1 Popular description

The publication of the entire sequence of the human genome is imminent. The sequencing project has been on its way for several years but with the event of Celera entering the field, the speed has picked up enormously. Though a huge feat, the sequencing of the human genome is not the finishing project of human genomic disease. Once the sequence is known comes the project of trying to interpret the results and to find the genes that lie hidden in the sequence. This mapping project is perhaps an even more difficult one since it requires subtle interpretations and much knowledge about the construction of genes in eukaryotes in general and humans in particular.

Mapping the genome will require new tools as well as brilliance. One interesting feature of the human genome is its variation between individuals, approximately one base in every 200-1000 differs between two human individuals. These variations are called Single Nucleotide Polymorphisms (SNPs) and they do not, generally cause any deleterious effects. Some genetic diseases, however, are believed to be caused by variations at single sites in the genome. One known such case is the disease Sickle-cell anaemia, where a mutation in a single nucleotide changes one amino acid in the protein haemoglobin, the protein responsible for carrying oxygen from the lungs to other parts of the body in the red blood cells. This change of one amino acid totally distorts the structure of the protein rendering it much less capable of transporting oxygen. It is believed that other diseases may have similar causes. To investigate this matter a scientist would need a tool that is sensitive enough to find a variation at one nucleotide in the human genome, which consists of, in total, approximately 3 billion nucleotides. The tools that are used are called probes and today two types of probes are regularly used for this kind of analysis: PCR-probes and hybridisation probes. The hybridisation probes are generally not sensitive enough to allow for finding single nucleotide variations since one mismatch in the hybridisation is not enough to give selectivity. The PCR-probes, though sensitive and selective enough, have another drawback. When investigating i.e. the cause of a genetic disease it might be necessary to investigate several sites. The most convenient way of doing this would be to be able to apply more than one probe at once. This, however, gives rise to new complications since the reactions used in the experiment might lead to the putting together of two different probes. If PCR-probes are used these "cross-reaction" products are not easily distinguished from the correct products of the detection reactions.

A padlock probe is a new kind of probe that eliminates the drawbacks of the two probe types mentioned above. It is a molecule designed to hybridise to the target sequence in such a way that the 3'- and 5'-ends of the probe meet at the nucleotide that is being investigated. This nucleotide in the target sequence hybridises with the nucleotide at the extreme 3'-end of the probe. A ligase is then added, which connects the two ends of the probe

making it circular and wrapped around the target sequence. If there is a variation in the nucleotide of interest, there will be a mismatch and the two ends of the probe will not be perfectly aligned. This one mismatch is enough to prevent the ligase from being able to connect the probe ends. Once the ligation reaction is finished any remaining probes are washed away. The fact that the probe is wrapped around the target means that it will not come loose even if the washing temperature is above the melting-temperature of the target-probe hybrid, which allows for more rigorous washing.

If more than one probe is used at the same time there is still a risk of cross-reactions occurring but the products of these reactions will not be circular molecules and will thus be easy to distinguish from the "correct" results. The probe consists of three distinct parts, the two end segments and a linker segment connecting the end segments. To enhance detection and to amplify the signal, the linker segment can be equipped with sequences that gives it certain such features such as primer sequences for replication reactions to amplify the signal or a zip code, enabling easy detection of probes on a micro-array.

We have written a program that, when given the target sequence, the position of the nucleotide of interest, what zip code to insert and certain experimental conditions, designs a padlock probe that can then be produced chemically and put to work in the investigative field.

The program begins with designing the end-segments, giving them sequences that enable them to hybridise with the target sequence in the correct fashion. It then inserts a zip code and checks whether the probe can hybridise to itself. If the probe can,and this hybridisation includes the zip code,the zip code is replaced by another one until a probe is formed that can function properly.

The program is run in a directory, defining a project, in which there is a file containing possible zip codes. Since it should be possible to apply (and design) several probes at once, the used zip codes are stored in a "history file" which is checked at the beginning of every insertion of a zip code. A zip code that has been used previously is not used again in the same project.

Checking for internal hybridisations is necessary since any such events would severely damage the performance of the probe. If the 3'-end of the probe could hybridise with another region, this would lead to a probe that would probably be totally useless.

The program is written in C++ using object-orientation. Its design is modular, making it easy to add or replace functions of the program, something that will probably be necessary since new detection methods or other features are likely to be developed in the future.