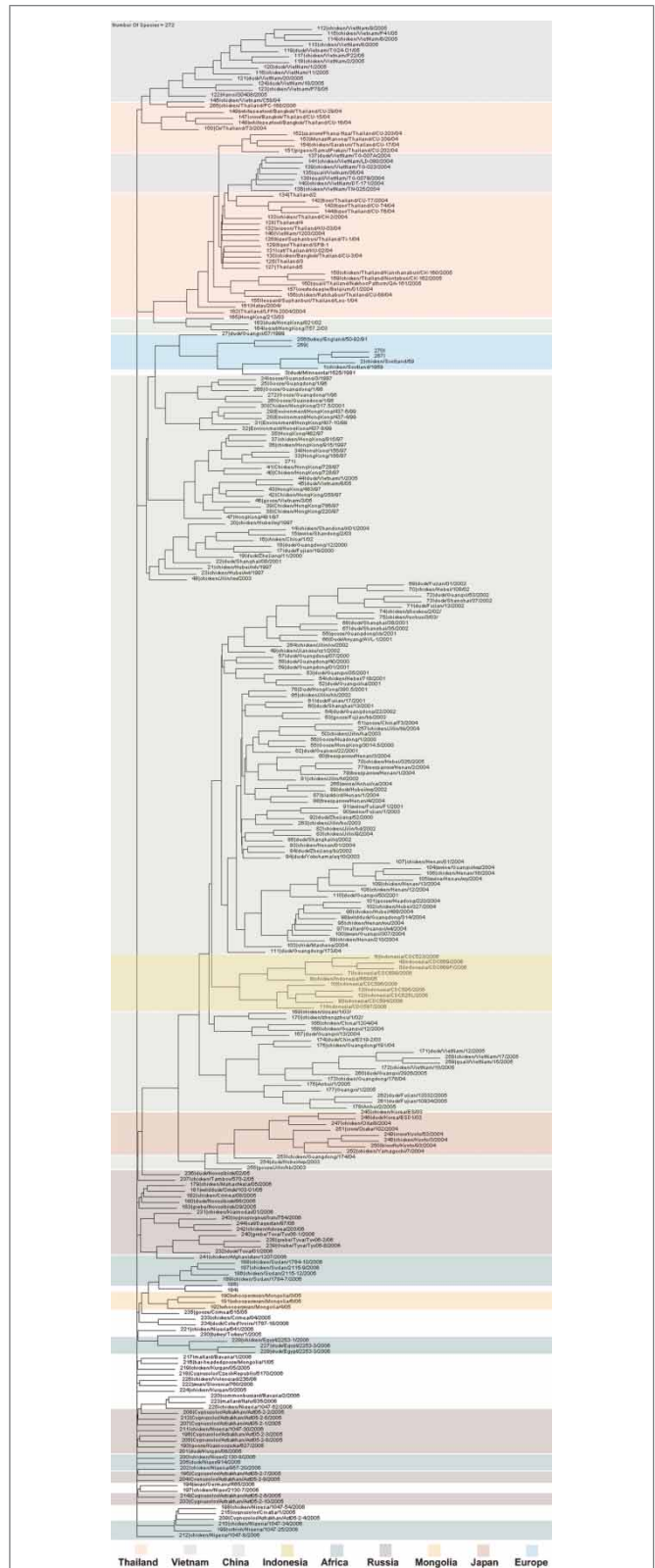


Figure 3. Phylogenetic tree constructed for a total of 272 HA sequences of H5N1 virus strains. The HA sequences were downloaded from the NCBI Influenza Resource database.

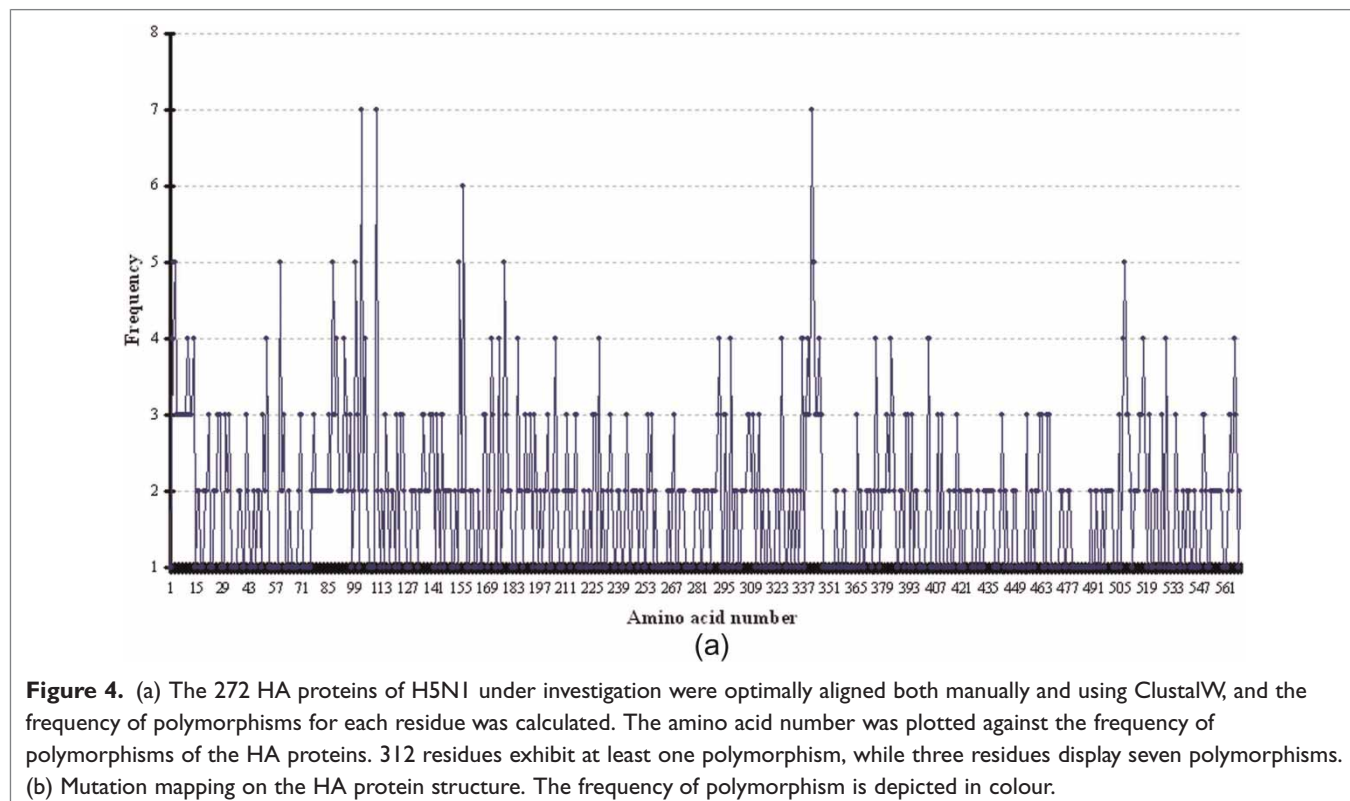
from Thailand, Vietnam, Indonesia, Japan, Mongolia, Russia, Europe to Africa. Interestingly, the host range across these H5N1 clades ranges from chickens to humans and includes several mammalian species. The clustering of strains does not show any bias towards the host species — for example, the Thailand clade includes isolates from chickens, cats and tigers.

Multiple sequence alignments using the ClustalW program reveal that HA is highly polymorphic. A total of 312 positions exhibit polymorphisms over two-thirds of the protein length (Figure 4a). Mapping of these positions on the protein databank (PDB) file using 2FK0 as the template is depicted in Figure 4b. Polymorphisms in the various residues of H5N1 HA occur individually and not in tandem — that is, two or more polymorphic residues may change independently in different HA variants; however, it is noteworthy that the polymorphisms are concentrated in the receptor-binding domain of HA. These data can facilitate monitoring of the receptor-binding specificity of modern influenza viruses, especially H5N1. The reasons why H5N1 HA mutants segregate among more than one host species, and why there are

differences in causing widespread disease, may be due in part to the differences in occurrence and distribution of cellular receptors for H5N1 (ie α 2,3 or α 2,6 sialic acid-linked receptors). The differences in the infectivity of various H5N1 strains in different hosts is ultimately determined by the compatibility and binding between the virus strain and the host cellular receptor. The polymorphism in the receptor-binding region also reflects the host immune response against the virus and the high mutating capacity of the virus that spawns escape mutants. These polymorphisms contribute to a greater diversity in viral virulence and to the expanded host range of H5N1. Such molecular-based surveillance can aid the designing of potential influenza vaccines, as well as a better understanding of the mechanisms of virus evolution and inter-species transfer.

Genetic determinants of host susceptibility to infection

Advances in genomics and understanding of pathogen variability, as well as diversity of the human



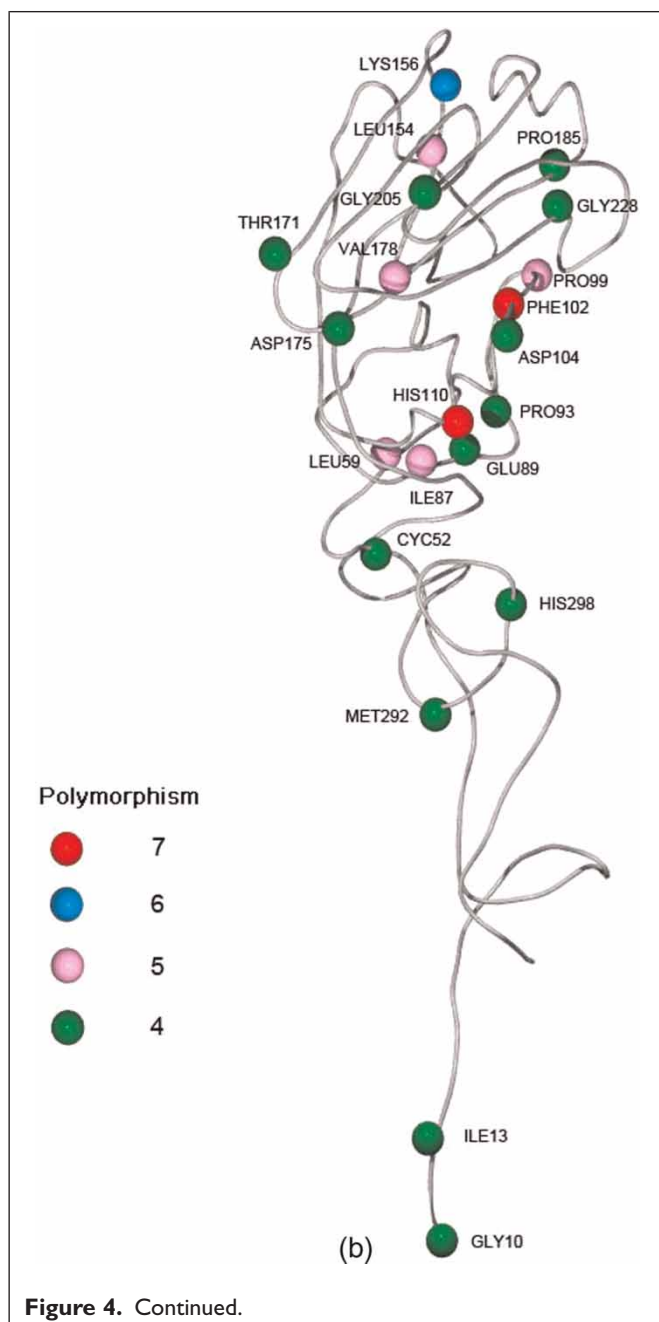
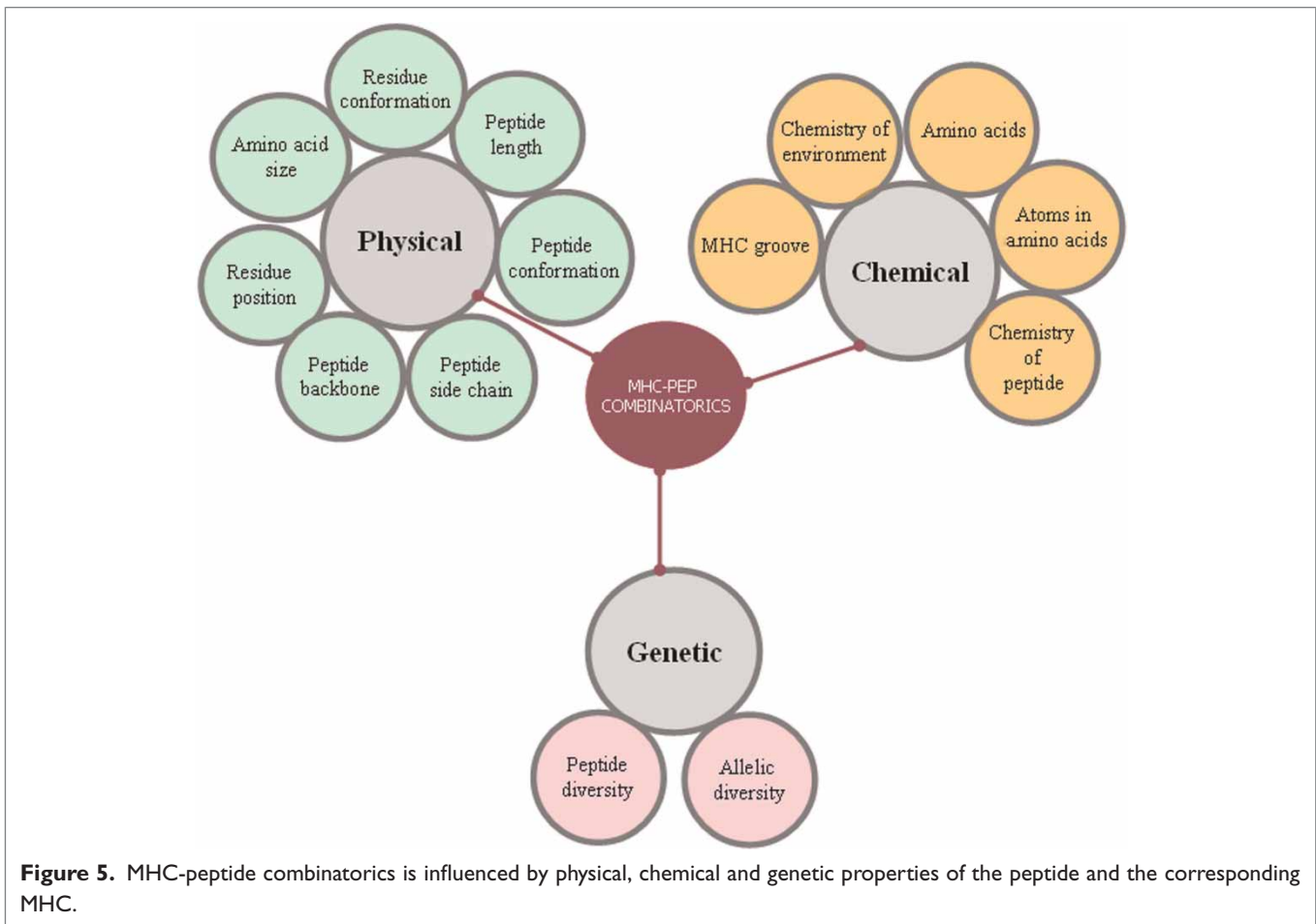


Figure 4. Continued.

immune systems have led to new trends in vaccine development that focus on epitope-based vaccines. An epitope is a small peptide fragment from an infectious agent that can induce a host immune response to eliminate the pathogen. Such a vaccine strategy shows promise in dealing with host and pathogen diversity. Compared with traditional vaccines, epitope-based vaccines are more specific, safe and more easily produced and controlled. The keys

to the success of such approaches are the prediction models for rapidly scanning pathogen genomes to identify effective T-cell epitopes. A recent review article focuses on different methods available for MHC-peptide binding prediction for epitope-based vaccine design.⁴⁴

The virulence of the pathogen and the susceptibility of the host determine the occurrence and severity of an infectious disease. The highly polymorphic glycoproteins of the MHC are the key proteins involved in the host immune response. MHC class I glycoproteins are expressed on the surface of every nucleated human cell and play important roles in viral infections. They present endogenous peptides derived from the cell itself to cytotoxic T cells. Since human viruses use their host's cellular machinery for replication, infected cells present viral proteins on their surfaces by using human leukocyte antigen (HLA) class I glycoproteins. This co-presentation of viral peptides elicits a cell-mediated immune response that destroys the virally infected cell. Conversely, HLA class II glycoproteins expressed on antigen-presenting cells display antigenic peptides derived from the pathogen. T cells recognise these antigenic peptides as foreign and initiate an immune response to the antigen. Each antigenic peptide must fit into the peptide-binding cleft; both peptide size and composition determine the fit. Each peptide is typically nine to 14 amino acids long, and its sequence is determined by the pathogen. At the host level, depending on the particular surface of the peptide-binding cleft, some antigenic peptides may be preferentially presented, while others may not be presented at all. The great diversity of clefts across the human population translates into the ability to recognise and generate an immune response to virtually any pathogen. The type of antigenic peptide displayed in the cleft is an important factor in the immune response generated. Thus, several physical, chemical and genetic factors determine whether a given peptide will fit into the peptide-binding cleft and elicit an immune response (Figure 5). Furthermore, class I and class II MHC molecules are the most polymorphic human proteins; some of these have over 200 allelic



variants. This extreme polymorphism is driven and maintained by the long-standing battle for supremacy between our immune system and infectious pathogens. The polymorphisms within both HLA class I and class II glycoproteins occur almost exclusively in the region of the glycoprotein that constitutes the peptide-binding cleft.

Conclusions and future prospects

The pharmaceutical industry is currently grappling with a tremendous number of potential drug targets against infectious diseases identified through sequence data for the human genome and many important pathogens. The challenge ahead is to delineate the factors involved in host–pathogen interactions, and to translate the enormous discovery potential of human and pathogen genomes into real products for therapeutic intervention. One of

the major bottlenecks in bringing new drugs to market is our incomplete understanding of the genes and proteins central to host–pathogen interactions and the mechanisms underlying certain human diseases. Other limitations that need to be addressed include patient heterogeneity and pathogen polymorphisms in clinical trials, the existence of multiple molecular targets and the shortage of experimental models of therapeutic efficacy with good predictive validity and objective surrogate measurements of disease progression. Integration of computational biology, together with experimental approaches, can accelerate our ability rapidly and reliably to identify target proteins that can be harnessed for therapeutic intervention.⁴⁵ The effective application of such complementary analytical methods will blaze trails for the exploitation of available genetic and molecular information. When supplemented with ‘individualised therapy’ based

on a patient's genetic profile, these developments may lead to fewer serious adverse effects and improved responses to drug treatments and vaccine regimens. Finally, the ultimate goal must be to bridge theoretical biological, genetic and molecular phenomena, and cellular and organism biology, as well as medicinal chemistry in the relentless search for new cures in the battle against pathogens that plague humankind.

References

- Goldstein, D.B. and Cavalleri, G.L. (2005), 'Genomics: Understanding human diversity', *Nature* Vol. 437, pp. 1241–1242.
- International HapMap Consortium (2005), 'A haplotype map of the human genome', *Nature* Vol. 437, pp. 1299–1320.
- Brookes, A.J. (1999), 'The essence of SNPs', *Gene* Vol. 234, pp. 177–186.
- Hoehe, M.R., Timmermann, B. and Lehrach, H. (2003), 'Human inter-individual DNA sequence variation in candidate genes, drug targets, the importance of haplotypes and pharmacogenomics', *Curr. Pharm. Biotechnol.* Vol. 4, pp. 351–378.
- Maitland, M.L., DiRienzo, A. and Ratain, M.J. (2006), 'Interpreting disparate responses to cancer therapy: The role of human population genetics', *J. Clin. Oncol.* Vol. 24, pp. 2151–2157.
- Masood, E. (1999), 'As consortium plans free SNP map of human genome', *Nature* Vol. 398, pp. 545–546.
- Vitkup, D., Sander, C. and Church, G.M. (2003), 'The amino-acid mutational spectrum of human genetic disease', *Genome Biol.* Vol. 4, p. R72.
- Wang, Z. and Moulton, J. (2001), 'SNPs, protein structure, and disease', *Hum. Mutat.* Vol. 17, pp. 263–270.
- Zanotti, G. and Berni, R. (2004), 'Plasma retinol-binding protein: structure and interactions with retinol, retinoids, and transthyretin', *Vitam Horm* Vol. 69, pp. 271–295.
- Seelinger, M.W., Biesalski, H.K., Wissinger, B., Gollnick, H. *et al.* (1999), 'Phenotype in retinol deficiency due to a hereditary defect in the retinol binding protein synthesis', *Invest. Ophthalmol. Vis. Sci.* Vol. 40, pp. 3–11.
- Yao-Borengasser, A., Varma, V., Bodles, A.M., Rasouli, N. *et al.* (2007), 'Retinol binding protein 4 expression in humans: Relationship to insulin resistance, inflammation, and response to pioglitazone', *J. Clin. Endocrinol. Metab.* Vol. 92, pp. 2590–2597.
- Hash, R.B. (2001), 'Hereditary hemochromatosis', *J. Am. Board Fam. Pract.* Vol. 14, pp. 266–273.
- Kim, S. and Misra, A. (2007), 'SNP genotyping: Technologies and biomedical applications', *Annu. Rev. Biomed. Eng.* Vol. 9, pp. 289–320.
- Delrieu, O. and Bowman, C. (2006), 'Visualizing gene determinants of disease in drug discovery', *Pharmacogenomics* Vol. 7, pp. 311–329.
- Giacomini, K.M., Brett, C.M., Altman, R.B., Benowitz, N.L. *et al.* (2007), 'The pharmacogenetics research network: From SNP discovery to clinical drug response', *Clin. Pharmacol. Ther.* Vol. 81, pp. 328–345.
- McLeod, H.L. and Evans, W.E. (2001), 'Pharmacogenomics: Unlocking the human genome for better drug therapy', *Annu. Rev. Pharmacol. Toxicol.* Vol. 41, pp. 101–121.
- Evans, W.E. and Johnson, J.A. (2001), 'Pharmacogenomics: The inherited basis for interindividual differences in drug response', *Annu. Rev. Genomics Hum. Genet.* Vol. 2, pp. 9–39.
- Shah, R.R. (2005), 'Mechanistic basis of adverse drug reactions: The perils of inappropriate dose schedules', *Expert Opin. Drug Saf.* Vol. 4, pp. 103–128.
- Rane, A. (2001), 'Postgenomic prospects of success in drug development and pharmacotherapy', *Pharmacogenomics J.* Vol. 1, 6–9.
- Bosch, T.M., Meijerman, I., Beijnen, J.H. and Schellens, J.H. (2006), 'Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer', *Clin. Pharmacokinet.* Vol. 45, pp. 253–285.
- Lai, E. (2001), 'Application of SNP technologies in medicine: Lessons learned and future challenges', *Genome Res.* Vol. 11, pp. 927–929.
- Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L. *et al.* (2001), 'Integrating genotype and phenotype information: An overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base', *Pharmacogenomics J.* Vol. 1, pp. 167–170.
- Ustav, M. and Stenlund, A. (1991), 'Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames', *EMBO J.* Vol. 10, pp. 449–457.
- Gloss, B., Bernard, H.U., Seedorf, K. and Klock, G. (1987), 'The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones', *EMBO J.* Vol. 6, pp. 3735–3743.
- Wilson, R., Fehrmann, F. and Laimins, L.A. (2005), 'Role of the E1-E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31', *J. Virol.* Vol. 79, pp. 6732–6740.
- Fehrmann, F., Klumpp, D.J. and Laimins, L.A. (2003), 'Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation', *J. Virol.* Vol. 77, pp. 2819–2831.
- Zhang, B., Li, P., Wang, E., Brahmi, Z. *et al.* (2003), 'The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon-gamma', *Virology* Vol. 310, pp. 100–108.
- Snijders, P.J., Steenbergen, R.D., Heideman, D.A. and Meijer, C.J. (2006), 'HPV-mediated cervical carcinogenesis: Concepts and clinical implications', *J. Pathol.* Vol. 208, pp. 152–164.
- Stanley, M. (2008), 'Immunobiology of HPV and HPV vaccines', *Gynecol. Oncol.* Vol. 109, pp. S15–S21.
- Chow, V.T. and Leong, P.W. (1999), 'Complete nucleotide sequence, genomic organization and phylogenetic analysis of a novel genital human papillomavirus type, HLT7474-S', *J. Gen. Virol.* Vol. 80, pp. 2923–2929.
- de Villiers, E.M., Fauquet, C., Broker, T.R., Bernard, H.U. *et al.* (2004), 'Classification of papillomaviruses', *Virology* Vol. 324, pp. 17–27.
- Delius, H., Saegling, B., Bergmann, K., Shamanin, V. *et al.* (1998), 'The genomes of three of four novel HPV types, defined by differences of their L1 genes, show high conservation of the E7 gene and the URR', *Virology* Vol. 240, pp. 359–365.
- Forslund, O. and Hansson, B.G. (1996), 'Human papillomavirus type 70 genome cloned from overlapping PCR products: Complete nucleotide sequence and genomic organization', *J. Clin. Microbiol.* Vol. 34, pp. 802–809.
- Ho, L., Chan, S.Y., Chow, V., Chong, T. *et al.* (1991), 'Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree', *J. Clin. Microbiol.* Vol. 29, pp. 1765–1772.
- Chan, S.Y., Ho, L., Ong, C.K., Chow, V. *et al.* (1992), 'Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind', *J. Virol.* Vol. 66, pp. 2057–2066.
- Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z. *et al.* (2007), 'Intrinsic disorder and functional proteomics', *Biophys. J.* Vol. 92, pp. 1439–1456.
- Shimizu, K., Hirose, S. and Noguchi, T. (2007), 'POODLE-S: Web application for predicting protein disorder by using physicochemical features and reduced amino acid set of a position-specific scoring matrix', *Bioinformatics* Vol. 23, pp. 2337–2338.
- Hirose, S., Shimizu, K., Kanai, S., Kuroda, Y. and Noguchi, T. (2007), 'POODLE-L: A two-level SVM prediction system for reliably predicting long disordered regions', *Bioinformatics* Vol. 23, pp. 2046–2053.
- Uversky, V.N., Roman, A., Oldfield, C.J. and Dunker, A.K. (2006), 'Protein intrinsic disorder and human papillomaviruses: Increased

- amount of disorder in E6 and E7 oncoproteins from high risk HPVs', *J. Proteome Res.* Vol. 5, pp. 1829–1842.
40. Ma, H.C., Chen, J.M., Chen, J.W., Sun, Y.X. *et al.* (2007), 'The panorama of the diversity of H5 subtype influenza viruses', *Virus Genes* Vol. 34, pp. 283–287.
41. Chow, V.T., Tambyah, P.A. and Goh, K.T. (2008), 'To kill a mocking bird flu?', *Ann. Acad. Med.* Vol. 37, pp. 451–453.
42. Wallace, R.G., Hodac, H., Lathrop, R.H. and Fitch, W.M. (2007), 'A statistical phylogeography of influenza A H5N1', *Proc. Natl. Acad. Sci. USA* Vol. 104, pp. 4473–4478.
43. Narasaraju, T., Sim, M.K., Ng, H.H., Phoon, M.C. *et al.* (2009), 'Adaptation of human influenza H3N2 virus in a mouse pneumonitis model: Insights into viral virulence, tissue tropism and host pathogenesis', *Microbes Infect.* Vol. 11, pp. 2–11.
44. Zhao, B., Sakharkar, K.R., Lim, C.S., Kanguane, P. *et al.* (2007), 'MHC-peptide binding prediction for epitope based vaccine design', *Int. J. Integr. Biol.* Vol. 1, pp. 127–140.
45. Sakharkar, M.K., Sakharkar, K.R. and Pervaiz, S. (2007), 'Druggability of human disease genes', *Int. J. Biochem. Cell Biol.* Vol. 39, pp. 1156–1164.