through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150, 1217–1228

HLA complex genes in type 1 diabetes and other autoimmune diseases. Which genes are involved?

Dag E. Undlien, Benedicte A. Lie and Erik Thorsby

The predisposition to develop a majority of autoimmune diseases is associated with specific genes within the human leukocyte antigen (HLA) complex. However, it is frequently difficult to determine which of the many genes of the HLA complex are directly involved in the disease process. The main reasons for these difficulties are the complexity of associations between several HLA complex genes might be involved, and the strong linkage disequilibrium that exists between the genes in this complex. The latter phenomenon leads to secondary disease associations, or what has been called ‘hitchhiking polymorphisms’. Here, we give an overview of the complexity of HLA associations in autoimmune disease, focusing on type 1 diabetes and trying to answer the question: how many and which HLA genes are directly involved?

D.E. Undlien* B.A. Lie E. Thorsby Institute of Immunology, The National Hospital and University of Oslo, N-0027 Oslo, Norway. *e-mail: d.e.undlien@rh.uio.no

Type 1 diabetes, rheumatoid arthritis, multiple sclerosis and Graves’ disease are all autoimmune diseases, to name but a few, probably resulting from an inappropriate immune response to self-proteins. Apart from being characterized by loss of immunological tolerance to self, these diseases share other important characteristics. First, they are all complex diseases where multiple genes and environmental factors act in concert in their aetiology. Second, at least one of the genetic factors involved in the development of these diseases is located in the HLA complex on chromosome 6 (Ref. 1). The genes within this complex that are involved in disease are frequently difficult to identify. In this article, we focus on progress made in establishing which HLA genes are directly involved and discuss methodological aspects that are important to consider when studying the effects of this gene complex. We focus directly on type 1 diabetes because studies of this disease have implications for the understanding of other HLA associated diseases.

Type 1 diabetes

Diabetes mellitus is a group of heterogeneous disorders characterized by hyperglycaemia (high blood sugar). The incidence of these disorders is increasing worldwide.
Type 1 diabetes (previously called insulin-dependent diabetes mellitus; IDDM) can affect people in all age groups. It is a multifactorial disease involving multiple genes, with unidentified environmental factors contributing to the pathogenesis. Through what is believed to be an autoimmune process, the disease results in the destruction of insulin-producing β-cells in the pancreas, with concomitant reduction in endogenous insulin production. Epidemiological evidence for a genetic component to the disease comes from, among others, studies of twins that show a concordance rate in monozygotic twins of 30–70% (i.e. if one twin is affected, the second twin has a 30–70% chance of developing the disease), which is higher than in dizygotic twins (approximately 10%) and in siblings (6%) [14]. The incidence of the disease has large geographical variations, which might be caused by differences in frequencies of HLA encoded genetic susceptibility factors between different ethnic groups [15]. It is clear that non-HLA genes are also involved. Many linkage studies are, however, in disagreement and molecular mechanisms of non-HLA associations are mostly unknown [16]. The emerging picture is that there is one gene complex harbouring the major predisposing genes, the HLA complex, plus several minor predisposing genes spread throughout the genome. The HLA complex can explain 35–50% of the familial clustering of the disease [16]. However, the HLA complex contains at least 128 genes [17], and it is difficult to determine which gene or genes are responsible.

The HLA complex

The HLA complex is the major histocompatibility complex (MHC) in human and its complete sequence was presented in 1999 (Ref. 9). This, a milestone in our understanding of this complex, should facilitate further studies aimed at dissecting the complex nature of HLA associations with several complex diseases including type 1 diabetes. The HLA complex is a densely packed gene cluster containing at least 128 genes (not including the extended class I region where the number of genes is not known in detail) located on the short arm of chromosome 6 (Fig. 1). Approximately 40% of the genes within this cluster have an immune function.

Originally, the HLA complex was considered to extend from HLA-DR to HLA-E. However, there are related genes on both the telomeric and centromeric side of these borders, leading to the concept of the extended HLA complex. This includes the extended class I and II regions where the centromeric border is the HSET gene and the provisional telomeric border is the haemochromatosis gene HFE. Two microsatellites discussed in this article are also depicted.

Type 1 diabetes, celiac disease and Graves’ disease (see Box 3 for nomenclature). In such diseases both the frequency of the DQA1*0501 and DQB1*0201 alleles (encoding the DQ2 molecule), the DRB1*03 allele (encoding DR3) and the B*08 allele (encoding B8) will tend to be increased among patients compared with controls. It is then very difficult to determine which of these alleles are directly involved in the predisposition and which are only secondarily associated because of LD with the directly associated allele.
Box 1. Function of the classical HLA class I and II molecules

HLA molecules are located in the cell membrane and are very similar in their three-dimensional structure.

The HLA class I molecules (HLA-A, -B and -C)
These consist of an $\alpha$-chain encoded by the respective polymorphic class I genes and a $\beta_2$-microglobulin encoded by a non-polymorphic gene on chromosome 15. Class I molecules are expressed on almost all nucleated cells. Their most important function is to present peptide fragments predominantly derived from endogenous, cytosolic proteins to CD8$^+$ cytotoxic T lymphocytes. HLA class I molecules are synthesized and transported to the endoplasmic reticulum where they encounter peptides resulting from degraded cytosolic proteins (including viral proteins if the cell is infected). Different HLA molecules bind to different sets of peptides. Peptides that fit into the peptide-binding groove of the particular HLA class I molecule can bind non-covalently, and the complex of HLA class I molecule and peptide is then transported to the cell membrane where the HLA class I peptide complex can encounter CD8$^+$ cytotoxic T cells. This can result in cytotoxic lysis of the virally infected cell by the T lymphocyte.

The HLA class II molecules (HLA-DR, -DQ and –DP)
These are made up of an $\alpha$- and a $\beta$-chain encoded by two separate polymorphic genes in the HLA complex; for example, DQA1 encodes the $\alpha$-chain and DQB1 encodes the $\beta$-chain of the DQ molecule (DRA1 is an exception as this gene is virtually non-polymorphic). Under normal circumstances class II molecules are expressed only on specialized antigen-presenting cells such as dendritic cells, B lymphocytes, macrophages and thymic epithelial cells. Their main function is to present peptides to CD4$^+$ helper T lymphocytes. Extracellular proteins are endocytosed or phagocytosed and degraded in endosomes/lysosomes. The HLA class II molecules are routed to the endosomal compartment where they can bind to peptide fragments if the peptides ‘fit’ in the peptide-binding groove of the particular HLA class II molecule.

The problem of LD will probably not be unique to HLA. Clusters of genes of similar function are found at several places in the genome. It is probable that in several complex diseases there might be loci linked with each other and involved in disease pathogenesis.

HLA-DRB1, -DQA1 and -DQB1 genes are directly involved in the predisposition to type 1 diabetes
Type 1 diabetes was initially associated with some HLA class I alleles, namely the alleles encoding the B8 and B15 molecules. Later, a stronger association with the HLA class II alleles encoding DR3 and DR4 was found, and it was concluded that the class I associations observed previously were secondary to LD with these high-risk DR alleles. Similarly, B7 was found to be decreased among type 1 diabetes patients, and later this negative association was found to be even stronger for DR2 which is in LD with B7 (Ref. 16). Finally, in 1983, Owerbach et al. observed the distribution of a restriction-fragment length polymorphism (RFLP) in an HLA class II subregion, later identified as the DQ subregion, was significantly different in type 1 diabetes patients compared with controls matched for alleles at the DR locus. This implied that the observed associations between particular DR alleles and type 1 diabetes could be secondary to LD with high-risk DQ alleles. This was further supported by the finding that type 1 diabetes strongly correlates with HLA alleles not encoding aspartic acid at position 57 of the DQ$\alpha$-chain. Aspartic acid in this position was later shown to be important for the shape of the peptide-binding groove of HLA-DQ molecules, and hence, have implications for which antigenic peptides can be bound by these molecules.

It has become generally accepted that the genes encoding the peptide-presenting DQ molecules (i.e. the DQA1 and DQB1 genes) are directly involved in the development of type 1 diabetes. First, the HLA associations with diabetes are consistently strongest for the DQ subregion (or, in rarer instances, the DR subregion). Second, transracial studies show that similar alleles of DQA1 and DQB1 are associated with the disease in all populations. Thus, despite the fact that the composition of the haplotypes can vary considerably between various ethnic groups, particular combinations of alleles of the DQ genes are the common denominators of susceptibility to type 1 diabetes. The DQ involvement in type 1 diabetes is complex, as some DQ haplotypes confer susceptibility, others protection and still others are neutral (Table 1). Third, there is a clear correlation between particular DQ ($\alpha_1,\beta_1$) heterodimers and type 1 diabetes, irrespective of whether the $\alpha$- and $\beta$-chains are encoded by genes on the same chromosome or haplotype (cis) or on different chromosomes (trans).

Even though there is general agreement that DQ genes are directly involved in type 1 diabetes, there is substantial evidence that the DQ genes cannot explain all the genetic susceptibility encoded within the HLA complex. This has been made particularly clear by studies of the contribution of DR4 subtypes to the development of the disease where different DR4 subtypes were found to be distributed differently on DQB (DQA1*03-DQB1*0302) haplotypes between patients and controls (Fig. 2). Studies of other haplotypes also indicated that DR
The history of the discovery of the HLA molecules and the HLA complex genes are reflected in the nomenclature for factors in the HLA system. Originally, HLA specificities were defined serologically by antibodies or by T-cell specificities and only later were they genomically defined. The table gives some examples. The serological specificity DR4 can be encoded by several different alleles, reflecting the much higher resolution of DNA-based HLA typing.

<table>
<thead>
<tr>
<th>HLA locus</th>
<th>Serological specificity</th>
<th>Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB1</td>
<td>DR4</td>
<td>DRB1*0401</td>
</tr>
<tr>
<td></td>
<td>DR4</td>
<td>DRB1*0402</td>
</tr>
<tr>
<td></td>
<td>DR4</td>
<td>DRB1*0403</td>
</tr>
<tr>
<td></td>
<td>DR4</td>
<td>DRB1*0404</td>
</tr>
<tr>
<td></td>
<td>DR4</td>
<td>DRB1*0408</td>
</tr>
<tr>
<td></td>
<td>DR3</td>
<td>DRB1*0301</td>
</tr>
<tr>
<td>DQA1</td>
<td></td>
<td>DQA1*0501</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DQA1*0401</td>
</tr>
<tr>
<td>DQB1</td>
<td>DQ2</td>
<td>DQB1*0201</td>
</tr>
<tr>
<td></td>
<td>DQ8</td>
<td>DQB1*0302</td>
</tr>
<tr>
<td>B</td>
<td>B8</td>
<td>B*0801</td>
</tr>
</tbody>
</table>

The names of genomically defined alleles consist in general of the locus name followed by an asterisk and 4-5 digits. The first two digits frequently reflect the previous serological nomenclature.

**Box 3. HLA nomenclature**

**Linkage disequilibrium**

Linkage disequilibrium (LD) refers to the nonrandom assortment of alleles at neighbouring loci at the population level. Thus, when the occurrence of pairs of specific alleles at different loci on the same haplotype is not independent, the deviation from independence is termed linkage disequilibrium (e.g. the alleles encoding DQ2 occur together with the alleles encoding DR3 more frequently than expected by chance).

Several statistical measures of LD have been developed; one of the simplest being:\n\[
D = \frac{f_{AB}}{f_A f_B} - 1
\]

where $A$ is a particular allele at locus 1 and $B$ is a particular allele at locus 2.

Today, LD is generally taken to mean allelic association caused by tight linkage. In more general terms, LD can refer to any form of allelic association, including associations caused by admixture. If, for example, two genetically different populations are mixed and these two populations have different allele frequencies at a given locus and also have different frequencies of a particular disease, the gene in question will tend to be associated, but neither linked to nor involved in the pathogenesis of the disease. This phenomenon of spurious associations caused by admixture is also referred to as population stratification.

**Methods to look for new susceptibility genes eliminating secondary associations caused by LD with linked directly involved genes**

- **Stratified case-control analysis**
  Comparing cases and controls that are matched for known disease susceptibility genes in the case of type 1 diabetes, DQ (DQA1, DQB1) and DR (DRB1) genes.

- **Homozygous parent test**
  Linkage test using affected sib pair (ASP) analysis only including parents being homozygous for DR and DQ genes in the analysis.

- **Homozygous parent transmission disequilibrium test (TDT)**
  TDT only includes transmissions from DR-DQ homozygous parents in the analysis (Fig. 3).

- **Conditional extended transmission disequilibrium test (CETDT)**
  CETDT is an extended TDT analysis conditioned (stratified) on parental DR-DQ genotypes.

**Linkage test using affected sib pair (ASP) analysis**

Comparing cases and controls that are matched for known disease susceptibility genes in the case of type 1 diabetes, DQ (DQA1, DQB1) and DR (DRB1) genes.

- **Homozygous parent test**
  Linkage test using affected sib pair (ASP) analysis only including parents being homozygous for DR and DQ genes in the analysis.

- **Homozygous parent transmission disequilibrium test (TDT)**
  TDT only includes transmissions from DR-DQ homozygous parents in the analysis (Fig. 3).

- **Conditional extended transmission disequilibrium test (CETDT)**
  CETDT is an extended TDT analysis conditioned (stratified) on parental DR-DQ genotypes.
genes are probably directly involved in predisposition to the disease.\(^{24}\) The evidence for a primary role of some DRB1 alleles is, however, not as conclusive as for DQA1 and DQB1 because the possibility exists that DRB1 genes might only be a marker for other HLA complex gene(s) directly involved in the predisposition. However, the genes surrounding DRB1 do, in general, display weaker associations with type 1 diabetes than DRB1 when comparing DQ-matched cases and controls. Also, the DRB1 associations with type 1 diabetes seem to be constant through different ethnic groups. In addition, the DRB1 and DQB1 genes are virtually identical, and the three-dimensional structure of DR and DQ molecules is thought to be very similar.

Further evidence for a direct involvement of MHC class II genes encoding peptide-presenting molecules comes from animal models. In the non-obese diabetic (NOD) mouse, an animal model for type 1 diabetes, a role for the mouse homologues of the DQ molecules (I-A) and the DR molecules (I-E) has been established. Both DQ and DR alleles are probably directly involved in predisposition to the disease.\(^{24}\) The evidence for a primary role of some DRB1 alleles is, however, not as conclusive as for DQA1 and DQB1 because the possibility exists that DRB1 genes might only be a marker for other HLA complex gene(s) directly involved in the predisposition. However, the genes surrounding DRB1 do, in general, display weaker associations with type 1 diabetes than DRB1 when comparing DQ-matched cases and controls. Also, the DRB1 associations with type 1 diabetes seem to be constant through different ethnic groups. In addition, the DRB1 and DQB1 genes are virtually identical, and the three-dimensional structure of DR and DQ molecules is thought to be very similar.

Further evidence for a direct involvement of MHC class II genes encoding peptide-presenting molecules comes from animal models. In the non-obese diabetic (NOD) mouse, an animal model for type 1 diabetes, a role for the mouse homologues of the DQ molecules (I-A) and the DR molecules (I-E) has been established by genetic and transgenic models, providing functional evidence that the DRB1, DQA1 and DQB1 genes are directly involved in type 1 diabetes. Mice transgenic for the susceptible DR3- and DQ8-encoding genes develop diabetes, whereas identical mice made transgenic for the genes encoding the protective DQ6 molecule do not.\(^{25}\)

### Analytical methods to look for additional disease susceptibility genes in the HLA complex

Several lessons should be learnt from the history of HLA associations in type 1 diabetes. First, one must expect to find associations with virtually all genes in the HLA complex because of the high-risk DR and DQ alleles. The important question is: are there other HLA complex genes that have effects independent or additional to the involved DR and DQ genes? To address this question, LD must be taken into account (Box 2). One has to make sure that the potential associations observed are not secondary to LD with alleles at the primary susceptibility loci (DRB1, DQA1, DQB1) before one can conclude that such disease associations are caused by independent additional genetic risk factors in the HLA complex.

A method frequently used to try to eliminate secondary associations caused by LD with DQ or DR alleles is to calculate some measurement of LD, of which there exist several methods.\(^{26}\) Between alleles at the established disease loci (i.e., DR and DQ) and the locus being tested for disease association. Absence of statistically significant LD between DR and/or DQ genes and the associated gene being studied has frequently been taken as evidence for an association not being caused by LD with alleles at the established disease-associated loci (i.e., DR and DQ). Our belief is that this is not sufficient to eliminate subtle effects of LD. Several weak and statistically insignificant LDs might still confound the data significantly.\(^{27}\) Application of this procedure probably underlies several unreplicated reports of independent disease associations with additional genes in the HLA complex.

#### Methods to eliminate the effects of LD

There are methods that efficiently eliminate LD with the disease-involved class II genes. The most widely used is a stratified case-control analysis where patients and controls are matched for the genes involved in disease, that is, DR and DQ. A problem has been that it requires a very large number of random controls to get a sufficient number of controls carrying the same HLA class II alleles as the patients. In addition, case-control analysis is received with a great deal of scepticism, mainly because associations might be caused by population stratification. However, case-control analysis is statistically powerful and avoids the need to collect families. In addition, there are an increasing number of controls available, and methods aimed at detecting population stratification are being developed.\(^{28}\) Hence, provided

---

**Table 1. Some HLA-DRB1-DQA1-DQB1 haplotypes and their impact on type 1 diabetes susceptibility**

<table>
<thead>
<tr>
<th>DRB1-DQA1-DQB1 haplotype</th>
<th>04-03-0302</th>
<th>03-0501-0201</th>
<th>01-0101-0501</th>
<th>0102-0602</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Susceptible</td>
<td>Neutral(^a)</td>
<td>Protective</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)This haplotype might be weakly susceptible in some populations, and in others it is weakly protective relative to other HLA haplotypes depending on haplotype frequencies in the background populations.

**Fig. 2.** DR4 subtypes in type 1 diabetes. The risk encoded by DQ8 haplotypes (DQA1*03–DQB1*0302) is influenced by which DRB1*04 subtype is carried. For example, DQ8 haplotypes carrying the DRB1*0402 or 0405 alleles are susceptible, whereas those carrying DRB1*0403 or 0406 are protective.
The transmission of from DR3-DQ2 gene distinct from DR-DQ. disease susceptibility disequilibrium (LD) with a is in linkage from the expected 50%.

significant difference in linkage disequilibrium (LD) with a disease susceptibility gene distinct from DR-DQ.

Are HLA complex genes other than DR and DQ involved in the predisposition to type 1 diabetes?

In the past few years, there have been several studies applying methods able to eliminate the effects of LD to DR and DQ genes. These studies have demonstrated that there are indeed additional genes in the HLA complex involved. Among the first to show this were Robinson et al. who applied the homozygous parent test to families with type 1 diabetes. They found evidence for heterogeneity in terms of genetic risk on different DR3 and DR4 haplotypes implying the presence of genetic susceptibility factors in addition to DR and DQ on these haplotypes. The results observed for DR4 haplotypes can probably be explained by the well-established effect of different DR4 subtypes on the risk of type 1 diabetes. The results for DR3 haplotypes, however, suggest the presence of additional HLA linked type 1 diabetes susceptibility genes besides the DR and DQ genes. Because the method used was based upon linkage analysis (ASP), the region where such a novel susceptibility gene could be located is large.

Susceptibility genes in the HLA class I–extended class I region

We have systematically assessed microsatellite markers spanning the extended HLA complex using both a large number of HLA class I matched cases and controls as well as the homozygous parent TDT in large family datasets from Norway, Denmark and the UK. We found clear evidence that a gene in LD with a particular allele at microsatellite D6S2223 modifies the risk encoded by the DR3-DQ2 haplotype both in case-control analysis and in TDT analysis (Fig. 3). The transmission of D6S2223*3 from DR3 homozygous parents to their diabetic child was significantly less than the expected 50%. This strongly implies that there is a gene in LD with marker D6S2223 that influences the pathogenesis of type 1 diabetes. A study by Herr et al. using data partly overlapping the UK data that we used, applied a CETDT analysis of markers spanning the HLA complex and found only one marker outside the class I region that showed significant evidence for independent association to type 1 diabetes, namely D6S2223. In this study, they also showed using microsatellite markers that the region of maximum association, identified the region harbouring the major type 1 diabetes susceptibility genes DR and DQ with great precision. On the basis of this study and analysis of multilocus haplotypes (B.A. Lie et al., unpublished), we believe that the putative susceptibility gene is located in the extended class I region in the vicinity of D6S2223.

Some reports have also suggested that HLA class I genes could be associated with type 1 diabetes independently of DR and DQ genes. If these associations are confirmed, they will need to be correlated to the associations at D6S2223 to see whether these associations are independent of LD between D6S2223 and class I genes.
Susceptibility genes in the class III region

There have been two studies suggesting that the class III marker D6S273 is associated with development of type 1 diabetes independently of the DR and DQ genes\(^\text{30,33}\). However, different allelic loads were found associated in the different populations and it has not been established clearly whether this marker is associated independently of the association observed for D6S2223 (Ref. 30). Further studies are needed to clarify this.

Additional susceptibility genes in the class II region

There have been contradictory reports on a potential role of the genes encoding the DP molecules (DPA1, DPB1)\(^\text{36,37}\), and this issue is not yet settled. The peptide transporters TAP1 and TAP2 have been analyzed in several studies, and, although some initial reports suggested an independent role of the TAP2 gene, later studies suggested that this was a false positive result caused by statistically insignificant LDs confounding the data. Hence, current evidence suggests that the antigen-processing genes (TAP1, TAP2, LMP2, LMP7) in the class II region are not directly involved in type 1 diabetes\(^\text{27,38}\).

Animal models

It has been clearly established in the NOD mouse that there are other genes linked to the MHC (in addition to the I-A and I-E genes) that affect disease susceptibility\(^\text{39,40}\) (Fig. 4).

Conclusion

The HLA complex is characterized by a high gene density, genetic complexity and strong LD. Several common complex genetic diseases, in particular most autoimmune diseases, have a major genetic component encoded within this complex. Current evidence suggests that, for type 1 diabetes, there are at least four genes involved. The class I genes QA1, QB1, DRB1 are the most important ones, but in addition there is at least one unidentified gene most probably located telomeric to the class I/extended class I region. So far, LD has hampered the detection of all genetic susceptibility factors within the HLA complex. However, the availability of increasing numbers of families, cases and controls, as well as methodological improvements to eliminate the ‘unwanted’ secondary effects of LD give reason to believe that we will soon be able to identify all the HLA complex genes directly involved in the predisposition to develop type 1 diabetes and other autoimmune diseases. The currently running disease component of the 13th International Histocompatibility Workshop has this as one of its main aims (http://www.IHWG.org). Such studies will also be beneficial for the analysis of other clusters of disease-associated genes in genetically complex diseases.

References

16. Farid, N.R. et al. (1979) HLA-D-related (DRw) antigens in juvenile diabetes mellitus. Diabetes 28, 552–557
Alternative splicing: increasing diversity in the proteome world

Brenton R. Graveley

How can the genome of Drosophila melanogaster contain fewer genes than the undoubtedly simpler organism Caenorhabditis elegans? The answer must lie within their proteome. It is becoming clear that alternative splicing has an extremely important role in expanding protein diversity and might therefore partially underlie the apparent discrepancy between gene number and organismal complexity. Alternative splicing can generate more transcripts from a single gene than the number of genes in an entire genome. However, for the vast majority of alternative splicing events, the functional significance is unknown. Developing a full catalog of alternatively spliced transcripts and determining each of their functions will be a major challenge of the upcoming proteomic era.

A major undertaking of the post-genomic era will be the description and functional characterization of the full complement of proteins (i.e., the proteome) expressed by an organism. DNA recombination, RNA editing and alternative splicing make this task more difficult than it first appears as these processes increase the number of proteins that can be synthesized from each gene. As a result, the number of proteins in the proteome is by no means equivalent to the number of genes, but can exceed it by literally orders of magnitude. The mechanism most widely used to enhance protein diversity, with regard to the number of genes affected and the breadth of organisms it occurs in, is alternative splicing (Box 1).

Alternative splicing can generate multiple transcripts encoding proteins with subtle or opposing functional differences that can have profound biological consequences. In this article, I will discuss the prevalence of alternative splicing and the amount of diversity it can create, provide examples of several complex alternative splicing events and, finally, discuss the issue of how much of the observed alternative splicing actually represents errors or ‘noise’ in the system. Readers should refer to an excellent recent review by Smith and Valcárcel[1] for more information about the biochemical mechanisms of alternative splicing.

It’s all around us
All eukaryotes contain introns (see Glossary) in at least some of their genes although the number can vary considerably from organism to organism. For example, only 250 out of the 6000 genes in Saccharomyces cerevisiae contain introns[4], whereas most of the estimated 35 000 human genes[4] are thought to contain introns. But how many of these genes encode transcripts that are alternatively

References: