343P INTRODUCTION TO PHARMACOGENESIS: TEACHING THE PRINCIPLES OF INTERACTION OF MEDICINES WITH MOLECULES, ORGANS AND THE HUMAN BODY WITH THE HELP OF CD-ROM TECHNOLOGY

[1]A.Sj. Koster, [2]N.J. de Mol, [3]G. Storm, [1]D. van Heuven-Nolsen, [1]F.P. Nijkamp, [4]G. van den Brink and [3]J. Perquin, Departments of [1]Pharmacology, [2]Medicinal Chemistry, [3]Biopharmacy and [4]Pharmacotherapy, Faculty of Pharmacy, Utrecht University, P.O. Box 80.082, 3508 TB Utrecht, the Netherlands.

In our Faculty, the principles of interaction between drugs and biomolecules (enzymes, receptors, plasma proteins), organs (elimination, pharmacodynamics) and the animal body (pharmacokinetics, pharmacotherapy) are treated in an integrative course at the end of the first year. New educational insights, in combination with a large number of students (over 200 every year), made it necessary to redesign this course in such a way that interactive and self-directed, self-paced individual learning is stimulated. Large-scale lectures will be replaced by 5 small-scale meetings between teacher and students. The students have to prepare themselves for these meetings using written assingments, which can be worked out with the help of a specially designed multimedia programme: "Introduction to Pharmacogenesis".

The programme consists of 10 modules (physicochemical properties, protein binding, biopharmacy, absorption, distribution, elimination, receptor action, aspecific action, pharmacokinetics, pharmacotherapy), each consisting of screens (in total 130). Measurable qualities and experimental findings are described under FACTS, theoretical aspects and models under CONCEPTS. The concepts will help to integrate a large number of facts into a unified way of thinking. A self-test is included in each module. In every session these test questions are presented in other random sequences. In addition, an entrance- and final-test of the programme is supplied. All screens can be viewed in audio-mode (visuals explained by spoken word) or text-mode. By multisensorial stimulation (eyes and ears) the impact of the message is enhanced. All screens are directly linked to screens with related subjects. This makes it possible to wander through the programme and to discover unsuspected relations between subjects. Intuitive learning is facilitated in this way. A total of 10 simulations, related to drug transport, metabolism, excretion, concentration in plasma, dose response curves, and therapeutic plasma levels, are included. More than 450 keywords can be searched alphabetically or by direct links from the screens in text-mode. Pharmacochemical, pharmacological, pharmacokinetic and therapeutic data of ca. 20 medicines are included in a compendium.

The newly designed course has been tested with a relatively small number of 1st-year pharmacy students (50). Our first experience is that students can indeed be stimulated to individual self-directed learning. However, carefully written assignments and small-scale meetings between students and teachers are essential to guide all students succesfully through the course. Although the CD-ROM programme is specifically designed for the 1st-year curriculum, several aspects will make it a useful addition to teaching programmes in other disciplines.

"Introduction to Pharmacogenesis" was produced in collaboration between the Faculty of Pharmacy, Utrecht University and OMI Multimedia/NOB Interactive. The CD-ROM will run on a 486 DXII PC, 66 MHz processor, 4 MB internal memory, 2-speed CD-ROM player, SVGA video card (1 MB, 640x480, 256 colours). Windows 3.x or Windows95 and Quicktime for Windows (version 2.02 or higher) are required. A Pentium 133 MHz PC with 6-speed CD-ROM player is recommended.

344P GENOMICS AND PHARMACOLOGY

J Craig Venter, The Institute for Genomic Research, Rockville, Maryland, USA

The Institute for Genomic Research (TIGR) is a nonprofit research institute with interests in structural, functional, and comparative analysis of genomes and gene products. Researchers at TIGR developed Expressed Sequence Tag (EST) sequencing and were first to apply high-throughput, automated techniques to the collection of EST data. TIGR has contributed more than 170,000 sequences, from 300 cDNA libraries constructed from 37 distinct tissues, to the public EST collection. To identify the unique transcripts within the nearly 700,000 ESTs generated world-wide, we assembled them as elements of shotgun sequence assembly project, producing nearly 65,000 distinct tentative human consensus (THC) sequences, a first approximation of human genomic coding potential.

TIGR was first to sequence and publish a complete microbial genome. The *Haemophilus influenzae* genome (1.83 Mb) contains 1741 predicted coding regions; nearly 60% can be assigned a putative biological role based on matches to proteins from other species. We have since sequenced five additional genomes. *Mycoplasma genitalium* with 470 genes, has the smallest genome of any free-living organism. *Methanococcus janaschii*, a deep-sea, hyperbaric, thermophilic methanogen is the first sequenced *Archaea*. *Helicobacter pylori* is the causative agent of gastric ulcers and a contributes to development of stomach cancer. *Archaeoglobus fulgidus* is the first sulfate-reducing organism and second *Archaea* sequenced. *Borrelia burgdorferi*, which causes Lyme disease, is the first spirochete sequenced and may provide insight into the eukaryotic evolution.

TIGR is also involved in eukaryotic sequencing. *Plasmodium falciparum* is the most deadly of the four *Plasmodium* species known to cause malaria. TIGR is sequencing chromosomes 2, 10, 11, and 14 as part of an effort to complete the 30 Mb genome within five years. Our approach to sequencing *Arabidopsis thaliana*, a plant model organism, relies on shotgun sequencing of bacterial artificial chromosome (BAC) clones and identification of new targets using an innovative BAC end sequencing strategy.

TIGR human sequencing focus on chromosome 16p. One of six NIH funded large-scale pilot Human Sequencing Centers, we will finish nearly 30 Mb of genomic sequence before March 1999. We plan to be a major contributor to the sequencing of the human genome and have worked to develop laboratory protocols and software to facilitate sample management, quality control, closure, finishing, and annotation of BAC projects.

Sequence data is the first step in developing an understanding of fundamental processes underlying life. Determination of the function each gene is the target for the next generation of whole-genome analysis, "functional genomics." A number of techniques have been developed, including cDNA microarraying and the Affymetrix GeneChip™ technology, that may allow whole-genome expression monitoring. The challenge for the future is to develop these techniques and integrate their results with existing data to develop a picture of cellular metabolism and gene function. TIGR aims to be one of the leaders in developing these functional genomic approaches to address fundamental biological questions.

345P THE DISCOVERY OF LEPTIN AND ITS POTENTIAL IMPACT ON THE TREATMENT OF DIABETES AND OBESITY

J M Friedman, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021, USA

The obese (ob) and diabetes (db) genes are recessive mouse mutations that result in massive obesity as part of a syndrome that resembles morbid human obesity. The cloning of these mutant genes has identified two proteins, leptin and its receptor, Ob-R, both of which are important in the control of body weight.

Leptin is an adipocyte hormone encoded by the ob gene that plays an important role in regulating the size of the adipose tissue mass. The plasma concentration of leptin is increased in obese subjects and decreases after weight loss. Injections of recombinant leptin at physiologic levels reduce body weight and food intake of normal and ob mice in a dose-dependent manner but have no effect on db mice. These data identify leptin as an afferent signal in a negative feedback loop that maintains constant stores of body fat. The complete insensitivity of db mice to leptin and the identical phenotype of ob and db mice suggested that the db locus encodes the leptin receptor.

The db gene was found to be identical to a leptin receptor (Ob-R) which was functionally cloned by Tartaglia and colleagues from choroid plexus. However, because this receptor was normal in C57BL/KS db/db mice, the possibility was raised that this db mutation affected an alternatively spliced form. The Ob-R gene was found to encode at least five different splice variants. One of the splice variants, Ob-Rb, is expressed at a high level in the hypothalamus and at a lower level in other tissues. In hypothalamus, Ob-Rb is expressed in the arcuate, VMH and LH nuclei, each of which as been known to play a role in the regulation of body weight. The Ob-Rb transcript is mutant in C57BUKs db/db mice. The mutant protein is missing the cytoplasmic region and is defective in signal transduction. Further studies have revealed that the Stat3 transcription factor is activated specifically in hypothalamus within 15 minutes of a single injection of leptin in ob and wild type but not in db mice. Thus leptin appears to have rapid, direct effects on hypothalamus after peripheral administration.

Consistent with a CNS site of action, low dose infusions of leptin (5 ng/h) intracerebroventicularly reproduce the effects of much larger doses given peripherally. Studies of the metabolic response to leptin and the mechanisms by which CNS signals affect glucose and fat metabolism indicate that ICV leptin elicits a metabolic response that is qualitatively different from that which results from food restriction. The nature of the efferent signals from hypothalamus that coordinate these effects are under study.

Leptin infusion into obese animals has indicated that several obese rodents including Diet Induced Obese (DIO), AY and New Zealand Obese (NZO) mice are leptin resistant. DIO mice are partially resistant to peripheral leptin while AY mice do not respond to subcutaneous leptin and respond poorly to ICV leptin (at 100 fold higher doses). NZO mice do not respond to peripheral leptin but respond normally to ICV leptin. The decreased potency of subcutaneous and ICV leptin in DIO, NZO and AY mice indicate that obesity in these strains tested is likely to be the result of insensitivity to leptin. In NZO mice, leptin resistance may be the result of decreased transport of leptin into the CSF. In AY mice, leptin resistance probably results from defects in the signaling pathway downstream of the leptin receptor in the hypothalamus.

The basis for the partial leptin resistance of DIO mice is unknown. Further studies of the basis of leptin resistance in these and other obese strains may have important implications for understanding the pathogenesis of obesity and other mutational disorders

---

346P NEURODEGENERATIVE DISEASES AND TRIPLET REPEAT EXPNATIONS: MECHANISMS AND APPROACHES TOWARDS PHARMACOLOGICAL TREATMENT

J-L Mandel, Institut de Génétique et Biologie Moléculaire et Cellulaire, CNRS/INSERM/Université Louis Pasteur de Strasbourg

Unstable expansions of trinucleotide repeats have been found associated, since 1991, to 12 neurological diseases. These diseases are often characterised by some form of anticipation (increased severity or increased risk to develop the disease, in succeeding generations), linked to the tendency of pathological alleles to further expand. Triplet diseases can be divided into 4 classes: (a) Massive expansions of untranslated CGG/CCG repeats associated to abnormal DNA methylation cause the fragile X mental retardation syndrome, and mild mental retardation linked to FRAXE; (b) Massive expansion of an untranslated CTG repeat causes myotonic dystrophy; (c) Moderate expansions of CAG repeats coding for polyglutamines are found in 8 purely neurodegenerative diseases; (d) Massive expansion of an intronic GAA repeat is found in Friedreich's ataxia (FA). The genes and mutations have been identified by positional cloning approaches, and this has provided important diagnostic tools. The goal is now to develop therapeutic approaches for these diseases. This requires an understanding of pathological mechanisms and the creation of mouse models. Genomic approaches may also be used to understand some of the factors that underlie the variability of phenotypic expression in these diseases.

The abnormal methylation in Fragile X causes the loss of expression of the FMR1 gene, and it is thus most important to understand the function of the RNA binding protein encoded by this gene. In myotonic dystrophy the situation is less clear, but is certainly not a simple loss of function of the target gene coding for a protein kinase.

Huntington's disease (HD) is caused by an expansion of a CAG/polygln repeat in huntingtin, a protein of unknown function. A similar mechanism was demonstrated in spinobulbar muscular atrophy (SSMA), spinocerebellar ataxia (SCAs) types 1, 2, 3, 6 and 7, and in dentatorubro-pallidoluysian atrophy (DRPLA) These diseases are characterised by adult-onset neuronal death in selected but different regions of the nervous system and by an inverse correlation between length of the polyglutamine expansion and age of onset. The age of onset for patients carrying the same expansion can vary by 20 years, and it is important to determine whether other genes may modulate the clinical severity. The proteins implicated in these diseases show no common features apart from the polyglutamine tract, and their distribution shows no obvious correlation with the site of neurodegenescence. The mechanism by which expanded polyglutamine causes neuronal death is mysterious, but is not a loss of function. Recent observations suggest that aggregation of polyglutamine containing proteins within the nucleus may constitute a common pathological mechanism. We have characterized a monoclonal antibody that recognizes the expanded polyglutamine-containing proteins in HD and in SCAs 1, 2, 3 and 7; normal proteins are not detected under the same conditions. The efficiency of detection is dependent on polyglutamine length, and thus parallels disease severity. Using this antibody, we have recently cloned the SCA2 locus, and showed that it is most sensitive to increased number of glutamines. This mAb was also tested on cell lines from patients with early onset forms of schizophrenia, manic depressive illness or rare familial forms of Parkinson's disease, but no new pathological protein has yet been detected. These negative results, but also the contrasting suggestion that some uncloned CAG repeats may be associated with psychiatric diseases, have to be taken with caution. Cell lines and transgenic mice that carry and express a mutated protein have been constructed, and some show pathological effects, allowing an investigation of the disease mechanisms that involve a gain of toxic property.

FA is a recessive neurodegenerative disease affecting dorsal root ganglia and the spinocerebellar tracts. It is caused, in most patients, by an intronic GAA repeat expansion in a gene coding for a small mitochondnal protein. Inactivation of the homologous gene in yeast and the finding of specific biochemical defects in patients implicate accumulation of iron in mitochondria. A very similar phenotype is caused by vitamin E deficiency resulting from mutations affecting the α-tocopherol transfer protein.

347P COMPARATIVE GENOMICS AND THE IDENTIFICATION OF HUMAN DISEASE GENES

Stephen G Oliver, Department of Biomolecular Sciences, UMIST, PO Box 88, Manchester M60 1QD

Classically, genetics starts with the phenotype of a mutant and uses this to define and, latterly, to isolate and sequence the gene whose mutation is responsible for that phenotype. Systematic genome sequencing reveals genes which have no associated phenotype and demands that we develop systematic approaches with which to elucidate the function of these novel genes. This problem is one of the major challenges for modern biology since systematic genome sequencing is revealing new genes at a rate which far exceeds that of the classical approach. The complete sequence of the *Saccharomyces cerevisiae* genome has revealed ca 6 times as many genes as were discovered by the former, classical, methods. However, what is most surprising is that the functions of at least one-third of the genes are completely unknown. The facility with which we can manipulate the yeast genome, as well as its convenience as an experimental organism, makes it an excellent choice with which to pursue a systematic approach to functional analysis.

The availability of a comprehensive data set containing both sequence and functional data will make the *S. cerevisiae* an important tool for navigating our way around the human genome. Of the genes so far implicated in human heritable diseases some 30-40% have yeast homologues. The study of the physiological role of these genes and their interaction with other yeast genes, or gene families, will promote the understanding of the corresponding human disease syndromes and may suggest routes whereby the latter may be treated by chemical or genetic approaches.

Many important medical conditions, either actual diseases (such as early onset diabetes) or the predisposition to particular conditions (such as colon cancer or heart disease) are controlled by several genes. The discovery of the complete set of genes involved in a particular condition is a difficult and lengthy process, even when modern molecular diagnostic techniques may be applied. The recognition of sequence or functional homology between the protein product of a single human gene implicated in the control of a particular disease and its yeast counterpart may provide an important key to significant improvements in diagnosis or therapy. Both *in silico* searches of the yeast genome sequence and *in vivo* studies which can quickly uncover interactions between genes or their products (eg the identification of extragenic suppressors or synthetic phenotypes, control-of-expression studies, and two-hybrid analyses) may reveal the relevant set of genes whose structural or functional homologues may determine the human disease syndrome.

The recognition of such gene sets should illuminate the molecular and biochemical basis of the human disease as well as providing important reagents for the rapid identification of the corresponding set of human genes.

348P GLOBIN GENE EXPRESSION AND DRUG DISCOVERY

Frank Grosveld, Dept. of Cell Biology and Genetics, Erasmus University, Rotterdam.

The effects of thalassaemia and sickle cell anaemia are greatly ameliorated if the foetal gamma globin genes which are normally silenced around the time of birth, remain active in the adult. Hence it is important to study the developmental regulation of the globin genes to obtain a handle on the proteins that are involved in gamma globin gene expression.

We have followed two approaches towards this goal:

1. Reverse genetics in transgenic mice to understand the switching process at the molecular level. This should eventually lead to that rational design of small molecules that would interfere with gamma globin gene expression.

2. A random screening approach based on a transcriptional assay to find lead compounds that influence gamma globin gene expression.

349P DISCOVERING NEW TARGETS FOR ANTI-BACTERIAL AGENTS BY IDENTIFICATION OF GENES ESSENTIAL FOR PATHOGENICITY

David W Holden, Department of Infectious Diseases, Imperial College of Science, Technology & Medicine, Hammersmith Hospital, Ducane Road, London W12 0NN

Analysis of complete genome sequences (genomics) has become a very powerful tool for investigating the basis of bacterial pathogenesis and for suggesting new molecular targets for drug design. Although genomics can assign functions to a significant proportion of bacterial genes directly, other approaches are required to analyse the functions of many genes whose products have weak or no similarity to proteins of known or suspected function.

Conventional genetic analysis has provided important information about the functions of bacterial genes required for pathogenesis. For example, some of the stages of bacterial pathogenesis can be simulated by using tissue-cultured host cells, and genetic screens using different host cell types have been developed for the identification of virulence genes whose products interact with host cell components. However, such *in vitro* screens cannot be expected to identify all the virulence genes of a pathogen because they do not reflect the complex environment a pathogen encounters as it grows in its host.

To overcome this problem we developed a negative selection technique called signature-tagged mutagenesis (STM). STM allows a large number of insertional mutant strains to be tested simultaneously in a living host which selects against strains attenuated in virulence. The approach was initially validated using the Gram-negative pathogen *Salmonella* and a murine model of typhoid fever by the isolation of many genes known to be important for virulence, and by the discovery of a new pathogenicity island which encodes a type III secretion system essential for *Salmonella* virulence.

The broader applicability of STM was assessed using the Gram-positive pathogen *Staphylococcus aureus* in a murine model of bacteraemia. Most of the attenuated mutants appear to be affected in cell surface metabolism or unknown functions. Several of the affected gene products have significant potential as targets for new antibacterial drugs. The work establishes a basis for genome-wide screens for virulence genes involving a variety of different infection models, such as endocarditis, soft tissue abscesses and pneumonia, and for further phenotypic characterisation of the mutant strains, to determine specific functions of *S. aureus* virulence genes.

350P ROLES OF STRUCTURAL BIOLOGY IN DRUG DEVELOPMENT AND DESIGN

Shoshana J Wodak, Unite de Conformation de Macromolecules Biologiques, P2 CP160-16, Universite Libre de Bruxelles, 50 Ave Roosevelt, 1050 Bruxelles, Belgium, and EMBL-EBI, Wellcome Genome Campus, Hinxton Hall, Hinxton, Cambridge, CB10 1SD, UK.

Highly specific inter-molecular interactions are central to many biological processes such as enzymatic reactions, signal transduction, physiological regulation and gene expression. Such interactions take place between macromolecules (eg protein-protein and protein-nucleic acid interactions) and between macromolecules and small molecules. Understanding these interactions at the atomic level is important in order to be able to design efficient drugs whose role is to specifically block or foster well defined molecular recognition events, with as few side effects as possible.

Our ability to determine atomic resolution structures of biological macromolecules has markedly improved over the last 15 years, due to progress in molecular biology and technical advances in X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and other structure investigation techniques. These have brought about an exponential growth in the number of known structures, which presently reaches 10,000, providing us with an unprecedented wealth of information on the structural properties of macromolecules and their complexes.

We wish to illustrate ways in which this information may be, and has been, useful in the design and development of drugs. There is the lucky but rare case where the structures of the therapeutically relevant proteins or protein-ligand complexes are known. Detailed analysis of the atomic interactions in these structures can then be used as a guide for designing more potent ligands (Blundell, 1996). This involves the use of molecular modelling tools, including increasingly sophisticated computational procedures for protein-ligand docking (Lybrand, 1995). Still, in the vast majority of the cases, the structure of the therapeutically interesting proteins, primarily receptors, is not known. Here, the fast growing body of structural and sequence information makes it increasingly likely that a protein with a structure resembling that of the protein of interest be among the known lot (Brenner & Chothia, 1997). An approach to the problem then consists of identifying such resembling structure (Lemer, Rooman *et al.*, 1995), and using it as a template to generate an approximate model for the desired protein. This is commonly performed with the help of the so-called homology modelling tools (Sanchez & Sali, 1997).

A promising new approach is the peptide mimetic design (Damewood, 1996). This approach can be considered when the aim is, for example, to block intermolecular interactions between a peptide ligand and a protein or between two proteins. It consists of identifying from the available structural and biochemical information the smallest peptide segment, or minimal protein substructure, capable of binding with similar affinity and specificity as the original system (Cunningham & Wells, 1997). Such 'mini proteins' that retain function can then be used as a template from which small molecule drugs, which mimic the peptide functionalities, can be derived. Though the concept itself has been around for some time, the generation of stable protein substructures with desired functional properties, has become practical enough only recently, owing to the combination of structural information with protein engineering strategies (Nygren & Uhlen, 1997)

Blundell T (1996) *Nature*, 384, 23-26
Brenner S, Chothia C, Hubbard TJP (1997) *Curr Opin Struct Biol*, 5, 369-376
Cunningham BC & Wells JA (1997) *Curr Opi. Struct Biol* 7, 457-462
Damewood JR (1996) in *Reviews in Computational Chemistry* (Lipkowiz KB & Boyd DB Eds) 9, 180
Lemer C, Rooman M & Wodak SJ (1995) *Proteins Struct Func. Genet* 23, 337-355
Lybrand T *Cur. Opin Struct Biol* (1995) 5, 224-228
Nygren PA & Uhlen M (1997) *Curr Opin Struct Biol* 7, 453-467
Sanchez, R & Sali A (1997) *Curr Opin Struct Biol* 7, 206-214