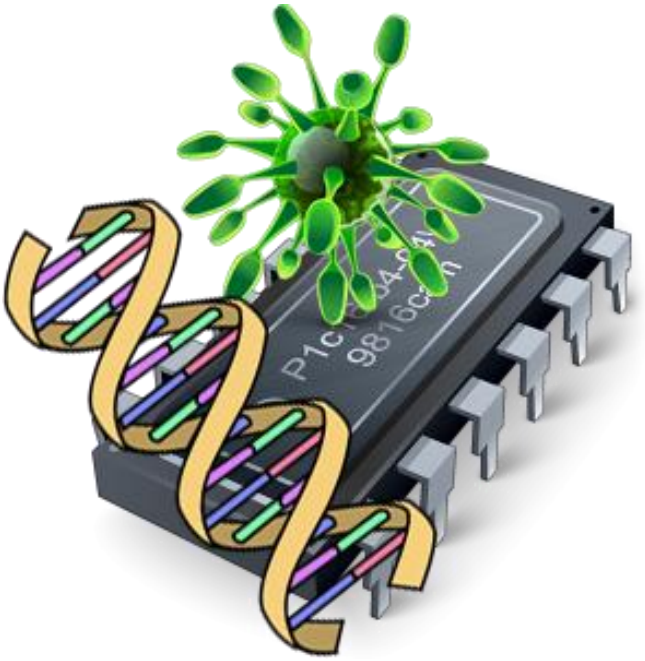

Bioinformatics

What is it about and where should be the limits?
Seminar by Christoph Kröger

Table of contents

What is Bioinformatics?

Field of activity

Since scientists have had access to genome information, biology has undergone a change of method: Biology, once a phenomenological and descriptive science has now emerged as a field of enquiry that is able to analyse and explain.

This paradigm shift could be compared to the change in chemistry in the last century, when the periodical system of elements was introduced. It reduced the seemingly endless number of substances in the world to just a limited number of elements.

We have the same complexity, if we try to look at all kinds of different people in our near or far neighbourhood. Only the difference is that the data pool in bioinformatics is much larger than it is in the case of chemistry. It is in particular because of the diverging genetic information that is apparent in even two very similar humans.

Primarily the task of bioinformatics is to facilitate and accompany the process of sequencing. This means to decode the genomes with labour machines and biological techniques to make the extraction of DNA and later the analysis of it more efficient, easier and automated. Currently there are about over a million base pairs processed on a daily basis.

To handle these huge numbers of information, complex methods for data mining are required in order to interpret this difficult-to-understand genomic text.

For example: finding parts in gene which have a 3-dimensional structure or to reconstruct and make recommendations for the functional use of this gene's.

Especially experiments involving space structuring and in areas which contain room structures and functions of molecules it is necessary to work with these algorithms. The data bases are so huge that it is almost impossible to make a complete analysis by hand.

The total encryption of the human gene could not be accomplished without bioinformatics' support and could not be analysed without this algorithms. Some people compare this encryption to major human inventions like the moon landing or even the invention of the wheel.

Development of the gene database

Margaret Belle Oakley Dayhoff (* 11th. March 1925 in Philadelphia; † 5. February 1983) is usually named as the founder of the field of bioinformatics. After studying mathematics at Washington Square College, New York University, she successfully graduated in 1945 with distinction, i.e. „Magna Cum Laude".

She received her Ph.D. at University Columbia University in quantum chemistry. Her doctoral thesis was about the „usage of computer systems for mass data processing in theoretical chemistry" and was followed by a professorship for physiology and biophysics at Georgetown University („Georgetown University Medical Center") where she continued her research.

The main topic in her researches since 1955 was the comparison of amino acid sequences of homologous proteins of different species. She analysed the amino acid sequences of DNA strings of many species and checked them for homology to determine the respective relations.

Protein homology is defined as being derived from a mutual "ancestor ". (Fig 1).

If the shared ancestry dates back millions of years and the successor genes' have developed independently (genetic drift), it is more than likely that only a very low degree of analogy pertains. (Fig 2)

Sequence identity for an amino acid sequence of a protein produced gene is greater than 10%, then homology is given.

If the genes' identity exceeds the 30% threshold, i.e. the identity of the genes' amounts to a third, another reason than mutual evolution is highly implausible. In this case the genes are classified as homologous.

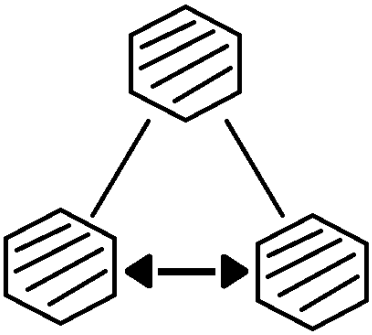

Fig. 1 - homology

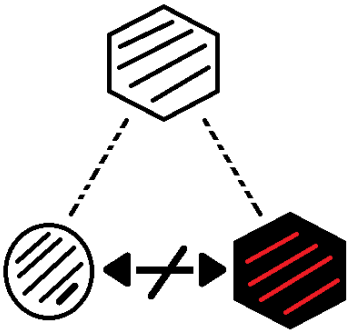

Fig. 2 - no homology any more

The Dayhoff heritage

In 1965 appeared the „Atlas of Protein Sequence and Structure"– a collection of all – at the publishing moment known – protein sequences. 1984 were the foundation of the Protein Information Resource database. All information of the Atlas was transferred into the PIR-database. This was also assumed in 2002 into the UniProt-database. Uniprot (universal protein) is the biggest bioinformatics database for proteins of all living species and viruses. It contains information about the protein function and –structure and also linked to other topic relevant database entries.

It combines the data from the swiss-prot, trEMBL and the PIR database and is regularly republished and updated.

Entries in this database are long strings which are grouped. The groups existing of letters which are standing in for the different combinations of amino acids.

```
MATAASPRKL ELEQFTSSCS PSCPQHPARF QYTMADFAGT VFLFFVQVLP
```

Fig. 3 - Example of an amino acid string (Swiss Cheese)

What exactly the meaning of these fragments is and how one has to interpret them should be explained in a different place.

Possibilities of bioinformatics

Reconstruction of creatures through smallest item only

Hitherto it has been assumed that the development of humanity dates back approximately about 110.000 years. Accordingly the anatomical modern human spread over central Africa and the Middle East. In the period between 120.000 and 35.000 years the Neanderthal spread over the Eurasian continent. Partly he did it simultaneous with the anatomic human.

Until today the opinion prevails that anatomic human established himself against the Neanderthal and is responsible for the latter's extinction.

But parallel to the anatomic human and the Neanderthal another species developed. The so-called „Denisova-human", named after a region in the Altai-mountains in the southern Siberia.

The Denisova was only reconstructed out of the small rest of a fingertip which was found in the Denisova-Hills. Scientists succeeded in reconstructing the Denisova from just a fingertip.

For reconstructing the whole human they took about 30mg pulverised material of the fingertip. The insight to be gained from this is that for less than 50.000 years there were 2 archaic kinds of humans were living on the Eurasian continent.

Fig. 4 - Fingertip from Denisova Human

To correct this part of the history we have to say that about 110.000 years ago the modern human began to spread over the Arabian Peninsula and Middle East where they mixed with the Neanderthal.

This new 'group' then immigrated through several travelling waves to south coast of Asia and here they encountered the Denisova human with which they mixed. Subsequently this group again settled over to New Guinea and Australia where they stayed and remained isolated.

If we compare the genetic information of the Denisova fingertip and the genes of the aborigine's, we are able to still find 5% of Denisova-Gene in the aborigine's. It is not only in Australia that we can find the genetic traces of the Denisova-human, all along the route those primordial emigrates took we can – until today – find Denisova genes.

The successful application of bioinformatics constitutes a breakthrough in this area because in the past, it had not been possible to reconstruct the Denisova human evolutionary history – because even nobody knew about him - as direct evidence like bones, completely preserved skeletons or other body parts have never been found. With the advances made in the field of bioinformatics, however, it is now possible to reconstruct and retrace the human evolutionary process based on as little as 30mg of DNA-powder.

Comparison of hereditary information of different organisms

A common example of the application of bioinformatics is to determine the degree of similarity found in two given sequences.

Bioinformatics offers us two possible ways to do this.

1.) Optimal solution: Needleman-Wunsch-Algorithmus which is based on dynamic programming

2.) Efficient solution: BLAST, the most commonly used bioinformatics algorithm in the world

Needleman-Wunsch-Algorithm

The Needleman-Wunsch-Algorithmus calculates the optimal global Alignment or the optimal global Similarity-Score.

It is used to compare two Nucleotide or amino acid sequences – for example compare cheddar and a Swiss cheese. The similarity score is a degree of similarity of 2 sequences; the higher the score the more similarities the sequences have – defined by a given Scoring-Modell.

The algorithm uses the method of dynamic programming which allows any Scoring-Modell as basis for the development (for example the CYK-Algorithm also uses the method of dynamic programming).

For the purpose of further explanation the two sequences below are given. Obviously, we can identify two major kinds of differences.

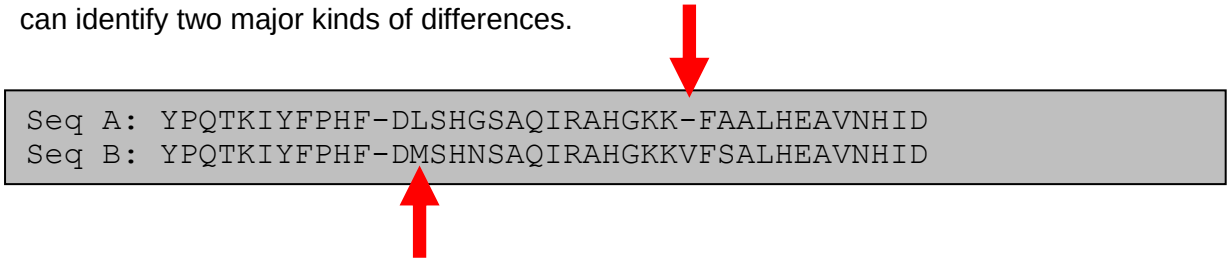

Fig. 5 - Example of comparing sequences

These two kinds are representing the two reasons for differences in amino acids and further more for the evolution. The first one is a mismatch like the letter ‘L’ from Seq A and the ’M‘ from sequence B. In evolution this is a normal mutation which quite frequently happens in the evolution. The other kind of mismatch we can find if we look at the second minus symbol in Seq A. This minus is called “gap”. Gaps are really necessary to find relative strings. If we would not insert the gap in the first sequence the rest of Seq A and Seq B would not match any more. To describe it figuratively, it makes the strings more transparent on the search.

Besides the algorithm it‘s of course possible to align small datasets manually, but this way needs more accuracy.

The gaps refer to a Deletion or an Insertion. In genetics, a deletion is a mutation (a genetic aberration) in which a part of a chromosome or a sequence of DNA is missing. Deletion is the loss of genetic material. Any number of nucleotides can be deleted, from a single base pair to an entire piece of chromosome. Deletions can be caused by errors in chromosomal crossover during meiosis. This causes several serious genetic diseases.

An insertion is the addition of one or more nucleotide base pairs into a DNA sequence. This can often happen in microsatellite regions due to the DNA polymerase slipping.

Insertions can be anywhere in size from one base pair incorrectly inserted into a DNA sequence to a section of one chromosome inserted into another.

On a chromosome level, an insertion refers to the insertion of a larger sequence into a chromosome. This can be due to unequal crossover during meiosis.

After the process of comparison it is possible to ascertain if this is an evolutionary or functional relation.

More information about this topic can be found in “Pairwise Sequence Alignment - Cordula Eichhorn” or “Mehrfachsequenzabgleich - Timo Schmidt”

BLAST

BLAST is a collection of the worldwide most commonly used algorithms for analysis of biological sequences. Blast is used with experimental found DNA or Protein-sequences to look them up in a protein database. The result of the search is a row of local alignments. This means a comparison of parts found in database to the original one. Furthermore, Blast rates the significance of the matches.

Blast is the most efficient solution to compare two strings. Efficient in this context means: economy regarding resources, computing time and memory capacity required to solve a defined problem.

Blast was designed by Stephen Altschul, Warren Gish, David J. Lipman, Webb Miller and Eugene Myers at National Institutes of Health.

The application systematically reviews the sequence entries in the database and singles out the matches.

Blast classifies the input sequence information to different types of words for which it searches in databases like UniProt. The words are compared to database entries thus identifying which sequences indicate the highest degree of similarity. The crucial factor here is the target threshold.

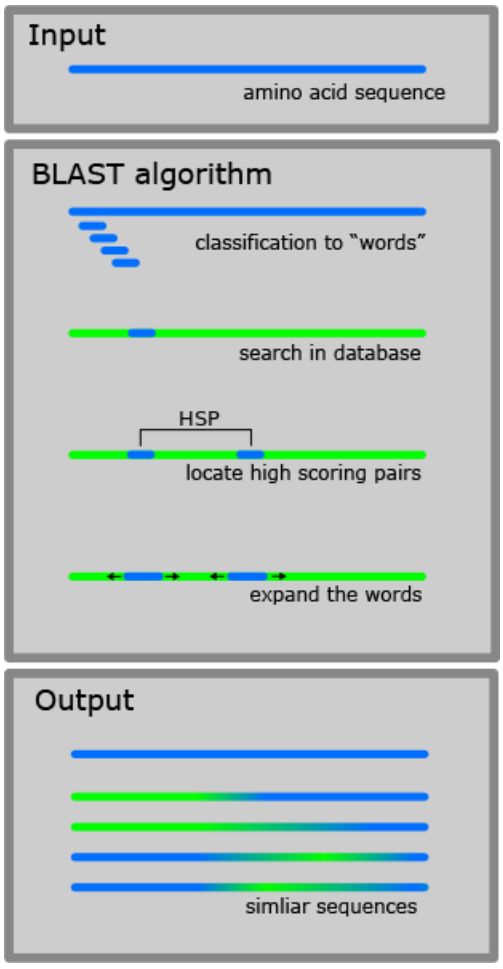

Fig. 6 - BLAST algorithm

After locating the high scoring pairs of the initially searched sequence – those which match 100% – The search tool applies anther function: the words are expanded. ‘Expanding’ here refers to a process of inserting gaps in order to single out those sequences whose codes does not entirely match the initial sequence.

Finally, BLAST determines the degree of similarity and displays a juxtaposition of both sequences. Additionally researchers can perform similarity searches against various other databases like UniProt in order to have access to a greater pool of information.

Reconstruction of evolutionary history

A phylogenetic Tree is a tree that shows the evolutionary relationships between different species. In such a diagram an edge represents the most recent mutual ancestor, whereas the different branches stand for the evolutionary pathway of the respective species. The results received through BLAST can be translated into such a tree by expressing the degree of similarity of homologous sequences.

The edge length represents the estimated time in which the species have separated or the number of mutation during this development.

Sequence analysis is a common method to create phylogenetic trees. One possibility to create these trees is, after analysing the sequences, to explain ancestry by the law parsimony. This principle, also known as the ‘law of succinctness’ or ‘Occam’s razor’ , states that amongst competing hypotheses we should chose the explanation which makes the fewest assumptions. For example, it is possible to explain the relationship between humans and apes with much less explanation than between human and bat.

Alternatively, there is the method of neighbour-joining. In this technique all sequences are compared to an alignment. The ones with the highest similarity are matched together. In the next round they are handled as one species – until a complete tree is created.

Orthologic analysis is another possible way to create phylogenetic trees. The relationship between species is said to be ‘orthologic’ if they descend from a single common ancestor but have evolved in two different species.

If many genomes are known, and if the single genes have successfully been characterised the orthologic genes may be marked in the DNA string. Everything that is not orthologic stems from an insertion or a deletion of a gene – depending on the time order. Now, to create a tree it is necessary to analyse the event that has led to the insertion or deletion of genes.

Functional or evolutionary relationship?

Decisive for creating the phylogenetic tree is the differentiation between functional or evolutionary relationships. The evolutionary relationship in this case – also known as Homology - is called in the biological classification and in the comparative anatomy the fundamental accordance of organs, organ systems, corpus structures, physiological process or behaviour of different Taxa based in a mutual evolutional source.

Homologue attributes are based on mutual ancestor. They are equal regarding their phylogenetic ancestry.

The origin attributes may develop in different directions and used is different functions.

In biology, ‘analogy’ describes the similarity of functions or structures in organs, proteins, genes or behaviour of different taxa that have mutual ancestors who do not have these attributes. This implies that the same attributes are found in different species. However, this similarity is not due to identical genetic information, passed on from a shared ancestor, but is a result of similar demands generated by the species’ living conditions.

Analogue organs are not only similar in their function; in some cases they are also physically or anatomically similar. Nevertheless, if considering the phylogenetic background, they appear to be different and show that they have developed independently from each other.

Analogue organs result from the „System theory of evolution“ through an interplay of convergential selection pressure or development corridors.

Example: Whale – the fin of a whale has the same shape like a fish’s, but phylogenetically the whale’s fin is a extremity of the former terrestrial mammal.

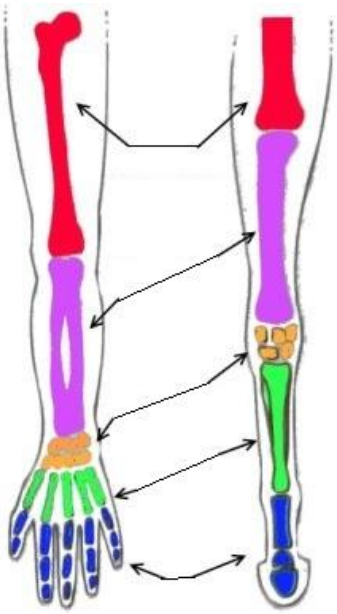

Fig. 7 - evolutionary relation between human and horse

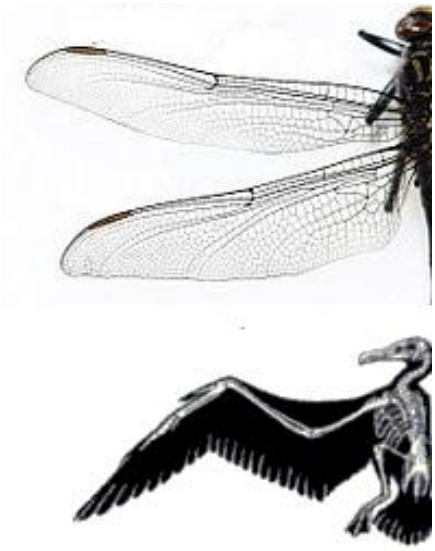

Fig. 8 - functional relationship between insect and a bird

Drug Design

Until the early 19th century Drug discovery relied on a method of trial-and-error. Scientist simply tried combinations of substances and tested them, until they found a ‘cure’ that had the desired effect. Today the development of drugs requires many toxological tests.

For a drug to permit into the market a number of legal constraints have to be met. This includes, notoriously, mandatory animal experiments in the majority of countries. Drugs that are intended for human use additionally have to run through a test in clinical studies involving the below phases:

1.) Short time tests; enquiring into the quality and side effects which is usually done on a small scale of approx. 10-15 probands.

2.) Quality and quantity testes; enquiring into the direct effects and side effects. Also to determine the optimal dosage, approx. 100-500 probands.

3.) Quantity proof of effect for this drug against a placebo – usually conducted with approx. 1.000 probands.

4.)After a preliminary permission has been granted the drug has to be re-checked on long-term studies.

And this was just the test phase and not the development or research phase of the drug!

This means for pharmaceutical company “early fail = cheap fail”.

Bioinformatics has made significant contributions to a more economic pathway of drug design and testing. Currently, there are several companies worldwide, which have already

specialised in electronic tests on mice brains. Electric impulses are sent through the mouse's brain and about 100.000 nerve cells are fitted to a chip at half the size of a credit card.

The electronic signals are like a concert which creates sounds changes on every different substance of drugs or chemicals.

To fit the chips to the mouse's brain and to run these chips the cells from mouse embryos are taken and transferred to these chips. The embryos have to be killed, but the cells could be „copied" on up to 40 Chips.

Personalised Medicine / healing of diseases

In the past drugs for treating individuals with a low clotting factor was based on human blood. This, of course, is an approach susceptible to the danger of transmitting AIDS. An approach based on genetic information could certainly avoid this threat.

Incurable diseases like Multiple Sclerosis or Parkinson's disease could be detected at the embryonic stage of a human being's development and consequently prenatal actions could be taken.

Having said this, our discussion has now been directed to the issue of how one ought to assess the potential and desirability of personalised medicine.

Summary of possibilities

The summary presented below is an attempt to reiterate and establish the following points:

Reconstruction and "re-creation" of creatures through their smallest item:
Decoding the entirety of an organism's genetic information with the help of a sample from a DNA strings.

Comparison of hereditary information of different organisms:
For analytical and historical reasons with BLAST-Algorithm or Needleman-Wunsch-Algorithm.

Construction of evolutionary trees:
Visualising evolutionary history and enables predictions about future development.

Questions about functional or evolutionary relationships:
To set the focus and start defining ranges for species, finally determine the grade of relation.

Drug Design:
Facilitates and shortens the test-phase considerably since promising pathways are more easily detected.

Where should be limits of the bioinformatics?

A steak from a test tube and the cultured liver

Scientifics in the USA and Britain are now working on artificially grown meat which would make factory farming, with all its negative side effects, unnecessary. They view their research as a contribution to the worldwide sacristy of food. Yet they have to fight against the disgust factor that is attached to the artificially produced meat.

Certainly, thinking of a Frankenstein-Cutlet produced in a bio reactor sounds not really delicious and is far from appetising. Nevertheless, one only has to remember the problems of the growing world population - today already there is no room for rearing animals in Singapore or New York – and lab-grown meat turns into an increasingly appealing alternative.

The research project was originally initiated by the NASA, searching for a suitable protein supply on long-time flights to different planets. The flight to Mars takes approx. about 6 months and on board is no spare space for slaughtered cattle.

Not only does this alternative method of food production answer the needs of a growing world population, but it is also a way of achieving a reduction in the CO2 emission. This means, instead of having huge herds standing on grassland and creating masses of CO2, it is certainly a more efficient alternative to create more food in laboratories at the same time also offering new possibilities to fight climate change.

If we think of solving all of these problems through creating artificial ribs, backs and necks in test tubes, we have to consider some ethical aspects: when shall we start to talk about a ‘meaningful life' in contrast to 'mere existence‘. To find an answer we have to decide at what point we want to regard a creature as having dignity and when a collection of muscles reaches the status of a ‘genuinely living’ and, above all feeling, creature.

This approach assumes a defined level of self- awareness of the creature. From the religious point of view the “soul” would be the significant factor requiring us to treat this creature with respect. The soul, then, confers a status of dignity; the creature thus has a undeniable value that is usually interpreted as the individual’s basic rights.

However, on the other hand, organ transplants really important and needed. The problem is just the low number of available suitable organ donations. Furthermore, it is extremely difficult to predict with 100% certainty that the donated organ will not be rejected as an alien by the patient’s body.

An available solution would be to grow artificially created organs, based on the cells of the invalid person. The development of this kind of treatment is unfortunately not as far advanced as it would to be in order to be accepted as part of conventional treatment. The first artificially created urinary bladders haven been transplanted to 7 people already. This is more or less easier than other organs like a heart, kidney or a lung, because the urinary bladder does not participate in the metabolism. Furthermore, the development of urinary bladder costs about several thousand dollars.

But the idea of having the possibility to create all kinds of lab-grown organs, just lying next to each other in test tubes leads us to the question of when this collection of organs becomes a working system? And which requirements are necessary in order to call this collections of organs a living being?

When to call it „a living creature“

The question of how and when to define about a creature a ‘truly living’ has always been discussed in many different ways and the debate has been approached from several different positions. This chapter is intended to give an overview of some of the prevalent opinions regarding the definition of “living being” so that readers may come to their own conclusion.

Utilitarianism is a theory in normative ethics and has many in different subbranches. According to this school of thought, the proper course of action is the one that maximizes overall happiness. The amount of pain in the world should be kept to a minimum.

In Utilitarianism, the bases for living are feelings which are founded on impulses of an nervous systems. This means as far as a creature is able to handle impulses of their nervous system they have to be a living creature which must be handled with dignity and aim of preventing pain from it.

Does the accumulation of organs and a functioning nervous system suffice to regard an organism as 'truly living’?

In the renowned Stanford Encyclopaedia of Philosophy SEP, a renowned philosophy Georg & Gomez-Lobo noted in the entry on the ethics of embryonic stem cell research that:

Human embryos are said to be “whole living member[s] of the species homo sapiens … [which] possess the epigenetic primordia for self-directed growth into adulthood, with their determinateness and identity fully intact.

This statement now is explicitly concerned with human embryos only, but if applied to other creatures this argument describes the self-determined will to live and grow into adulthood or, in the case non-human animals to reach maturity, as the decisive factor. The American moral philosopher Christine Korsgaard (Harvard) considers the capacity to make conscious, self-reflexive decisions about actions as that which makes us a person; thus distinctively human. Following this line of thought, research involving embryonic stem cell research is morally impermissible because the required experiments disrespects the wishes of a potentially autonomous agent.

However, at this proponents of stem cell research and related fields voice their criticism: If the dignity of a person is derived from his or her status as an autonomous and self-reflective agent, then this does clearly not apply to the human embryo and even less to an accumulation of lab-grown organs. What they lack is an awareness of themselves as agents acting not only in the world but also on the world. If this ability cannot be found in an organism, we can hardly speak of this organism as being (self-) conscious, consequently, it is not the dignity of a living being that is being violated by the practise of stem cell research.

Nevertheless, even if it is agreed that this position cannot deliver a compelling argument it does not follow that there are no compelling arguments at all. A German politician once said that “if life is a present from God, then we are not allowed to pack it again”. This statement is definitely influenced by a religious way of thinking. Compared to the above argument we can observe a shift in emphasis here: Whereas the first position offered a ‘functional’ explanation for the respect we feel for a fellow (human) being, this second position assumes that there is an unconditional value in creation per se. Meddling or interfering in this god-given plan would question the divine authority and is therefore to be rejected.

From a scientific point of view, as has been taken by NASA, life is to be thought of as “a chemical system, able to do Darwin’s evolution.” So here we have life in the original biological sense, i.e. the Latin ‘bios’ which means ‘to live’. Should we agree with argument and regard all biological life as being of same value? Alternatively, is it possible to find a definition that could allow us to distinguish between living in the biological sense, that is, as semi-independent existence of an organism in contradistinction to ’living’ as a specific mode of experience and consciousness?

What kinds of limitations are required? – Concluding remarks

Is it desirable to continue work in this field? Given the possibilities that are now available to us thanks to bioinformatics it is urgent question ask if we actually want to remove these barriers and control organ supply.

As we have seen, what we can do is to treat disease more efficiently, alleviate pain and bring back quality of life for a patient simply because bioinformatics enabled personalised medicine. Some scientist assume that in future it should even be possible to derive adult stem cells from the patient himself which means that ethically difficult question of embryonic stem cells can be avoided.

My concern is that medical progress is not considered in it’s proper context, stem cell research with all its attendant possibilities may conceal the vulnerability of life. Increased success rates in treatment of congenital abnormalities makes life with disease a rare case and I fear that this means we will take a life free from these adversities for granted and are not willing to put up with them should they occur. This, again, is a problem that is more concerned with the attitude behind the acting agent. So the locus for discussion is, in my opinion, not the direct question of ‘no or go’ but an informative discussion should be sought on an entirely different plane.

Austro-British Philosopher Karl Popper remarked that the value of life lies in its limits.

Sources

- Mount: Bioinformatics: Sequence and Genom Analysis, Cold Spring Harbour Laboratory Press 2004 (2. Auflage),
- Zvelebil / Baum: Understanding Bioinformatics, Taylor & Francis 2007
- http://plato.stanford.edu/entries/cloning/
- http://plato.stanford.edu/entries/stem-cells/
- Erwin Schrödinger: Was ist Leben? Piper Verlag, München 2001
- http://www.iep.utm.edu/Ethics/
- http://bioethics.georgetown.edu/
- http://rs-ebern.de/httpdocs/Lehrer/Fachschaften/KathRel/LDL/Wann%20beginnt%20menschliches%20Leben%20.pdf
- http://philpapers.org/browse/meta-ethics
- http://www.templeton.org/reason/Essays/korsgaard.pdf
- Bild der Wissenschaft 05/2012
- http://www.medicom.de/gesundheits-ratgeber/news/vom-traum-organe-im-reagenzglas-zu-zuechten
- http://www.ratschlag24.com/organzchtung-vorstufe-des-auges-im-reagenzglas-entwickelt-_107623
- http://www.manager-magazin.de/lifestyle/genuss/0,2828,749411,00.html
- http://www.bbc.co.uk/news/magazine-17113214