UvA-DARE (Digital Academic Repository)

NADPH oxidases and mutation analysis

Meischl, C.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

Discussion
NADPH oxidases and mutation analysis
Changing perspectives of gene analysis: some considerations with specific reference to chronic granulomatous disease

The practice, but also the relevance of gene (or mutational) analysis is in a period of transition. Thanks mainly to the technological surge created by the "Human Genome Project" the practical aspects of gene analysis such as sequencing power have seen dramatic progress but, possibly more important, the knowledge relevant for gene analysis has evolved considerably also. As I will review in the following, some aspects of gene analysis that in the past were of particular scientific interest have lost in relevance, while other aspects are now entering the limelight. As a result of this change in perspective, on the other hand, certain aspects of gene analysis have to be given renewed attention. Ultimately, as I will discuss in the conclusion, gene analysis has to live up to its promise by being the basis of an individualized treatment of disease.

Gene analysis as tool of unraveling mechanisms of (patho)physiology
Historically, the primary interest of gene analysis has been the identification of genes responsible for certain pathological conditions. Ideally, once the primary defect, the mutation at the DNA level, had been identified, one could then start to unravel the whole pathophysiological cascade, on a molecular, cellular and (patho)physiological level in a logical chain of cause and result. Today, with with the current state of knock-out and knock-in technology in mice (and other species), the (patho)physiological aspect of gene analysis has lost in importance. However, mice are not men, and results obtained in the mouse are not automatically transferable to the human organism. In that respect, detection of mutations in the corresponding human gene and demonstration of pathophysiological relevance remain the ultimate proof of relevance. The (patho)physiological aspect of gene analysis can be expected to return to the center stage once genetic linkage- (SNP-) based studies will start to identify the first candidate genes implicated in the pathogenesis of multifactorial disease (see below).

Gene analysis and disease phenotype
Normal physiology results from the continuous and adaptive homeostasis of a myriad of interrelated parameters by a multitude of interconnected and complementary signaling networks. Phenotype, therefore, very rarely is the result of one determinant alone, because the expression even of dominant gene products is subject to modulation by multiple regulatory influences. The main players of these regulatory mechanisms, the knots in the net, are proteins whose differential expression and finely regulated activity allow for adaptation to a continuously changing environment. Since the structure - and thus function - and expression of all proteins are defined by the sequence information contained in the genome, every participant of the network(s) is prone to genetically determined variability affecting both expression and activity. While there are several modes of genetic variability, the most important form - both in numbers and overall effect - are the single nucleotide polymorphisms (SNPs), stable, inherited, biallelic, single-base-pair differences that are present in the human genome at a frequency of 1 to 10 per 1,000 nucleotides. SNPs, by definition, occur at a frequency of less than 1% in a population. As a result of the efforts of the "SNP Consortium" - a collaboration of 14 major pharmaceutical companies and the
Wellcome Trust, as well as members of the Human Genome Project – there are almost 2 million SNPs in public databases, and perhaps twice that number of SNPs in commercial databases, such as that of Celera Genomics. While most SNPs are in non-coding sequences or silent, i.e. do not result in an amino-acid change in the affected gene product, many SNPs have functional consequences if they occur in the coding or regulatory regions of genes. The functional consequences of single nucleotide changes – be it polymorphism or mutation - can range from very subtle to complete eradication of function, depending on the protein and the intramolecular or regulatory domain affected.

Chronic granulomatous disease (CGD) is a severe congenital immunodeficiency syndrome characterized by recurrent and often life-threatening bacterial and fungal infections. At the molecular level CGD is due to a defect in the superoxide-generating NADPH oxidase and, in consequence, to the impaired microbicidal activity of phagocytic leukocytes. In the coding and regulatory regions of the genes coding for the four subunits of the NADPH oxidase several SNPs have been described: four in CYBB (coding for gp91phox), eight in CYBA (coding for p22phox), one in NCF1 (coding for p47phox) and twelve in NCF2 (coding for p67phox). No functional significance of these polymorphisms has been established.

Point mutations differ from SNPs in being relatively recent events which then may be passed on through the generations, but are comparable to SNPs as to possible functional consequences. Still, since recent mutations – in contrast to most SNPs – have not yet been subject to selection by evolutionary pressure, they carry the potential for more deleterious effects on protein function. There are proteins, such as cytochrome b5 reductase or glucose 6-phosphate dehydrogenase, where one can establish a correlation between the protein domain affected by the mutation and the functional phenotype. For the NADPH oxidase, and especially for the gp91phox subunit, such a correlation has not been established. The great majority of mutations in CYBB lead to a complete lack of protein expression, and in the few cases of X-linked CGD with measurable protein expression and function, a close correlation between these two parameters was found. Only three cases of X-linked CGD are known in which the superoxide production is disproportionately more reduced than the expression of gp91phox. Two of these mutations affect the FAD binding site and the third one the NADPH binding site.

Because of the strong selective pressure in favor of optimal performance of the regulatory systems and because of the redundancy of most of these systems, the polymorphic variability of proteins in most cases will have only subtle effects on the general clinical phenotype. However, there is a broad agreement today that the sum total of these subtle differences in protein function defines the biochemical basis of our individuality. Susceptibility to cancer and autoimmune disease, to adverse drug reactions or certain infections: these are just a very few examples of clinical phenotypes defined by the interplay of a multitude of networks.

The importance of the regulatory background of a certain component also becomes evident when the function of this component becomes compromised. One striking example is the clinical heterogeneity of CGD in patients with a comparable lack of NADPH oxidase function. Clearly, under those circumstances the effectiveness of complementary immune effector systems is especially important in determining the clinical phenotype. One first step in mapping this cell-biological background in CGD has been made by analyzing the association of certain SNPs in seven candidate genes (myeloperoxidase (MPO), mannose binding lectin (MBL), Fcγ receptors IIa, IIIa, IIIb, TNF-α and IL-1 receptor antagonist) with a number of clinical complications.
observed with high frequency in CGD patients. Among other findings, genotypes of MPO and FcγRIIIb were strongly associated with an increased risk for gastrointestinal complications, and patients with variant forms of both MBL and FcγRIIa had the highest risk of developing an autoimmune/rheumatologic disorder. Clearly, under an utopian perspective, this kind of analysis represents the ultimate future of gene analysis. Huge efforts are under way to establish a comprehensive map of SNPs in a large group of individuals. But while the present studies have as principal goal the elucidation by linkage analysis of genes involved in the pathophysiology of multifactorial diseases such as asthma or diabetes mellitus, future gene analysis will try to establish a map of SNPs in relevant genes to predict an individual's propensity to develop certain diseases or complications secondary to disease or treatment. In that way (preventive) treatment of, for example, different forms of senility can be initiated early on, or the risk of an antibiotic treatment or anesthesia can be assessed before beginning the treatment. However, these are perspectives beyond the possibilities of the current knowledge or technological capacities. They will require the interplay of high-throughput sequencing and highly evolved systems biology.

Gene analysis as tool of genetic investigation

Gene analysis and the mechanisms of mutations

DNA, as is well known, functions as medium for the storage and transfer of genetic information. The principal challenge for such sort of system is, of course, to transfer the vast amounts of information error-free during mitosis and meiosis and to assure the integrity of the information in the periods between two cycles of replication. While the enzymatic systems responsible for DNA replication and maintenance meet this challenge astonishingly well, they are not infallible. It has been estimated that in humans in about one third of the 10^7 cells that divide each second, spontaneous mutations arise. Although such mutations are indispensable for adaptation to environmental changes and - in the long run - for the evolutionary process, mutations that disrupt the proper functioning of a cell need to be repaired immediately. There are several different types of mutations which, in a simplified scheme, can be assigned to three different categories. The first category consists of mutations that are the result of a certain basal error rate of the replication machinery, such as small deletions or insertions. The second category comprises mutations that are introduced by chemical or physical mutagens, such as environmental agents or different types of radiation. This category includes many point mutations or chromosomal rearrangements, that elude or are beyond the scope of the DNA-repair mechanisms. The third category encompasses changes in DNA sequence due to the action of transposable elements.

Several different mechanisms for the generation of gene deletions or insertions have been proposed involving the misalignment of short direct repeats during replication. According to these models the replication machinery would slip on the template strand of DNA during second-strand synthesis from one repeat motif to a vicinal one and resume replication there. Depending on the direction of slippage, the result would be the deletion or insertion of one repeat motif plus intervening sequence. Results of the analysis of one of the two main databases of mutations associated with human phenotypes, the Human Gene Mutation Database (HGMD), are in agreement with this model. The majority (70%) of mutations recorded in the HGMD are single-base-pair substitutions. A closer analysis reveals that 62.5% of these mutations are transitions
and 37.5% transversions. There is, therefore, a highly significant excess of transitions as compared to random expectation (33%). Several mechanisms contribute to this preponderance of transitions, the most prominent of which is the hypermutability of the CG dinucleotide to TG and CA. In the majority of these CG dinucleotides the C appears to be methylated in eukaryotic genomes to form 5-methylcytosine (5mC), which has a propensity to undergo deamination to form thymine. While uracil, formed by the deamination of (nonmethylated) cytosine, is recognized and excised by DNA glycosylase, no comparably effective DNA repair mechanism exists for the normal DNA constituent thymine$^{16}$. The observed frequency of CG to TG and CG to CA mutations varies between human genes. For example, it is less than 10% in the β-globin and HPRT genes, but it is greater than 50% in the ADA gene$^{16}$. In X-linked CGD a frequency of 34% has been calculated$^{11}$. Transposable elements are heterogeneous group of mobile DNA sequences that are dispersed throughout the genome and represent in the human 45% of the genomic sequence. They can be subdivided into DNA transposons that move as pieces of DNA, cutting and pasting themselves into new genomic locations, and the predominating retrotransposons that transpose via an RNA intermediate, with the original transposon remaining immobile$^{13}$. LINE-1 (L1) elements are the master retrotransposons in mammalian genomes because, besides duplicating themselves, they likely have been responsible for the genomic expansion of of non-autonomous retrotransposons such as Alu elements and processed pseudogenes$^{18}$. The insertion step of the retrotransposition of L1 elements is initiated by an L1-encoded endonuclease specific for a loose consensus site that is found at a large number of genomic sites. Still, more than half of known recent human L1 insertions have occurred in only three genes: CYBB (three events$^{19-21}$), factor VIII (three events$^{22-25}$), and dystrophin (three events$^{26}$)$^{27}$, which that raises the possibility of insertional hotspots.

The above examples serve to illustrate the inherently non-random character of mutations. Since this non-randomness is the result of the local sequence environment, thorough analysis of mutation patterns will possibly not only lead to elucidation of the underlying mutational mechanisms but also to a more directed gene analysis, based on the knowledge of mutation-prone sequence elements$^{16}$. Malignant tumors are characterized by somatically acquired genetic changes, which are essential for their progression and survival. These changes can involve a wide range of genomic alterations, including point mutations, insertions, deletions, and translocations. Some of these alterations can be directly linked to specific genes that contribute to the development and progression of cancer.

**Mutations in upstream and downstream control elements of genes**

The precise, sequence-specific regulation of mRNA synthesis is the primary mechanism underlying differential gene expression, both in prokaryotes and eukaryotes. This precise and finely modulated control of adaptive gene transcription is obtained by the intricate interplay of the RNA polymerase II transcription complex with various transcription factors bound to the promoter region immediately upstream of the transcription start site or bound to enhancers located at variable distances up- or downstream of the regulated gene. This pattern of genomic organization creates domains containing genes or gene clusters that have distinct patterns of expression, both during development and in differentiated cells. These domains are separated by insulators, a class of DNA sequence elements with the common ability to protect genes from inappropriate signals generated in their surrounding environment$^{25}$. Gene analysis, both in patient-derived DNA and in experiments based on reporter constructs, has considerably enlarged our knowledge of the regulation of myeloid-specific expression of the gp91phox and, to a lesser extent, p47phox and p67phox components of the NADPH oxidase.
Expression of gp91phox is regulated by the complex interplay between a maturation-stage dependent repressor, CCAAT displacement protein (CDP), and various ubiquitously expressed or myeloid-specific activators. CDP binds to at least five sites within the proximal promoter of gp91phox\textsuperscript{30,31} and represses expression of gp91phox prior to terminal phagocyte differentiation\textsuperscript{31}. The DNA-binding activity of CDP decreases during terminal phagocyte differentiation, thus permitting the binding of transcriptional activators and induction of gp91phox expression\textsuperscript{31}. Possibly the matrix attachment region-binding protein SATB1, another repressor that binds to at least seven sites within the gp91phox promoter and shows a similar pattern of expression, adds an element of redundancy to the maturation-stage dependent repression of gp91phox\textsuperscript{31}. CDP exerts its repression by competing with several transcriptional activators for binding to the gp91phox promoter\textsuperscript{32}. These activators include IFN regulatory factors (IRF)-1\textsuperscript{33} and -2\textsuperscript{34}, the CCAAT box-binding factor CP1, and YY1\textsuperscript{35}. In addition, gp91phox promoter mutations have been identified in several CGD patients. These mutations are clustered in the region around bp -55 region, and each ablates a binding site recognized by the Ets family members PU.1 and Elf-1\textsuperscript{36,37}. It is to be assumed that the lack of gp91phox-protein expression in a patient with a 220-bp deletion directly upstream of the transcription start site (unpublished results) is also the consequence of the ablation of the affected transcription factor binding sites. Furthermore, gene analysis on DNA of a CGD patient with gp91phox expression restricted to eosinophils\textsuperscript{38} has led to the characterization of eosinophil-specific regulation of gp91phox gene expression by transcription factors GATA-1 and GATA-2\textsuperscript{38}.

Less is known about the transcriptional regulation of the other NADPH oxidase components. Expression of p47phox requires the transcription factor PU.1\textsuperscript{40} and is modulated by interferon \(\gamma\)`\textsuperscript{41}, while the transcription factor AP-1\textsuperscript{42}, together with myeloid-specific transcription factors such as Sp1/Sp3, PU.1 and HAF1, is responsible for the transcriptional regulation of p67phox\textsuperscript{43}. Three PU.1-binding sites in the proximal 120 bp of the promoter region regulate transcription of p40phox in myeloid cells\textsuperscript{44}. No data, to our knowledge, are available about the promoter sequences and transcription factors involved in the regulation of p22phox expression. No mutations in insulator elements leading to disease, CGD or otherwise, have yet been described. Still, this is another area of research where future gene analysis, more complete in its assessment of mutational differences, is expected to contribute to the explanation of the more graded interindividua nuclances that cannot be traced back to the mutational knock-out of genes that so far have dominated the endeavour of gene analysis.

Mutations in the 5' UTR of genes
In eukaryotes, about one gene in ten is regulated at the level of protein translation\textsuperscript{45}. While global control of translation affects all cellular mRNAs and is mediated by posttranslational modifications of initiation factors and ribosomal proteins, genespecific (negative) regulation often takes place in the 5' untranslated (5' UTR) region, by interference with the so-called scanning mechanism of translation initiation. This model of translation initiation proposes that a 40S ribosomal subunit, carrying Met-tRNA\textsubscript{Met} and an as yet incompletely characterized set of initiation factors, binds to the 5' end of the mRNA and migrates linearly until it reaches the first AUG codon, whereupon a 60 S ribosomal subunit joins and the first peptide bond is formed\textsuperscript{46}. The two main categories of negative regulators that interfere with this mechanism are elements of secondary RNA structure (stem-loop structures) and upstream open
reading frames. A few disease cases resulting from mutations that affect these regulatory motifs have been described\(^45\).

Mutations in the iron-responsive element (IRE) in the 5' UTR of the mRNA coding for L-ferritin have been found to be responsible for the autosomal dominant disorder called hereditary hyperferritinemia/cataract syndrome (HHCS)\(^47\). The IRE is located in a stem-loop structure and is a binding site for an iron regulatory protein (IRP) that stabilizes the stem-loop and thereby inhibits ferritin synthesis in the case of reduced iron availability. The mutations disturb to various extents the IRE-IRP interaction and consequently lead to varying degrees of hyperferritinemia\(^48\).

Hereditary thrombocythemia (sometimes called familial essential thrombocythemia or familial thrombocytosis) is caused by mutations in the 5' UTR of the mRNA coding for thrombopoietin (TPO). The synthesis of this potent regulator of platelet formation normally is strongly inhibited by the presence of seven upstream open reading frames that interfere with translation initiation from the physiological start site. So far, in four families mutations in the 5' UTR of TPO mRNA have been characterized that all eliminate the inhibitory uORFs by different mechanisms. The result in all four instances is an increased translational efficiency, overproduction of TPO and thrombocythemia\(^49\)-\(^52\).

Finally, in several kindreds with predisposition to melanoma a mutation has been found that creates a new uORF in the gene coding for cyclin-kinase dependent kinase inhibitor-2A (CDKN2A)\(^53\). This gene product constrains cells from progressing through the G1 restriction point, a constraint that is abolished by the introduction of the new inhibitory uORF. Individuals carrying this germ-line mutation are predisposed to melanoma through loss of heterozygosity. While more cases of translational pathophysiology will quite certainly be described in the years to come, the true challenge, in this area also, will be to elucidate the contribution of 5' UTR variability, including mutations, to interindividual differences in phenotype.

**Mutations affecting splicing**

A typical mammalian gene is composed of several relatively short exons that are interrupted by much longer introns. To generate correct, mature mRNAs, the exons must be identified and joined (spliced) together precisely and efficiently, in a process that requires the coordinated action of five small nuclear (sn)RNAs and more than 60 polypeptides\(^54\)-\(^57\). Classically, three DNA-sequence elements, the 5' splice site, the 3' splice site and the (intronic) branch site have been recognized to help define the exon-intron borders on the pre-mRNA. However, in recent years the importance of alternative splicing, the splicing of identical pre-mRNAs into distinct mature RNAs by combining different sets of exons, has become evident. Alternative splicing is responsible for much of the complexity of the proteome, which partly explains the unexpected finding that the human genome might consist of only 31,000-39,000 genes\(^55\)-\(^58\). Because a single primary transcript can have several regions that each undergo alternative splicing, the resulting combinatorial effects of alternative splicing can be very pronounced, and genes that code for tens to hundreds of different isoforms are common. An extreme example is the *Drosophila Dscam* gene, the pre-mRNA of which is alternatively spliced and can potentially generate 38016 different protein isoforms\(^59\).

The complexity thus created is rendered even more elaborate by the fact that the different splice forms of a gene have to be expressed in a tissue- and/or developmental stage-specific manner. It is not surprising, therefore, that there are cis-acting elements other than the classical splicing signals to help regulate and organize these multiple
layers of differential gene expression. Several examples of intronic and exonic cis-elements that are important for correct splice-site identification and are distinct from the classical splicing signals have been described. These elements can act by stimulating or repressing splicing, and they seem to be especially relevant for regulating alternative splicing. Exonic splicing enhancers (ESEs), in particular, appear to be very prevalent, and might be present in most, if not all, exons, including constitutive ones. No well-defined consensus for these DNA-sequence motifs has yet been established, and there may be, in fact, numerous functionally different classes of ESEs.[55; 60; 61]

Alternative splicing, thus, is regulated possibly by trans-acting factors that interact with the regulatory DNA motifs and are expressed in a tissue- and/or developmental stage-specific manner. Furthermore, since the activities of these proteins are affected by phosphorylation, such post-translational modification of these proteins could represent an important mechanism of linking extracellular signaling to alternative splicing.

Mutations affecting the classical splicing signals, either destroying existing splicing signals or creating new ones, have been estimated to represent at least 15% of the point mutations that result in human genetic disease.[57] This estimate is valid also for CGD, where numerous mutations affecting splice sites and leading to exon skipping in all four oxidase subunits have been described.[11]

At present the effect of a mutation on the mRNA or on the encoded protein is usually predicted from the primary sequence, rather than by experimentally analyzing mRNA expression and splicing patterns. Therefore, point mutations that occur in introns or exons and that affect the classical consensus splicing signals are considered to be splicing mutations, whereas point mutations in the coding regions or introns that do not create ectopic splice-site consensus sequences are usually regarded as missense, nonsense or silent mutations.[55; 57]

Nonsense mutations are usually assumed to result in unstable mRNAs or proteins or in truncated proteins, whereas missense mutations are presumed to identify amino acids that are important for the structure or function of a protein. Translationally silent mutations are normally classified as allelic polymorphisms and are considered to be neutral. These assumptions might be correct in some cases, but when they are not supported by characterization at the mRNA level, they could be misleading, because mutations that affect (exonic) sequences important for splicing are likely to have a profound effect on the translated gene product.[55; 57]

Future comprehensive gene analysis, when analyzing newly discovered mutations in coding sequences, should, therefore, be based on both DNA and mRNA information. Not only will this working protocol preclude the misclassification of "silent" mutations as innocuous but it should also help in establishing a comprehensive database of non-classical splicing signals and in avoiding erroneous information to enter the structural and functional analysis of proteins.

Mutations in the 3' UTR of genes

Recent years have seen the accumulation of evidence strongly implicating the 3' untranslated region (3' UTR) in the regulation of gene expression. This regulation occurs at the levels of nuclear export, polyadenylation, subcellular targeting and of the rates of translation and degradation of mRNA. The differential control of translation, for instance, allows for the mRNA to be translated at a place and time different from that of transcription. Such spatial and temporal separation of transcription and translation has been found important in gametogenesis, embryogenesis and the targeting of specific mRNAs into neuronal dentrites.[62] For
example, 3' UTR sequences are responsible for the initial silencing and later translational activation of several maternal mRNA molecules during *Xenopus* and mouse oocyte maturation and early development, and for the concomitant deadenylation and readenylation of their poly-A tails. So far, only a few mutations in the 3' UTR have been identified as cause for human disease. Myotonic dystrophy (DM), for instance, a dominantly inherited multi-system disorder, is caused by an expansion of trinucleotide (CTG) repeats in the 3' UTR of a cAMP-dependent protein kinase gene (*DMPK*). These mutations lead to enhanced binding of the affected transcripts by an RNA-binding protein (CUG-BP), retention in the nucleus and impaired kinase synthesis. The reduced kinase activity, in turn, may contribute to the accumulation of the hypophosphorylated nuclear form of CUG-BP and the resulting retention of unaffected mRNAs containing CUG repeats in their 3' UTR, possibly explaining the dominant mode of inheritance and the multiple systems affected.

One main class of AU-rich elements (AREs), pentanucleotide sequences controlling the cytoplasmic half-life of various mRNA molecules, is located in the 3' UTR of transcripts encoding oncoproteins, cytokines and growth and transcription factors. There is evidence that defective function of the AREs leads to abnormal stabilization of the affected transcripts and may form the basis of several human diseases such as mantle cell lymphoma, neuroblastoma and immune and inflammatory diseases. α- and β-Thalassemia are two more well-known examples of human disease caused, in some instances, by mutational mechanisms involving the 3' UTRs of the affected transcripts. Mutations in the stop codon of the α-globin gene lead to translation continuing into the 3' UTR of the gene and to interference of the translational machinery with the interaction between the so-called α-complex and the three C-rich elements in the 3' UTR. This complex, in concert with the poly(A)-binding protein, protects the α-globin transcript from deadenylation, and, consequently, from degradation. The disturbed interaction between α-complex and C-rich elements may, therefore, explain the markedly decreased half life of the α-globin mRNA responsible for the disease phenotype. While the 3' UTR clearly plays a role in the stabilization of the β-globin transcript, the exact structural determinants responsible for defining the half-life of the mRNA have yet to be determined.

Clearly, it is to be expected that future gene analysis will describe more pathologies caused by mutations in the 3' UTR of affected genes. While these findings certainly will enlarge our knowledge of the mechanisms by which the 3' UTR affects gene expression, they may possibly also demonstrate the contribution of this region to the inter-individual differences in gene expression that lie at the basis of phenotypic variability.

**Gene analysis and analysis of protein structure/function relationships**

As is the case with the (patho)physiological aspect of gene analysis (see above) the once important aspect of gene analysis as principal contributor to the analysis of protein structure/function relationship has lost in relevance. Originally, amino-acid substitutions predicted from mutations found in patients have been used to draw conclusions about the structure and/or function of the protein in questions. Missense mutations that led to a lack of protein expression were said to code for structurally essential amino acids, while those missense mutations that retained protein expression but compromised function were considered to affect functionally relevant amino acid residues. In the case of gp91phox, structural domains that have been localized by
mutational and consecutive biochemical analysis were the binding sites for heme74, FAD75,76, NADPH77 and p47phox78. Recombinant DNA technology, site-directed mutagenesis and the different forms of crystallography have, however, greatly reduced the importance of this aspect of gene analysis. Today, almost any mutation can be introduced into a gene of interest deliberately, and the resulting mutant protein can then be produced in large amounts and subjected to the relevant analytical assays. Recombinant DNA technologies have the additional advantage of avoiding the ambiguities of interpretation that are caused by the possible presence of exonic splicing signals (see above). In fact, random mutagenesis in recombinant cDNA of CYBB has produced many more mutants leading to expression of gp91phox than would have been expected, given the scarcity of missense mutations with retained protein expression in CGD patients (R. van Bruggen, personal communication). While this has not been subject of further investigation, one possible explanation for this phenomenon is that recombinant DNA technology circumvents the need for pre-mRNA splicing.

Given the amounts of money and effort invested in the structural-genomics projects, that are now already well in the phase of preparation, it is to be expected that the number of crystallized proteins with detailed structural information will increase almost exponentially. An even larger group of proteins will be amenable to advanced structural modelling based on the increasingly comprehensive data base of folding motifs. Clearly, the role of gene analysis in investigating the structure/function relationships of proteins will then come to an end.

Concluding remarks - the presence and future of gene analysis
The technological foundation and the knowledge relevant for gene analysis, as we have seen, have changed profoundly and continue to do so. As a result, gene analysis today should adopt certain standards in order to maximize scientific profit and progress of gene analysis. This standardization extends both to the technological side, characterized by an increasing degree of automatization, and to the procedural aspect. Comprehensive gene analysis, therefore, should routinely encompass the whole sequence present on the mRNA, i.e. including 5’ and 3’ UTRs, and be done both on genomic and mRNA level. Optimally, the raw sequence information thus obtained should be made publicly accessible to advance a comprehensive characterization of human variation.

Once the technological and knowledge-related bases have been created, gene analysis will produce an ultimately individualized genetic map that will be essential for optimal prevention and treatment of disease.

REFERENCES


NADPH oxidases and mutation analysis


38. Kuribayashi F, Kumatori A, Suzuki S, Nakamura M, Matsumoto T, Tsuji Y. Human peripheral eosinophils have a specific mechanism to express gp91-phox, the large subunit of cytochrome b558. Biochem Biophys Res Commun. 1995;209:146-152.


NADPH oxidases and mutation analysis


NADPH oxidases and mutation analysis
