UPTEC X 02 033
JUN 2002

ISSN 1401-2138

NICLAS RIML

The use of bioinformatics within academia and small pharmaceutical companies in Sweden

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 02 033 | **Date of issue 2002-06**

Author

Niclas Riml

Title (English)

The use of bioinformatics within academia and small pharmaceutical companies in Sweden

Title (Swedish)

Abstract

In recent years, bioinformatics has come to play an important role in basic and pharmaceutical research. An investigation of the use of bioinformatics within academia and pharmaceutical companies by Bosson & Riml (2002) displays not only the usage, but also determines the maturity of bioinformatics as a technology and presents critical factors needed for further development. This report is a part of that study and focuses on the use of bioinformatics within one academic department and two small pharmaceutical companies. The main results were that there exist differences in the usage and type of tools primarily used.

Keywords:

Bioinformatics, Technology life cycle model, Market maturity, Scenario analysis

Supervisors

Magnus Isaksson, Patrik Nylander
Accenture, Stockholm

Examiner

Per Kraulis
Stockholm Bioinformatics Center

Project name	Sponsors
	Accenture
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **99**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

The use of bioinformatics within academia and small pharmaceutical companies in Sweden

Niclas Riml

Sammanfattning

Bioinformatiken är ett område som under de senaste åren har kommit att påverka den biologiska forskningen allt mer. Bioinformatik är ett samlingsnamn för metoder där datorer och mjukvara används för att lösa biologiska frågeställningar. Behovet av att använda datorer inom biologin har ökat, bland annat beroende på att det idag produceras stora mängder biologisk data samt att de aktuella biologiska frågeställningarna blivit mer avancerade.

Detta examensarbete har utrett användandet av bioinformatik hos två mindre läkemedelsföretag samt en akademisk institution. Arbetet är en del av en större undersökning som finns bifogad i appendix I. Syftet med detta arbete var att identifiera förhållanden som är karakteristiska för de tre organisationernas bioinformatikanvändande. Undersökningsmetoden som användes var intervjuer och materialet presenteras i tre fallstudier. De resultat som framkom var att användandet av bioinformatik skiljer sig mellan företagen och institutionen. Den institution som undersökt använde ett bredare spektrum av bioinformatiska verktyg. En annan skillnad var typen av bioinformatiska verktyg som i huvudsak användes. Institutionen använde freeware i stor utsträckning medan företagen framförallt använde kommersiell programvara.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala universitet juni 2002

TABLE OF CONTENTS

1 Introduction

During the last decades, there has been an enormous development within both biology and computer science creating new possibilities for biological research. The introduction of computer science into the biology area has made it possible to foresee the folding of proteins, understand the regulation of cell activity and more, taking research to a new level. The use of software for solving biological issues is called bioinformatics.

One of the largest areas of application for bioinformatics is within pharmaceutical research. The pharmaceutical industry is characterized by high development costs and extensive lead-times. The cost and time of developing one new drug today averages $500 million and 12-15 years[1]. Furthermore, each prescribed drug only generates annual revenue of approximately $400 million[2]. This implies a high pressure for drug companies on putting successful products on the market in order to stay alive, however only one out of every five thousand initiated research project result in a market drug[3].

In bioinformatics, the pharmaceutical industry sees a possibility of shortening the lead-time and at the same time producing better drugs. Bioinformatics can be applied in all first four parts of the drug development pipeline (see figure 1.1)

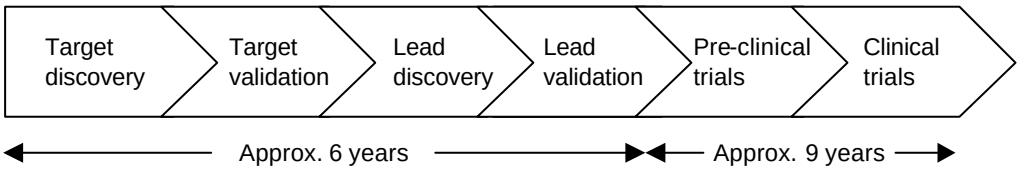

Figure 1.1 Drug development pipeline **Source: Accenture, internal material**

Bioinformatics methods were first developed within academic research and most of the improvements done within this field have so far been made in the academic world. The area emerged around the mid-1970s when the first automated protein and DNA sequencing technologies became available creating a need for a computer-aided way to gather and analyze data. However, it was first in the mid-1990s when the Internet opened the possibility

[1] *Why Do Prescription Drugs…* (2000). Pharmaceutical Research and Manufactures of America, p. 2

[2] *Life Science Informatics* (2001). UBS Warburg, p. 9

[3] *Why Do Prescription Drugs…* (2000). Pharmaceutical Research and Manufactures of America, p. 2

for an easy way of sharing software and databases (i.e. knowledge) between research teams all over the world that the field really took off.[4]

1.1 Outlining the Report

The information presented in this report is based on a study performed by Bosson and Riml (2002) that investigated the usage of bioinformatics within Swedish academia and pharmaceutical companies. The empirical method for this study was case studies in the form of interviews. This study, which was produced in collaboration with Accenture, is presented in Appendix I.

This report describes case studies made with the department of Genetics and Pathology at Uppsala University and two small pharmaceutical companies Active Biotech AB and Medivir AB. It also summarizes some results specific for these study objects. To get a complete picture of the study it is recommended to read Appendix I.

1.2 Purpose and Questions of Issue

The purpose of this report is to elucidate certain conditions prevailing for small pharmaceutical companies in relation to an academic department. The questions that are addressed in this report concern bioinformatics organization and strategy, bioinformatics tools and databases, problems with bioinformatics, and future possibilities for the field.

[4] Persidis, Aris, *Nature Biotechnology* (1999), Vol. 17, p. 828

2 Case Studies

Each section consists of five parts reflecting the material achieved during the interviews. This is done for easy comparison between the objects of study. The parts are Bioinformatics Organization and Strategy, Bioinformatics Tools and Databases, Problems with Bioinformatics, Bioinformatics in the Future, and Case Summary.

2.1 Department of Genetics and Pathology, Uppsala University

The facts presented concerning the Department of Genetics and Pathology is based on interviews held with Kalle Näslund, Ph.D. student, and Fredrik Granberg, Ph.D. student, at the Department of Genetics and Pathology 2002-03-13. Additional information was found on the department's homepage, www.genpat.uu.se.

The Department of Genetics and Pathology consists of three units, medical genetics, clinical genetics, and pathology. The unit for medical genetics conducts research concerning the structure and function of genes of higher organisms. The focus lies on heredity, organization, and regulation of the genome. The unit for clinical genetics studies the human genome and the use of clinical diagnostics. Finally, the unit for pathology investigates the underlying causes that make diseases appear. The overall objective of the department is to bring basic and clinical research together to produce a better knowledge of mechanisms causing diseases.

The department uses bioinformatics in a wide range of areas including sequence analysis, transcriptomics, and structural genomics.

2.1.1 Bioinformatics Organization and Strategy

The Department of Genetics and Pathology does not have an overall strategy for the use of bioinformatics. All groups within the department work fairly independently from each other and there are no formally formed inter-group co-operations to support bioinformatics utilization. Knowledge is exchanged when needed over informal channels. Further, the department does not have anyone overall responsible for collecting and storing the information produced within various projects. There have been discussions concerning such a position, but nothing concrete has yet happened.

Collaborations with commercial companies are rare within the department. The existing ones often concern development of company products, where the department collaborates with the company to further develop biotech products. An example has been a collaboration with

Pyrosequencing. On the academic level, there are several co-operations with other departments, but none merely concerning bioinformatics.

2.1.2 Bioinformatics Tools and Databases

The software most often used at the Department of Genetics and Pathology is freeware. The main reasons are money and attitude. For many applications, there exists high quality commercial software solutions, but money limits the possibilities to purchase these programs since they often are too expensive. In a worst-case scenario, this implies that a project must be prematurely closed since needed software investments cannot be afforded. Attitude further restricts software investments since conservatism sometimes rule in the laboratories. It is often considered better to invest in a machine rather than a computer program since it is easier for the people in the laboratories to directly see in what way a machine could affect their daily work.

However, it is not always that freeware dominates over commercial programs. For some standard research applications, bioinformatics software, both freeware and commercial, has been used for a relatively long time. Within some of these areas, there are several products available. Here competition over the years has produced quality commercial software to a relatively low cost, making purchasing more common.

There are few local databases at the department as of yet. The intention is to store more data locally. One project is to build an in-house nucleotide sequence database with the objective to lessen the time spent on conducting queries on distant databases. The local database would rely on available services, at for example NCBI, to automatically update the local replica with new information.

2.1.3 Problems with Bioinformatics

Annotation is seen as one of the larger problems for bioinformatics. Different standards for naming sequences have been used throughout the years, creating inconsistent data structures in sequence databases. This has resulted in multiple names for the same sequences making homology searches more difficult. Annotation standards does exist, but has not consequently been used. The problem is greater for older data.

The annotation problem has created a need for databases with better search engines, a uniform classification system and correctly sequenced data. This is manifested through the

department's need of constructing in-house databases for specific research projects in order to achieve the characteristics just mentioned.

To be able to construct and develop new databases and search algorithms the cross-competence gap between biology and computer science needs to be lessened. The department therefore has a need for people with both biological knowledge and software development competence.

2.1.4 Bioinformatics in the Future

The amount of biological data produced will demand better and cheaper software products in the future. Also, a paradigm shift is foreseen as researchers discover the full potential of bioinformatics, radically changing the way biological research is performed. The paradigm shift will imply a turn in focus from laboratory work towards more computer-aided research.

In the future two categories of bioinformatics users within the academia are predicted to exist. For standard applications users, more programs will be available to a lesser cost implying that new software, to a greater extent, will be bought. On the other hand, biologists with the need for specific advanced applications will continue to develop software themselves. This relates to the researchers need for credibility when publicizing and for their understanding of results.

2.1.5 Case Summary

Below is a short summary of the case presented. The summary describes aspects from the first two areas: Bioinformatics Organization and Strategy, and Bioinformatics Tools and Databases.

Table 2.1 Summary of empirical data for the department of Genetics and Pathology

Description of study object	Academic department
Organizational form	No group
Bioinformatics collaborations	
With companies	None
With academia	None
Bioinformatics tools	
In-house developed	Some
Freeware	Mostly
Commercial	Some
Local replicas of large databases	None (some planned in the future)

2.2 Active Biotech AB

The facts presented concerning Active Biotech AB is based on interviews held with Mats Hansen, Knowledge Manager and Ph.D., and Ann-Sofie Sjögren, Applied Bioinformatics and Ph.D., at Active Biotech AB 2002-04-03. Additional information was found on the company's homepage, www.activebiotech.com.

Active Biotech AB was founded in 1996 when Pharmacia took a decision to no longer convey research within the areas of focus at Active Biotech AB. Active Biotech AB focuses its research on three areas of disease; autoimmunity/inflammatory, cancer, and infection. The goal is to be a leading player in these areas through global co-operation agreements and organic growth on selected markets. The company group has approximately 350 employees.

The research at Active Biotech AB foremost use methods within proteomics in the target identification stage and consequently this area accounts for most of the use of bioinformatics tools. Moreover, bioinformatics tools are also used in cheminformatics and structural genomics.

2.2.1 Bioinformatics Organization and Strategy

Active Biotech does not have a formal bioinformatics department. Actually, it was first in 1999 that a discussion concerning bioinformatics was raised within the company. The responsibility for coordinating bioinformatics within Active Biotech lies under the knowledge management manager in the Scientific Affairs group. Within the same group, the company has created a position called Applied Bioinformatics. The idea is to support the different research projects by helping them apply bioinformatics tools where possible. In addition to the Applied Bioinformatics position, there exists an informal network with the responsibility to raise an awareness of bioinformatics within Active Biotech. The intentions are also to use the network to solve bioinformatics problems that arise, both practical and strategic.

One of the main objectives with bioinformatics at Active Biotech is for it to be a tool that makes it possible to take strategic decisions about research projects at an earlier stage, before heavy investments are done.

Active Biotech has a number of collaborations with academic departments. None of these, however, concern bioinformatics, mainly because the company does not strive towards a continuous development of bioinformatics tools. If certain bioinformatics problems arise that cannot be handled in-house Active Biotech would sooner hire consultants to solve the problems than enter collaborations with academic departments.

2.2.2 Bioinformatics Tools and Databases

The tools used at Active Biotech are mostly commercial. The largest investment in software has been the purchase of the Genomics Computer Group (GCG) system. As supplement, some freeware and software accessible via the web are used. Almost no software is developed in-house since the company does not have the knowledge or resources to accomplish this.

The large international databases that are of importance for the research at Active Biotech are copied and stored locally to prevent competitors to monitor what kind of searches are performed. The local copies are updated through license agreements where needed.

2.2.3 Problems with Bioinformatics

One problem seen at Active Biotech is that the available commercial software today is not able to handle all the questions asked by researchers. There is definitely room for improvement on the software development side within bioinformatics.

According to Active Biotech, a common misunderstanding is the general view that bioinformatics is the solution to all problems. This overrates the expectations on bioinformatics. At Active Biotech they emphasizes the importance of realizing that bioinformatics is just a part of the research process, a tool, and not the solution to all problems.

Active Biotech has not seen a lack of competence within bioinformatics, but on the other hand, they have not tried to hire a great number of bioinformaticians.

2.2.4 Bioinformatics in the Future

Active Biotech will probably not start developing its own software in the future. Instead, the company will look to keep their in-house systems well structured so that introduction of new tools, developed elsewhere, will be facilitated. In order to help this process the intention is to bring the IT-department closer to the research process.

A trend believed to become more obvious in the future is that more and more of the information today freely available in databases through the Internet will become private and accessible only through purchasing licenses.

The future will probably reduce the belief that bioinformatics holds the solution to all problems. Also, it will be important to reduce the competence-gap between computer scientists and biologists. Therefore, education is considered to hold the key for future success.

2.2.5 Case Summary

Below is a short summary of the case presented.

Table 2.2 Summary of Empirical Data for Active Biotech

Description of study object	Small company
Organizational form	Informal network
Bioinformatics collaborations	
With companies	None
With academia	None
Bioinformatics tools	
In-house developed	None
Freeware	Some
Commercial	Mostly
Local replicas of large databases	Some

2.3 Medivir AB

The facts presented concerning Medivir AB is based on interviews held with Björn Ursing (Ph.D.), Bioinformatics Manager, Jonas Ekstrand (Dr.Med.Sc), Associate Director, and Peter Lind, Cheminformatics Manager, at Medivir AB 2002-03-22. A complementary interview was held via telephone with Katarina Jansson, Research Scientist Computational Chemistry and Cheminformatics, 2002-03-26. Additional information was found on the company's homepage, www.medivir.se.

Medivir AB is a pharmaceutical R&D company that focuses its research on infectious diseases and autoimmune disorders. Medivir AB develops compounds into new pharmaceuticals based on proteases and polymerases as target enzymes. The company is originally a spin-off from Astra and has approximately 170 employees.

Medivir's research focuses on developing compounds active against different proteases and polymerises, and therefore the bioinformatics tools used at Medivir mainly lie within cheminformatics and structural genomics. To a lesser extent, bioinformatics is used within the areas of sequence analysis and transcriptomics.

2.3.1 Bioinformatics Organization and Strategy

Medivir has not yet developed a clear strategy for how bioinformatics should be brought into use within the organization. To a great extent, this has to do with how research has been performed at the company and within what areas. Recently, however, new needs have arisen from the company's decision to widen its' research. The new scope includes making target identification and validation a part of the research process thus increasing the need for bioinformatics. As a result, the company has hired a person responsible for facilitating and structuring the incorporation of bioinformatics into the organization. So far the new position has resulted in the development of an internal web service that aim at making the staff use the most appropriate bioinformatics tools. The website contain links to efficient tools available online together with information about the links contents.

In the areas of cheminformatics and structural genomics, in contrast to other bioinformatics areas, Medivir has longer experience and a more structured approach. The group responsible for these areas is a separate department called Cheminformatics and Computational Chemistry. The department's main task is to manage the chemical compound information stored in cheminformatics databases and to accelerate the screening process by using information about the structure of the target protein.

Medivir has never had any formal collaboration with the academia or companies dealing with bioinformatics. However, an informal connection with CGB exists through a part-time employee. There exist collaborations with academic departments concerning structure determination of proteins, but Medivir has recently decided to move structure determination in-house.

2.3.2 Bioinformatics Tools and Databases

Medivir holds structure as well as cheminformatics databases in-house. The in-house databases are built on commercial software that has been further developed within the company to fit the specific needs at Medivir. The information stored in the cheminformatics databases is to a certain extent bought from companies specialized in selling data for specific groups of leads. In addition, Medivir has online access to international structure databases. The company does not keep local gene sequence databases though, since the needs for this up till now has been limited.

Medivir uses commercial software in the structural genomics area. Supplementing this are some locally developed structural genomics software tools.

2.3.3 Problems with Bioinformatics

A problem seen at Medivir with many bioinformatics tools available on the market is that the solutions offered are seldom innovative. The tools are often based on miscellaneous freeware developed elsewhere, put together and presented via a new interface. In addition to this, the tools are often expensive.

Medivir sees a problem in the difficulty of creating functional structure databases for easy structure comparison. It is not possible to adequately represent 3D-structures with data strings. This makes comparison between structures difficult and better solutions in this area is needed.

More generally, Medivir sees a problem with the quality of the information stored in the available databases used for biological research. This creates problems and the need for verification of test results. In addition, Sweden lacks competent bioinformaticians to fill the needs of both the industry and the academia.

2.3.4 Bioinformatics in the Future

Medivir believe that the company will start developing some smaller tools in the coming future. In relations to Medivirs new line of research, new databases have to be designed and implemented. They will probably not purchase software but utilize freeware, mainly because of the high costs associated with commercial products.

An expectation at Medivir is that the wider incorporation of bioinformatics in the research process will increase the quality of the research conducted, through making it possible to discard unfit target candidates at an earlier stage.

2.3.5 Case Summary

Below is a short summary of the case presented.

Table 2.3 Summary of empirical data for Medivir

Description of study object	Small company
Organizational form	No formal group
Bioinformatics collaborations With companies With academia	 None Informal
Bioinformatics tools In-house developed Freeware Commercial	 Some Some Mostly
Local replicas of large databases	None

2.4 Presentation of Case Summaries

To summarize the case studies and to make it easier to get an overview of the material presented in this chapter, table 2.4 displays the information from all three Case Summary chapters at the end of each case.

Table 2.4 Summary of all the Case Summary chapters

Name	**Department of Genetics and Pathology**	**Active Biotech AB**	**Medivir AB**
Category	Academic department	Small company	Small company
Organization	No group	Informal network	No group
Collaborations Companies Academia	 None None	 None None	 None Informal
Tools In-house Freeware Commercial	 Some Mostly Some	 None Some Mostly	 Some Some Mostly
Replicas of databases	None	Some	None

3 Conclusions and Discussion

The case studies in this report have clearly shown that there exist differences between academic departments and small pharmaceutical companies usage of bioinformatics.

At the academic department, bioinformatics is a more commonly used tool compared to the small companies since the department performs research within more areas. The usage is not steered and everybody is free to personally decide what tools are to be used. The small companies on the other hand have a limited need for bioinformatics and therefore also a limited need for a specific bioinformatics group responsible for introducing bioinformatics.

None of the study objects have outspoken collaborations concerning bioinformatics. This has in part to do with the limited need for bioinformatics within the organizations, but can also be seen as a restraining factor for bioinformatics as collaborations are needed for further progress.

The use of bioinformatics tool differs between the study objects. Within the academic department, most tools used are freeware complemented by some in-house developed tools as well as commercial. The companies on the other hand mostly use commercial software.

The problems found in the study are several. There is said to exist a lack of competent bioinformaticians in Sweden. The products are often too expensive and of low quality. Further, general problems concerning common standards for annotation and software are identified. For further development of Swedish bioinformatics it is of importance that these issues are dealt with.

References

Articles
Persidis, Aris, (1999) "Bioinformatics", *Nature Biotechnology*, vol. 17, pp.828-830

Printed reports
Life Science Informatics (2001). UBS Warburg
Bioinformatics in Sweden (2002). Bosson, Oskar & Riml, Niclas, M. Sc. Degree Project, performed at Accenture
Why Do Prescription Drugs Cost So Much… (2000). Pharmaceutical Research and Manufactures of America.

Interviews
Ekstrand, Jonas, *(Dr.Med.Sc.) Associate Director*, Medivir AB, 2002-03-22
Granberg, Fredrik, *Ph.D Student*, Department of Genetics and Pathology, Uppsala University, 2002-03-13
Hansen, Mats, *(Ph.D.) Knowledge Manager*, Active Biotech AB, 2002-04-03
Jansson, Katarina, *Research Scientist Computational Chemistry and Cheminformatics*, Medivir AB, (telephone) 2002-03-26
Lind, Peter, *Cheminformatics Manager*, Medivir AB, 2002-03-22
Näslund, Kalle, *Ph.D Student*, Department of Genetics and Pathology, Uppsala University, 2002-03-13
Sjögren, Ann-Sofie, *(Ph.D.) Applied Bioinformatics*, Active Biotech AB, 2002-04-03)
Ursing, Björn, *(Ph.D.) Bioinformatics Manager*, Medivir AB, 2002-03-22

Homepages
www.activebiotech.com, 2002-04-01
www.genpat.uu.se, 2002-03-18
www.medivir.se, 2002-03-20

Bioinformatics in Sweden

- a Study of the Present and Needs for Future Development

Master's degree project

Authors: Oskar Bosson
Niclas Riml

Supervisors: Magnus Isaksson
Patrik Nylander

Executive Summary

The former president of the Wallenberg foundation, Jan S Nilsson, describes Swedish bioinformatics as having a "…great potential for development" and being "…important for maintaining Sweden's competitiveness as a research nation and keeping pharmaceutical companies in Sweden".[5]

This report describes the bioinformatics technology and the Swedish bioinformatics industry.
An initial background study of the industry gave an opening for a SWOT-analysis, which led to identification of positive and negative factors affecting the future of the Swedish industry. The strengths identified were the well-developed relations between the academia and start-ups, product development through partnerships, and wide knowledge within both computer science and biology. The weaknesses found were the currently small market, the demand for a rapid expansion, and the existing cross-competence gap between computer science and biology. Further, the opportunities seen were the large market potential and strong academic research generating new competence and start-ups. Finally, the threats recognized were the fact that Europe is lagging behind the industry development in the US and the lack of database management knowledge in Sweden.

The main material was gathered through case studies, performed to view how bioinformatics is used within the Swedish academia and pharmaceutical companies and to enable an analysis of the maturity of the technology as well as possible future scenarios. The Technology Life Cycle Model, as described by Afuah & Tucci[6], was used for the maturity analysis. The material shows an existing uncertainty of roles, products, and standards on the market. Further, the products on the market do not always show high quality, the costs and prices connected to bioinformatics are high, the users are mainly lead or high income users, and there exists competition for resources between bioinformatics and older technologies. These characteristics point towards the technology residing in the earliest phase of its development, the fluid phase, although some indications suggest a commencing transition into the next phase, the transitional phase.

In the light of these facts, and with support from the model, four critical factors important for further development of the technology were determined (see Table 1).

[5] Nilsson, Jan S., Interview (2002)

[6] Afuah, Allan & Tucci, Christopher L., *Internet Business Models and Strategies* (2000), pp. 73-75

Table 1 Identified critical factors needed for further development of the bioinformatics technology

Identified critical factors
• Need for people with the knowledge to bridge the cross-competence gap between computer science and biology • Establishment of bioinformatics collaborations and forums to lessen market and product uncertainties • Development of international standards such as annotation and data format standards • Clarification of how bioinformatics add value to the research process to motivate investments

To face the challenges put forward by the critical factors we propose actions that should be taken by the actors of the bioinformatics sphere for a favorable development (see Table 2).

Table 2 Proposed actions for actors connected to the Swedish bioinformatics sphere for a favorable future development

Actor	Short-term actions	Long-term actions
Academia	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Review their bioinformatics strategy and seek collaborations within university structures • Strive towards establishing international standards through collaborations	• Create good educational programs and secure a high level of competence • Seek memberships in forums
Bioinformatics companies	• Hire foreign competence • Actively seek collaborations with pharmaceutical companies and the academia • Strive towards establishing international standards through collaborations • More explicitly demonstrate how bioinformatics adds value	• Initiate forums with pharmaceutical companies
Foundations		• Keep supporting non-government funded research areas
Government	• Educate available computer scientists in biology	• Assign sufficient funds for education and research
Pharmaceutical companies	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Define if bioinformatics is a core business activity or not. If not, facilitate spin-offs and collaborations with bioinformatics companies. • Actively seek collaborations with the academia • Strive towards establishing international standards through collaborations	• Initiate forums with bioinformatics companies

We have further concluded that the entry of a third party on the bioinformatics market could help catalyze the technology's development. A third party could connect the right market actors thereby creating bioinformatics solutions with a wider scope.

TABLE OF CONTENTS

1 Introduction

When Watson and Crick discovered the genetic code, i.e. DNA, they must have realized that their findings would revolutionize science. But could they, however, foresee the enormous impact it would have on drug development only a few decades later. The knowledge of how life functions at the lowest level, together with the introduction of information technology into the life science area, has made it possible to foresee the folding of proteins, understand the regulation of cell activity and more, taking drug developing to a new level.

1.1 Aspects on Pharmaceutical Research

Illness has sadly always been one of the basic facts of life. Ways of trying to cure illness have existed for just as long. This urge for finding cures has throughout the centuries developed into an industry totally focused on drug development - the pharmaceutical industry. The pharmaceutical companies have been very successful in delivering drugs for numerous diseases. The enormous progress made within the industry, especially within the last decades, has created expectations from society, demanding drugs and vaccines for all possible diseases.

These expectations together with a huge market potential have driven pharmaceutical companies to invest billions of dollars every year on research and development. The cost and time of developing one new drug today averages SEK 5000 million and 12-15 years[7], however only one out of every five thousand initiated research projects result in a market drug[8]. Furthermore, each approved prescribed drug only generates annual revenue of approximately SEK 4000 million[9]. This implies a high pressure for drug companies on putting successful products on the market in order to stay alive.

It is understood that if drug companies could reduce the time spent on research and development of new drugs, and at the same time increase the number of successful projects, there would be a lot of money to be saved. If the average time of developing a new drug could be cut down by one third it would mean additional revenue of approximately SEK 20000 million, per produced prescription drug[10], this given that time saved in the

[7] *Why Do Prescription Drugs…*, Pharmaceutical Research and Manufactures of America, (2000) p. 2

[8] *ibid.*

[9] *Life Science Informatics,* UBS Warburg, (2001), p. 9

[10] 15 years · $^1/_3$ · SEK 4000 = SEK 20000

development process for a drug implies longer time on the market protected by patent. Further, more efficient research would result in cures for various diseases reaching the market faster.

The research and development process for developing a new drug can roughly be divided into six steps called target discovery, target validation, lead discovery, lead validation, pre-clinical trials, and clinical trials (see figure 1.1). Target being what a drug acts on, and lead the active substance of a possible drug.

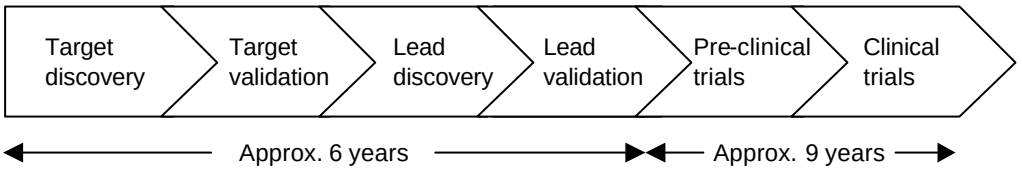

Figure 1.1 Drug development pipeline. **Source: Accenture, internal material**

Traditionally one of the bottlenecks within the first four steps of pharmaceutical research has been to efficiently identify targets and screen for leads, and to obtain relevant information about these compounds. In recent years, new technologies have made it possible to quickly obtain useful information about large amounts of candidates for targets and leads. This has shifted the bottleneck to managing, structuring, and analyzing the produced data.

The problem discussed above could be tackled in several ways. For example, a computer program could be designed to help scientists find disease specific targets and thereby reducing the amount of time spent on target discovery and validation. Another computer program could then help reveal the shape of this target, making it possible to produce leads so specific to this target, that one could almost exclude the possibility of side effects. Surely, these kinds of programs would greatly lessen the time spent on pre-clinical research and trials. Presumably, it would also diminish the amount of non-useful leads put up for clinical studies since the substance would already have been virtually tested in a computer for possible affinity to its target and potential side effects. This scenario is not just a high-flown plan, but a present-day reality.

1.2 Bioinformatics

The term informatics is used to describe methods utilizing the incomparable power of computers and software to analyze data material. The applications supported by informatics

span a wide range of areas, such as image analysis and data mining. Within the life science area, informatics methods are often referred to as bioinformatics. In the journal Nature Biotechnology, Persidis defines bioinformatics as:

> *"the computer-assisted data management discipline that helps us gather, analyze, and represent this [life science] information in order to educate ourselves, understand life's processes in the healthy and disease states, and find new or better drugs"*[11]

This area of science emerged around the mid-1970s when the first automated protein and DNA sequencing technologies became available creating a need for a computer-aided way to gather and analyze data. However, it was first in the mid-1990s when the Internet opened the possibility for an easy way of sharing software and databases (i.e. knowledge) between research teams all over the world that the field really took off.[12] Most of the improvements done within this field have so far been made in the academic world, and still much of present-day improvements are made here. Another important factor affecting the field of bioinformatics was the introduction of high-throughput analyzing technologies. The result was a great increase of the speed at which biological information could be obtained. The ever-increasing amount of information has aptly been called the "tidal wave of data"[13] and has really put pressure on the development of new technologies to store and analyze the enormous amounts of information. We have at present-day only seen the beginning of this growth; the amount of biological data is now doubled every 12 months[14].

The possible use of bioinformatics covers a wide range of applications. With the help of a computer and the right software it is now possible to perform a lot of experiments *in silico*, that is in a theoretical framework in the computer, which can reduce the time earlier spent on practical laboratory work. Some of these *in silico* applications can, for example, make it possible to try predicting protein products from a gene (see figure 1.2), visualizing 3-D-structures of proteins knowing only the polypeptide sequence or measuring the affinity between a tailored molecule and a protein.

---

[11] Persidis, Aris, *Nature Biotechnology* (1999), Vol. 17, p. 828

[12] *ibid.*

[13] Reichhardt, Tony, *Nature* (1999), 399, p. 517-520

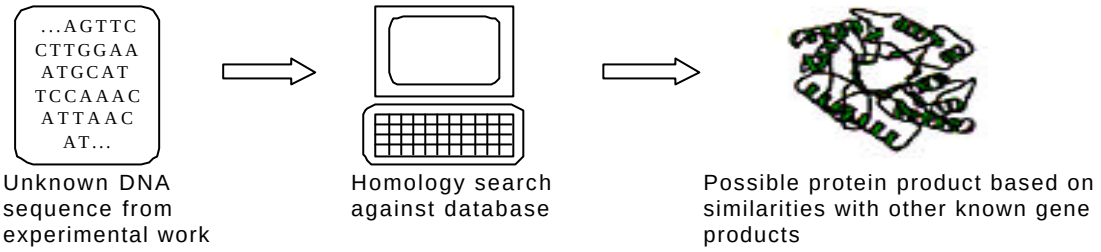

Figure 1.2 Bioinformatics. A schematic description of how bioinformatics is used to predict protein products from a gene.
Source: Authors' research

All these applications are widely transferable into the drug development pipeline. Therefore, bioinformatics has naturally rendered a lot of interest from pharmaceutical companies as well as the biotech market in general. Bioinformatics has the power to completely alter the ways drug research is performed, and has been identified as a foundation of modern biotechnology[15].

1.3 Questions of Issue

The potential for bioinformatics to radically change the research process within today's pharmaceutical companies and the basic research within the academic world is immense. More efficient research would not only result in greater revenues for the pharmaceutical companies, but also create a value for the society as a whole, boosting basic research and making more effective drugs available faster and to a lesser cost.

There could be endless possibilities for making research more efficient by the help of bioinformatics, but how is bioinformatics used and organized within companies and academia and how mature is bioinformatics as a technology? What does the future hold for the technology and what kinds of visions exist within this field? This study hopes to provide answers to these kinds of questions.

1.4 The Purposes of the Study

The purposes of this study are to describe how academia and pharmaceutical companies in Sweden utilize bioinformatics and to identify factors critical for further development of the technology and industry.

[14] *Life Science Informatics*, UBS Warburg, (2001), p. 4

[15] Persidis, Aris, *Nature Biotechnology* (1999), vol. 17, p. 830

1.5 Definitions and the Scope of the Study

1.5.1 The Swedish Bioinformatics Industry

According to Savotti *et al.*, bioinformatics "…cannot be considered either a discipline or an industrial sector"[16]. With this in mind, we find it necessary to make our way of looking upon the Swedish bioinformatics industry clear. When discussing the Swedish bioinformatics industry in this report we refer to companies active in Sweden, i.e. companies that offer their products and services to the Swedish market. According to Figure 1.3, this implies that companies A-C are considered, but not company D.

So, what defines a bioinformatics company? To elucidate this we first present our definition of bioinformatics:

> *a software solution that helps manage, structure, analyze, and/or present life science data in order to address biological issues*

Our definition of a bioinformatics company is thus a company that offers a software solution to manage, structure, analyze, and/or present life science data. We want to point out that the mere collection of data is not enough. This implies that companies that only provide machines or technologies that produce biological data are not to be considered bioinformatics companies, if they do not offer a freestanding complementary software product like the one just described.

[16] Saviotti, Paolo P. *et al*., *Nature Biotechnology* (2000), vol. 18, p. 1247

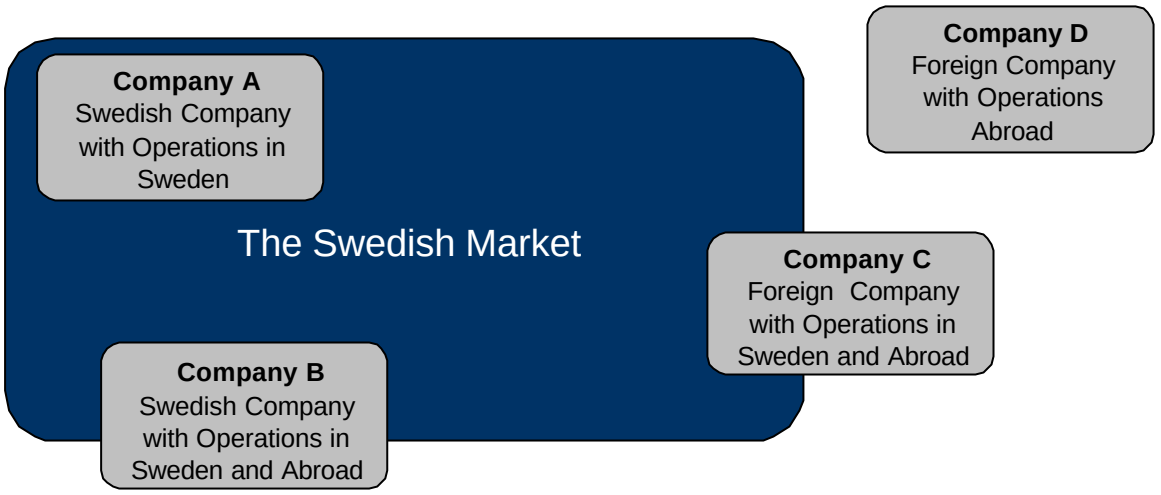

Figure 1.3 A graphical representation of the Swedish market.
Source: Authors' research

1.5.2 Case Studies

The focus of the case studies is Swedish pharmaceutical companies and academia. With Swedish pharmaceutical companies, we mean companies with research and development located in Sweden. Accordingly, company A fits within this description while B and C might, depending on if they have research and development (R&D) facilities in Sweden or not (see figure 1.3). Only companies fitting the definition above were taken into consideration when selecting the objects of study.

Furthermore, we have chosen to categorize the pharmaceutical business in Sweden into large and small companies. This categorization is based on size and annual turnover, where more than 400 employees and/or an annual turnover greater than SEK 500 million represent a large company.

To narrow our scope of study regarding the actual use of bioinformatics at pharmaceutical companies, we have decided to focus on the first four steps of the drug development pipeline, i.e. target discovery, target validation, lead discovery, and lead validation (see figure 1.4). This was done since the majority of bioinformatics implementations are to be found within these areas.

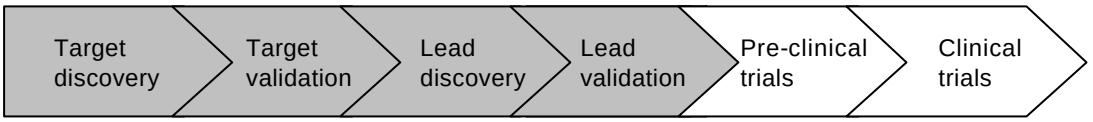

Figure 1.4 The scope of the study concerning pharmaceutical companies.

1.6 About the Assigner

Accenture, Stockholm, as a part of their scholarship program, sponsors this report. Accenture is one of the world's largest providers of management and technology consulting and a leader in its field. They furnish the market with knowledge and competence regarding consulting, technology, outsourcing, and alliances. The company employs more than 75,000 people in 47 countries. (For further information, please visit www.accenture.com)

1.7 Disposition

The subsequent part of this report begins with a short description of the methods and models used in the report. This is followed by a general discussion of the Swedish bioinformatics industry and the technologies available. In the light of these facts, a number of case studies are presented where it is shown how and where bioinformatics technologies are used within pharmaceutical companies and academia. To analyze the case studies, the maturity of the technology is discussed and displayed, critical factors for further development identified and scenarios for the future presented. At the end, the results are summed up and the conclusions presented (see figure 1.5).

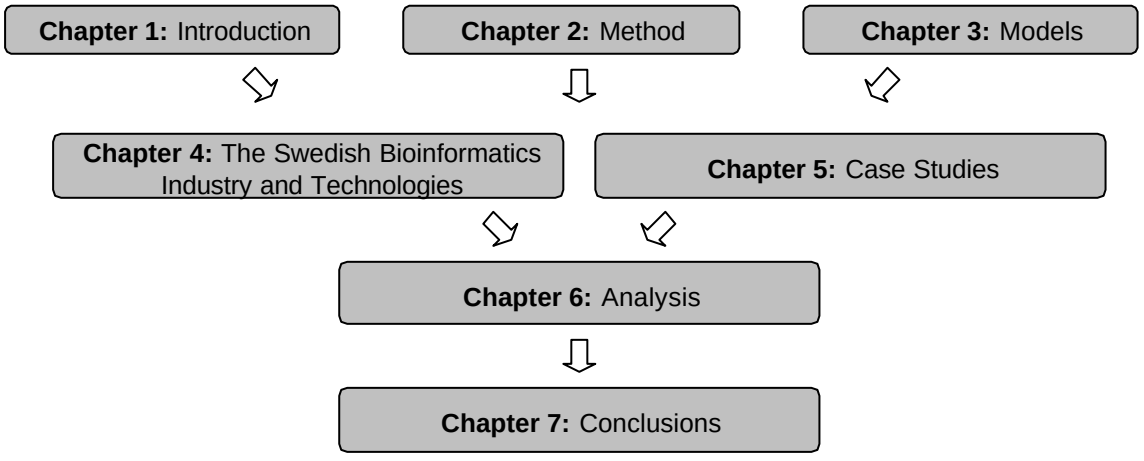

Figure 1.5 Disposition of the report.

2 Method

2.1 Choosing Objects to Study

When performing a study an important decision to make is how the study should be carried out. Mainly, there are two different approaches: quantitative (surveys) and qualitative (interviews). In this report, the qualitative approach was chosen since the questions of issue were of such nature that they would have been hard to address through a survey.

Case studies were conducted with different academic departments and pharmaceutical companies active on the Swedish market through interviewing key persons within the selected organizations. Three different categories of bioinformatics users were chosen: large companies, small companies, and academic departments. The first two categories were defined by number of employees and annual turnover (see chapter 1.5). The goal was to find a way of observing common features and differences among and within all categories as well as pointing out possible characteristics of commercial drug development and how commercial interests affect this research.

Two large and two small companies were chosen as case companies. They were selected since they could fulfill the requirements determined (i.e. R&D in Sweden, right size, and willingness to participate). Additionally, two academic departments that utilize bioinformatics were included.

2.2 Conducting the Interviews

The main sources of information for the case studies are interviews. Altogether 16 interviews have been conducted; all with people within the chosen study objects that have a sound knowledge of the business and of how bioinformatics is used in their respective field.

The interview sessions have been carried out following an open questionnaire (see Appendix I), where the questions asked were meant to induce a discussion regarding a certain topic rather than demanding specific answers. The interviews were recorded and typed out. A copy was then sent to the interviewee for revision and approval before used in this report.

2.3 Outlining the Case Study

Each case study in this report is presented in five parts reflecting the material achieved during the interviews. This was done for easy comparison between the objects of study. The parts are Bioinformatics Organization and Strategy, Bioinformatics Tools and Databases, Problems

with Bioinformatics, Bioinformatics in the Future, and Case Summary. The headings define the contents of each part.

The Case Summary consists of a table which format is shown in Figure 2.1.

Description of study object	Large company, Small company or Academic department
Organizational form	Formal group, Informal group or No group
Bioinformatics collaborations With companies With academia	 Formal, Informal or None Formal, Informal or None
Bioinformatics tools In-house developed Freeware Commercial	 Mostly, Some or None Mostly, Some or None Mostly, Some or None
Local replicas of large databases	Many, Some or None

Figure 2.1 Displaying the format of the case summary table.

The Description of Study Object relates to the case category. The Organizational Form describes how the usage of bioinformatics is organized within the study object. Bioinformatics Collaborations illustrates collaborations concerning bioinformatics with companies or academic departments. Bioinformatics Tools display to what extent the company or department use in-house developed tools, freeware or commercial software respectively. The grading relates to each company or department's usage of the respective tool types (i.e. in-house developed tools, freeware or commercial) in proportion to the other tool types. It is not a comparison with the usage within the other study objects. At the end, the amount of replicas of international sequence or structure databases is presented.

2.4 Discussion Concerning the Study

When choosing a qualitative approach in the form of case studies, we were well aware of the faults of this way of tackling a problem. Since only a small number of people are interviewed at each study object, the information gathered tends to be colored by these people's thoughts. Nevertheless, a case study renders a good picture of the object of study but makes it hazardous to draw any general conclusions concerning the questions of issue. Knowing this we felt that by choosing two representatives from each of the categories earlier described, it would be possible to present some common features of these categories. We do not claim that these features are general, but hope that they point out important aspects for the respective categories.

A reader of this report should have in mind that the picture presented and conclusions drawn are colored by our experiences and background. Therefore, we do not declare this report to present the absolute truth but a version of it.

3 Models

In this report, two different models are used to present and analyze the data. The SWOT model is used for description and analysis of the Swedish bioinformatics industry, as this model offers an easy and well-structured method to present this kind of information. The Technology Life Cycle Model proposes a way of determining the maturity of a technology related to market environment and to identify conditions to be fulfilled for sustained development. For this report, it is of interest to see how mature bioinformatics is as a technology and what lies ahead.

3.1 SWOT Model

The SWOT model is usually used to display the critical organizational strengths and weaknesses and the opportunities and threats facing a company. The objective is often to show where a business should focus its attention in order to succeed. Strengths and weaknesses focus on present aspects while opportunities and threats try to describe future trends and conditions.[17] In this report the SWOT model is used to describe an industry, namely the Swedish bioinformatics industry, instead of a company. However, the structure of the model and its components are of such nature that using it to describe an industry does not imply any difficulties.

3.2 Technology Life Cycle Model

When a technology change occurs on a market, this affects the strategy of a firm since the change alters the competitive landscape. In order to describe and understand this process the Technology Life Cycle Model can be used[18].

According to the Technology Life Cycle Model, three phases exist in a technology's life cycle called the fluid, transitional, and specific phases (see figure 3.1). The first phase is called the emerging or fluid phase and is characterized by product and market uncertainty since the technology is still undeveloped. Neither customers nor producers know quite what to put into the product. Also, the undeveloped technology faces competition from older established ones. To solve the technological and market uncertainties that arise, firms interact with their local environment of suppliers, customers, complementors, and competitors. The quality of available products is low and the prices and costs are high since economics of scale

[17] Kotler, Philip *et al*., *Principles of Marketing*, p. 94, 1999

[18] Afuah, Allan. & Tucci, Christopher L., *Internet Business Models and Strategies*, p. 73-75, 2001

and learning have yet to set in. Further, market penetration is low and customers are mainly lead users (customers whose needs are similar to those of other users except that they have these needs months or years before most of the market place[19]) or high-income users. In this phase, firms must decide how and where in the value chain they want to exploit the technological change.

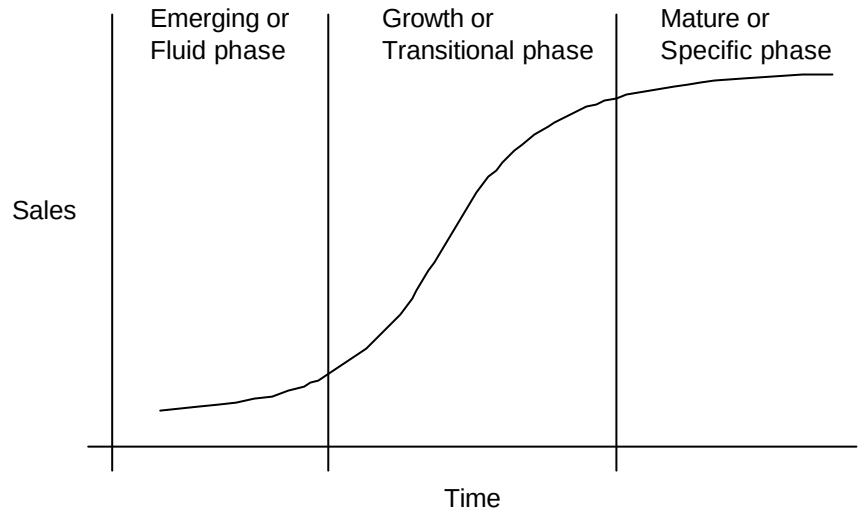

Figure 3.1 An overview of the Technology Life Cycle Model, and its three phases.
Source: Afuah Allan & Tucci C. L., Internet Business Models and Strategies, p. 74

As standardization of components, market needs, and product design features takes place and a standard or common framework emerges signaling reduction in market uncertainty, experimentation, and major changes, the technology enters the growth or transitional phase. During this phase, the customer base increases to a mass-market level. The firms supplying the market should at this time decide where it excels or where it wants to excel, and try to reinforce or build this skill.

The mature or specific phase is characterized by proliferation of products built around the common frameworks or standards that exist. The products offered to the market are highly defined and similar. Demand growth fades away and most output is to satisfy replacement needs. In this phase the strategy for the firm should be to defend its position and watch out for the next technological change.

[19] von Hippel, Eric, *The Sources of Innovation* (1998)

Table 3.1 summarizes the features for each phase suggested by the Technology Life Cycle Model.

Table 3.1 The different phases of the Technology Life Cycle Model and important features for each phase.

Source: Authors' summary of Afuah and Tucci's model

Emerging or Fluid Phase	Growth or Transitional Phase	Mature or Specific Phase
• Product and market uncertainty • Competition between old and new technology • Low product quality • High costs and prices • Customers largely lead users or high-income users • Lack of standards	• Standardization of components, markets needs, and product design • Development of standard or common framework • Customer base increases	• Products built around the common standard proliferate • Demand growth fades away

4 The Swedish Bioinformatics Industry and Technologies

To introduce the reader to the Swedish bioinformatics industry this chapter begins with describing the development of this branch and continues with a short SWOT analysis of the industry. Following this are descriptions of molecular biology methods wherein bioinformatics technologies are used. The latter descriptions are meant to present how bioinformatics is used and what possibilities the technology has.

4.1 Historical View of the Development of Bioinformatics in Sweden

Bioinformatics tools have been used within the Swedish academia for a relatively long time. In the early days, the mid eighties, bioinformatics led a fairly secluded life. The scientists developing and utilizing bioinformatics often kept to themselves, and so the knowledge was confined. As more and more biological data became available, bioinformatics tools developed and so did the need for up-scaling these methods.[20] During this time, the pharmaceutical industry became interested in these methods. The company Pharmacia & Upjohn started negotiating with the Karolinska Institute (KI) regarding the establishment of a new department focusing on bioinformatics development. The result was the foundation of the Center for Genomics and Bioinformatics (CGB) in 1997, as a department at KI.[21]

At about the same time as CGB was founded, the other major Swedish pharmaceutical company Astra AB (now AstraZeneca AB) realized the importance of the bioinformatics field and decided to start their very own bioinformatics center in Lund. As it turned out, the merger with the British company Zeneca, in 1999[22], revised this decision before it was realized. Zeneca already had an established bioinformatics center and the newly formed company, AstraZeneca AB, decided that another center was not necessary. Another contributory cause was the lack of bioinformatics specialists in Sweden.[23]

The decision from AstraZeneca could have been a big setback for the development of bioinformatics in Sweden, and to a certain extent it was, but at the same time it was a wake-up call for the government that Sweden was missing competence in a growing field. The result was a request from the government to the Foundation for Strategic Research (SSF) to

[20] Andersson, Siv, Interview (2002)

[21] www.cgb.ki.se

[22] www.astrazeneca.com

[23] Pierrou, Stefan, Interview (2002)

fund a center for bioinformatics in Sweden. One center was thought to be enough. All universities were invited to present their case and the choice fell on Stockholm.[24] This led to the establishment of the Stockholm Bioinformatics Center (SBC) in January 2000, a center that focuses its efforts on developing new bioinformatics methods[25, 26].

Some of the other universities, though not assigned any funds from SSF, decided to keep on developing their own bioinformatics centers. Uppsala was one of them and their Linneaus Center for Bioinformatics has already earned some recognition. It has raised almost SEK 70 million, mainly from the Wallenberg Foundation, for the coming five years and the focus of this center will lie on the integration of mathematics and computer science into biology research.[27]

The Wallenberg Foundation has continued to play an important role for the development of Swedish bioinformatics research through contributions to SweGene and Konsortium Nord. These consortiums have been assigned SEK 1000 million over a five-year period, that is to be distributed to academic institutions focusing on bioinformatics research. One intention with this funding is to keep Swedish research competitive, another to keep pharmaceutical companies within the country.[28]

The development for bioinformatics within the academia has been a breeding ground for ideas and competence. Most bioinformatics companies established in Sweden are generated out of the academia. Ideas and software developed during research projects have laid the foundation for companies such as Spotfire and BioBridge Computing AB[29].

4.2 SWOT-Analysis of the Swedish Bioinformatics Industry

4.2.1 Strengths

The fact that most bioinformatics companies in Sweden originate from the academia has generated a natural connection between basic research and companies within the industry. These kinds of collaborations can take various forms, e.g. personnel exchange and

[24] Andersson, Siv, Interview (2002)

[25] Kraulis, Per, Interview (2002)

[26] www.sbc.su.se

[27] Andersson, Siv, Interview (2002)

[28] Nilsson, Jan S., Interview (2002)

[29] Servenius, Bo, Interview (2002)

sponsorship[30], and helps widen the knowledge base of ongoing projects and research. [31] It is also useful in the development of bioinformatics tools by creating an interface between cutting edge research and software developers.[32]

Bioinformatics companies often supply consulting skills together with their product, since a sold product in many cases must be uniquely designed to fit the needs of the purchasing company[33]. Consulting services can be considered a strength since it helps companies in developing products further, as well as deepening relations to partner companies.

Further, Swedish research is well spoken of internationally both within biology and computer science. These competences, that Sweden houses, must be considered a strength.

4.2.2 Weaknesses

A weakness for the bioinformatics industry in Sweden today is that the Swedish universities have had problems in merging computer science and biology. A competence gap between these two sciences exists, and until now, few people have had the knowledge to bridge this gap.[34] Also, there are few people in Sweden with the competence to lead and organize large industrial bioinformatics projects[35].

The market for bioinformatics products in Sweden is small. The majority of bioinformatics investments within pharmaceutical companies are internal (i.e. personnel, software, and hardware)[36]. Another reason for the low demand is the academia's limited resources and their need for very specialized software resulting in in-house development[37]. The number of companies developing and marketing bioinformatics products are in all cases less than ten (see Appendix II for a description of Swedish bioinformatics companies).

[30] Kraulis, Per, Interview (2002)

[31] Waleij, Martin, Interview (2002)

[32] Kraulis, Per, Interview (2002)

[33] Andersson, Siv, Interview (2002)

[34] Asp, Allan, Interview (2002)

[35] Andersson, Leif, Interview (2002)

[36] Robinson, Jack, Interview (2002)

[37] Wickenberg, Ulrika, Interview (2002)

The Swedish market is, as pointed out above, presently too small. Sooner or later introductions into foreign markets is necessary, something that is strengthened by Martin Waleij, CEO of BioBridge Computing AB. This puts pressure on newly founded companies for rapid expansion abroad. The weakness can to a certain extent be turned into a strength through establishing partnerships with global companies, something that is verified by the fact that Swedish bioinformatics companies are active outside Sweden through partners or subsidiaries. Partnerships can help small companies in the internationalization process and thereby creating a reputation on a wider market.

4.2.3 Opportunities

A rough estimate of the annual turnover for the Swedish bioinformatics business as a whole is SEK 400 million, whereof some SEK 20 to 30 million are spent externally on services and software. The potential market is definitely greater than today, one of the reasons being that many pharmaceutical companies have never bought software, but use free downloads from the Internet.[38] If bioinformatics companies could exploit parts of the internal bioinformatics investments, within pharmaceutical companies, the sales potential would definitely increase.

Swedish academic research is strong and Swedish scientists are conceived to lie in the biology knowledge frontier.[39] Within the academia's progress lies an opportunity for new knowledge to be turned into companies, an explicit objective of academic research[40], but also for this knowledge to be used within the existing industry. Further, the upcoming of academic centers in Sweden focusing on bioinformatics will hopefully deepen the academic knowledge. So far, these centers have had shifting focuses. A wide focus can turn out to be an opportunity for the future creating a wide knowledgebase.[41]

4.2.4 Threats

An important aspect related to the Swedish bioinformatics industry is that the European biotech market is lagging behind the American.[42] One reason for this is the fact that it has been easier for US companies to raise money than it has been for European.[43] The fact that

[38] Robinson, Jack, Interview (2002)

[39] Lake, Staffan, Interview (2002)

[40] *Biotekniken inför 2000-talet,* SOU 2000:103, p. 25

[41] Servenius, Bo, Interview (2002)

[42] *Integration,* Ernst & Young (2001), p. 21

[43] *Integration,* Ernst & Young (2001), p. 9

the European biotech market, and thus also the bioinformatics sector, raises less money affects the conditions for progress within the sector.

Competence is hard to develop, especially within such a knowledge-intensive area as bioinformatics. A competence that Sweden is lacking is how to build and maintain large life science databases. This has made Sweden dependent on information kept in large databases abroad and also on the knowledge of how to maintain them.[44]

4.2.5 SWOT Summary

The results of the SWOT analysis are summarized in table 4.1.

Table 4.1 SWOT analysis of the Swedish bioinformatics industry

Strengths • Relations with the academia • Product development through partnerships and consulting services • Wide knowledge within biology and computer science	**Weaknesses** • Cross-competence gap • Small market • Demand for rapid expansion
Opportunities • Large market potential • Strong academic research generating new competence and start-ups	**Threats** • Lagging behind USA • Lack of database management competence

4.3 Technologies

Bioinformatics can be described in different ways. One way of looking at bioinformatics is viewing it through the various molecular biology methods wherein bioinformatics solutions greatly improve the potential of the method. This is the approach taken below. The frame for the different methods described is based on the UBS Warburg report "Life Science Informatics", but revised to adhere to our view of bioinformatics. The molecular biology methods within this frame are sequence analysis, pharmacogenomics, transcriptomics, structural genomics, and proteomics. In addition to these descriptions, a discipline in the outskirts of bioinformatics but within informatics, called cheminformatics, is presented.

[44] Kraulis, Per, Interview (2002)

Below, Figure 4.1 puts the methods mentioned in proportion to the central dogma of molecular biology. The central dogma is the most basic way of describing the central mechanisms of life. The starting point of the central dogma is the DNA molecule. DNA contains all information needed to produce living organisms, organized as genes. These genes can be transcribed into mRNA, a messenger that transports the information from the gene to the machinery that translates the mRNA into proteins. Proteins are the workforces of a cell, responsible for cell function.

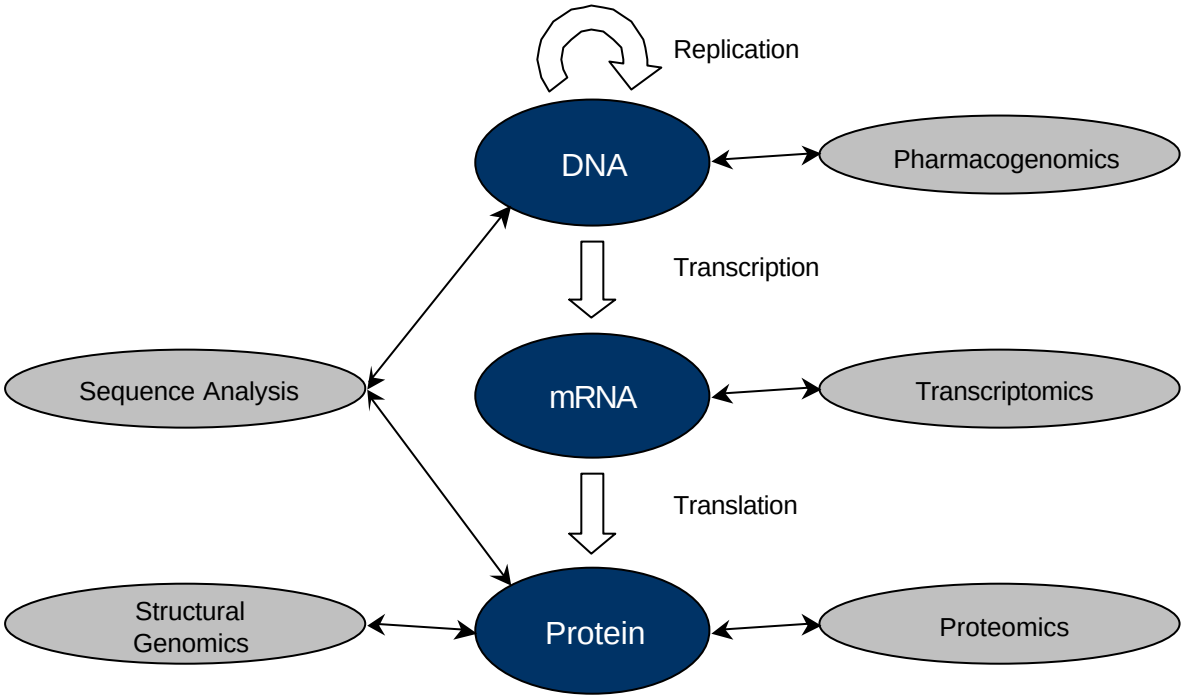

Figure 4.1 The central dogma of molecular biology and related molecular biology methods wherein bioinformatics solutions greatly improve the potential of the method.
Source: Authors research and analysis

The DNA-molecule has been referred to as the code of life. The reason for this is that the combination of bases that the molecule is built up of (called A, C, T and G) serves as a template for protein production. The genome is the total DNA set-up for a species, and as one might expect, the genome differs in size between different species; the human genome, for example, contains approximately 3 billion base pairs[45].

[45] Lewin, Benjamin, *Genes VII* (2000), p. 546

However, only a small fraction of these bases (approximately 1,1 %)[46] have a coding function. It is only these parts of the DNA-molecule that are called genes. A gene is a well-defined sequence; it has a start and an end site. The total number of human genes has been discussed, but the most common view is that the human genome contains approximately 30.000 genes[47].

All organisms existing today have the same origin. Consequently, many of the proteins in our body are similar to proteins in other species, both in function and structure. This implies that the genes coding for homologous proteins have similar basic sequences, although located in species totally different. This makes much of the knowledge about sequences general for all living life forms.

4.3.1 Sequence Analysis

In molecular biology today, DNA sequencing is a basic method for studying an organism. DNA sequencing reveals the particular DNA-sequence for the part of the genome investigated. The result is a long array of bases (e.g. AATCGTCGACG…). The further study of this sequence is called gene sequence analysis, and can roughly be described as a method that tries to identify genes in the sequence and predict the function for found genes based on previous knowledge of other genes. Raw sequence data can, with the help of computer software, be analyzed for similarity (homology) with any known gene stored in a compatible database. If a gene shows homology with another gene there is a great possibility that these genes have similar gene products, i.e. code for similar proteins.

However, there are some complications related with this comparative identification of genes. In some cells (eukaryotic cells), gene sequences are built up by coding regions, called exons, and non-coding regions, called introns. This makes the homology search more difficult, since genes therefore cannot be directly compared. Furthermore, the possibility of multiple start and end sites within a gene complicate the analysis even more (see figure 4.2).

[46] Venter, Craig *et al*., *Science* (2001), vol. 291 p. 1305

[47] www.hugo-international.org

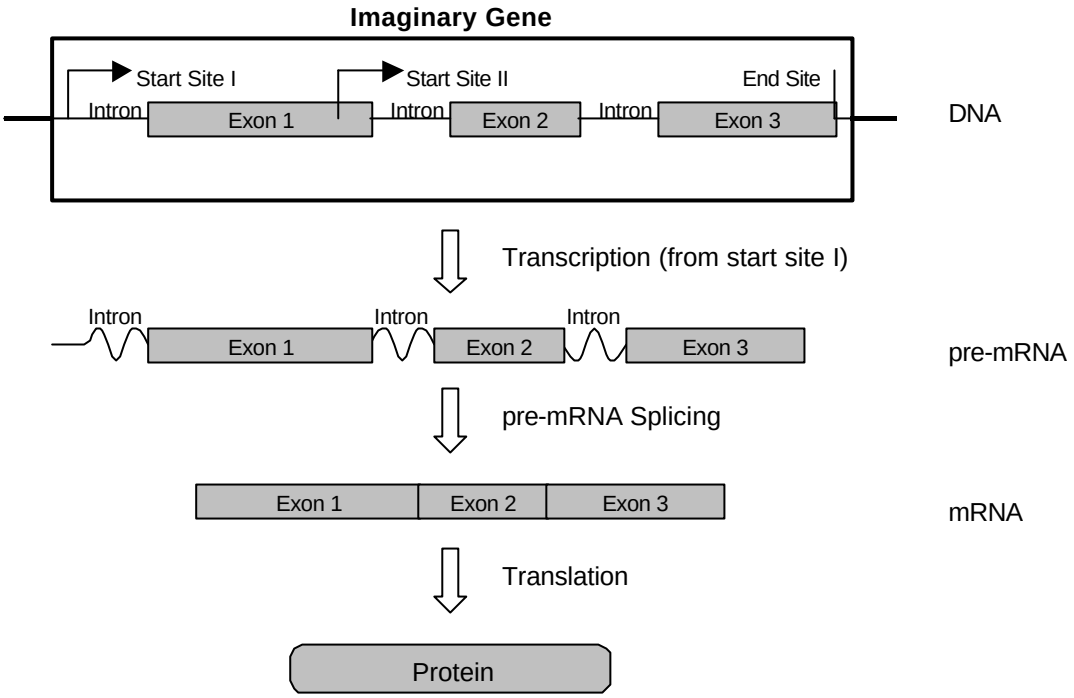

Figure 4.2 How a protein is produced in eukaryotes. A view of what could cause difficulties when comparing sequences on a DNA level (i.e. introns and multiple start sites).
Source: Authors' research and analysis

There are a variety of gene sequence analysis tools available. Many of these are accessible for free over the Internet, at least for the academia. Research-groups have built up large databases with vast amounts of sequence data from all kinds of organisms and tools to search these databases. In addition to this freely available software, there are commercial programs available on the market.

To connect gene sequence analysis to the drug development pipeline, figure 4.3 displays in which part it is commonly used. Sequence analysis lays the foundation for many of the other technologies later described and is the basic tool for mapping genes.

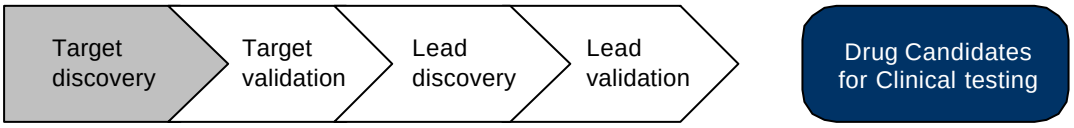

Figure 4.3 Main area within the drug development pipeline where gene sequence analysis can be applied.
Source: Authors research and analysis

The underlying method for comparing sequences can also be applied when comparing protein sequences. The practice of gene sequence analysis and comparative analysis between proteins is called comparative genomics.[48]

4.3.2 Pharmacogenomics

Pharmacogenomics studies how people react differently to drugs, because of their genetic setup, and how this knowledge can be used to tailor drugs for a group of people with certain genetic characteristics. These studies are based on the appearance of single nucleotide polymorphisms, SNPs, in the DNA.

Every organism within every species has a unique DNA setup (with the exceptions for identical twins). These differences are mainly due to SNPs. An SNP is a variation of a specific single base in the DNA. SNPs can be the cause of unique diseases and can explain how some people respond well to drugs that might be toxic to others. Researchers are working on mapping the location of every known SNP in the human genome. Hopefully this will make it possible to identify SNPs tied to specific physiological conditions, e.g. the response to drugs or the emergence of diseases (see figure 4.4).

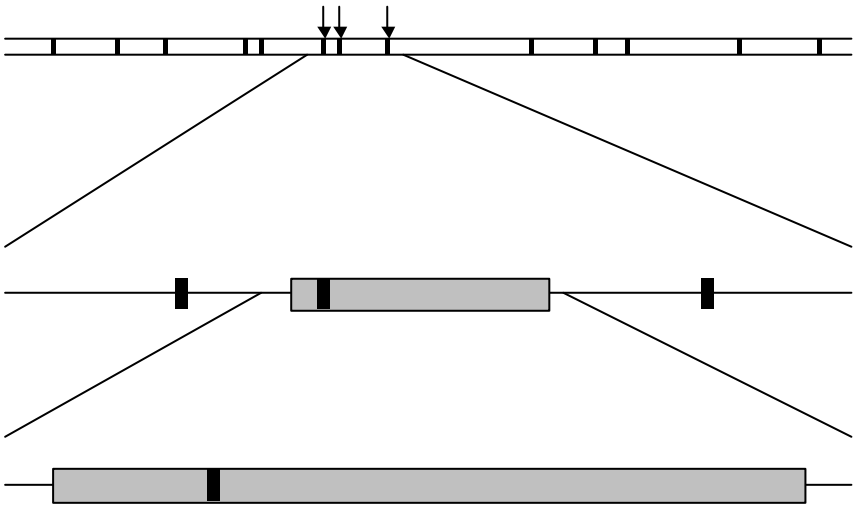

Figure 4.4 A possible use of SNP analysis.

Source: Authors' research and analysis.

[48] *Life Science Informatics*, UBS Warburg (2001), pp. 44-45

If the location of a SNP linked to a condition is known, it can be assumed that the underlying genetic cause is situated nearby this SNP. The result is a limited part of the genome to study further. When the genetic cause of the condition hereby has been identified, it can help to predict structural differences in the protein products affected. This makes it possible to design drugs to specifically fit parts of a population with a certain condition. SNP analysis can be applied in target discovery and validation in the drug development pipeline (see figure 4.5).

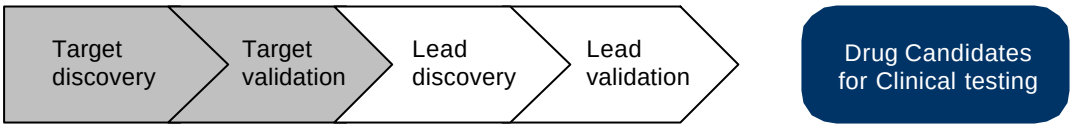

Figure 4.5 Areas within the drug development pipeline where SNP-analysis can be applied.
Source: Authors research and analysis

4.3.3 Transcriptomics

The regulation of a cell is an extremely complex process that scientists have only begun to elucidate. Often the expression of one gene affects the expression of several others in a long complex chain reaction. The study of gene expression is referred to as transcriptomics, and focuses on mRNA-expression (see figure 4.1).

In order to investigate which genes are expressed and under what conditions, several methods exist. The most commonly used today is probably microarray technology, which measures the quantity of different mRNA:s in a cell at a certain time and under a certain condition. This information can be used to, for example, identify improperly expressed genes in diseased tissue, something that can be used for target discovery. The technology can also be used to analyze the efficiency of new drugs (i.e. lead validation) by studying how a drug effects the expression of genes in a cell. Transcriptomics has a potential use in all four steps in the part of the drug development pipeline studied (see figure 4.6).

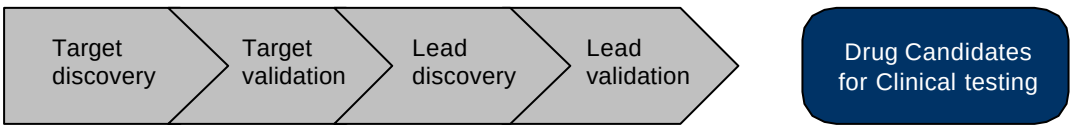

Figure 4.6 Areas within the drug development pipeline where transcriptomics can be applied.
Source: Authors research and analysis

The microarray technology has made it possible to collect vast amounts of data on gene expression. The bottleneck within transcriptomics thereby lies in analyzing and managing data. One solution to this problem is to use software that digitizes microarray images, making it possible to let a computer search for patterns and trends in the data collected. Software can also assist in the visualization of gene expression in cells, making comparison between levels of expression in cells affected by different drugs possible.

4.3.4 Structural Genomics

Structural genomics is the study of the three-dimensional (3-D) shape of proteins and how the shape correlates to function. The knowledge of protein 3-D structures is central to understanding the mechanisms by which proteins work. Knowing only the polypeptide sequence gives no clue of what part of the protein is involved in its function, since the protein, after being synthesized, is folded into a specific 3-D structure. This folding puts different parts of the protein, which can be located far apart in the polypeptide sequence, into near vicinity of each other, thus creating 3-D environments where reactions can take place.

Obtaining 3-D structures for a protein is still a rather complex and time consuming process. There are two ways of tackling the problem, through wet lab approaches (i.e. x-ray crystallography, cryo-electron microscopy and nuclear magnetic resonance) or by using so called theoretical modeling.[49]

The wet lab methods have their limitations. X-ray crystallography and cryo-electron microscopy need for the protein to be in a crystallized phase, something that is not possible to obtain for all proteins. Nuclear magnetic resonance (NMR) has limitations concerning size and insoluble proteins, since the proteins need to be in a solution for the method to work. Consequently, some interesting proteins are almost impossible to analyze with these methods, for example membrane proteins that are in part insoluble and dependent of a membrane to maintain their structure.

Theoretical modeling, on the other hand, is a way of predicting the 3-D structure based only on sequence data. The technology is fairly new, but improving fast. Theoretical modeling relies heavily on comparison with known protein structures for similarities, and therefore the success of the method has a strong connection to the number of known protein structures.

[49] *Life Science Informatics*, UBS Warburg (2001), pp. 69-70

Structural genomics is dependent on the use of bioinformatics. Both the wet lab and theoretical modeling methods described above render extensive information and require computational power and software to store and extract the information needed for further analysis. Furthermore, bioinformatics software is required to help visualize these structures. The use of structural genomics is important in the target validation and lead discovery stages of the drug development pipeline (see figure 4.7). When a possible target has been identified, and its 3-D structure determined, the visualization of the protein could help the scientists in finding and designing a lead able to block the function of the target protein. The leads' interaction with the target can be visualized and analyzed in a theoretical framework, and possible modifications of the lead can be made at an early stage, making a more rational drug design possible[50].

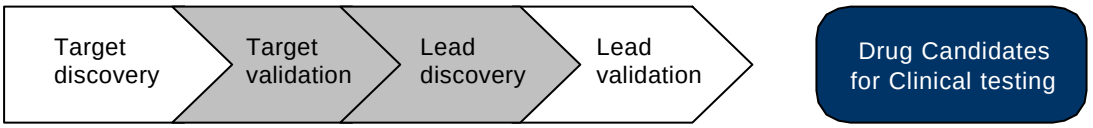

Figure 4.7 Areas within the drug development pipeline where structural genomics can be applied.
Source: Authors research and analysis

4.3.5 Proteomics

According to Figure 4.1, proteomics relates to the expression of proteins within a cell. A complex network of signals regulates protein expression. All through the different stages of a person's development, the levels of different proteins in different cells fluctuate. By learning more about this regulation and fluctuation and about how different proteins are affected by different signals and each other, knowledge can be acquired about how loss of function in these networks can lead to different diseases. This study of protein expression and protein function is called proteomics.

Proteomics is used is to display differences between healthy and unhealthy cells or tissues. By analyzing what proteins are expressed in a healthy cell and comparing this to the proteins expressed in an unhealthy cell, scientists can discover what proteins are lacking or are over-

[50] *Life Science Informatics*, UBS Warburg (2001), p. 73

expressed in a damaged cell. This way of analyzing cells can lead to the discovery of targets and help in analyzing the effect of new leads (see figure 4.8).

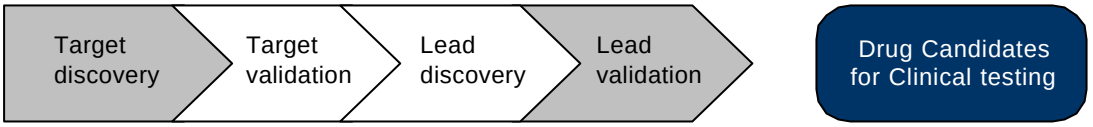

Figure 4.8 Areas within the drug development pipeline where proteomics can be applied.
Source: Authors research and analysis

To effectively perform a proteomics analysis the proteins expressed in a cell must be separated from each other in a way that makes it possible not only to see which proteins are expressed but also in what amount. There are many different variables that can be used in order to separate proteins. The technology most commonly used for protein separation is the 2-D gel. This technology separates the proteome (i.e. the total protein content) of a cell by size and charge. The result is a matrix of spots, where each spot represents a specific protein. Thereafter the protein can be identified using mass spectroscopy, a way of determining the weight of the purified protein and comparing it to known proteins in a database.

Bioinformatics can come to play an important part of proteomics by accelerating and improving 2-D gel analysis. This can be done through image analysis. Software programs that, for example, add contrast to images visualizing spots not seen by the naked eye can be constructed. These programs can also be used to identify patterns by using sophisticated algorithms that correct, smooth and triangulate to make room for experimental variants among the data collected and to store experimental results. Bioinformatics can also play a part in automating the process of identifying proteins in 2-D gels.[51]

4.3.6 Cheminformatics

As mentioned initially in this chapter cheminformatics lies in the outskirts of bioinformatics and is not by all considered to be a bioinformatics discipline. Cheminformatics is the managing and analyzing of information concerning chemical compounds. This includes, among other things, structural and molecular information. The area is to a great extent dependent on databases and database applications, since all compound information is stored

[51] *Life Science Informatics*, UBS Warburg (2001), p. 64

in large databases. The discipline combines computational chemistry with 3-D structural analysis and visualization tools.

Cheminformatics is mainly used in the lead discovery and validation processes in the drug development pipeline (see figure 4.9). Information about chemical compounds stored in cheminformatics databases can be used to virtually screen these compounds' affinity towards a target. If a high affinity compound is found, the scientist can modify this potential lead in the computer to obtain a better affinity for the target. The technology makes it possible to design a suitable lead practically from scratch. This presumes that structural genomics methods are used to achieve well-defined targets with defined structures.

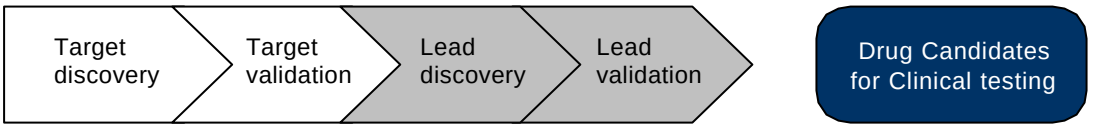

Figure 4.9 Areas within the drug development pipeline where cheminformatics can be applied.
Source: Authors research and analysis

5 Case Studies

One of the main objectives of this report is to present the usage of bioinformatics within academia and pharmaceutical companies in Sweden. As mentioned in chapter 1.5 the objects of study have been chosen from three different categories, the academia, large pharmaceutical companies, and small pharmaceutical companies. The participants are presented in Table 5.1.

Table 5.1 A presentation of the objects of study

Academia • Department of Genetics and Pathology, Uppsala University • Department of Molecular Evolution, Uppsala University
Large pharmaceutical companies • AstraZeneca AB • Biovitrum AB
Small pharmaceutical companies • Active Biotech AB • Medivir AB

The obtained material is displayed below. Presented first are the two academic departments in chapter 5.1 and 5.2. Following, the two large companies are described in chapter 5.3 and 5.4. The presentation is ended by descriptions of the two small companies in chapter 5.5 and 5.6.

Each section consists of five parts reflecting the material achieved during the interviews. This is done for easy comparison between the objects of study. The parts are Bioinformatics Organization and Strategy, Bioinformatics Tools and Databases, Problems with Bioinformatics, Bioinformatics in the Future, and Case Summary.

5.1 Department of Genetics and Pathology, Uppsala University

The facts presented concerning the Department of Genetics and Pathology is based on interviews held with Kalle Näslund, Ph.D. student, and Fredrik Granberg, Ph.D. student, at the Department of Genetics and Pathology 2002-03-13. Additional information was found on the department's homepage, www.genpat.uu.se.

The Department of Genetics and Pathology consists of three units, medical genetics, clinical genetics, and pathology. The unit for medical genetics conducts research concerning the structure and function of genes of higher organisms. The focus lies on heredity, organization, and regulation of the genome. The unit for clinical genetics studies the human genome and the use of clinical diagnostics. Finally, the unit for pathology investigates the underlying

causes that make diseases appear. The overall objective of the department is to bring basic and clinical research together to produce a better knowledge of mechanisms causing diseases.

The department mainly uses bioinformatics for sequence analysis, transcriptomics, and structural genomics.

5.1.1 Bioinformatics Organization and Strategy

The Department of Genetics and Pathology does not have an overall strategy for the use of bioinformatics. All groups within the department work fairly independently from each other and there are no formally formed inter-group co-operations to support bioinformatics utilization. Knowledge is exchanged when needed over informal channels. Further, the department does not have anyone overall responsible for collecting and storing the information produced within various projects. There have been discussions concerning such a position, but nothing concrete has yet happened.

Collaborations with commercial companies are rare within the department. The existing ones often concern development of company products, where the department collaborates with the company to further develop biotech products. An example has been a collaboration with Pyrosequencing. On the academic level, there are several co-operations with other departments, but none merely concerning bioinformatics.

5.1.2 Bioinformatics Tools and Databases

The software most often used at the Department of Genetics and Pathology is freeware. The main reasons are money and attitude. For many applications, it is considered to exist high quality commercial software solutions, but money limits the possibilities to purchase these programs since they often are too expensive. In a worst-case scenario, this implies that a project must be prematurely closed since needed software investments cannot be afforded. Attitude further restricts software investments since conservatism sometimes rule in the laboratories. It is often considered better to invest in a machine rather than a computer program since it is easier for the people in the laboratories to directly see in what way a machine could affect their daily work.

However, it is not always that freeware dominates over commercial programs. For some standard research applications, bioinformatics software, both freeware and commercial, has been used for a relatively long time. Within some of these areas, there are several products

available. Here competition over the years has produced quality commercial software to a relatively low cost, making purchasing more common.

There are few local databases at the department as of yet. The intention is to store more data locally. One project is to build an in-house nucleotide sequence database with the objective to lessen the time spent on conducting queries on distant databases. The local database would rely on available services, at for example NCBI (National Center for Biotechnology Information), to automatically update the local replica with new information.

5.1.3 Problems with Bioinformatics

Annotation is seen as one of the larger problems for bioinformatics. Different standards for naming sequences have been used throughout the years, creating inconsistent data structures in sequence databases. This has resulted in multiple names for the same sequences making homology searches more difficult. Annotation standards does exist, but has not consequently been used. The problem is greater for older data.

The annotation problem has created a need for databases with better search engines, a uniform classification system and correctly sequenced data. This is manifested through the department's need of constructing in-house databases for specific research projects in order to achieve the characteristics just mentioned.

To be able to construct and develop new databases and search algorithms the cross-competence gap between biology and computer science needs to be lessened. The department therefore has a need for people with both biological knowledge and software development competence.

5.1.4 Bioinformatics in the Future

The amount of biological data produced will demand better and cheaper software products in the future. Also, a paradigm shift is foreseen as researchers discover the full potential of bioinformatics, radically changing the way biological research is performed. The paradigm shift will imply a turn in focus from laboratory work towards more computer-aided research.

In the future two categories of bioinformatics users within the academia are predicted to exist. For standard applications users, more programs will be available to a lesser cost implying that new software, to a greater extent, will be bought. On the other hand, biologists

with the need for specific advanced applications will continue to develop software themselves. This relates to the researchers need for credibility when publicizing and for their understanding of results.

5.1.5 Case Summary

Table 5.2 presents a short summary of the case. The summary describes aspects from the first two areas: Bioinformatics Organization and Strategy, and Bioinformatics Tools and Databases.

Table 5.2 Summary of empirical data for the department of Genetics and Pathology

Description of study object	Academic department
Organizational form	No group
Bioinformatics collaborations With companies With academia	 None None
Bioinformatics tools In-house developed Freeware Commercial	 None Mostly Some
Local replicas of large databases	None (some planned in the future)

5.2 Department of Molecular Evolution, Uppsala University

The facts presented concerning the Department of Molecular Evolution is based on interviews held with Professor Siv Andersson, Hans-Henrik Fuxelius, programmer, and Carolin Frank, Ph.D. student, at the Department of Molecular Evolution 2002-03-27 and 2002-04-09. Additional information was found on the department's homepage, web1.ebc.uu.se/molev.

The department of Molecular Evolution at Uppsala University focuses on the study of microbial genomes. The main course of study address biological questions. In these studies, bioinformatics play a key role. The basic work performed at the department concerns gene sequence analysis and this work is the driving force for creating new types of tools. Approximately 20 people work at the department, which is led by professor Siv Andersson.

5.2.1 Bioinformatics Organization and Strategy

The work within the Department of Molecular Genetics is dependent on computational power. Three quarters of the personnel currently use computers to analyze biological data. This number however is not static as the needs for computer-aided analysis varies with time

depending on what stages the current projects are in. Most researchers at the department have programming experience and therefore the need for a specific bioinformatics group is small. However, the department has hired a programmer. The intention is to strengthen the programming knowledge and to support different projects.

The department does not have any collaboration with commercial companies or outspoken bioinformatics collaborations with other academic departments. However, the department has exchanged ideas with different actors, one of them being Prevas.

5.2.2 Bioinformatics Tools and Databases

Almost all of the bioinformatics tools used are developed within the department. In general, no bioinformatics tools are bought, and therefore, in addition to the tools developed in-house, complementing freeware is used. The reasons for this are limited resources and very specific needs. Something lacking at the department, however, is a common standard for the software produced.

The department houses databases, for data produced in-house, and has built up knowledge of how to build these databases and how to maintain them. Additionally, the department has local copies of specific parts of large international databases especially important for the research conducted.

5.2.3 Problems with Bioinformatics

One problem mentioned is the difficulty in finding persons with deep knowledge in both biology and computer science. A way of eluding this problem is to create groups with both competences present, and letting them work together. The department has a strategy to hire young personnel with great interest in mathematics and computer science, and a willingness to learn new things to create an innovative environment.

A problem for the research performed at the department is knowledge management, i.e. how to transfer knowledge between different persons and how to keep the knowledge within the organization. The staff turnover is high in the academic world and knowledge management is therefore a considerable problem. This is especially true for the bioinformatics area, where specific knowledge about a particular tool is tightly connected to the person or persons developing it.

The researchers that do not have programming experience lack most of the knowledge needed to make rational and relevant system demands, as well as the knowledge of what can be done using bioinformatics. The issue of being able to make sound system demands is also problematic for several of those with programming experience.

When comparing sequences, the varying quality of the data in databases creates a problem. One cause for this varying quality is that many of the sequence comparison methods used today are easy to use, resulting in usage without knowledge of how to interpret the results. This leads to both under- and over-interpreting of results and thus data of uncertain value. A second reason for the varying quality of the information in databases is the inconsequent annotation used for classification of data.

5.2.4 Bioinformatics in the Future

There are no indications that the use of bioinformatics within the department will decrease in the future. Professor Siv Andersson foresees a trend towards a development in two different levels. The front line usage of bioinformatics will surely be much more advanced through the emergence of more and more experts within the field. The use within the other level, the every-day-usage, will develop as well, and probably reach the knowledge level the bioinformatics centers such as SBC, CGB and Linnaues Center holds today. The government's support to these bioinformatics centers is considered of great importance, as these centers fuel the progress in the bioinformatics field.

5.2.5 Case Summary

Table 5.3 presents a short summary of the case.

Table 5.3 Summary of empirical data for the department of Molecular Evolution

Description of study object	Academic department
Organizational form	No group
Bioinformatics collaborations With companies With academia	 Informal None
Bioinformatics tools In-house developed Freeware Commercial	 Mostly Some None
Local replicas of large databases	Some

5.3 AstraZeneca AB

The facts presented concerning AstraZeneca AB is based on interviews held with Rolf Johansson (Ph.D.), Director Support Office, at AstraZeneca AB Södertälje 2002-03-20, and Stefan Pierrou (Ph.D.), Team Leader Cell & Molecular Biology, at AstraZeneca AB Lund 2002-04-03. Some additional information was collected from the company's homepage, www.astrazeneca.com.

AstraZeneca AB is one of the world's leading pharmaceutical companies. The company has a strong research base that over the years has provided medicines designed to fight disease in different medical areas; cancer, cardiovascular, central nervous system, gastrointestinal, infection, pain control, and respiratory. The research is divided into seven different research areas (called RA:s). The company has worldwide operations and more than 54 000 employees.

AstraZeneca AB performs all steps of the research required, from target discovery to finished drug. Within this chain of operations, bioinformatics and computer science help the process in a number of ways. AstraZeneca AB thus use all of the molecular biology methods described in chapter 4.3 (i.e. sequence analysis, pharmacogenomics, transcriptomics, structural genomics, and proteomics) and use bioinformatics tools to solve problems in these areas.

5.3.1 Bioinformatics Organisation and Strategy

Within AstraZeneca, bioinformatics is organized as a network, supported by a so-called EST-function (Enabling Science and Technology) with the overall responsibility for the bioinformatics knowledge base within the company. The EST-function for informatics is shared among three sites – Boston, Manchester, and Mölndal – where Manchester formally is the main site. In addition to the EST-function, every separate Research Area (RA) keeps a bioinformatics group. These local groups are interconnected throughout the company through informal networks. Using these networks the groups can exchange knowledge and retrieve software produced at different sites. This structure is a manifestation of AstraZeneca's desire to share information throughout the organization in order to boost progress.

Within each RA, the bioinformatics groups' main task is to solve bioinformatics problems locally. When a project needs help a request is sent to the bioinformatics group, where a decision is made if resources can and will be allocated.

AstraZeneca has several collaborations with leading biotech and bioinformatics companies, such as Affymetrix and GeneLogic. Collaborations with academic departments also exist, and are often in the form of sponsored Ph.D. or Postdoc programs. These collaborations often aim at developing specific bioinformatics tools.

5.3.2 Bioinformatics Tools and Databases

The software used within the company is mostly produced in-house, since few commercial companies are thought to offer products good enough to deal with the issues that need to be addressed. Thus, it is not so much money, but quality that is the issue if AstraZeneca is interested in purchasing software. Not much academic software (freeware) is used either. Academic software is often very specialized and not directly applicable for the applications needed within AstraZeneca. If academic software is used it often has to be modified in some way to fit into the structures at AstraZeneca.

AstraZeneca has a goal to have access to all databases containing information that can improve the research process. Therefore, the company has online access agreements with a number of commercial literature databases. This model of accessing information online reduces the amount of information that needs to be handled locally. To prevent competitors from tracking queries made by AstraZeneca, the company has chosen to keep local replicas of most sequence and structure databases within the company's firewalls. Additionally, AstraZeneca has two or three local databases accessible to all RA:s containing high throughput screening and chemical structure data. The framework for most of the databases kept locally is developed in-house.

5.3.3 Problems with Bioinformatics

One major problem is the lack of good commercial software that offers an overall solution to the issues within the research process. Much of the software offered to the market today is not complete and must be revised and brought together with other entities to form a whole. This creates a need for pharmaceutical companies to keep competence of how to construct usable bioinformatics solutions in-house.

Another problem for bioinformatics is the problem of annotation and standardization of data. This restrains the progress and creates duplication of work. AstraZeneca does presently not have an annotation standard within the company, but it might be on its way. Further causing duplication of work is the development of tools. Rarely the same software can be used at

separate departments since the need varies between departments. Therefore, much software must be modified to fit the need of each department.

Sweden has a shortage of competence in the bioinformatics area. At AstraZeneca it is believed that one of the major challenges for bioinformatics is to present concrete results of how bioinformatics can improve research. As long as there are few experienced bioinformaticians, this will prove to be a difficult task. However, AstraZeneca points out the importance of realizing that bioinformatics is not the sole solution to all problems related to research.

5.3.4 Bioinformatics in the Future

In the future, AstraZeneca aims at, to a greater extent, integrate information from all aspects of research. To do this interdisciplinary competence is needed. One requirement for AstraZeneca to know to what extent bioinformatics can affect their research process is that annotation standards exist and are implemented. One project aiming at solving some of the problems with annotations is the Gene Ontology project[52], an academic project that AstraZeneca supports financially.

It is believed that a wider usage of bioinformatics will develop in the future. Probably, bioinformatics tools will be commonly used in all parts of the research process. A wider usage and the ever increasing amount of data to be analyzed will demand better ways of structuring information and better algorithms for analyzing data.

[52] For more information, please visit www.geneontology.org

5.3.5 Case Summary

Table 5.4 presents a short summary of the case.

Table 5.4 Summary of empirical data for AstraZeneca AB

Description of study object	Large company
Organizational form	Formal group (supported by informal network and EST)
Bioinformatics collaborations With companies With academia	 Formal Formal (Ph.D. and postdoc sponsorships)
Bioinformatics tools In-house developed Freeware Commercial	 Mostly Some Some
Local replicas of large databases	Many

5.4 Biovitrum AB

The facts presented concerning Biovitrum AB is based on interviews held with Staffan Lake (Ph.D) Senior Scientist, and Per Johansson (Ph.D.), Bioinformatics Scientist at Biovitrum, 2002-03-06. Some additional information was obtained via e-mail from Sarah Hunter, Bioinformatics Group Coordinator, and from the company's homepage, www.biovitrum.se.

Biovitrum AB is a pharmaceutical R&D company performing research in three different lines of business, metabolic diseases, recombinant proteins, and blood plasma. Biovitrum AB is today one of Europe's largest pharmaceutical R&D companies and employs over 850 people.

Biovitrum AB is a spin off from Pharmacia Corporation (formerly Pharmacia & Upjohn), established in 2001. Pharmacia Corporation is still a minority shareholder, but only 10 % of Biovitrum AB's annual turnover is generated by business with Pharmacia Corporation. The company generally develops leads and sells these leads as they approach the phase of clinical testing. Among the molecular biology methods described in chapter 4, Biovitrum actively uses sequence analysis, transcriptomics, and structural genomics. Some research is also performed within pharmacogenomics, but on a secondary basis. Within all these methods, Biovitrum consider bioinformatics to play an important part.

5.4.1 Bioinformatics Organization and Strategy

Biovitrum's main objective concerning the use of bioinformatics is to actively incorporate bioinformatics in all parts of the research process. To achieve this the company continuously educates its personnel within the bioinformatics area. The company also makes a point of making the available information stored in biological databases part of the whole R&D process.

Biovitrum does not have a bioinformatics department in a formal sense. Rather, the company has a central bioinformatics group with an overview of all research areas, responsible for integrating bioinformatics wherever useful. The bioinformatics group members are all formally part of different R&D teams with the responsibility to introduce bioinformatics as a possible solution to problems. The underlying reason for this organizational form is a desire within Biovitrum to create a closer relationship between bioinformaticians and wet-lab scientists.

Biovitrum does not engage in any partnerships specifically addressing bioinformatics issues. The collaborations entered have a wider scope and often focus on biotechnology problems, whereof bioinformatics can be a part. These partnerships mainly include biotech companies but also academic departments are involved in some projects.

5.4.2 Bioinformatics Tools and Databases

A great deal of the bioinformatics software used at Biovitrum is freeware. Only a small portion of the software is actually bought and paid for. The people in the bioinformatics group are responsible for bioinformatics investments, and handle the process of integrating new software within Biovitrum. This is often done by downloading or buying newly derived algorithms and making them accessible via, for example, a web interface. Further, to get the most out of new software, the bioinformatics group educates the research groups that will use the new application.

When integrating new software the foremost important issue for Biovitrum is that the software can solve present-day problems. According to Biovitrum, commercial software lies a few steps behind the academic software. Therefore, Biovitrum to a greater extent chooses tools that have been developed within the academia to supply their needs.

The use of bioinformatics tools within Biovitrum's different research projects varies, but there are some general features. The research processes within Biovitrum almost always start at a gene level; interesting genes are identified with the help of bioinformatics tools, where after the gene's products are characterized according to function.

As a consequence of the gene level approach described above, database processing is central within the use of bioinformatics at Biovitrum. The company strives towards having local copies of databases with biological information, a reason being efficiency. In-house databases increase the speed of proposed queries. To always keep these local copies of databases up-to-date Biovitrum in several cases employ services that automatically update them. Biovitrum also holds private databases containing information about their own research. Some of these have been developed in cooperation with Prevas.

5.4.3 Problems with Bioinformatics

Generally, there are many improvements needed within all the technologies that Biovitrum uses today. All tools used today could be improved in one way or an other. From a wider point of view, one competence that Sweden is lacking today is good knowledge of how to build and maintain biological databases.

Another problem is the processing of biological data. The algorithms used today are not optimized and contain too many simplifications. Biovitrum therefore identifies theoretical knowledge as being a problem and not so much computational power.

5.4.4 Bioinformatics in the Future

Biovitrum has an objective to implement bioinformatics in every process within the company, and the company is currently looking into how this should be done. As a part of this project, the bioinformatics group is investigating general trends concerning bioinformatics and the implications they could have on how research is done. Biovitrum has not investigated to what extent bioinformatics could help shorten the drug development pipeline, but the company predicts that a more efficient utilization would result in more efficient drugs.

The bottlenecks within Biovitrum's R&D process have changed since the introduction of bioinformatics. The use of bioinformatics has solved several problems in the target identification and validation stages. A consequence of this has been that the investments have

shifted towards streamlining lead identification. This is an area where much improvement is needed, and bioinformatics can be a useful tool in accomplishing this. Already some of the obstacles within the lead identification process have been solved, thereby further shifting the bottleneck towards pure chemistry issues such as lead synthesis.

A trend that has been foreseen at Biovitrum is possible outsourcing of parts of the bioinformatics studies to smaller companies or the academia. This could help in the development of new tools that Biovitrum predicts will emerge within all areas of bioinformatics. In spite of the predicted progress in the field of bioinformatics, Biovitrum emphasizes that the potential of bioinformatics should not be overestimated.

A general viewpoint at Biovitrum is that much of the innovation within the bioinformatics field occurs in academia. The Center for Genomics and Bioinformatics and the Stockholm Bioinformatics Center (see chapter 4.1) are both considered world leaders in bioinformatics research. At Biovitrum, it is believed that the government should recognize the importance of such groups.

5.4.5 Case Summary

Table 5.5 presents a short summary of the case.

Table 5.5 Summary of empirical data for Biovitrum AB

Description of study object	Large company
Organizational form	Informal group
Bioinformatics collaborations	
With companies	None
With academia	None
Bioinformatics tools	
In-house developed	None
Freeware	Mostly
Commercial	Some
Local replicas of large databases	Many

5.5 Active Biotech AB

The facts presented concerning Active Biotech AB is based on interviews held with Mats Hansen (Ph.D.), Knowledge Manager, and Ann-Sofie Sjögren (Ph.D.), Applied Bioinformatics, at Active Biotech AB 2002-04-03. Additional information was found on the company's homepage, www.activebiotech.com.

Active Biotech AB was founded in 1996 when Pharmacia took a decision to no longer convey research within the areas of focus at Active Biotech AB. Active Biotech AB focuses its research on three areas of disease; autoimmunity/inflammatory, cancer, and infection. The goal is to be a leading player in these areas through global co-operation agreements and organic growth on selected markets. The company group has approximately 350 employees.

Active Biotech AB foremost uses proteomics methods in the target identification stage within their research, and consequently this area accounts for most of the use of bioinformatics tools. Moreover, bioinformatics tools are also used in cheminformatics and structural genomics.

5.5.1 Bioinformatics Organization and Strategy

Active Biotech does not have a formal bioinformatics department. Actually, it was first in 1999 that a discussion concerning bioinformatics was raised within the company. The responsibility for coordinating bioinformatics within Active Biotech lies under the knowledge management manager in the Scientific Affairs group. Within the same group, the company has created a position called Applied Bioinformatics. The idea is to support the different research projects by helping them apply bioinformatics tools where possible. In addition to the Applied Bioinformatics position, there exists an informal network with the responsibility to raise an awareness of bioinformatics within Active Biotech. The intentions are also to use the network to solve bioinformatics problems that arise, both practical and strategic.

One of the main objectives with bioinformatics at Active Biotech is for it to be a tool that makes it possible to take strategic decisions about research projects at an earlier stage, before heavy investments are done.

Active Biotech has a number of collaborations with academic departments. None of these, however, concern bioinformatics, mainly because the company does not strive towards having a continuous development of bioinformatics tools. If certain bioinformatics problems arise that cannot be handled in-house Active Biotech would sooner hire consultants to solve the problems than enter collaborations with academic departments.

5.5.2 Bioinformatics Tools and Databases

The tools used at Active Biotech are mostly commercial. The largest investment in software has been the purchase of the Genomics Computer Group (GCG) system. As supplement, some freeware and software accessible via the web are used. Almost no software is developed in-house since the company does not have the knowledge or resources to accomplish this.

The large international databases that are of importance for the research at Active Biotech are copied and stored locally to prevent competitors to monitor what kind of searches are performed. The local copies are updated through license agreements where needed.

5.5.3 Problems with Bioinformatics

One problem seen at Active Biotech is that the available commercial software today is not able to handle all the questions asked by researchers. There is definitely room for improvement on the software development side within bioinformatics.

According to Active Biotech, a common misunderstanding is the general view that bioinformatics is the solution to all problems. This overrates the expectations on bioinformatics. At Active Biotech they emphasizes the importance of realizing that bioinformatics is just a part of the research process, a tool, and not the solution to all problems.

Active Biotech has not seen a lack of competence within bioinformatics, but on the other hand, they have not tried to hire a great number of bioinformaticians.

5.5.4 Bioinformatics in the Future

Active Biotech will probably not start developing its own software in the future. Instead, the company will look to keep their in-house systems well structured so that introduction of new tools, developed elsewhere, will be facilitated. In order to help this process the intention is to bring the IT-department closer to the research process.

A trend believed to become more obvious in the future is that more and more of the information today freely available in databases through the Internet will become private and accessible only through purchasing licenses.

The future will probably reduce the belief that bioinformatics holds the solution to all problems. Also, it will be important to reduce the competence-gap between computer scientists and biologists. Therefore, education is considered to hold the key for future success.

5.5.5 Case Summary

Table 5.6 presents a short summary of the case.

Table 5.6 Summary of Empirical Data for Active Biotech

Description of study object	Small company
Organizational form	Informal network
Bioinformatics collaborations	
With companies	None
With academia	None
Bioinformatics tools	
In-house developed	None
Freeware	Some
Commercial	Mostly
Local replicas of large databases	Some

5.6 Medivir AB

The facts presented concerning Medivir AB is based on interviews held with Björn Ursing (Ph.D.), Bioinformatics Manager, Jonas Ekstrand (Dr.Med.Sc), Associate Director, and Peter Lind, Cheminformatics Manager, at Medivir AB 2002-03-22. A complementary interview was held via telephone with Katarina Jansson, Research Scientist Computational Chemistry and Cheminformatics, 2002-03-26. Additional information was found on the company's homepage, www.medivir.se.

Medivir AB is a pharmaceutical R&D company that focuses its research on infectious diseases and autoimmune disorders. Medivir AB develops compounds into new pharmaceuticals based on proteases and polymerases as target enzymes. The company is originally a spin-off from Astra and has approximately 170 employees.

Medivir's research focuses on developing compounds active against different proteases and polymerases, and therefore the bioinformatics tools used at Medivir mainly lie within cheminformatics and structural genomics. To a lesser extent, bioinformatics is used within the areas of sequence analysis and transcriptomics.

5.6.1 Bioinformatics Organization and Strategy

Medivir has not yet developed a clear strategy for how bioinformatics should be brought into use within the organization. To a great extent, this has to do with how research has been performed at the company and within what areas. Recently, however, new needs have arisen from the company's decision to widen its research. The new scope includes making target identification and validation a part of the research process thus increasing the need for bioinformatics. As a result, the company has hired a person responsible for facilitating and structuring the incorporation of bioinformatics into the organization. So far the new position has resulted in the development of an internal web service that aim at making the staff use the most appropriate bioinformatics tools. The website contain links to efficient tools available online together with information about the links contents.

In the areas of cheminformatics and structural genomics, in contrast to other bioinformatics areas, Medivir has longer experience and a more structured approach. The group responsible for these areas is a separate department called Cheminformatics and Computational Chemistry. The department's main task is to manage the chemical compound information stored in cheminformatics databases and to accelerate the screening process by using information about the structure of the target protein.

Medivir has never had any formal collaboration with the academia or companies dealing with bioinformatics. However, an informal connection with CGB exists through a part-time employee. There exist collaborations with academic departments concerning structure determination of proteins, but Medivir has recently decided to move structure determination in-house.

5.6.2 Bioinformatics Tools and Databases

Medivir holds structure as well as cheminformatics databases in-house. The in-house databases are built on commercial software that has been further developed within the company to fit the specific needs at Medivir. The information stored in the cheminformatics databases is to a certain extent bought from companies specialized in selling data for specific groups of leads. In addition, Medivir has online access to international structure databases. The company does not keep local gene sequence databases though, since the needs for this up till now has been limited.

Medivir uses commercial software in the structural genomics area. Supplementing this are some locally developed structural genomics software tools.

5.6.3 Problems with Bioinformatics

A problem seen at Medivir with many bioinformatics tools available on the market is that the solutions offered are seldom innovative. The tools are often based on miscellaneous freeware developed elsewhere, put together and presented via a new interface. In addition to this, the tools are often expensive.

Medivir sees a problem in the difficulty of creating functional structure databases for easy structure comparison. It is not possible to adequately represent 3D-structures with data strings. This makes comparison between structures difficult and better solutions in this area is needed.

More generally, Medivir sees a problem with the quality of the information stored in the available databases used for biological research. This creates problems and the need for verification of test results. In addition, Sweden lacks competent bioinformaticians to fill the needs of both the industry and the academia.

5.6.4 Bioinformatics in the Future

Medivir believe that the company will start developing some smaller tools in the coming future. In relations to Medivirs new line of research, new databases have to be designed and implemented. They will probably not purchase software but utilize freeware, mainly because of the high costs associated with commercial products.

An expectation at Medivir is that the wider incorporation of bioinformatics in the research process will increase the quality of the research conducted, through making it possible to discard unfit target candidates at an earlier stage.

5.6.5 Case Summary

Table 5.7 presents a short summary of the case.

Table 5.7 Summary of empirical data for Medivir

Description of study object	Small company
Organizational form	No group
Bioinformatics collaborations With companies With academia	 None Informal
Bioinformatics tools In-house developed Freeware Commercial	 Some Some Mostly
Local replicas of large databases	None

5.7 Presentation of Case Summaries

To summarize the case studies and to make it easier to get an overview of the material presented in this chapter Table 5.8 display the information from all the Case Summary chapters at the end of each case.

Table 5.8 Summary of all the Case Summary chapters

	Academic departments		**Large pharmaceutical companies**		**Small pharmaceutical companies**	
Name	**Department of Genetics and Pathology**	**Department of Molecular Evolution**	**AstraZeneca AB**	**Biovitrum AB**	**Active Biotech AB**	**Medivir AB**
Category	Academic department	Academic department	Large company	Large company	Small company	Small company
Organization	No group	No group	Formal group	Informal group	Informal network	No group
Collaborations Companies Academia	 None None	 Informal None	 Formal Formal	 None None	 None None	 None Informal
Tools In-house Freeware Commercial	 None Mostly Some	 Mostly Some None	 Mostly Some Some	 None Mostly Some	 None Some Mostly	 Some Some Mostly
Replicas of databases	None	Some	Many	Many	Some	None

6 Analysis

Chapter 5 has presented the usage of bioinformatics within academia and pharmaceutical companies in Sweden fulfilling the first part of the study's purpose. The second part of the purpose was to identify critical factors for further development of the technology and the industry. This analysis will first determine the maturity of the technology to then extract the critical factors. Further, proposed actions to meet the challenges put forward by the critical factors are presented followed by possible scenarios for the future.

6.1 Bioinformatics – the Maturity of the Technology

According to the Technology Life Cycle Model (see chapter 3.2), a technology's life cycle consist of three different phases, related to the maturity of the technology. Each of these phases state different conditions that need to be fulfilled to connect the industry to the correct phase. Below an analysis is performed to evaluate bioinformatics according to the conditions presented for each phase to determine the current maturity of the technology.

6.1.1 Emerging or Fluid Phase

At the beginning of an industry's life cycle, it is foreseen that there exists product and market uncertainty. The Swedish bioinformatics industry shows visible evidence of such market uncertainty. The conducted interviews display doubt about who leads the technology development and what roles academia, pharmaceutical companies, and bioinformatics companies should have respectively. The department of Molecular Evolution develops and has a need for very specialized products restricting their potential as a bioinformatics tools customer. The department of Genetics and Pathology on the other hand have more general needs and therefore a larger potential of becoming a bioinformatics tools customer. These differences make it evident that no one role for the academia exists. Further, the fact that much of present day bioinformatics investments are internal, and that many applications are based on freeware or in-house solutions, has made it difficult for bioinformatics companies to find their niche in the value chain.

Also, the interviews indicate an existing product uncertainty. There seems to be differences in what to expect of a product and what services should be included in the offer. The pharmaceutical and bioinformatics companies express different views, something that is demonstrated through the in-house adoptions sometimes needed when a pharmaceutical company buys a product.

Another proposed condition for the fluid phase is competition between the new technology and older technologies. In the case of bioinformatics, this competition can be anticipated through opposition against relying too heavily on bioinformatics for solving research problems, something stated by AstraZeneca and Active Biotech, and uncertainty in what way bioinformatics truly can improve the research process. At the department of Genetics and Pathology, competition between bioinformatics and old technologies such as every-day lab equipment and analyzing machines becomes apparent when prioritizing between investments.

The third condition for the fluid phase concerns low product quality. Here the case material presents a divided picture. Both good and bad products exist, but generally, most products leave much to be desired. Significant for bioinformatics products presently is that they mostly are too narrow in their applications, only solving specialized problems.

Interconnected with the product quality and product uncertainty are high costs and consequently high product prices according to the Technology Life Cycle Model. Mentioned in the interviews, with for example Medivir and the department of Genetics and Pathology, are high product prices. Additionally, the products bought in several cases need to be further developed in-house to fit the companies demands which renders extra costs.

The Technology Life Cycle Model predicts that the customers of a technology in the fluid phase mainly will consist of lead or high-income users. Within bioinformatics, two categories of lead users can be identified: academic departments developing their own tools and large pharmaceutical companies. Departments such as the Stockholm Bioinformatics Center (see chapter 4.1) and the department of Molecular Evolution (see chapter 5.2) develop front-edge tools to serve internal needs, and can therefore be considered lead users. However, these departments do not have sufficient funds to purchase externally produced software and are consequently not customers. The other category of lead users, i.e. large pharmaceutical companies, also constitutes the high-income users. However, our case studies reveal differences between, in this case, AstraZeneca and Biovitrum. AstraZeneca is a customer of bioinformatics products as well as a developer, whereas Biovitrum is a small customer without advanced in-house development, utilizing front-edge freeware. The smaller companies in the case studies started their bioinformatics investments later and are minor customers. The predictions of the model correlates fairly well considering customers, but noteworthy is that some customers also act as developers of in-house tools.

Finally, the model recognizes the fluid phase by a lack of standards. Several of the case studies show evidence of need for standardization. The annotation problem is the most evident, but also lack of standards for handling and presenting experimental data constitutes a problem.

6.1.2 Growth or Transitional Phase

The transitional phase is characterized by standardization of components, market needs, and product design. In the case studies, there has been little evidence of standardization. The products offered to the market are few, specialized and seldom compatible with other products available. The needs also differ, as the internal work methods diverge between users. Consequently, there is no standard for product design. On the other hand, the problem has been identified and efforts to deal with it exist.

The second condition proposed for the transitional phase is the development of a standard or common framework. To a certain extent, the case study material confirms efforts to develop data classification systems for the technology through initiatives such as the Gene Ontology project (see chapter 5.3.4), which deals with the annotation problem. Another initiative, although not mentioned in the case studies, is the I3C organization[53], which deals with software standardization issues.

An increasing customer base is the third and last condition identifying the transitional phase. The Swedish bioinformatics market has a small customer base but it has the potential to grow. So far, however, there have been few signs for this. The internal bioinformatics investments still widely dominate over external within academia and pharmaceutical companies, and the number of bioinformatics companies present on the Swedish market is small (see Appendix II).

6.1.3 Mature or Specific Phase

When a technology reaches the specific phase this will be signified by proliferation of products built around a common standard and a fading demand growth. The case studies demonstrate nothing that points towards proliferation of any standard or product. Further, the demand growth for products seems to be increasing more than fading.

[53] For more information, please visit www.i3c.org

6.1.4 Determining the Maturity of Bioinformatics

The discussion carried out above clearly shows that bioinformatics is an immature technology. The technology has nothing in common with the characteristics proposed for the specific phase according to the Technology Life Cycle Model. Neither a fading demand growth nor common standards exist. Some similarities can be found with the transitional phase, such as the initiatives that exist to develop common standards and working methods.

When comparing the theoretical characteristics of the fluid phase with the actual characteristics of bioinformatics technology it becomes evident that they show great likeness. There exists an uncertainty of roles, products, and standards on the market. Further, the products on the market do not always show high quality, the costs and prices connected to bioinformatics are high, the users are mainly lead or high income users, and there exist a competition between bioinformatics and older technologies. Bioinformatics is therefore found to be in the fluid phase of the technology life cycle, although some signals (see chapter 6.1.2) indicate that the first steps towards entering the transitional phase have been taken (see figure 6.1).

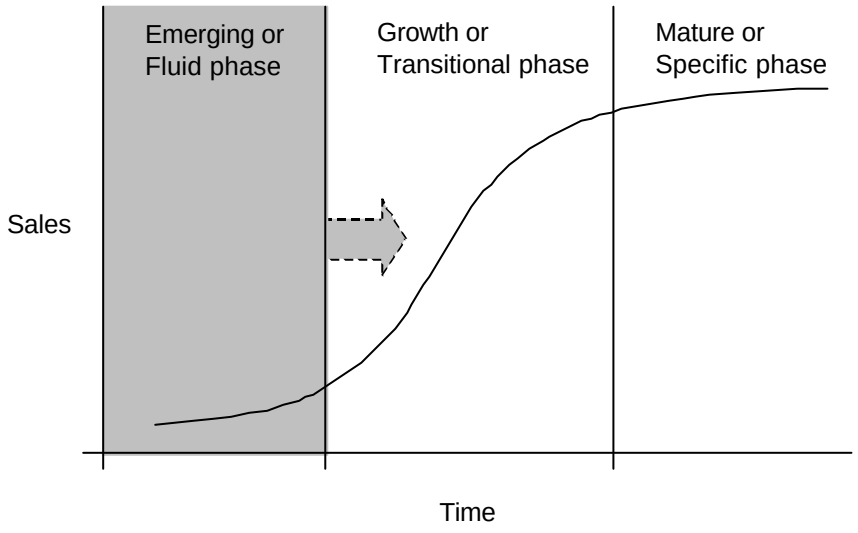

Figure 6.1 The current phase of the bioinformatics technology in Sweden.
Source: Authors research and analysis

6.2 Bioinformatics – Demands for Entering the Next Phase

It has been determined that the bioinformatics technology belongs to the fluid phase according to the Technology Life Cycle Model, but that some steps towards entering the transitional phase have been taken. What is needed for this progress to continue? Below, four

critical factors important for further development of the technology are presented and discussed.

6.2.1 Bridging the Cross-Competence Gap

One of the most frequently discussed issues during the case studies was the existing lack of people with the knowledge to bridge the gap between biology and computer science. The introduction of bioinformatics as a modern biology tool has created new conditions, making **solid biology knowledge insufficient** for taking part in the latest progress. A good biologist can come up with relevant questions or topics of study and can interpret the results, but is usually unable to comprehend the difficulties and possibilities when constructing software to solve questions or interpret data. Further, good computer scientists might be very skilled programmers, but without fully understanding the questions or what results the questioner is looking for, he or she will find great difficulty in constructing good analytical software.

The *bridging of the cross-competence gap* is the first identified critical factor.

6.2.2 Establishing Collaborations and Forums

A problem for the bioinformatics industry today is the prevailing product and market uncertainty. To clarify roles and relations between actors on the market and what to expect in a product it is significant to realize the value of *establishing collaborations and forums*. Collaborations so that useful algorithms and products can be developed in cooperation between actors, and forums for discussing more general issues such as trends for the technology and standards. Only as these partnerships develop the actors can carry through changes necessary for lessening the market and product uncertainties, increasing product quality, and solving other important issues for the industry.

6.2.3 Establishing Standards

According to the Technology Life Cycle Model, a crucial factor for entering the transitional phase is the development of common standards. This definitely applies to bioinformatics, a technology with great diversity in standards with none prevailing.

The problems with standards within bioinformatics are several. To start with there are no overall used standards for naming studied genes and proteins, a procedure called **annotation**. The annotation of a gene or a protein is based on molecular function, biological process, and

cellular component[54]. This results in some genes or proteins being given a number of names, since discovered by different researchers using different annotation systems. Further, in the case that a standard does exist, it is often species specific, making comparison with other species very difficult. Inter-species comparison is a useful tool when trying to elucidate mechanism or function for a newly discovered gene or protein. The annotation problem creates an uncertainty in what findings have already been done and also where information about these findings can be found.

Lack of data format standards is another problem. This involves lack of standards for how data is stored, how databases are structured, and how images are represented. Without standards for these applications, comparison of experimental data and sharing of data between different software is difficult. An example where lack of standards for representing images creates difficulties is within structural genomics (see chapter 4.3.4), the study of how the 3-D structure of proteins correlates to function. If structures are to be compared, they must be stored in the same data format. It is not enough to have software able of visualizing different data formats, at least not if more than a visual comparison is to be made. For a full comparison, a defined data format that saves the structures coordinates in a specific way is needed. Today, no such prevailing standard exist.

Additionally, there are no standards for how experiments should be carried out and consequently **how the results and information should be stored and interpreted** Examples are microarray (see chapter 4.3.3) and proteomics (see chapter 4.3.5) assays. In these types of experiments, two steps are central, the quantification of the experimental results (i.e. to represent the amount of protein or mRNA with a number) and the normalization of the results (for example defining a number that correlates to normal expression and putting present expression level in relation to this number). Presently, there are no standards for how these two central steps should be performed, making comparison of experimental results gained earlier or elsewhere very difficult.

The *establishment of standards* is a key issue for further development of bioinformatics.

[54] http://www.geneontology.org/#annotations, 2002-06-12

6.2.4 Demonstrate how Bioinformatics Add Value

When conducting the interviews, it became evident that pharmaceutical companies hold an uncertainty in what to expect from bioinformatics. A clear connection to *how bioinformatics add value* to the research process is needed for companies to motivate and increase their bioinformatics investments. Adding value can involve a number of things. **Firstly**, if it can be demonstrated that bioinformatics shorten the research process this can economically motivate investments and lead to development. **Secondly**, if bioinformatics is proven to be connected with an end product (i.e. drug) of higher quality, this would motivate not only the pharmaceutical companies but also bring fourth a social value in bioinformatics investments. **Thirdly**, if it was confirmed that the use of bioinformatics leads to better decisions about what research projects to fulfill, and what projects to dismiss, early in the development pipeline it would strengthen the technology's position. Only

6.2.5 Identified Critical Factors

The critical factors identified above are summarized in Table 6.1.

Table 6.1 Identified critical factors needed for further development of the bioinformatics technology

Identified critical factors

• Bridging of the existing cross-competence gap
  - *Merely knowledge in biology or computer science insufficient*
• Establishment of collaborations and forums
  - *Collaborations for product development*
  - *Forums for discussion of general issues*
• Development of standards
  - *Annotation*
  - *Data formats*
  - *How experimental results are presented*
• Clarification of how bioinformatics add value to the research process
  - *Shortening of the research process?*
  - *Higher quality end product?*
  - *Better decision-making?*

6.3 Bioinformatics – Proposed Actions

In order to meet the challenges put forward by the critical factors we here propose short-term and long-term actions for the actors in the Swedish bioinformatics sphere.

6.3.1 Actions for Bridging the Cross-competence Gap

For bridging the cross-competence gap, people with sufficient skills in both biology and computer science are needed. In the short-term, there are three possible approaches. **Firstly**,

the creation of groups consisting of both competences will facilitate knowledge exchange and understanding of the other science respectively. This is an approach that can be implemented presently and without further education within companies as well as academia. **Secondly**, academic departments, pharmaceutical companies, and bioinformatics companies can fill the cross-competence gap by hiring foreign competence. Money will be an important factor in attracting this competence. **Thirdly**, by educating unemployed or otherwise available computer scientists in biology the government could swiftly create a pool of bioinformaticians accessible for both companies and academia.

In the long run, **the gap must be bridged through education**, creating people that excel in both areas, or at least people with deep knowledge in one of the areas and understanding of the other. The recent establishment of academic bioinformatics centers and different educational programs[55] has definitely boosted progress. The government has an important role to play in assigning sufficient funds to universities and centers to secure a coming generation of Swedish bioinformaticians. Additionally, freestanding foundations must continue to support non-government supported initiatives. The academia has a responsibility to make the most of these funds by creating good educational programs and securing a high level of competence within the academia. If the academia is to have a good chance of doing this the value of not draining the academic bioinformatics knowledge base must be recognized.

Table 6.2 summarizes the proposed actions for bridging the cross-competence gap.

Table 6.2 Summary of proposed actions for bridging the cross-competence gap

Actor	Short-term actions	Long-term actions
Academia	• Create groups consisting of both biologists and computer scientists • Hire foreign competence	• Create good educational programs and secure a high level of competence
Bioinformatics companies	• Hire foreign competence	
Foundations		• Keep supporting non-government funded research areas
Government	• Educate available computer scientist in biology	• Assign sufficient funds for education and research

[55] See for example www.his.se/his/utbildning/program/bioinf.html and www.ibg.uu.se/bioinformatik

Actor	Short-term actions	Long-term actions
Pharmaceutical companies	• Create groups consisting of both biologists and computer scientists • Hire foreign competence	

6.3.2 Actions for Establishing Collaborations and Forums

In the short term, it is essential that the academia, bioinformatics companies, and pharmaceutical companies actively seek partnerships.

Initially, **pharmaceutical companies must determine if bioinformatics tools development should be a core business activity**. If not, decisions must be made to facilitate the spin-off of tool development and actions taken to initiate collaborations with tool developers to secure future bioinformatics needs.

Whatever decision pharmaceutical companies make, **bioinformatics companies must initiate collaborations** with pharmaceutical companies in order to get a clear view of what products are sought after and to define their role in the value chain. It is important that these collaborations are supported by, or involves, the pharmaceutical companies' management that holds knowledge of future investments and strategies.

The academia must review their bioinformatics strategy. Presently, most departments act only on their own behalf, not knowing if their solutions have already been produced elsewhere. A better way would be to have central functions at the universities that act as resource for bioinformatics issues. These should have knowledge within areas such as database management, programming and biology. Some of the benefits of this organizational form would be lesser duplication of work and economies of scale. Inter-university collaborations of this kind could come to play an important part in the development of bioinformatics usage within the academia.

Collaborations between companies, both bioinformatics and pharmaceutical, and the academia are also important. For companies it is a way of keeping up with the latest development and influencing this development to fit their needs. For academic departments this could mean financial support and could initiate development of saleable and more widely applicable tools. Companies ought to have great interest in collaborations with the academia and should therefore be the ones that initiate them.

For the long-term development, it is of interest for the actors on the market to join together in forums. The bioinformatics and pharmaceutical companies should initiate these forums and make efforts to include representatives from the academia. The academia should have interest in joining these forums since it offers an opportunity to influence the future of the technology. For starters, a forum could serve as a way to exchange ideas and views and discussing present and future trends. In the longer run, such a forum could develop into a trade association working for the bioinformatics industry's interests towards external actors, for example the government.

Table 6.3 summarizes proposed actions for establishing collaborations and forums.

Table 6.3 Summary of proposed actions in order to establish collaborations and forums

Actor	Short-term actions	Long-term actions
Academia	• Review their bioinformatics strategy and seek collaborations within university structures	• Seek memberships in forums
Bioinformatics companies	• Actively seek collaborations with pharmaceutical companies and the academia	• Initiate forums with pharmaceutical companies
Pharmaceutical companies	• Define if bioinformatics is a core business activity or not. If not, facilitate spin-offs and collaborations with bioinformatics companies. • Actively seek collaborations with the academia	• Initiate forums with bioinformatics companies

6.3.3 Actions for Establishing Standards

Development of standards is important and concerns not only Sweden but is a problem of global interest. Therefore, the goals set up must be internationally agreed on.

In the short-term, it is important that the leading academic departments, bioinformatics companies, and pharmaceutical companies strive towards common standards. These three actors must seek and initiate collaborations in order to agree on annotation standards, how data should be stored, how databases are structured, and how images are represented. It is essential that these collaborations remain in the long run since the technology will continue to develop, thereby creating a need for upgrading standards. It is especially important that the larger pharmaceutical companies and leading academic departments realize the importance of their actions, since they presently have great influence on the industry.

Table 6.4 summarizes the proposed actions for establishing standards.

Table 6.4 Summary of proposed actions for establishing standards

Actor	Short-term actions	Long-term actions
Academia	• Strive towards establishing international standards through collaborations	
Bioinformatics companies	• Strive towards establishing international standards through collaborations	
Pharmaceutical companies	• Strive towards establishing international standards through collaborations	

6.3.4 Actions for Demonstrating how Bioinformatics Add Value

So far, it has proven to be a challenge to clearly define in what ways bioinformatics adds value to the research process. It is essential in the short-term that bioinformatics companies are specific when explaining how their software solution shortens the research process, enhance end product quality, lead to better decisions or solves a problem thereby meeting the challenge of explicitly demonstrating how bioinformatics add value. Products too often seem to be marketed towards researchers with arguments addressing their needs. To achieve a wider approval of their bioinformatics solutions, **bioinformatics companies should concentrate their marketing efforts towards management**, with arguments designed for this group such as business cases and proof of productivity enhancements.

Academic departments can also contribute in clarifying the value of bioinformatics through publication of research results.

Table 6.5 summarizes the proposed actions for demonstrating how bioinformatics add value.

Table 6.5 Summary of proposed actions for demonstrating how bioinformatics add value

Actor	Short-term actions	Long-term actions
Bioinformatics companies	• More explicitly demonstrate how bioinformatics adds value	

6.3.5 Summary of proposed actions

The actions proposed in order to meet the challenges put forward by the critical factors are summarized in Table 6.6.

Table 6.6 Summary of proposed actions

Actor	Short-term actions	Long-term actions
Academia	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Review their bioinformatics strategy and seek collaborations within university structures • Strive towards establishing international standards through collaborations	• Create good educational programs and secure a high level of competence • Seek memberships in forums
Bioinformatics companies	• Hire foreign competence • Actively seek collaborations with pharmaceutical companies and the academia • Strive towards establishing international standards through collaborations • More explicitly demonstrate how bioinformatics adds value	• Initiate forums with pharmaceutical companies
Foundations		• Keep supporting non-government funded research areas
Government	• Educate available computer scientists in biology	• Assign sufficient funds for education and research
Pharmaceutical companies	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Define if bioinformatics is a core business activity or not. If not, facilitate spin-offs and collaborations with bioinformatics companies. • Actively seek collaborations with the academia • Strive towards establishing international standards through collaborations	• Initiate forums with bioinformatics companies

6.4 Bioinformatics – Possible Future Scenarios

The future for bioinformatics is uncertain, but interesting to discuss. A way of envisioning possible developments is through scenario analysis. By viewing the current status of the market, the critical factors summarized in Table 6.1, and the proposed actions outlined in Table 6.6 possible future scenarios can be outlined.

It has been stated in chapter 6.1 that there does not exist a clear distinction between being a customer and a developer, a market uncertainty associated with the fluid phase in the technology life cycle. Most customers on today's Swedish bioinformatics market develop

tools in-house in one way or the other. To be able to map the current status of the market the uncertainty of the customer/developer role was used. A quadrant referred to as the *customer-developer quadrant* was constructed (see figure 6.2) in which market actors were placed according to current conditions. The quadrant is also used to visualize future changes in the relationships between being a customer or a developer when presenting the scenarios.

However, the results from the case studies made it evident that a different classification than the one used for the case studies was needed for the scenario analysis. Academic departments have therefore been divided into two different categories, *developing* and *non-developing academic departments* referring to whether advanced bioinformatics tools are developed within the department or not. The pharmaceutical companies have also been divided into two categories called *big pharma* and *pure research companies*. Big pharma represent companies that carry out all parts of the drug development pipeline (see figure 1.1), while pure research companies relate to companies that only perform parts of the drug development pipeline. The case studies showed that a company performing all parts of the development pipeline differed from the others in its bioinformatics usage in that they had more advanced in-house development and were larger customers. Further, *bioinformatics companies* are also represented (by a spot) in the quadrant to show how they are affected by market changes. Figure 6.2 shows the present state of the customer-developer quadrant.

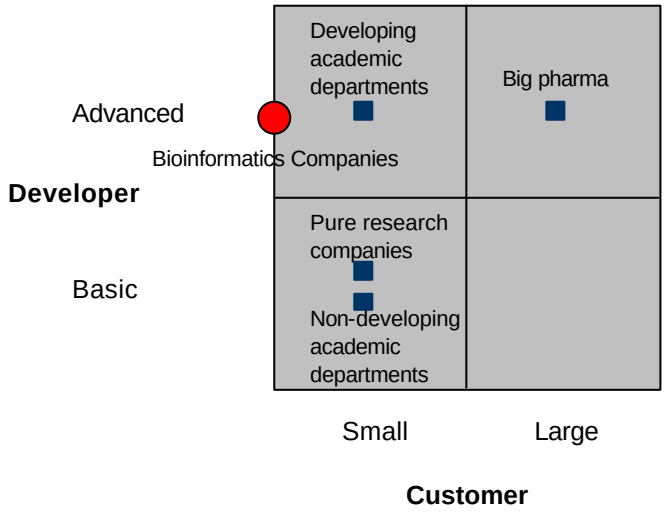

Figure 6.2 A graphical representation of case study objects state as customers and developers of bioinformatics, the customer-developer quadrant.
Source: Authors' research and analysis

6.4.1 Scenario 1: A Low Profile

In a proposed low profile scenario, big pharma choose to continue to keep most of their bioinformatics investments in-house. This implies that these companies will depend on having strong competence in-house. With such competence, collaborations with other actors on the market will be of minor importance since big pharma will not have the same need for engaging in bioinformatics collaborations. Keeping bioinformatics investments in-house could also affect the cross-competence gap negatively in the sense that big pharma will drain the universities knowledgebase when hiring competent bioinformaticians to secure internal needs. All the above prevents the establishment of a freestanding bioinformatics industry and lessens the need for developing more than just company standards. With no international standards, it will be difficult to develop widely compatible software, making it hard for pure research companies and non-developing academic departments to integrate externally produced software in a wider sense. They will remain small customers since the software offered to the market is non-compatible with each other and demands in-house adoptions, something that will be too costly to be done repeatedly. This implies that it will be hard for pure bioinformatics companies to enlarge their customer base.

Since there will be a limited number of commercial software available the use of freeware will continue to play an important role not only within the academia but also within companies. Using in-house competence to adapt freeware to internal structures within big pharma, pure research companies, and non-developing academic departments will be common. Developing academic departments will not be noticeably influenced by this scenario, but will continue as today.

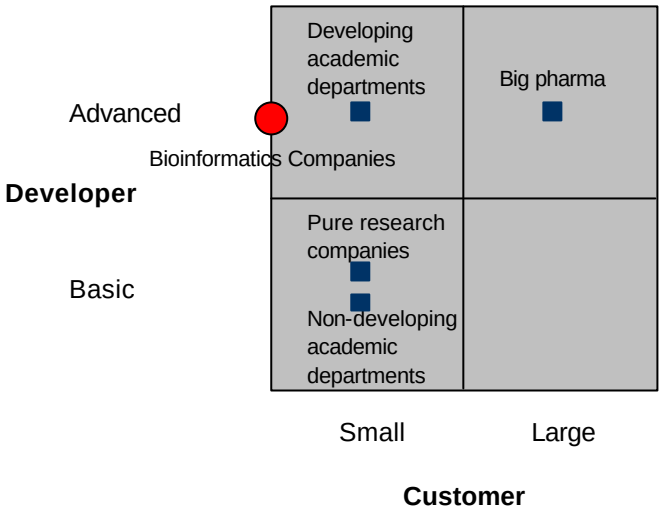

Figure 6.3 A low profile scenario will imply no or small movements within the customer-developer quadrant.

Without wider compatible tools, bioinformatics will continue to be a tool among others within the research process. A low profile scenario will imply no or small movements within the customer-developer quadrant (see figure 6.3).

6.4.2 Scenario 2: A Slow Start

For the development of bioinformatics, it is of importance that standards are determined, as has earlier been observed. In a slow start scenario, big pharma and the developing academic departments will strive towards agreeing around and developing standards through engaging in collaborations of different kinds. Both these actors therefore play an important role for this scenario. As collaborations are formed and standards slowly emerge, it will become easier for companies to adopt their internal systems to common annotation and software standards. This facilitates the possibility for pure research companies to become larger customers of bioinformatics products (see figure 6.4). This also applies for non-developing academic departments.

Big pharma, however, will in an initial stage continue to develop advanced software in-house, at least until their needs can be supplied by the growing market. Initially, big pharma will therefore continue to drain the market and universities in search for people to fill their in-house needs. As big pharma observes that the growing market can fulfill their needs there will be a gradual downsizing of these companies' internal development of bioinformatics tools. At this point, it will be of increasing importance for pharmaceutical companies to enter into strategic collaborations with developing academic departments and bioinformatics

companies to keep up to date with the latest innovations and trends. This development will free knowledge, and could help fill the cross-competence gap.

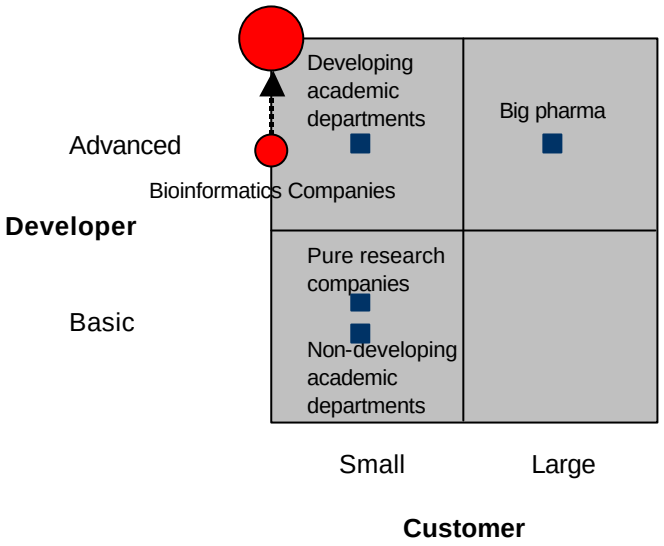

Figure 6.4 A slow start scenario will imply initial movement by pure research companies and non-developing academic departments towards becoming larger customers. At a later stage, represented by a dashed arrow, big pharma will lessen their in-house development, moving towards becoming pure customers. The bioinformatics companies will slowly increase in size, number, and level of attainment.

The slow start scenario suggests the establishment of a bioinformatics market evolving at a slow pace mainly because of big pharma's will of keeping internal advanced development of tools. As the market grows, more bioinformatics companies will emerge and the existing ones will have an opportunity to grow in size. This development will also lead to production of more advanced tools of higher quality, making it possible to more explicitly clarify how bioinformatics add value to the research process.

6.4.3 Scenario 3: A Rapid Expansion

As in the slow start scenario, standards will be of importance for a rapid expansion scenario to happen. The differences will be the pace of the development of standards and the strategy of big pharma.

Common annotation and software standards will evolve fast, as big pharma and developing academic departments agree on basic frameworks through discussions and collaborations. This will facilitate the establishment of a fast growing bioinformatics market, since standards will open the possibility for producing compatible wide range software. As this happens, both

the pure research companies and the non-developing academic departments will grow as customers, as a result of more and better available products.

A contributory cause to the market development will be big pharma´s strategic decision to gradually phase out the advanced in-house development of bioinformatics tools. When decreasing the in-house competence within bioinformatics, it will be important for pharmaceutical companies to enter into strategic collaborations with developing academic departments and bioinformatics companies. Another consequence of the phasing out of in-house development could be spin-offs in the form of new bioinformatics companies.

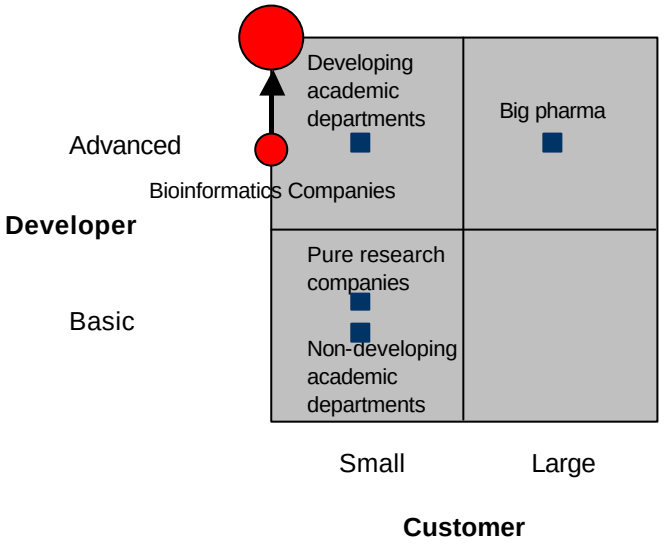

Figure 6.5 A rapid expansion scenario will imply movement by pure research companies and non-developing academic departments towards becoming larger customers. Also, big pharma lessen their in-house development, moving towards becoming pure customers. This could lead to spin-offs in the form of bioinformatics companies, a development represented by the dashed arrow. Bioinformatics companies will grow fast in size, number, and level of attainment.

In a rapid expansion scenario, the bioinformatics market will grow at a fast pace and cause large structural changes. Bioinformatics companies will grow fast, in size, number, and standard of attainment (see figure 6.5). The attitude towards bioinformatics will also change, giving bioinformatics a more important role in the research process.

6.4.4 Discussion Concerning the Scenarios

All three scenarios present possible future outcomes for bioinformatics. However, we believe that the most probable scenario is the second one, a slow start. This mainly due to the fact that the technology is still immature and the changes needed have only just started. It is likely

that big pharma will continue to keep advanced development of tools in-house for quite some time. Also, the development of standards has so far been a slow process and both these factors restrain the growth of a freestanding industry. Additionally, it will take time to generate enough people with competence to bridge the cross-competence gap.

6.5 Window of Opportunity

So what is needed for a faster development than the one presented in scenario two? We believe that the entry of a third party could be the catalyst needed for a faster development of the technology and industry.

A third party with the right circle of contacts could identify demands, map suppliers, and build a network consisting of business actors. The most important issue for a third party would therefore be to engage in partnerships with the right business actors. **Firstly**, pharmaceutical companies must be part of the network for enabling identification of prevailing and future demands. **Secondly**, it is necessary that bioinformatics software producers become part of the network. The ambition should be to engage as wide a spectrum of producers as possible to be able to present broad, but customized, solutions. **Thirdly**, venture capital companies are needed since they possess funds necessary for development of companies, which will enhance the network's possibilities. **Fourthly**, leading academic researchers should be included since they add front-edge knowledge that could be used for, for example, consulting services.

The role for the third party would be to become the single point of contact between customers and producers in the bioinformatics market. Pharmaceutical companies presently concentrate on in-house solutions for bioinformatics issues. This is due to the hardship in finding bioinformatics software producers providing broad solutions and a difficulty within the pharmaceutical companies to define their needs. A third party has the possibility to **view the pharmaceutical companies from without and help in defining needs**. Through their network they could then provide knowledge to meet these needs. Bioinformatics software producers on the other hand often seem too small to keep up to date with the needs of the pharmaceutical companies and the solutions provided are often narrow. An active third party can help overcome these obstacles by **introducing bioinformatics companies to the right customers and possible strategic partners**. Further, management capabilities that are sometimes lacking in these smaller companies could be introduced. Also, the third party can

use their channels to present solutions and anchor ideas at a higher management level than presently done by bioinformatics companies.

From these observations, we see an opportunity for a third party to catalyse the technology's development, consisting of two parts. **Number one**, offer pharmaceutical companies management an evaluation of their internal bioinformatics usage and needs. This evaluation can be used to, together with the company, form a specification of requirements. To address the specification, a group of actors form the network that hold complementary competences can be put together. **Number two**, offer producers of bioinformatics software help in probing the market and defining the needs of pharmaceutical companies. Also, help these companies in finding the right customers and strategic partners, and in promoting their products. Additional aid in raising venture capital could also be included in the offer. The benefit of this way of catalysing the development is that it enables a more dynamic way of developing bioinformatics software than can be done internally or by sole bioinformatics software producers. It can also facilitate the creation of more extensive solutions and widen the technology knowledge base through co-operations.

7 Conclusions

This report has described the Swedish bioinformatics market and portrayed the usage of bioinformatics within Swedish academia and pharmaceutical companies. The study has determined bioinformatics as an immature technology, with great potential for improvement. For continued development of the technology four critical factors have been identified:

- Bridging of the cross-competence gap between biology and computer science
- Establishment of collaborations and forums between actors on the market
- Development of standards
- Clarification of how bioinformatics add value to the research process

In our view, the future of bioinformatics in Sweden depends on the progress of these factors.

What can be done for favorable development of these critical factors? Bioinformatics has already affected the way research is performed and will continue to do so in the future. The question is how much and in what way. We believe that the critical factors and decisions concerning these factors are vital for the future development. Table 7.1 summarizes actions proposed for a favorable future development.

Table 7.1 Proposed actions for actors connected to Swedish bioinformatics sphere for a favorable future development

Actor	Short-term actions	Long-term actions
Academia	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Review their bioinformatics strategy and seek collaborations within university structures • Strive towards establishing international standards through collaborations	• Create good educational programs and secure a high level of competence • Seek memberships in forums
Bioinformatics companies	• Hire foreign competence • Actively seek collaborations with pharmaceutical companies and the academia • Strive towards establishing international standards through collaborations • More explicitly demonstrate how bioinformatics adds value	• Initiate forums with pharmaceutical companies
Foundations		• Keep supporting non-government funded research areas
Government	• Educate available computer scientists in biology	• Assign sufficient funds for education and research
Pharmaceutical companies	• Create groups consisting of both biologists and computer scientists • Hire foreign competence • Define if bioinformatics is a core business activity or not. If not, facilitate spin-offs and collaborations with bioinformatics companies. • Actively seek collaborations with the academia • Strive towards establishing international standards through collaborations	• Initiate forums with bioinformatics companies

We have further concluded that the entry of a third party on the bioinformatics market could help catalyze the technology's development. A third party could connect the right market actors thereby creating bioinformatics solutions with a wider scope.

Acknowledgements

We would foremost like to thank our supervisors at Accenture, Magnus Isaksson and Patrik Nylander, for their valuable guidance and encouragement. We would also like to thank Matti Nikkola, responsible for degree projects at the Biology Education Center in Uppsala, and our examiner Per Kraulis, at Stockholm Bioinformatics Center, for all help. Finally, we would like to thank Leif Andersson, Bengt Anell, Siv Andersson, Allan Asp, Jonas Ekstrand, Carolin Frank, Hans-Henrik Fuxelius, Fredrik Granberg, Mats Hansen, Sarah Hunter, Katarina Jansson, Per Johansson, Rolf Johansson, Staffan Lake, Peter Lind, Jan S Nilsson, Kalle Näslund, Stefan Pierrou, Anna Polgren, Jack Robinson, Bo Servenius, Ann-Sofie Sjögren, Björn Ursing, Martin Waleij, Alexandra Waluszewski, Frida Wennerström, Ulrika Wickenberg, and everybody else who made this report possible.

Oskar Bosson and Niclas Riml, June 2002, Uppsala

References

Articles

Afuah, Allan N. & Utterback, James M., (1997) "Responding to Structural Industry Changes: A Technological Evolution Perspective", *Industrial and Corporate Change*, vol. 6:1, pp. 183-202

Persidis, Aris, (1999) "Bioinformatics", *Nature Biotechnology*, vol. 17, pp.828-830

Reichhardt, Tony, (1999) "It's sink or swim as a tidal wave of data approaches", *Nature*, 399, pp. 517-520

Saviotti, Paolo P. *et al.*, (2000) "The changing marketplace of bioinformatics", *Nature Biotechnology*, vol. 18, pp. 1247-1249

Venter, Craig, *et al*., (2001) "The Sequence of the Human Genome", *Science*, vol. 291, pp. 1304-1351

Printed reports

Bioinformatics in Stockholm and Uppsala (2002), Centre for Medical Innovations at Karolinska Institutet. p. 29

Biotekniken inför 2000-talet, SOU 2000:103, p. 25

Executive Summary (2001). Virtual Genetics Laboratory AB

Integration, Ernst & Young's Eighth Annual European Life Sciences Report 2001. Ernst & Young

Internal seminar (2001). Accenture

Life Science Informatics (2001). UBS Warburg

Why Do Prescription Drugs Cost So Much…(2000). Pharmaceutical Research and Manufactures of America

Published books

Kotler, Philip *et al.*, (1999) *Principles of Marketing*, Prentice Hall Europe, Second European Edition

Lewin, Benjamin, (2000) *Genes VII*, Oxford University Press

Interviews

Andersson, Leif, *Professor*, Department of Pet Genetics, SLU and Founder of Arexis AB, (e-mail) 2002-03-08

Andersson, Siv, *Professor*, Department of Molecular Evolution, Uppsala University, 2002-02-21 and 2002-03-27

Asp, Allan, *Product Manager*, Pyrosequencing AB, (e-mail) 2002-03-05

Ekstrand, Jonas, *(Dr.Med.Sc.) Associate Director*, Medivir AB, 2002-03-22

Frank, Carolin, *Ph.D. Student*, Department of Molecular Evolution, Uppsala University, 2002-04-09

Fuxelius, Hans-Henrik, *Programmer*, Department of Molecular Evolution, Uppsala University, 2002-03-27

Granberg, Fredrik, *Ph.D Student*, Department of Genetics and Pathology, Uppsala University, 2002-03-13

Hansen, Mats, *(Ph.D.) Knowledge Manager*, Active Biotech AB, 2002-04-03

Hunter, Sarah, *Bioinformatics Group Coordinator*, Biovitrum AB, (e-mail) 2002-03-19

Jansson, Katarina, *Research Scientist Computational Chemistry and Cheminformatics*, Medivir AB, (telephone) 2002-03-26

Johansson, Per, (*Ph.D.) Bioinformatics Scientist*, Biovitrum AB, 2002-03-06

Johansson, Rolf, *(Ph.D.) Director Support Office*, AstraZeneca AB, Södertälje, 2002-03-20

Kraulis, Per, *Senior Scientist*, Stockholm Bioinformatics Center, 2002-02-15
Lake, Staffan, *(Ph.D.) Senior Scientist*, Biovitrum AB, 2002-03-06
Lind, Peter, *Cheminformatics Manager*, Medivir AB, 2002-03-22
Nilsson, Jan S., Former President of the Wallenberg Foundation, 2002-05-14
Näslund, Kalle, *Ph.D Student*, Department of Genetics and Pathology, Uppsala University, 2002-03-13
Pierrou, Stefan, *(Ph.D.) Team Leader Cell & Molecular Biology*, AstraZeneca AB, Lund, 2002-04-03
Polgren, Anna, *Scientific Secretary*, Affibody AB, (e-mail) 2002-02-19
Robinson, Jack, *Managing Director*, Virtual Genetics Laboratory AB, (e-mail) 2002-02-18
Servenius, Bo, *(Ph.D. & Associate Professor) Bioinformatics Scientist*, AstraZeneca AB, Lund, 2002-04-03
Sjögren, Ann-Sofie, *(Ph.D.) Applied Bioinformatics*, Active Biotech AB, 2002-04-03)
Ursing, Björn, *(Ph.D.) Bioinformatics Manager*, Medivir AB, 2002-03-22
Waleij, Martin, *CEO*, Biobridge Computing AB, (telephone) 2002-02-25
Wickenberg, Ulrika, *M.Sc.*, Department of Genetics and Pathology, Uppsala University, 2002-02-27

Homepages

web1.ebc.uu.se/molev, 2002-05-06
www.activebiotech.com, 2002-04-01
www.amershambiosciences.com, 2002-06-13
www.astrazeneca.com, 2002-05-06
www.biobridge.se, 2002-02-25
www.biovitrum.se, 2002-03-11
www.cbg.ki.se, 2002-02-26
www.genpat.uu.se, 2002-03-18
www.globalgenomics.com, 2002-06-13
www.hugo-international.org, 2002-05-07
www.linnaeus.bmc.uu.se, 2002-02-26
www.medivir.se, 2002-03-20
www.mindtools.com/swot.html, 2002-03-01
www.prevas.se, 2002-02-25
www.pyrosequencing.se, 2002-02-25
www.sbc.su.se, 2002-02-26
www.spotfire.se, 2002-06-13
www.vivaldi.zool.gu.se/swegene, 2002-02-26

Appendix I – Questionnaire

General questions
Describe your background, work experience, and current position.

Describe the company/department. Within what areas are research performed?

Describe a typical research project. What parts of the drug development pipeline are included in the company's research?

Bioinformatics Organization and Strategy
Why does the company/department use bioinformatics?

How is the bioinformatics usage organized within the company/department?

How has bioinformatics altered the research process at your company/department?

Does your company/department engage in collaborations concerning bioinformatics with companies or academic departments?

Does your company/department outsource any bioinformatics projects?

Bioinformatics Tools and Databases
How is bioinformatics used within your company/department? What technologies are used?

Does your company/department use external or internal databases, or both?

Does your company/department develop bioinformatics tools in-house? Are any tools bought? Is freeware used?

Problems with Bioinformatics
What problems are seen with bioinformatics within your company/department, and in general?

In what way is information shared within different parts of your company/department?

Does your company/department have a standard for storing the data produced in the research process?

Bioinformatics in the future
In what way can bioinformatics change the research process within your company/department?

What needs are seen for the future?

How can the government stimulate a continuous development of bioinformatics in Sweden? Has the government so far done enough?

What are the visions and strategies for the future at your company/department?

Appendix II –The Bioinformatics Sphere in Sweden

Presented below is an overview of the bioinformatics sphere in Sweden, containing academic departments and commercial companies that in one way or the other produce bioinformatics tools. This overview is not all embracing in the sense that it undoubtedly contains all bioinformatics departments and companies in Sweden, but the major players are presented here.

Academic Centers

Center for Genomics and Bioinformatics

The Center for Genomics and Bioinformatics (CGB) was established as a new department at Karolinska Institutet (KI) in 1997. Presently, CGB harbors four different research units; transcriptomics, bioinformatics, genomics technologies, and clinical genomics. Each unit consists of several research groups that are interconnected in various projects. About 100 people currently work at CGB.[56]

The bulk of CGB's initial investment budget was directed towards establishing genomics technology and bioinformatics platforms not previously present at KI. This has resulted in the establishment of core facilities for all of KI in the areas of bioinformatics (KISAC), DNA sequencing (KISeq), microarrays (KIChip), and (in planning) SNP analyses (KISNiP).[57]

CGB as a department on KI Campus is a direct result of a research cooperation agreement between Pharmacia & Upjohn (P&U) and KI, reached in May 1997. The CGB is operated as any other academic department, while presently receiving about half of its funding from the P&U company. Some CGB scientists are also P&U employees allowing for the utilization of proprietary tools and databases (Celera, Incyte etc.).[58]

Linnaeus Center for Bioinformatics

The Linnaeus Center for Bioinformatics was founded in July 2001, and is located at BMC, Uppsala University. It is an inter-university collaboration between the Swedish University of Agricultural Sciences, SLU, and Uppsala Universitet, the Faculty of Science and Technology.

[56] www.cbg.ki.se

[57] *ibid.*

[58] *ibid.*

The main areas of research are genome analysis, structural biology, mathematics and computer science, medicine, pharmacy, and agriculture.[59]

The Wallenberg Foundation has provided most of the funds for the center.[60]

Stockholm Bioinformatics Center

Stockholm Bioinformatics Center (SBC) is a collaboration project between Stockholm University (SU), the Royal Institute of Technology (KTH), and Karolinska Institutet (KI). SBC was founded in January 2000 on an initiative from the Foundation for Strategic Research (SSF). The SSF provides the main funds during an initial five-year period. Additional support comes from the Knut and Alice Wallenberg Foundation as well as SU, KTH and KI.[61]

The two main goals of SBC are "…to create a "critical mass" of internationally competitive bioinformatics research and methods development…" and "…to provide a strong environment for the training of Ph.D. and Diploma students in bioinformatics, including advanced PH.D.-level courses…".[62] SBC currently houses nine research groups that cover different problem areas in bioinformatics.

Swegene

Swegene is a research and technology association, initiated by a consortium formed by Lund University, Lund Institute of Technology, Göteborg University and Chalmers University of Technology. Swegene's purpose is to promote research in transcriptomics at an internationally competitive level.[63]

The central activities within Swegene are divided into four technology research areas; genetic variability and biological diversity, phenotype profiling, structural biology, and bioinformatics. Swegene aims at using its resources to establish new technologies at freely accessible resource centers, available to scientists at the collaborating universities.[64]

[59] www.linnaeus.bmc.uu.se

[60] *ibid.*

[61] www.sbc.su.se

[62] *ibid.*

[63] vivaldi.zool.gu.se/swegene

[64] *ibid.*

Companies that produce and sell Bioinformatics Tools

Affibody AB

Affibody AB was founded in 1998 and is based on patented technologies developed at KTH and KI. The bases for Affibody's technology platform are small stable proteins called Affibodies™, which through combinatorial protein chemistry can be modified to bind to a specific target protein. One application for these small proteins is the use of these proteins to guide drugs to a specific area in the body. The company has 65 employees at the present time.[65]

With the use of the Affibody-proteins, Affibody offers an integrated overall solution for proteomics studies. The technical platform also includes specializes databases and enables high throughput analysis of proteoms. During the year 2001 Affibody acquired the bioinformatics company Visual Bioinformatics AB, thus further increasing the bioinformatics competence in the company. The bioinformatics field of business within Affibody AB are developing software packets for sales, and are also developing systems specifically design to fit into the proteomics projects within Affibody AB.[66]

Affibody AB has developed a software application called GeneWeaver, a program for gene expression analysis. GeneWeaver can be used for handling data from both sequencing-based and hybridization-based experiments. A project under development is a software application for analysis of proteomics information.[67]

Global Genomics

Professor Patrik Ernfors and Dr Sten Linnarsson founded Global Genomics AB in September 2000, in co-operation with Karolinska Innovations AB. The company's goal was commercialization of their gene profiling method. According to the company, no method available today can adequately, quickly, and to reasonable cost meet the needs of genomic research. It was in response to this challenge the company's product Tangerine was developed.[68]

[65] Polgren, Anna, Interview (2002)

[66] *ibid.*

[67] *ibid.*

[68] www.globalgenomics.com

Tangerine is a product that enables gene expression analysis services over the web. Customers submit RNA samples and receive the complete analysis, downloadable from the Internet. Tangerine is based on advanced in-house developed bioinformatics solutions.[69]

Spotfire Inc.

Spotfire is a provider of eAnalytic applications. The company offers capabilities for accessing and analyzing enterprise-wide data, housed in disparate formats and locations, and for transforming this data into important decisions. The company's European headquarters lies in Gothenburg, Sweden.[70]

Spotfire has four core products that are offered to the market; Spotfire DecisionSite, Spotfire DecisionSite Statistics, Spotfire DecisionSite Map Interaction Services, and Spotfire DecisionSite Developer. Additionally, Spotfire provides professional services including consulting, training and support programs to ensure that customers achieve maximum value from their Spotfire investment.[71]

Spotfire have partnerships with industry leaders such as Affymetrix, Celera, Incyte, MDL, CAS, and IBM.[72]

Virtual Genetics Laboratory AB

Virtual Genetics Laboratory views themselves as a "cross solution provider" combining skills in Life Sciences, Advanced Mathematics and Information Technology". The company was founded in 1997 and has at present time 17 employees. The products offered are search engines and data analysis tools, which has now entered the commercialization phase. As a support to these products, Virtual Genetics offers consultant services for support and customization of the products. They currently offer two products to the market, Virtual Adapt for Text Mining and Virtual Predict for Data Mining.[73]

---

[69] *Bioinformatics in Stockholm and Uppsala* (2002). Centre for Medical Innovations, p. 29

[70] www.spotfire.se

[71] *ibid.*

[72] *ibid.*

[73] *Excecutive summary* (2001). Virtual Genetics

Virtual Genetics has sold their products to Celera Genomics and AstraZeneca. The company has partners in Celera Genomics and Prevas Bioinformatics.[74]

Companies that produce and sell Bioinformatics Tools as a Byproduct

Amersham Biosciences

Amersham Biosciences was founded in 1997 through a merger of Amersham Life Sciences and Pharmacia Biotech. The company has a goal to be the world leader in providing integrated systems and solutions for disease research, drug development and manufacture.[75]

In Uppsala, Amersham Biosciences has started a center of excellence focused on bioinformatics. At the center, efforts are made for developing algorithms and data analysis tools for accurate differential analysis of several thousands of proteins simultaneously. Also, algorithms for cell segmentation and feature analysis are being developed together with the Center for Image Analysis at Uppsala University.[76]

Pyrosequencing AB

Pyrosequencing AB was founded in 1997. The founders were two professors at the Royal Institute of Technology who had discovered a way of sequencing DNA through detecting free pyrophosphate. After only two years of research and development, the product, PSQ96, was introduced to the market in November 1999. So far, the company has sold over 150 systems worldwide and today the company employs approximately 140 people.

The development of bioinformatics tools have so far taken place in collaboration with Prevas, but Pyrosequencing is looking to develop its own tools in the future and are therefore building their own software department to have all knowledge in-house.

Pyrosequencing has a strategic partner in Sumitomo Corp. of Japan.[77]

[74] *Excecutive summary* (2001). Virtual Genetics

[75] www.amershambiosciences.com

[76] *Bioinformatics in Stockholm and Uppsala* (2002). Centre for Medical Innovations, p. 26

[77] *Life Science Informatics* (2001) UBS Warburg, p. 141

Companies that perform Consulting Services within the Bioinformatics Area

BioBridge Computing AB

BioBridge Computing AB was founded in 2000. The founders were three theoretical physicists. The company's current activities focus on developing and marketing products concerning proteomics, where they work on products for data analysis of mass spectroscopy data. The company's aim is to "develop and market automated software in the field of bioinformatics with the aim to support high-throughput research for candidate drugs". BioBridge Computing AB at present time employ 11 people.[78]

BioBridge Computing AB offers two products to the market. Pepex, a fully automated peak extraction software for Maldi Tof mass spectrometers, and PIUMS, a fully automated software for identification of proteins using Maldi Tof mass spectrometry. Additionally the company offers custom software development, installation, and training.[79]

BioBridge Computing AB has both industrial and academic partners such as AstraZeneca AB and Swegene.

Prevas AB

Prevas Bioinformatics offers services within business development, software design and aftermarket products within the Life Science area. Since 1993, Prevas has been providing its services to companies within this market, but it was first in the year 2000 that the business area Prevas Bioinformatics was formed. The idea is to specifically meet the Life Science customers' needs. The business area focuses on customers active within biotechnical and pharmaceutical research as well as medical technology.[80]

One of the products offered by Prevas is the BioFrame software that helps structure information concerning development of projects.[81]

[78] www.biobridge.se

[79] *ibid.*

[80] www.prevas.se

[81] *ibid.*