

Molekularbiologie, Genetik, Bioinformatik: Neue Serviceangebote in medizinischen Bibliotheken am Beispiel „Helix Helper“

Nicola Gaedeke, Berlin

Im Vortrag wird die Entwicklung eines Netzwerkes aus Informationsspezialisten für den Bereich der Lebens-Wissenschaften vorgestellt. Das Netzwerk setzt sich vorwiegend aus Fachpersonal medizinischer Bibliotheken US-amerikanischer Hochschulen zusammen. Es gewährleistet einen „User Support Service“ für Wissenschaftler zur Anleitung und Unterstützung für die Benutzung biowissenschaftlicher Datenbanken, die Analyse von molekularbiologischen Daten und die Benutzung von Software der Bioinformatik. Der Bedarf eines solchen Services wurde im Jahre 2001 vom Natl. Center for Biotechnology Information (NCBI) ermittelt. Kurze Zeit später wurde das erwähnte Netzwerk vom NCBI initiiert und der „NCBI's Advanced Workshop for Bioinformatics Information Specialists“ entwickelt. Teilnehmer am Workshop werden durch „credits“ von der „Medical Library Association“ honoriert. Die Zahl an Medizinischen Bibliotheken, die einen solchen Service einrichten, wächst jährlich. Helix Helper, der „User Support Service“ für Biowissenschaftler an der Universität von Utah, bietet Unterstützung für die Wissenschaftler durch Fortbildungskurse, Beratung, Auftragsrecherchen und Sequenzanalysen an.

Medical libraries that are operating within a large research campus have begun to offer a Bioinformatics User Support Service for scientists in the field of the life sciences. The information specialists give support in database searches and the use of data-analysis-software. The National Center for Biotechnology Information itself is interested in the development of a network of Bioinformatic User Support People to ensure the efficient use of the complex resources offered by the institute. In January 2001 a group of medical library administrators, medical librarians, and NCBI staff discussed the need of an Advanced Course on NCBI tools and initiated NCBI's Educollab Project. Shortly after the meeting a team of librarians and scientists was formed to develop a 5 days course for health sciences and science librarians with the goal for the participant to be able to recognize and understand the various data types resulting from research in the biosciences and to identify the appropriate resources to answer the patrons/users question. Since August 2002 the course is being held yearly at NCBI. Credits of the Medical Library Association (MLA) can be earned.

The number at medical libraries, which offer such a service, grows annually. Helix Helper, the User Support Service for bio-scientists at the University of Utah, offers support for the scientists by further training courses, consultation, information retrieval and sequence analyses.

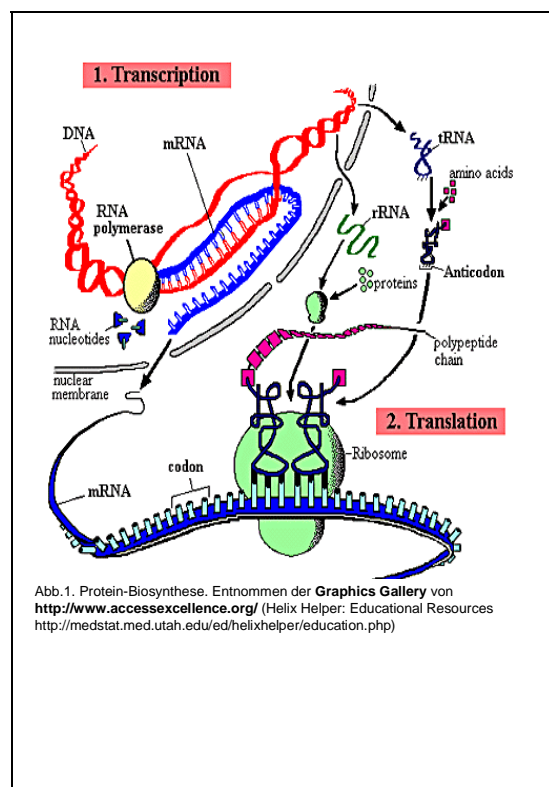
HelixHelper und Co

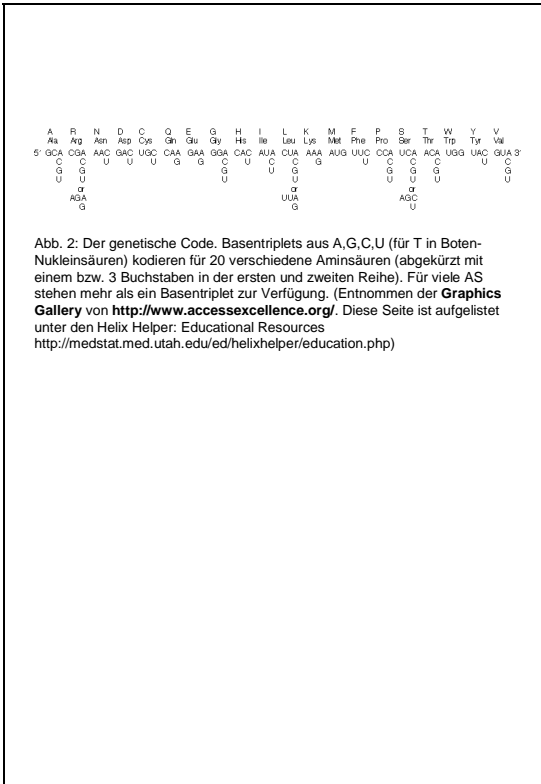
Moderne Methoden der Biologie ermöglichen die Entschlüsselung der gesamten Erbsubstanz eines Organismus. Das so genannte Genom besteht aus Desoxyribonukleinsäure (DNA), die sich aus vier verschiedenen chemischen Bausteinen, den Nucleotiden mit den Nucleotidbasen Adenin (A), Guanin (G), Cytosin (C), und Thymin (T), zusammensetzt. Die Reihenfolge, d.h. die Sequenz dieser Bausteine, repräsentiert die genetische Information des Organismus. Inzwischen sind die Genome von 197 einzelligen Organismen, und 26 Genome mehrzelliger Organismen vollständig bekannt. Viele Projekte zur Entschlüsselung weiterer Genome sind in Arbeit. Im Jahre 2001 veröffentlichten die Zeitschriften *Nature*¹ und *Science*² die vollständige Entschlüsselung des menschlichen Genoms. $3,5 \times 10^9$ bp (Basenpaare) wurden sequenziert. Im Vergleich dazu beträgt die Genomgröße des Bakteriums *Escherichia coli* $4,5 \times 10^6$ bp (Basenpaare). Der Prozess der DNA-Sequenzierung führt zu einer solchen Datenmenge, dass eine computer-

gestützte Aufarbeitung der Daten notwendig geworden ist.

Zusätzlich zu den genomischen Daten sind für den Wissenschaftler auch die Informationen über das Proteom, d.h. über alle vorhandenen Proteine in einem Organismus, von großer Bedeutung. Die chemischen Bausteine der Proteine sind Aminosäuren (AS). Wie schon bei den Genen, ist auch bei Proteinen die Reihenfolge der Bausteine ausschlaggebend für die Information des Moleküls. Die Aminosäuresequenz leitet sich von der genetischen Information ab. Jeweils drei Basen kodieren für eine Aminosäure (Abb. 1). Die Umsetzung der genetischen Sequenz in die Proteinsequenz ist in Abb. 2 dargestellt. Die Transkription des Genoms

und die Translation der Boten-Nukleinsäure sind die zentralen Schritte bei diesem Vorgang. Viele genetische





Krankheiten beruhen auf eine falsche Abfolge an Nucleotidbasen, und der daraus resultierenden falschen Aminosäuresequenz. Auch Proteinsequenzen, Proteinfamilien, veränderte genetische Information (Polymorphismen, Mutationen), Botenstoffe und genetische Krankheiten sind in Datenbanken hinterlegt. Für die Erkennung von kodierenden Sequenzen im Genom sind

Computerprogramme entwickelt worden. Viele Datenbanken und Sequenz-analyse-Programme stehen u.a. über das „World Wide Web“ zur Verfügung stehen. Der Bedarf von Seiten der Wissenschaftler im Bereich der Lebenswissenschaften nach einer Unterstützung im effizienten Umgang mit diesen „Tools“ steigt. Aufgrund dieser Entwicklung haben verschiedene medizinische Bibliotheken US-amerikanischer Hochschulen einen „Bioinformatics User Support Service“ zur Unterstützung der biowissenschaftliche Forschung eingerichtet.

„Bioinformatics User Support“ – Wirklich nötig?

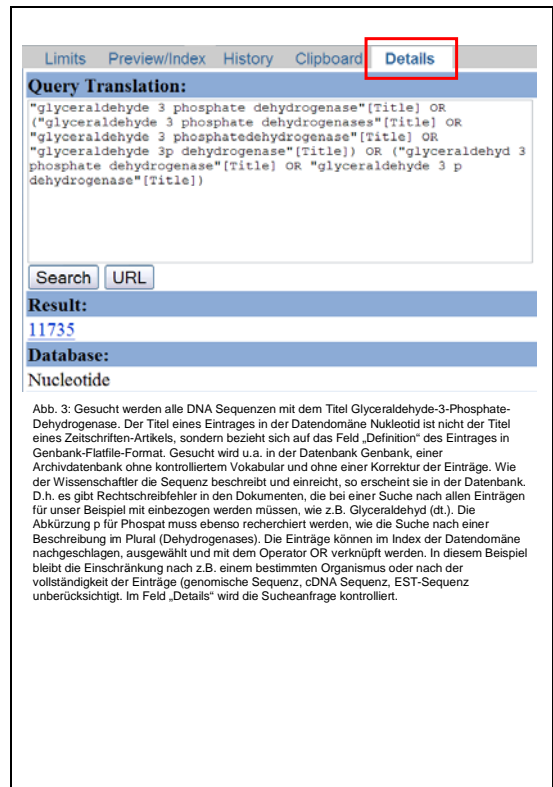
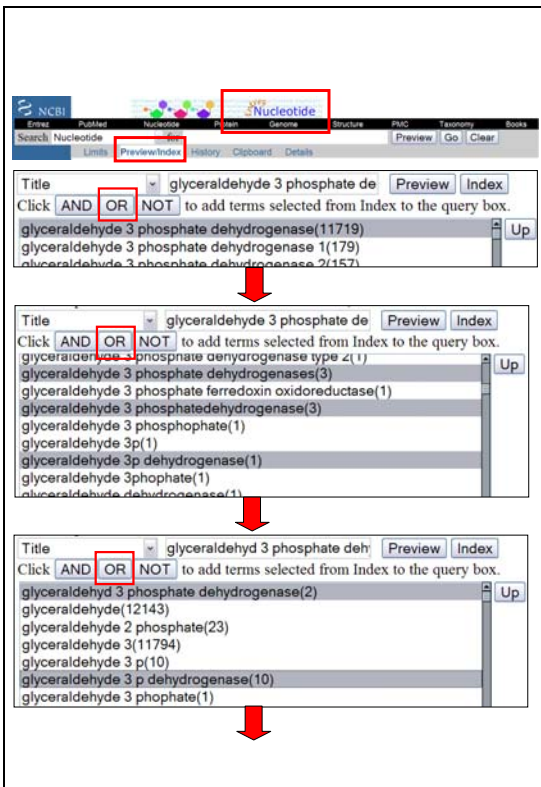
Drei Gründe sprechen für eine Erweiterung des Service-Angebots für Wissenschaftler aus dem Bereich der Lebenswissenschaften:

1. Die Datenmengen, die aus Sequenzierungsprojekten akkumulieren, wachsen aufgrund von

High-Throughput-Methoden exponentiell.

GenBank, das Sequenzdepot aller öffentlich zur Verfügung stehenden Nucleotid und Protein Sequenzen des Natl. Center for Biotechnology Information (NCBI), enthält z.Z. über 37,3 Millionen Sequenzeinträge (Aug 2004). Die Anzahl der Einträge in GenBank verdoppelt sich ca. jährlich. Bei Datenbanken dieser Größe ist es schwierig, die gewünschten Informationen herauszufiltern, wenn der Benutzer die Suchanfrage nicht präzise formuliert. Zusätzlich ist die Suchanfrage aufgrund eines fehlenden kontrollierten Vokabulars in biowissenschaftlichen Datenbanken erschwert. Abb. 3 veranschaulicht eine Suche in der Entrez Datendomäne der Nucleotidsequenzen unter Berücksichtigung von Rechtschreibfehlern für die Gensequenz von Glyceraldehyde-3-Phosphate-Dehydrogenase.

2. Für die Recherche in Datenbanken stehen viele verschiedene und komplexe Suchoberflächen wie z.B. Entrez (NCBI), BLAST (NCBI) und SRS (EMBL-EBI), zur Verfügung. Für eine Suche nach ähnlichen Sequenzen zu einer gegebenen Proteinsequenz gibt es auf der Suchoberfläche des Basic Local Alignment Search Tools (BLAST) allein über 20 Felder mit z.T. mehreren Einstellungsmöglichkeiten, um eine Suchanfrage zu präzisieren (Abb. 4).



NCBI Advanced Workshop for Bioinformatics Information Specialists: An Educational Collaboration.

<http://www.ncbi.nlm.nih.gov/Class/NAWBIS/>

Nicola Gaedeke¹, Kristine M. Alpi², Jennifer Lyon³, Donna Messersmith⁴, Mark Minie⁵, Janet A. Ohles⁶, David L. Osterbur⁷, Michele R. Tennant⁸, Renata Geer⁹

¹Spencer S. Eccles Health Sciences Library (Salt Lake City, UT), ²Weill Medical College (Cornell University, NY), ³Eskind Biomedical Library (Nashville, TN), ⁴The KEVRIC Company, ⁵Health Sciences Libraries (University of Washington, WA), ⁶Western Connecticut State University (Danbury, CT), ⁷Biological Labs Library at Harvard University (Boston, MA), ⁸Health Science Center Libraries (University of Florida, Gainesville, FL), ⁹National Center for Biotechnology Information (Bethesda, MD)

Abb. 7: Mitarbeiter des „NCBI-Educollab Projects“ zur Entwicklung Workshops für Informationsvermittler für Bioinformatik (Überschrift eines Posters zur Vorstellung des NAWBIS-Workshops auf der IFLA 2003 in Berlin).

Kurses (Abb. 7). Dieser Kurs, NCBI's Advanced Workshop for Bioinformatics Information Specialists (NAWBIS), behandelt die Datenbank-Recherche für biowissenschaftliche Datenbanken und die Anwendung der am NCBI entwickelten Software zur **A n a l y s e** molekularbiologischer Daten.

Der Kursaufbau ist modular. Der Kursinhalt soll hier aber nur in Beispielen wiedergegeben werden (mehr Informationen s. <http://www.ncbi.nlm.nih.gov/Class/NAWBIS/modules.html>):

Modul 1: Information Hubs (für **Molekularbiologische**

Datenbanken und Software) – in diesem Modul werden Datenbank- und Ressourcensammlungen vorgestellt und im Detail erläutert. Darüber hinaus lernt der Teilnehmer das effektive Suchen mit *Entrez*, der datenbank-übergreifenden Suchmaschine für über 20 der am NCBI zur Verfügung stehenden Datendomänen. (Eine Datendomäne kann aus mehreren Datenbanken bestehen, wie z.B. die Domäne

„Nucleotide“, über die man u.a. in den Datenbanken GenBank, Refseq und GSDB (Genom Sequence Data Base) recherchiert).

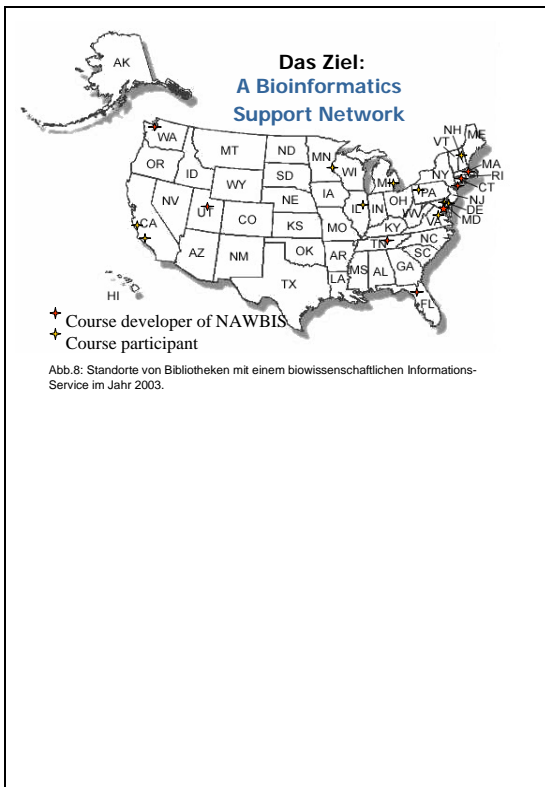
Modul 2: Sequenzähnlichkeitssuchen – dieses Modul lehrt die Anwendung des Basic Local Alignment Search Tools (BLAST), BLink (BLAST Link), sowie BLAST ähnlichen Programmen, wie z.B. von RPS (Reversed-Position-Specific)-BLAST und die Suche nach konservierten Proteindomänen.

Modul 3: Karten und Kartierung des menschlichen Genoms – dieses Modul beinhaltet die Analyse von Chromosomenkarten, sowie das Manipulieren und die Weiterbearbeitung dieser Karten. Zusätzlich werden die von den Karten verknüpften Ressourcen mit weiteren Informationen über das gesuchte Gen vorgestellt.

Modul 5: Ressourcen für klinisch arbeitende Genetiker – dieses Modul behandelt die Suche in Datenbanken und andere Ressourcen, die Informationen über genetische Krankheiten und deren klinischen Aspekte enthalten.

Die Zielgruppe des Kurses sind Bioinformatik-Informationsspezialisten, die in den USA in den medizinischen Bibliotheken arbeiten und dem Wissenschaftler einen Unterstützung durch aktive Informationsvermittlung sind. Im

diskutieren. Über eine kurze Umfrage bei den medizinischen Bibliotheken der USA wurde der Ist-Status eines solchen Services abgeklöpft und analysiert (Abb. 6). Kurze Zeit später wurde das NCBI-Educollab-Project ins Leben gerufen. Renata Geer, Mitarbeiterin des NCBI, rekrutierte schon amtierende Bioinformatik-Informationsspezialisten zur Entwicklung eines 40-stündigen



August 2002 fand der Kurs erstmalig in den Schulungsräumen des NCBI in Bethesda statt. Seitdem wird er jährlich in einer aktualisierten Form angeboten. Zurück am Arbeitsplatz in der medizinischen Bibliothek bietet ein aktives Netzwerk (Bioinformatic Support Network, BSN) bestehend aus Kursteilnehmern und Kursentwicklern, dem Informationsvermittler weiterhin Unterstützung auf dem Gebiet der angewandten Bioinformatik. Das Netzwerk erstreckt sich über die gesamte USA (Abb. 8) und Deutschland (vertreten durch mich).

„Helix Helper“, der „Bioinformatics User Support Service“ an der Spencer S. Eccles Health Sciences Library, Utah University (Salt Lake City)

<http://medlib.med.utah.edu/helixhelper/index.php>
Helix Helper, das Informationsportal für die Biowissenschaftler der Universität von Utah, wurde im Jahre 2001 etabliert. Hierdurch hat die Bibliothek sowohl durch die Mitarbeit in der Abteilung für Informationsvermittlung („Reference“), als auch über Öffentlichkeitsarbeit und ein umfangreiches Kursangebot ihr Service-Angebot für biowissenschaftliche Fragen erweitert. In den folgenden 2 Jahre hat sich ein eigenständiges Angebot entwickelt, dass sowohl im Bereich Informationsvermittlung, als auch im Bereich Fortbildungen („Educational

Services“) angesiedelt werden kann (Abb.9 und 10).

Weitere Beispiele für einen „Bioinformatics User Support Service“ anderer US-amerikanischer Bibliotheken sind:

1. Guide for Molecular Biology and Genetics - University of Pittsburgh Health Sciences Library System

<http://www.hsls.pitt.edu/guides/genetics/>

2. Molecular Biology Resources - Research Informatics Consult Service, Vanderbilt University

<http://www.mc.Vanderbilt.Edu/biolib/services/rics/molecular.html>

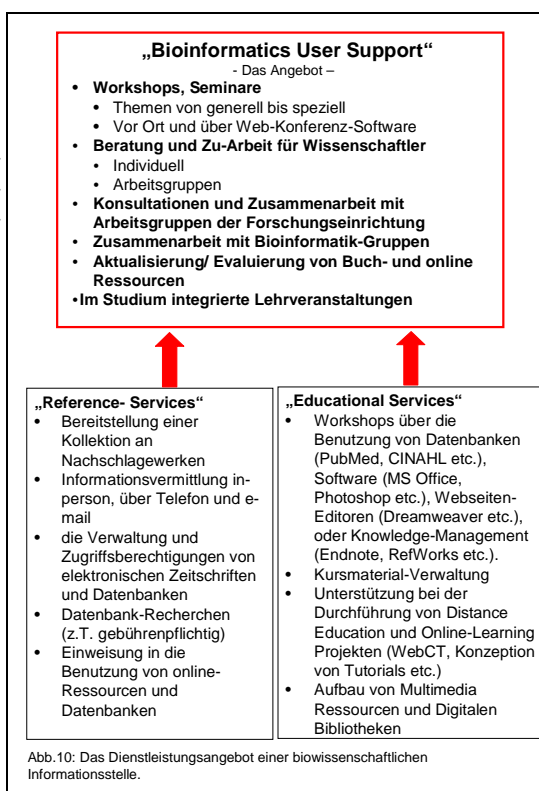
Der Aufbau eines „Bioinformatic User Support Service“ an der Bibliothek ... und überregional

A) Für die Einschätzung der Service-Notwendigkeit sollten folgende Fragen im Vordergrund stehen:

- Welche Abteilungen könnten von einem „Bioinformatics User Support“ profitieren?
- Welche Ressourcen kennen die Mitarbeiter in diesen Abteilungen?
- Welche Ressourcen werden bevorzugt verwendet?
- Welche Unterstützung/Service gibt es bereits für die Mitarbeiter?
- Wie würde sich der „neue“ Service von den anderen unterscheiden?

B) Für die Kommunikation eines Informations-Service für Molekularbiologie und (klinische) Genetik haben sich folgende Hilfsmittel (in Utah) bewährt:

- Webseite erstellen
- Konsultationen, Laborbesuche
- Aktive Beteiligung am Help-Desk/ Service Point
- Workshops entwickeln und anbieten
- Verankerung von Lehrveranstaltungen im



Kurrikulum naturwissenschaftlicher Studiengänge

C) Für die Etablierung eines überregionalen Service-Angebotes gewährleisten moderne Informationstechnologien, wie z.B. Konferenzschaltungen und/oder Video-Software, zusätzlich die Durchföhrung von Online-Workshops. Die Projektion und Kommunikation für einen Workshop ist selbst „über den großen Teich hinweg“ (von Berlin nach Salt Lake City) problemlos (Abb. 11).

Die Situation in Deutschland

Eine aktive Beratungsfunktion über die Recherche in Datenbanken übernehmen an manchen Instituten die Wissenschaftlichen Informationsdienste. Hier werden hauptsächlich Literatur Recherchen durchgeführt und eine Unterstützung in der Verwaltung bibliographischer Daten angeboten. Beispiele hierfür sind die Wissenschaftliche Informationsstelle im Lehrstuhl Bioinformatik der Biologisch-Pharmazeutischen Fakultät der Universität Jena (<http://pinguin.biologie.uni-jena.de/>) und die Informationsvermittlungsstelle für die Bio.-Med. Sektion der Max-Planck-Gesellschaft (<http://www.biochem.mpg.de/iv/>).

Aus meiner Sicht ist es an der Zeit, Wissenschaftler aus dem Bereich der

Mannheim 2004

Lebenswissenschaften einen umfangreichen Service für Fragestellungen bezüglich Datenbank-Recherchen und Daten-Analysen anzubieten. Informationsfachleute für den Bereich der Bioinformatik können die Fragestellung des Anwenders einordnen und Lösungswege aufzeigen.

Literatur:

- THE HUMAN GENOME. *Science*. 2001, 291:1145-1434.

- THE HUMAN GENOME. *Nature*. 2001, 409: 745-964.

- Planning Bioinformatics Education and Information Services in an Academic Health Sciences Library. W. John MacMullen, K.T.L. Vaughan, Margaret E. Moore College & Research Libraries, July 2004, 65(4): 320-33.

- Alpi K. Bioinformatics Training by Librarians and for Librarians: Developing the Skills Needed to Support Molecular Biology and Clinical Genetics Information Instruction. *Issues in Science & Technology Librarianship*. Spring 2003, 37.

- Yarfitz S, Ketchell DS. A library-based bioinformatics services program. *Bull Med Libr Assoc* 2000 Jan; 88(1): 36-48.

Dr. Nicola Gaedeke

- BioTools.info -

Gutzmannstr. 8

14165 Berlin

Tel.: +49 (0)30 - 84 72 39 94

E-Mail: ngaedeke@biotools.info

Web: <http://www.biotools.info>