1. Einleitung

1.1. Trypanosoma brucei

1.1.1. Schlafkrankheit

Die Schlafkrankheit (Afrikanische Trypanosomiasis) ist eine auf dem afrikanischen Kontinent grassierende Seuche, an der jährlich über 50.000 Menschen erkranken (Scientific-Working-Group-African-Trypanosomiasis, 2001). Ohne Behandlung ist die Schlafkrankheit tödlich. Ursächlich für die Schlafkrankheit ist die Infektion mit dem einzelligen Parasiten *Trypanosoma brucei* (Abb. 1.1 A). *T. brucei* wird durch den Stich einer infizierten Tsetsefliege (*Glossina spec.)* auf den Menschen übertragen (Abb. 1.1 B).

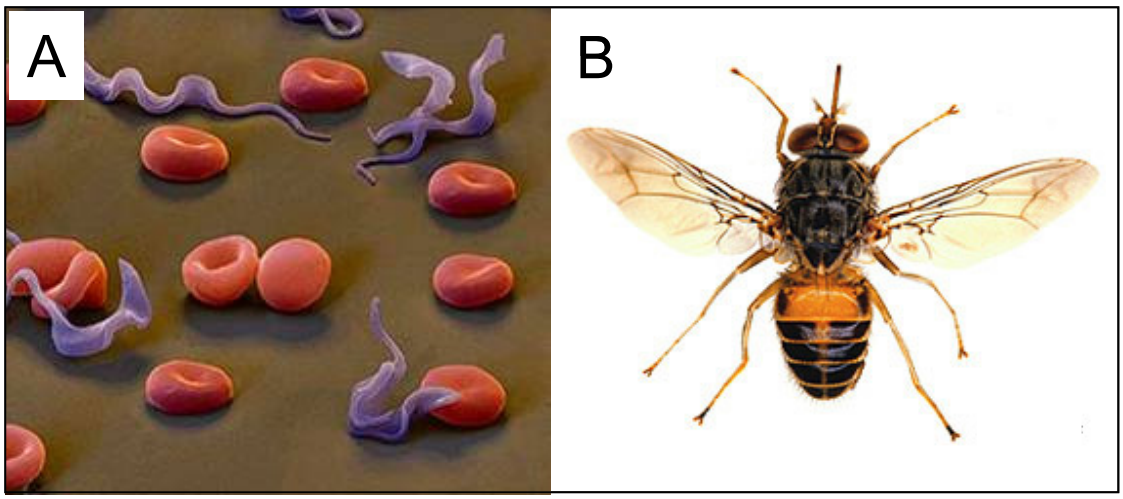

Abb. 1.1 *Trypanosoma brucei* und Tsetsefliege (*Glossina sp.*). (A) Elektronenrastermikroskopaufnahme von *T. brucei* (Blutbahnform) und Erythrozyten. Die Aufnahme wurde nachträglich eingefärbt. (B) Überträger der Schlafkrankheit. Die Trypanosomen werden durch den Stich der Tsetsefliege mit dem Speichelsekret auf den Menschen übertragen (Quelle: Dr. Steve Mihok).

Nachdem man die Schlafkrankheit in den siebziger Jahren schon besiegt geglaubt hatte, begann daraufhin eine neue epidemische Periode, die bis heute andauert. Abbildung (Abb. 1.2 A) gibt einen Überblick über die Zahl der gemeldeten Erkrankungsfälle. Der Hauptgrund für den Anstieg der Erkrankungsfälle liegt in der politischen Situation in vielen afrikanischen Ländern, die vielerorts zum Zusammenbruch des Gesundheitssystems geführt hat. Wegen ungenügender Erfassung, fehlender Diagnosemöglichkeiten und der Unzugänglichkeit vieler betroffener Gebiete liegt die tatsächliche Zahl der Infektionen nach Schätzungen der Weltgesundheitsorganisation (WHO) wahrscheinlich etwa zehnfach höher als die dokumentierten 50.000 Fälle (WHO-Report, 2001). Bedingt durch die Verbreitung der

Tsetsefliege, ist das Vorkommen der Schlafkrankheit auf das Gebiet südlich der Sahara und nördlich der Kalahari-Wüste beschränkt (Abb. 1.2 B). In den betroffenen Gebieten leben etwa 60 Millionen Menschen. Die Schlafkrankheit kommt in 36 Staaten vor, darunter sind 22 der ärmsten Länder der Erde, wie zum Beispiel Sudan, Uganda, Demokratische Republik Kongo und Angola. Im Verbreitungsgebiet der Tsetsefliege sind ebenfalls alle Rinder, Schafe und Ziegen von der Infektion bedroht. Bei der tierischen Form der Trypanosomeninfektion spricht man von der Nagana-Seuche. Durch die Nagana-Seuche ist in vielen betroffenen Gebieten die Rinderhaltung – und damit die Milch- und Fleischproduktion – unmöglich. Dadurch stellt die Trypanosomiasis neben dem Krankheitsrisiko ein erhebliches ökonomisches Problem für betroffenen Länder dar.

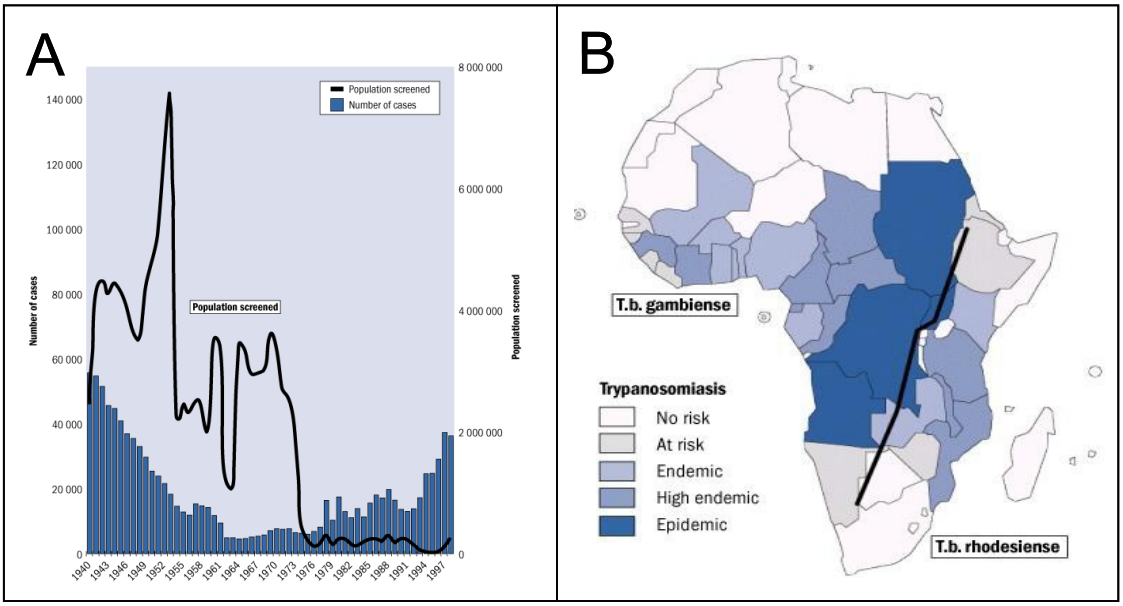

Abb. 1.2 (A) Anzahl der gemeldeten Fälle von Schlafkrankheit und medizinisch überwachte Bevölkerungszahl (1940-1998). (B) Verbreitungsgebiet der Afrikanischen Trypanosomiasis. Die Abbildung zeigt die geographische Verteilung von *Trypanosoma brucei rhodesiensiense* und *gambiense* (Stand 1999). Die ostafrikanische Form der Schlafkrankheit wird ausgelöst durch Infektion mit *Trypanosoma brucei gambiense*, die westafrikanische Form wird durch Infektion mit *T. brucei rhodesiense* hervorgerufen. In Sudan, Uganda, Dem. Rep. Kongo und Angola ist die Schlafkrankheit epidemisch (Quelle: WHO report on global surveillance of epidemic-prone infectious diseases).

1.1.2. Krankheitsverlauf und Symptomatik der Schlafkrankheit

Bei der menschlichen Form der Schlafkrankheit unterscheidet man zwischen der ost- und der westafrikanischen Form, die jeweils einer Subspecies von *T. brucei* zugeordnet werden. Die westafrikanische Form kommt in West- und Zentralafrika vor und wird durch *T. brucei gambiense* hervorgerufen. Die Infektion mit *T. brucei gambiense* verläuft chronisch über mehrere Monate bis Jahre. In der frühen Phase ist die Erkrankung schwer

diagnostizierbar, da die klinischen Symptome schwach sind. Die in Ostafrika vorherrschende Form der Schlafkrankheit wird durch den virulenteren Stamm *T. brucei rhodesiense* hervorgerufen. Sie zeichnet sich durch einen raschen und akuten Krankheitsverlauf aus, welcher unbehandelt meistens nach wenigen Wochen zum Tod führt. Eine dritte Unterart bildet *T. brucei brucei*, die Rinder infiziert, aber nicht humanpathogen ist.

Der generelle Verlauf der Trypanosomeninfektion wird in eine kutanmesenchymale, hämolymphatische und meningoenzephalitische Phase unterteilt. In der kutanmesenchymalen Phase verbleiben die Trypanosomen an der Einstichstelle und vermehren sich dort. Es entsteht ein schmerzhaftes Geschwür mit Anschwellung der Lymphknoten. Außerdem treten Blutungen und Ödeme des Gewebes auf. Nach 3-4 Wochen treten die Trypanosomen in das Blut- und Lymphsystem über. Diese hämolymphatische Phase geht anfangs mit hohem Fieber und starken Kopf- und Gliederschmerzen einher. Weiterhin treten Blutarmut, Gefäßentzündung und Lebervergrösserung auf. Typisch für eine Infektion mit *T. brucei rhodesiense* ist das Auftreten einer Herzmuskelentzündung, die sich auf die anderen Herzschichten und das Herzklappensystem ausbreitet. Es folgt die Erregerausbreitung in die Lymphknoten, wobei das Anschwellen der Nackenlymphknoten besonders charakteristisch ist (Winterbottom'sches Zeichen). In der meningoenzephalitischen Phase dringen die Trypanosomen in das Zentralnervensystem ein. Symptome wie Sprach- und Konzentrationsstörungen, Krämpfe, Verhaltensstörungen und ein erhöhtes Schlafbedürfnis treten auf. Die Patienten sind apathisch, verweigern die Nahrungsaufnahme und fallen ins Koma. Das Endstadium der Erkrankung tritt fast ausschließlich bei einer *T. brucei gambiense*-Infektion ein. Die Patienten sterben dann an den Folgen einer Meningoenzephalitis, Mangelernährung oder einer Sekundärinfektion, während Patienten mit einem *T. brucei rhodesiense*-Befall häufig vor Erreichen der meningoenzephalitischen Phase an Herzversagen sterben (Stich *et al.*, 2002).

1.1.3. Diagnose und Therapie

Zur Diagnose einer Trypanosomeninfektion stehen mikroskopische, serologische und molekularbiologische Verfahren zur Verfügung. Mikroskopisch können Trypanosomen im Blut, der Lymphflüssigkeit und im Liquor nachgewiesen werden. Die serologischen Nachweisverfahren [*Card Agglutination Test for Trypanosomiasis* (CATT) und *Card Indirect Antigen Test for Trypanosomiasis* (CIATT)] beruhen auf dem Nachweis von Antikörpern gegen die varianten Oberflächen-Glykoproteine (VSG, *variant surface glykoprotein*) der Trypanosomen im Blut von Patienten, und können zur Untersuchung von Patienten vor Ort

eingesetzt werden. Ein molekularbiologischer Nachweis auf Basis der Polymerasen-Kettenreaktion (PCR) ist ebenfalls möglich, ist aber für die Routineanwendung in den betroffenen Gebieten ungeeignet.

Zur Behandlung der Schlafkrankheit stehen verschiedene Chemotherapeutika zur Verfügung. Die Wirkung der Medikamente beruht auf der unspezifischen Blockierung oder Abtötung sich teilender Zellen. Für die Behandlung der Erkrankung in der frühen Phase der Infektion mit *T. b. rhodesiense* wird Suramin eingesetzt, das starke Nebenwirkungen hat. In der frühen Phase der Infektion mit *T. b. gambiense* verabreicht man Pentamidin. Das zur Zeit einzige Medikament auf dem Markt, daß bei fortgeschrittenener Erkrankung eingesetzt werden kann, ist das im Jahre 1949 entdeckte Arsen-Derivat Melarsoprol. Melarsoprol ist sowohl bei der Infektion mit *T. b. gambiense* als auch bei der Infektion mit *T. b. rhodesiense* effektiv. Die Nebenwirkungen von Melarsoprol sind drastisch, und beinhalten in 5-10% der Fälle Hirnschädigungen. Zudem wächst die Zahl der Infektionen mit resistenten Erregern, die in einigen Gebieten bereits 30% beträgt. Als Alternative zur Melarsoprol-Behandlung konnte bei *T. b. gambiense*-Infektion das 1990 entwickelte Medikament Eflornithin eingesetzt werden, allerdings ist die Produktion bereits 1999 wieder eingestellt worden (Gull, 2002).

1.1.4. Prophylaxe und Vektorkontrolle

Durch die prophylaktische Behandlung und die Reduktion der Verbreitung von Trypanosomen wird versucht, eine Erkrankung von Mensch und Tier zu verhindern. Ein Impfstoff exisitiert zur Zeit nicht und wird wegen des Phänomens der Antigen-Variation (Abschnitt 1.1.8) vorraussichtlich auch in naher Zukunft nicht entwickelt werden können (Ziegelbauer *et al.*, 1992). Möglichkeiten zum Schutz vor einer Trypanosomeninfektion beschränken sich daher auf Expositionsprophylaxe, wie dem Schutz der Haut durch das Tragen fester Kleidung und das Auftragen von Insektenabwehrmitteln.

Eine Reduktion der Verbreitung von Trypanosomen versucht man durch die Bekämpfung der Tsetsefliege in den betroffenen Gebieten zu erreichen. In den sechziger Jahren wurde dies durch das großflächige Versprühen des Insektizids DDT (1,1,1-Trichlor-2,2-bis(p-chlorphenyl)ethan) erreicht. Diese Maßnahme wurde eingestellt, nachdem festgestellt wurde, daß sich das hochgiftige DDT in der Nahrungskette anreichert. Heute werden alternative Insektizide eingesetzt, bei denen das Problem der Resistenzbildung allerdings ebenfalls besteht. Andere Strategien zur Vektorkontrolle bestehen im Aufstellen von Fallen oder dem Aussetzen steriler Männchen. Diese Technik hat die Ausrottung der Tsetsefliege auf der Insel

Sansibar ermöglicht und soll in weiteren Gebieten angewendet werden (Vinhaes und Schofield, 2003).

1.1.5. Phylogenetische Einordung

Trypanosoma brucei ist ein Protozoe aus der Ordnung der Kinetoplastiden (Mehlhorn H, 1995). Kinetoplastiden liegen phylogenetisch weit von den höheren Eukaryonten entfernt und stellen eine in der Entwicklungsgeschichte der Eukaryonten früh abgezweigte Gruppe dar (Cavalier-Smith, 1993). Alle Arten innerhalb der Familie der Trypanosomatidae leben parasitisch. Weitere Mitglieder der Kinetoplastiden-Familie sind die Humanparasiten *T. cruzi*, Erreger der Chagas-Krankheit, und *Leishmania sp.,* Erreger der Leishmaniose (Abb. 1.3).

System: Protozoa (Einzeller)
Stamm: Sarcomastigophorea
Unterstamm: Mastigophora (Flagellata)
Klasse: Zoomastigophorea (Zooflagellaten)
Ordnung: Kinetoplastida
Familie: Trypanosomatidae
Gattung: *Trypanosoma*
Gattung: *Leishmania*
5 weitere Gattungen

Abb. 1.3 Systematische Position der Gattung Trpanosoma innerhalb der Protozoen.

Die sehr entfernete Verwandtschaft der Kinetoplastiden mit anderen Eukaryonten spiegelt sich in zellulären und molekulargenetischen Besonderheiten der Trypanosomen wieder. Eine dieser Besonderheiten ist der Kinetoplast, das die Kinetoplasitden kennzeichnende Merkmal. Der Kinetoplast ist eine hochorganisierte DNA-Struktur des einzelnen Mitochondriums der Kinetoplastiden (Abb. 1.4).

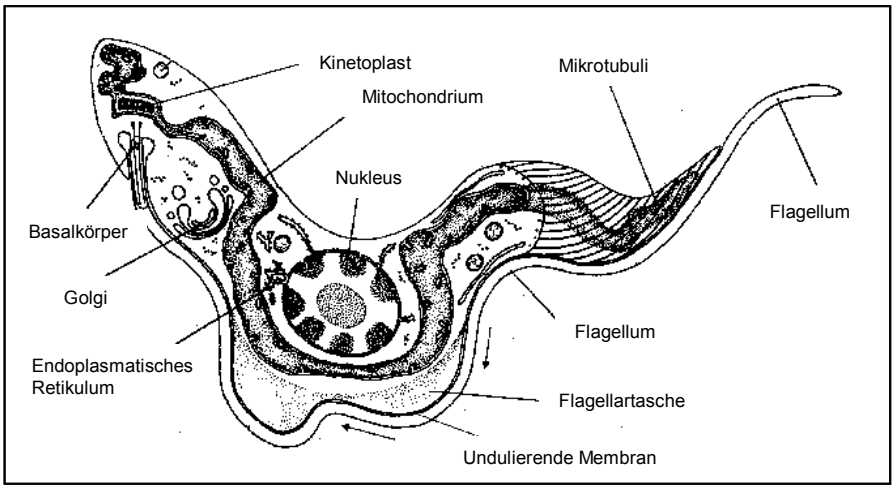

Abb. 1.4 Morphologie von *T. brucei*.

1.1.6. Lebenszyklus

Während ihres Lebenszyklus durchlaufen Trypanosomen verschiedene Umweltbedingungen: Vom Blut des Säugetier-Wirts über den Mitteldarm der Tsetsefliege zur Speicheldrüse und von dort wieder in den menschlichen oder bovinen Blutkreislauf (Abb. 1.5). Der Parasit paßt sich diesen Lebensumständen durch Veränderungen der Morphologie, des Stoffwechsels und der Oberflächenproteine an. Aufgrund dieser Merkmale lassen sich verschiedene Lebensformen definieren. Für die Einteilung der verschiedenen Lebensformen sind vor allem die Länge der Geißel und die Lage des Kinetoplasten zum Zellkern ausschlaggebend.

Die Blutbahnform von *T. brucei* ist zunächst länglich und spindelförmig (*long slender*) und besitzt eine am Basalkörperchen (Kinetosom) entspringende, nach vorn ziehende und frei endende Geißel (Flagellum) (Abb. 1.4). Die Trypanosomen der Blutbahnform sind komplett mit einem Mantel aus Oberflächenproteinen bedeckt. Die Energieversorgung der Blustromform erfolgt ausschließlich durch Glykolyse, deren Enzyme in den mit der Membran verbundenen Glykosomen lokalisiert sind. Diese speziellen Organellen stellen ebenfalls eine Besonderheit der Familie der Kinetoplastiden dar. Die mitochondrialen Funktionen sind in der Blutbahnform inaktiv. Mit dem Ansteigen der Parasitendichte differenzieren sich die *long slender* Trypanosomen mehr und mehr zur kurzen, gedrungenen Formen (*short stumpy*). Die *short stumpy* Form ist durch teilweise Aktivierung des Mitochodriums, Zellzyklusarrest, und anderer Merkmale an das Überleben in der Tsetsefliege vorbereitet (Matthews und Gull, 1998). Bei der Aufnahme der *short stumpy* Form durch die Tsetsefliege differenzieren sich die Trypanosomen zur sogenannten prozyklischen Form und beginnen sich zu teilen. In der prozyklischen Form ersetzt Procyclin das VSG-Protein als Oberflächenprotein. Die

Trypanosomen der prozyklischen Form betreiben außerdem einen oxidativen Stoffwechsel. Von der prozyklischen Form, die den Mitteldarm der Fliege besiedelt, wandern einige in die Speicheldrüsen des Insekts ein, lagern sich dort an und werden zu epimastigoten Formen. Die epimastigoten Formen teilen sich und differenzieren sich zu frei beweglichen, sich nicht teilenden metazyklischen Zellen, die für Säugetiere infektiös sind. Mit dem Stich des Säugetier-Wirtes durch die infizierte Fliege ist der Zyklus vollendet.

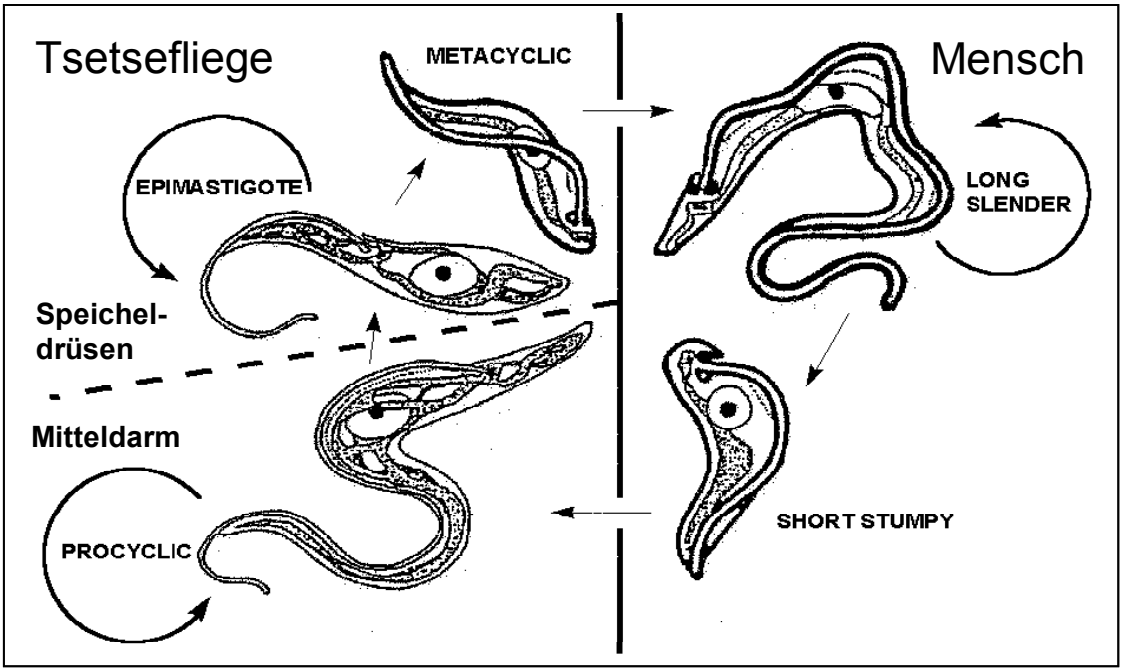

Abb. 1.5 Lebenszyklus von *T. brucei*. Metazyklische Trpanosomen gelangen durch den Stich der Tsetsefliege mit deren Speichelsekret in den Blutkreislauf des Menschen, wo sie sich zur Blutbahnform differenzieren. Die Blutbahnform von *T. brucei* ist zunächst länglich und spindelförmig (*long slender*). Die Blutbahnformen vermehren sich im Blutkreislauf des Menschen durch Längsteilung Mit dem Ansteigen der Parasitendichte differenzieren sich die *long slender* Blustromformen mehr und mehr zur kurzen, gedrungenen (*short stumpy*) Blutbahnform. Durch einen weiteren Stich des menschlichen Wirts durch die Tsetsefliege gelangen die *short stumpy* Formen mit dem Blutmahl in den Mittleldarm der Tsetsefliege, wo sie sich zur sogenannten prozyklischen Form differenzieren und sich zu teilen beginnen. Vom Mitteldarm der Fliege wandern einige der prozyklischen Formen, in die Speicheldrüsen des Insekts ein, lagern sich dort an und werden zu epimastigoten Formen. Die epimastigoten Formen teilen sich und differenzieren sich zu frei beweglichen, sich nicht teilenden metazyklischen Zellen, die für Säugetiere infektiös sind. Mit dem Stich des Säugetier-Wirtes durch die infizierte Fliege ist der Zyklus vollendet (Quelle: El-Sayed, 2000).

1.1.7. Transkription und Regulation der Genexpression

Neben den bereits erwähnten zytologischen Besonderheiten weisen Kinetoplasiden einige molekularbiologische Besonderheiten auf, die sie von höheren Eukaryonten unterscheiden. Einen außergewöhnlichen Vorgang, den man bei Trypanosomen entdeckt hat, ist das sogenannte *RNAediting*. Als Editing bezeichnet man die von *guide*RNA-Molekülen gesteuerte, post-transkriptionale Insertion oder Deletion von Uridin-Basen an spezifischen

Stellen der Kinetoplasten-mRNA (Stuart und Panigrahi, 2002). Ein anderer fundamentaler Unterschied zwischen Kinetoplastiden und anderen Organismen liegt in der Art ihrer Genregulation. Die meisten proteinkodierenden Gene der Trypanosomen liegen auf den Chromosomen in Clustern dicht hintereinander und sind nur durch kurze intergenische Regionen mit einer Länge von ungefähr 500 bis 1000 bp voneinander getrennt. Die Gencluster bilden Transkriptioneinheiten. Die Gene einer Transkriptionseinheit werden ähnlich wie bei Prokaryonten gemeinsam als Vorläufer-mRNA transkribiert (Tschudi und Ullu, 1988). Diese Art der Transkription wird als polycistronische Transkription bezeichnet (Abb. 1.6). Die Transkription der proteinkodierenden Gene erfolgt durch eine RNA-Polymerase, die ähnliche Eigenschaften wie die RNA-Polymerase II der höheren Eukaryonten besitzt. Es konnten bislang allerdings noch keine Promotoren für RNA-Polymerase II für proteinkodierende Gene im Genom von Trypanosomen gefunden werden. Es wird deshalb davon ausgegangen, daß die Transkription proteinkodierender Gene zufällig, oder an nur wenigen Promotoren initiert wird (Clayton, 2002).

Nach der Transkription werden die einzelnen mRNA-Moleküle von der Vorläufer-RNA durch einen als Trans-Spleissen bezeichneten Vorgang abgespalten. An die 5’ Enden der prä-mRNA-Moleküle wird eine RNA von ungefähr 40 Nukleotiden Länge, genannt *spliced leader* angehängt. Gleichzeitig wird das 3’ Ende polyadenyliert (Ullu und Tschudi, 1993). Im Gegensatz zu den proteinkodierenden Genen haben die Gene, die für die *spliced leader*-Vorläufer RNA (SLRNA) kodieren, eindeutig identifizierbare Promoterelemente (Luo *et al.*, 1999a) und werden vermutlich durch RNA-Polymerase II transkribiert (Mair *et al.*, 2000). Während bei den höheren Eukaryonten und auch bei Prokaryonten die Regulation der Genaktivität zu einem großen Teil über die Initiation der Transkription erfolgt, deuten viele Experimente darauf hin, daß die Regulation der Transkriptmenge bei Kinetoplastiden hauptsächlich durch mRNA-Degradation erfolgt. Als regulatorische Elemente sind Sequenzen in den 3’ untranslatierten Regionen (3’ UTRs) identifiziert worden (Clayton, 2002).

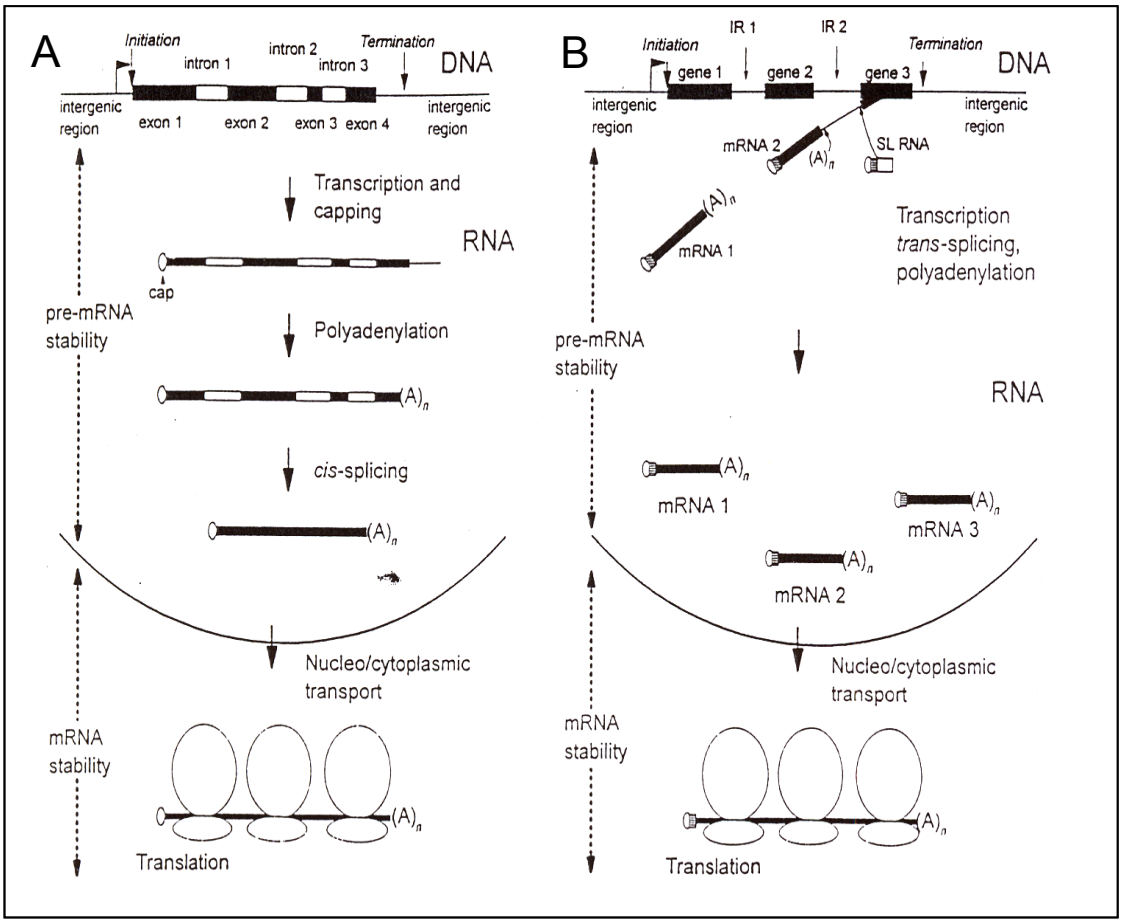

Abb. 1.6 Vergleich der Transkription in höheren Eukaryonten (A) und in Kinetoplastiden (B). Die meisten proteinkodierenden Gene der Trypanosomen sind in Transkriptionseinheiten zusammengelagert. Die Gene einer Transkriptionseinheit werden ähnlich wie bei Prokaryonten gemeinsam als Vorläufer-mRNA transkribiert (Quelle: Tschudi, 1988).

1.1.8. Antigenvariation

Trypanosoma brucei entgeht der Immunantwort des Wirts durch den zeitlich gesteuerten Wechsel des VSG-Oberflächenproteins. Dieses Phänomen wird als Antigen-Variation bezeichnet und stellt ein wesentliches Hindernis für die Entwicklung eines Impfstoffes dar (Borst und Ulbert, 2001; Gull, 2002).

Der molekulare Mechanismus der Antigen-Variation ist gut untersucht. *T. brucei* besitzt etwa 1000 verschiedene über das gesamte Genom verteilte VSG-Gene, von denen jeweils nur ein einziges aktiv ist (Chaves *et al.*, 1999). Das aktive Gen befindet sich an einer von vermutlich 20-40 speziellen Stellen im Telomer-Bereich (*expression sites*). Der Wechsel der VSG-Genexpression erfolgt zum größten Teil durch Genkonversion, also durch die replikative Transposition von VSG-Genen in aktive Expressionsstellen. Auch die Aktivierung einer anderen Expressionstelle ist möglich (Pays *et al.*, 2001).

Der Wechsel des aktiven VSG-Gens erfolgt mit einer Häufigkeit von ungefähr 10^{-2} - 10^{-6} Zellen pro Generation (Turner, 1997). Nach Einsetzen der humoralen Immunantwort des Wirts mit spezifischen Antikörpern gegen das „alte“ Oberflächenantigen wird die Parasitenpopulation stark dezimiert. Die Trypanosomen mit der „neuen“ Variante des Oberflächenproteins überleben und vermehren sich, so daß die Parasitendichte bald wieder stark ansteigt.

1.1.9. Genomgröße und Organisation

Der DNA-Gehalt, die Anzahl und Grösse der Chromosomen verschiedener Stämme und Isolate (*field isolates*) von *T. brucei* schwankt beträchtlich. Frühere Schätzungen des totalen DNA-Gehalts von Stamm 427, die auf DNA-DNA-Renaturierungexperimenten und Zytophotometrie beruhten, ergeben eine Größe von 35-40 Mb (haploid) (Borst *et al.*, 1982; Gibson *et al.*, 1992). Da die Chromosomen der Kinetoplastiden während der Mitose nicht kondensieren, ist eine zytologische Bestimmung der Chromosomenanzahl nicht möglich. Durch Pulsfeld-Gelelektrophorese konnten jedoch 11 diploide Chromosomenpaare im Größenbereich von 1-6 Mb nachgewiesen werden. Diese Chromosomen werden als Megabasenchromosomen bezeichnet. Der DNA-Gehalt der Megabasenchromosomen von Stamm TREU927/4 beträgt 53,4 Mb (diploid) (Melville *et al.*, 1998). Zusätzlich zu den großen Chromosomenpaaren besitzt *T. brucei* eine stark variierende Anzahl mittelgroßer Chromosomen mit einer Größe von 200–900 kb, sowie ungefähr hundert lineare, 50-150 kb große sogenannte Minichromosomen. Insgesamt sind schätzungsweise 10 Mb (25% des gesamten Genoms) auf diesen Chromosomen enthalten. Sie enthalten 177 bp große Wiederholungen (*repeats*), wobei jede dieser Wiederholungen an einer Stelle zweimal vorkommt (*tandem array*). Weiterhin enthalten sie inaktive, telomergekoppelte VSG-Gene. Die Funktion der mittelgroßen Chromosomen und der Minichromosomen ist ungeklärt. Die Anzahl der Gene von *T. brucei* wurde durch den Größenvergleich des Genoms mit *Saccharomyzes cerevisiae* auf etwa 12.000 geschätzt (El-Sayed *et al.*, 2000). Über das gesamte Genom von *T. brucei* sind Retroposon-Sequenzen verteilt, das größte Element, *ingi*, ist 5 kb groß und kommt in ungefähr 400 Kopien vor. Damit entfallen 2 Mb, oder 5% des Gesamtgenoms, auf die *ingi*-Sequenzen. Die meisten der untersuchten *ingi*-Elemente enthalten eine inaktive reverse Transkriptase.

1.1.10. Genomprojekte

Die von der WHO ins Leben gerufene TDR *(WHO special Program for research and training in tropical disease)* hat die afrikanische Trypanosomiasis aufgrund der Anzahl der Todesfälle, der zunehmenden Inzidenz und der verheerenden ökonomischen Auswirkungen in die Kategorie der Krankheiten eingeordnet, für deren Bekämpfung die Grundlagenforschung zur Entwicklung neuer Medikamente und diagnostischer Tests besonders wichtig ist (Remme *et al.*, 2002). Besonders wichig für die Identifizierung von Zielmolekülen für Therapie und Diagnostik ist die Genomanalyse (www.who.int/tdr/diseases/tryp/strategy.htm).

Durch die Sequenzierung des gesamten Genoms erhofft man sich, die für die Pathogenität, Virulenz, Antigenvariation und Medikamentenresistenz verantwortlichen Gene identifizieren zu können. Im Jahr 1995 wurde eine Genominitiative ins Leben gerufen, die die Sequenzierung des gesamten Genoms von *T. brucei* zum Ziel hat. Für die Sequenzierung wurde der Stamm *T. brucei brucei* TREU927/4 ausgewählt (Melville *et al.*, 1998). Finanziert wird die Sequenzierung durch das *National Institute of Allergy and Infectious Disease* (NIAID) in den USA und den *Wellcome Trust* in Grossbritannien. Bislang liegt die Sequenz der kleinsten Megabasenchromosomen - I und II - vor (El-Sayed *et al.*, 2003; Hall *et al.*, 2003). Der aktuelle Stand der Sequenzierung kann unter www.tigr.org/tdb/e2k/tba1/intro.shtml#seq abgerufen werden.

Die genomische Sequenz von *T. brucei* ist die Grundlage für zahlreiche weitere Untersuchungen zur Aufklärung der Funktion dieser Gene. Einem Teil der genetischen Sequenz kann man durch Homologien zu bekannten Genen anderer Organismen eine mutmaßliche Funktion zuordnen. Da die Kinetoplastiden aber nur sehr weitläufig mit eukaryontischen Modellorganismen verwandt sind, findet man nur wenige Homologien. Für die Entwicklung neuer Medikamente sind aber gerade die Gene interessant, die spezifisch für die parasitische Lebensweise sind. Die Wahrscheinlichkeit, solche Gene über Homologien zu finden, ist also eher gering.

Neben dem Vergleich von Gensequenzen können Vorhersagen über Proteinfunktionen auch durch Computer-Programme getroffen werden (Bork und Koonin, 1998; Marcotte *et al.*, 1999). Dieser bioinformatische Ansatz der Genomanalyse kann aber nur vage Hinweise auf mögliche Funktionen von Genen liefern. Bei der Sequenzierung von Chromosom I und II hat man die Erfahrung gemacht, daß von etwa 40% der Gene, die man identifiziert hat, die Funktion nicht ermittelt werden konnte, da keine ausreichenden Homologien vorliegen (Hall *et al.*, 2003) (El-Sayed *et al.*, 2003). Eine weiterführende experimentelle Analyse zur Ermittlung der Genfunktion ist daher unumgänglich.

1.2. Funktionelle Genomanalyse

1.2.1. Einführung

Einen wichtigen Hinweis auf die Funktion eines Gens können die Umstände geben, unter denen ein Gen exprimiert wird, wie zum Beispiel der Lebensform oder einer definierten experimentellen Bedingung. Die Regulation der Aktivität eines proteinkodierenden Gens kann auf der RNA-Ebene und auf der Proteinebene erfolgen. In vielen Fällen spielt dabei die Menge des Genprodukts eine wichtige Rolle für die Genfunktion. Die Menge des Genprodukts wird reguliert über das Gleichgewicht zwischen der Initiation und Effizienz der Transkription und Translation auf der einen Seite und dem Abbau der entstandenen Produkte auf der anderen Seite.

Zur Funktionsanalyse betrachtet man die Aktivität der interessierenden Gene in unterschiedlichen Zuständen, z.B. in unterschiedlichen Zelltypen oder Differenzierungsstadien oder bei Zellkulturen in verschiedenen Medien, usw. Da die Analyse der Funktion für jedes einzelne der im Rahmen eines Genomprojekts identifizierten Gene zu zeit- und arbeitsaufwendig ist, wurden zahlreiche Methoden entwickelt, mit denen sich Daten über Genexpression auf RNA- und Proteinebene im hohen Durchsatz gewinnen lassen. Zur quantitativen Analyse von Proteinen Hochdurchsatz-Genfunktionsanalyse auf Proteinebene werden 2D-Gelelektrophorese gekoppelt mit *matrix-assisted laser desorption/ionization time-of-flight mass spectrometry* (MALDI-TOF-MS) oder Proteinarrays eingesetzt (Kusnezow *et al.*, 2003).

1.2.2. Repräsentative Differenzanalyse

Durch subtraktive Hybridisierung von mRNA-Populationen aus zwei unterschiedlichen Proben können diejenigen mRNA-Spezies isoliert werden, die in einer Population überrepäsentiert sind. Dazu wird die mRNA durch reverse Transkription in doppelsträngige cDNA umgeschrieben, gemischt, denaturiert und anschließend renaturiert. Dabei dient eine Population als der sogenannte *Driver*, der im Überschuß gegenüber der sogenannten *Tester*-Population eingesetzt wird. Bei der Reassoziation der *Tester*- und *Driver*- Moleküle entstehen Hybride zwischen *Tester-Tester* , *Tester-Driver* und *Driver-Driver* Molekülen. Zur Anreicherung der *Tester-Tester*-Population werden entweder die *Driver-Driver*- und *Tester-Driver* Hybride entfernt oder die *Tester-Tester*-Hybride durch PCR selektiv amplifiziert. Die nach der Subtraktion resultierende *Tester-Tester* Population kann zur Herstellung einer Bibliothek verwendet werden. Heute werden vor allem die auf PCR basierende *Suppression*

Subtractive Hybridization (Diatchenko *et al.*, 1996) und *Representational Difference Analysis* (Hubank und Schatz, 1994) angewendet. Die RDA gehört zu den subtraktiven Hybridisierungsmethoden, die für die Identifizierung differenziell exprimierter Gene eingesetzt werden. Die RDA wurde ursprünglich entwickelt, um Unterschiede zwischen verschiedenen Populationen genomischer DNA zu untersuchen (Lisitsyn und Wigler, 1993). Diese Methode wurde so modifiziert, daß auch Unterschiede auf der RNA-Expressionsebene untersucht werden können (Hubank und Schatz, 1994)(Hubank und Schatz, 1999). Die RDA basiert auf sukzessiven subtraktiven Hybridisierungen, an die sich jeweils eine kinetische Anreicherung der Transkripte mittels PCR anschließt. Durch diesen Prozeß werden im *Tester*-Pool häufiger vorhandene Sequenzen angereichert. Gleichzeitig wird die Anreicherung der nicht unterschiedlich exprimierten Gene verhindert. Durch das Vertauschen der beiden Ausgangsproben lassen sich auch die Gene anreichern, die in der als Driver verwendeten Probe stärker exprimiert sind. In Abbildung 1.7 ist das Prinzip der RDA schematisch dargestellt. Die RDA beginnt mit dem Umschreiben der beiden mRNA-Populationen in cDNA. Die cDNA wird anschließend mit dem Restriktionsenzym *DpnII* verdaut. Die resultierenden Fragmente sind im Durchschnitt 256 (2^4) bp lang und haben an beiden Enden einen einzelsträngigen GATC-Überhang. Nach der Ligation eines Oligonukleotid-Adaptors wird die cDNA mittels PCR amplifiziert. Die resultierenden sogenannten *Amplicons* werden erneut mit *DpnII* verdaut, um die R-Bam-Adaptoren zu entfernen. An die *Amplicons*, die als *Tester* eingesetzt werden sollen, wird ein neues Oligonukleotid ligiert. Die *Driver*- und *Tester*- DNA wird gemischt, denaturiert und reassoziiert. Von den entstandenen drei Hybrid-Klassen lassen sich nur die *Tester*-*Tester* Hybride durch eine PCR exponentiell amplifizieren. Bereits nach dieser ersten Amplifikation können die angereicherten Fragmente auf einem Agarosegel sichtbar gemacht werden. Diese Fragmente sind nicht ausschließlich auf einen Unterschied in der Genexpression zurückzuführen, sondern werden auch durch eine zufällige Reassoziation von *Tester*-*Tester*-Hybriden verursacht. Aus diesem Grund schließt sich eine zweite und eventuell eine dritte subtraktive Hybridisierung an. Eine Möglichkeit, die differenzielle Expression der mittels RDA isolierten Gene zu verifizieren, ist die Verwendung der klonierten Fragmente als Probe für einen Northern-Blot (Kuang *et al.*, 1998). Obwohl dieses Verfahren erfolgreich eingesetzt wurden, ist der damit verbundene Zeit- und Arbeitsaufwand sehr groß. Mit der Array-Technologie ist es möglich, schnell die Expression vieler Gene in einer einzigen Hybridisierung zu untersuchen (Frohme *et al.*, 2000). Die verifizierten RDA-Fragmente wurden dann durch Sequenzierung und anschließende Datenbanksuche identifiziert.

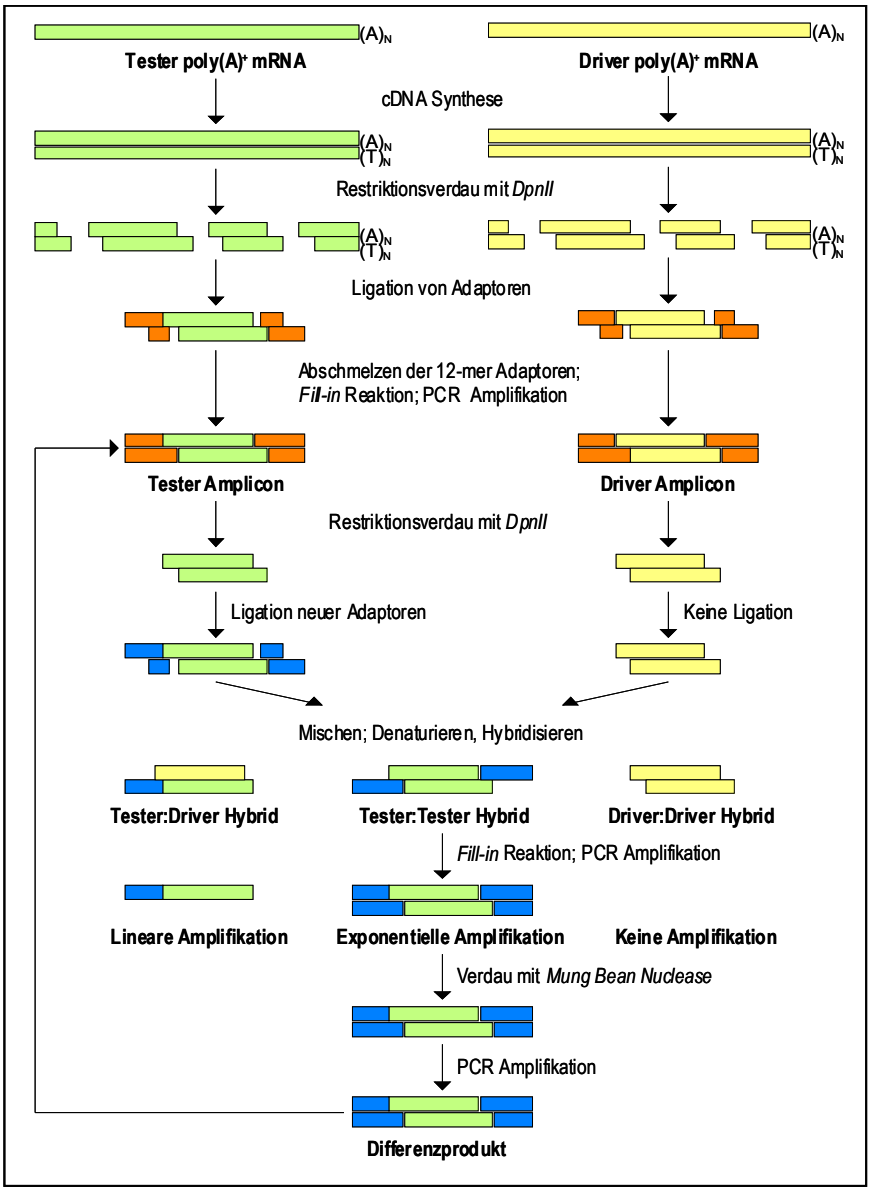

Abb. 1.7 Prinzip der RDA bis zur Herstellung des ersten Differenzproduktes. Um das zweite und dritte Differenzprodukt herzustellen, durchläuft das Produkt nochmals den gleichen Prozeß wie die *Tester*-Population bei der Herstellung des ersten Differenzproduktes. Es ändert sich in jeder weiteren Runde nur die Adaptorsequenz und das Mischungsverhältnis von *Tester* und *Driver* vor der subtraktiven Hybridisierung.

1.2.3. DNA-Microarrays

1.2.3.1. Prinzip der DNA-Microarray-Technologie

Die DNA-Array-Technologie basiert auf der Hybridisierung zwischen komplexen markierten cDNA-Populationen mit DNA-Sequenzen, die auf Nylonmembranen (Gress *et al.*, 1992) (Bernard *et al.*, 1996), Glas- (Schena *et al.*, 1995) (Southern *et al.*, 1992) oder

Siliziumoberflächen (Fodor *et al.*, 1991) in einem definierten und geordneten Muster (*array*) immobilisiert sind. Durch die großangelegte Sequenzierung exprimierter Gene (ESTs) und kompletter Genome wurde die Voraussetzung für die Herstellung genspezifischer DNA-Microarrays geschaffen. So sind für den eukaryotischen Modellorganismus *Saccharomyces cerevisiae* (Bierhefe) zahlreiche Micro- und Macroarrayanalysen durchgeführt worden (Hauser *et al.*, 1998; Lashkari *et al.*, 1997; Winzeler *et al.*, 1999). Für *T. brucei* ist es bislang nicht möglich, genomweite genspezifische Microarrays herzustellen, da die genomische Sequenz nur teilweise bekannt ist.

Ein DNA-Microarray besteht aus vielen einzelnen DNA-Sonden (*probes*), die an definierten Positionen (*spot*) eines Rasters (*array*) an eine feste Oberfläche gebunden sind. Die einzelnen DNA-Sonden auf dem Array dienen der Quantifizierung einer spezifischen Zielsequenz (*target*) in der zu untersuchenden Probe. Microarrays können durch das robotergesteuerte Aufbringen von vorsynthetisierten Nukleinsäuresequenzen, wie z.B. PCR-Produkten, auf einen Glasobjektträger mithilfe von Nadeln (*spotting*) hergestellt werden (Schena *et al.*, 1995). In Abbildung 1.8 ist der schematische Ablauf eines Microarray-Experiments dargestellt. Bei Transkriptionsuntersuchungen stellt man die Probe aus der Gesamt-mRNA der zu untersuchenden Zellen her. Um die Transkripthäufigkeit messen zu können, muß die Probe mit signalgebenden Molekülen markiert werden. Als signalgebende Moleküle können Fluoreszenzfarbstoffe (Schena *et al.*, 1995), Radioisotope (Gress *et al.*, 1992) oder Chemilumineszenz (Rajeevan *et al.*, 1999) verwendet werden. Die Markierung erfolgt durch den Einbau dieser Moleküle während einer cDNA-Synthese, oder durch die nachtägliche Kopplung an Aminoallyl-modifizierte Basen (Stratagene Inc., San Diego, CA). Die markierte Probe wird dann in einem Puffer gelöst und auf den Array aufgetragen.

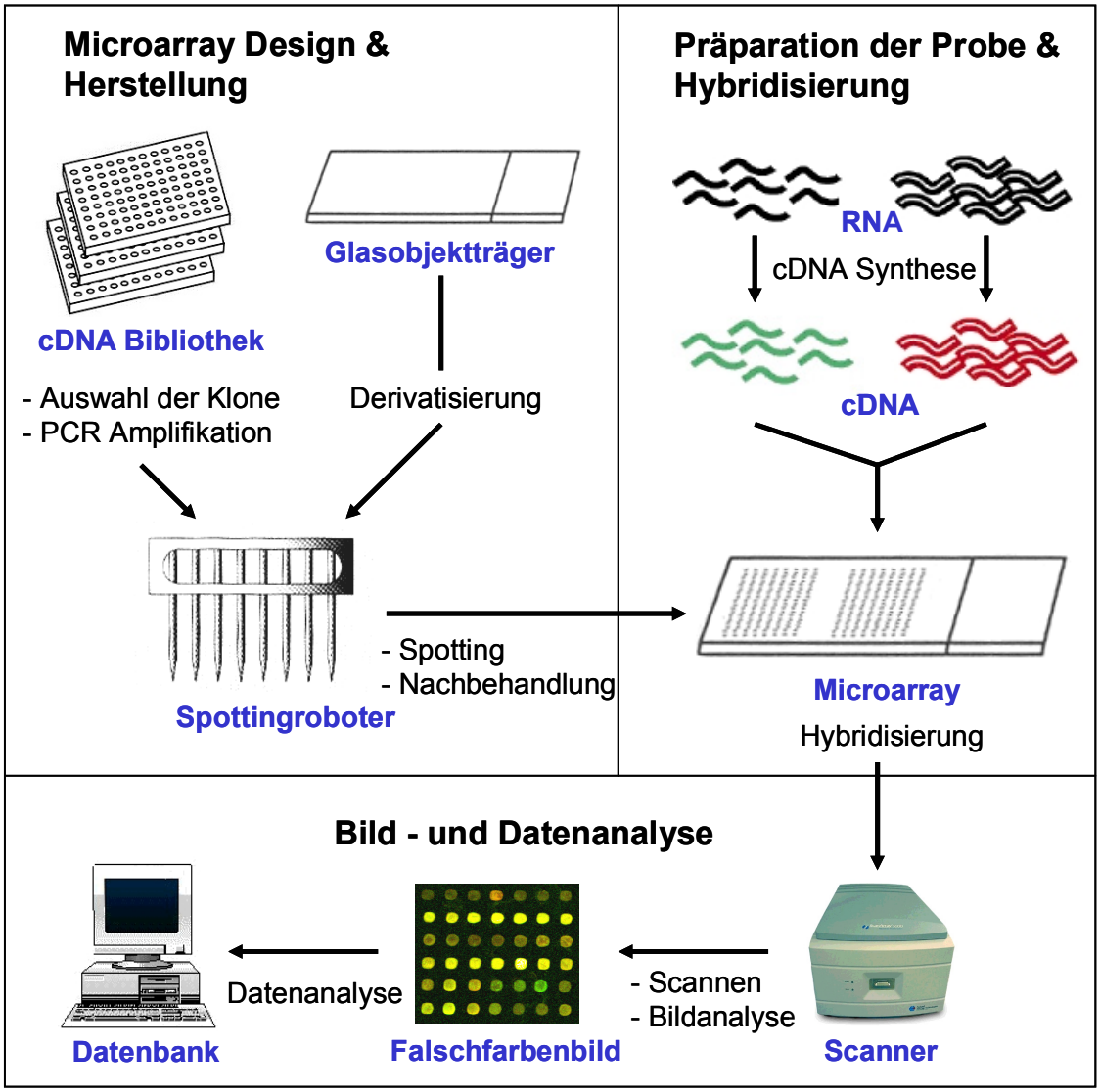

Abb. 1.8 Schematischer Ablauf eines Microarray-Experiments. Der linke Teil dieser Abbildung zeigt den Herstellungsprozeß der Microarrays. In der rechten Hälfte ist das Prinzip der kompetitiven Hybridisierung dargestellt. Nach der Hybridisierung werden die Hybridisierungsignale der beiden Proben gemessen und analysiert.

Durch die Spezifizität der Basenpaarung von Nukleinsäuren binden die markierten cDNA-Moleküle mit der auf dem Array immobilisierten komplementären Zielsequenz (Hybridisierung). Die Spezifizität der Hybridisierung zwischen den komplementären Nukleinsäuren wird durch die temperaturabhängige Dissoziation des Hybridkomplexes (Schmelzpunkt) bestimmt. Der Schmelzpunkt wird durch den GC-Gehalt in der jeweiligen Sequenz bestimmt, und ist definiert als der Punkt, an dem 50% der Basen dissoziiert vorliegen. Die Hybridisierungsreaktion ist eine Gleichgewichtsreaktion (Wetmur, 1991) zwischen der Assoziation der einzelsträngigen *target*-DNA in der Lösung und der auf dem Spot immobilisierten *probe*-DNA und der Dissoziation des Hybrids. Die Menge der an die immobilisierte Sonde gebundenen *target*-Moleküle hängt deshalb sowohl von der Menge der

immobilisierten Sonde als auch von der Menge der jeweiligen markierten cDNA in der Lösung ab. Nach der Hybridisierung werden die durch die gebundene markierte DNA erzeugten Hybridisierungssignale mithilfe eines Laserscanners oder einer CCD-Kamera gemessen. Dabei wird für jeden Microarray ein Bild generiert, daß die Hybridisierungssignale über die gesamte Fläche beinhaltet. Die vom Scanner erzeugten Bilder werden mit spezieller Software ausgewertet, um die Signale jedes Spots zu quantifizieren. Die Stärke des Hybridisierungssignals eines Spots repräsentiert die relative Menge der Transkripte in der Probe. Die Verwendung von Fluoreszenzfarbstoffen mit unterschiedlichen Anregungswellenlängen und Emissionsspektren für die Markierung erlaubt die gleichzeitige Hybridisierung von zwei Proben auf einem Microarray (Co-Hybridisierung oder kompetitive Hybridisierung). Dazu werden meistens die Cyaninfarbstoffe Cy3 und Cy5 verwendet. Dann wird durch einen Laserscanner ein Bild für jedes Fluorophor erzeugt. Zur Visualisierung und Auswertung der Hybridisierungsergebnisse werden die beiden Bilder in Falschfarben dargestellt und übereinander gelegt. Für Cy5-Signale verwendet man dabei meist Rot, für Cy3-Signale wird Grün verwendet. Das Verhältnis der Hybridisierungssignale Cy5- und Cy3-markierter Proben dient als Parameter zur Beurteilung der differenziellen Genexpression. Ein grüner Spot deutet darauf hin, daß ein Transkript in der Cy3-markierten cDNA Population häufiger vorkommt als in der Cy5-markierten cDNA. Ein roter Spot bedeutet dagegen, daß das Cy5-markierte Transkript stärker exprimiert wird. Ist kein Unterschied in der Expression eines Transkripts zwischen Probe und Referenz vorhanden, wird der Spot durch Mischung der beiden Farbkanäle gelb dargestellt (Abb. 1.9).

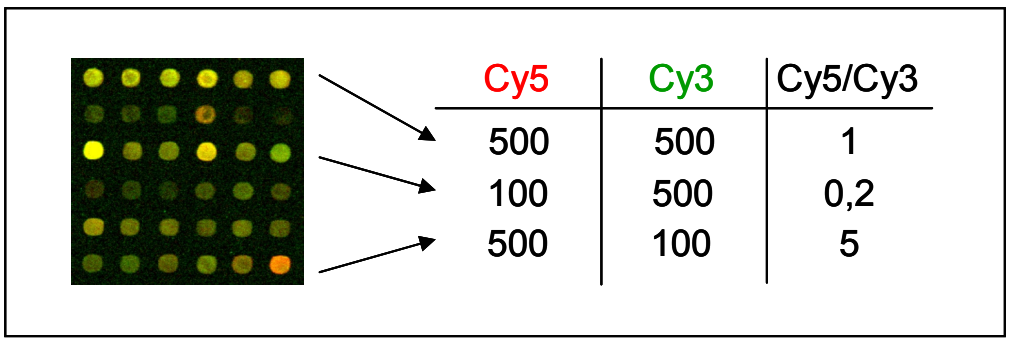

Abb. 1.9 Quantitifizierung der relativen Genexpression bei einem Microarray-Experiment. Für jeden Spot wird die Fluoreszenzintensität der Cy3- und Cy5-markierten cDNA-Proben separat quantifiziert. Das Verhältnis der Cy5/Cy3 Fluoreszenzintensitäten ist ein Maß für die relative Genexpression in den Ausgangsproben.

1.2.3.2. Datenbearbeitung

Einführung

Durch technische Limitierungen und durch biologische Variation sind sowohl die Messung der Transkriptmenge in der Probe als auch die Messung der Transkriptmenge der Kontrolle mit einem Fehler behaftet. Die Vielzahl von Faktoren, die zur Abweichung des Messwertes vom tatsächlichen Wert führen, können in systematische Fehler und Zufallsfehler unterschieden werden. Zufallsfehler können durch geringfügige Abweichungen bei der Herstellung der Probe oder der Herstellung der Arrays und der Hybridisierung entstehen. Zufallsfehler sind nicht reproduzierbar und führen zu einer zufälligen Streuung der Messwerte um den tatsächlichen Wert. Um zu verhindern, daß man aufgrund zufälliger Messfehler ein falsches Resultat erhält, empfiehlt sich die mehrfache Wiederholung von Microarrayexperimenten (Lee *et al.*, 2000). Systematische Messfehler betreffen die Messung aller Gene, bzw. Spots in gleicher Weise und sind reproduzierbar. Systematische Messfehler können durch unterschiedliche Einbauraten der Fluoreszenzfarbstoffe in die Sonde, unterschiedliche Empfindlichkeit des Scanners für die verwendeten Fluorophore oder ungenaue Kalibrierung des Scanners entstehen. Zur besseren Kontrolle über Farbstoff-spezifische Artefakte empfiehlt sich das Vertauschen der für Kontrolle und Probe verwendeten Fluorophors (Dobbin *et al.*, 2003). Parameter, die zu systematischen Fehlern führen, können alle Expressionswerte eines Arrays oder nur einen Teil betreffen. So kann zum Beispiel durch schlechte Einstellung des Laserscanners eine Abhängigkeit der Signalintensitäten von der räumlichen Anordnung der Spots in der *x*- und/oder *y*-Richtung des Objektträgers entstehen. Ebenso kann die Signalintensität von der für die Herstellung des jeweiligen Spots verwendeten Nadel abhängig sein, bedingt durch unterschiedliche Abgabevolumina oder geringfügige Abweichungen des Durchmessers der resultierenden Spots der jeweiligen Nadel. Da der systematische Messfehler berechnet werden kann, lassen sich die Messdaten durch Datentransformationen korrigieren.

Normalisierung

Aufgrund der technisch bedingten Messfehler können die Signalintensitäten trotz gleicher Transkriptmenge voneinander abweichen. Die für den roten und grünen Kanal erhaltenen Signalintensitätswerte müssen deshalb durch verschiedene Datentransformationen vergleichbar gemacht werden, bevor man anhand der gewonnenen Daten die tatsächliche differenzielle Expression beurteilen kann. Diese Datentransformationen bezeichnet man als Normalisierung. Für die Normalisierung der Daten sind verschiedene Methoden

vorgeschlagen worden: Sehr verbreitet ist die globale Mittelwert-Normalisierung, der die Annahme zugrunde liegt, daß die Mehrzahl der auf dem Array dargestellten Gene unter den untersuchten Bedingungen unverändert bleibt (Quackenbush, 2002). Ein Normalisierungsfaktor kann zum Beispiel durch die Division der Mittel- oder Medianwerte des roten und des grünen Kanals ermittelt werden.

In Fällen, bei denen die Anwendung der globalen Normalisierung nicht sinnvoll ist, kann ein Normalisierungsfaktor durch die Verwendung von Referenzstandards ermittelt werden, die den Proben in gleicher Menge zugegeben werden. Durch die Signalintensitäten der Referenzstandards erstellt man eine Eichkurve, die zur Normalisierung der Daten verwendet wird. Eine weitere Möglichkeit zur Normalisierung besteht darin, die Messwerte von sogenannten Haushaltsgenen zum Erstellen einer Eichkurve heranzuziehen. Von Haushaltsgenen nimmt man an, daß sie konstitutiv exprimiert werden. Man hat jedoch festgestellt, daß diese Annahme nicht immer zutrifft. Aus diesem Grund wird eine sehr große Anzahl von Haushaltsgenen benötigt (>75), (DeRisi *et al.*, 1996) um eine zuverlässige Normalisierung vornehmen zu können. Die vorgenannten Methoden sind nur bei linearer Abhängigkeit der Signalintensitäten beider Fluorophore geeignet.

Häufig findet man jedoch, daß systematische Fehler zu einer nicht-linearen Abhängigkeiten der Daten führen. Bei Effekten, die zu einer nicht-linearen Verzerrung der Daten führen, kann eine lokal-gewichtete Regression (Cleveland und Devlin, 1988) der Daten durchgeführt werden, bei der sowohl intensitäts- als auch ortsabhängige Effekte korrigiert werden können (Dudoit *et al.*, 2002).

Datenfilterung

Da bei der Messung von niedrigen Signalintensitäten nahe des Hintergrundsignals der Zufallsfehler größer werden kann als das Signal, nehmen die Quotienten niedriger Siganlintensitäten oft extreme Werte an und führen zu falsch-positiven Ergebnissen. Aus diesem Grund werden Spots mit niedrigen Signalintensitäten häufig vor der Datenanalyse herausgefiltert. Eine andere Möglichkeit besteht darin, für jeden Spot durch verschiedene Parameter einen Qualitätsfaktor zu berechnen, durch den der relative Differenzwert dividiert wird. Sowohl für die Filterung der Daten als auch die Division mit einem Qualitätsfaktor, können zur Festlegung eines Schwellenwertes verschiedene Parameter wie Signalintensität, Signal/Hintergrundverhältnis, oder Größe und Umriß des Spots herangezogen werden.

Für die Definition der Datenqualität gibt es zur Zeit keine objektiven Kriterien. Die Normalisierungs-Arbeitsgruppe der *Microarray Gene Expression Data* (MGED) Organisation

(http:/www.mged.org) arbeitet zur Zeit an der Festlegung von Standards für die Microarray-Analyse und der Festlegung objektiver Qualitätskriterien (Ball *et al.*, 2002).

1.2.3.3. Datenanalyse

Einführung in die statistische Datenanalyse

Microarrrays werden für verschiedene experimentelle Ansätze verwendet, wobei die Fragestellung des Experiments die Versuchsplanung und die Art der Datenanalyse bestimmt. Die grundsätzlichen Zielsetzungen der meisten Microarray-Studien lassen sich grob in drei verschiedene Kategorien einteilen: Dem Vergleich der Expressionsprofile verschiedener Proben und Suche nach differentiell exprimierten Genen *(class comparison)*, der Suche nach Untergruppen gemeinsam regulierter Gene (*class discovery)* oder zur Einteilung von Untersuchungsmaterial in durch Expressionsmuster typisierte Klassen (*class prediction*) (Simon *et al.*, 2003). Bei Ansätzen, bei denen mehrere Proben verglichen werden, versucht man, Gene mittels verschiedener Clustering-Algorithmen nach ihrem Transkriptionsprofil Gruppen zuzuordnen um Rückschlüsse auf ihre Funktion anzustellen (Miller *et al.*, 2002). Zur Identifikation differentiell exprimierter Gene bedient man sich zumeist des direkten Vergleichs der Probe mit der Kontrolle. Die Schwierigkeit bei der Analyse der Daten liegt dann darin, eine Definition für die differenzielle Expression zu finden. Die einfachste Möglichkeit besteht in der Festlegung eines Schwellenwertes für differenzielle Expression, etwa eines absoluten Wertes wie Faktor 2 oder 2 Standardabweichungen vom Mittelwert (Gray *et al.*, 1998; Schena *et al.*, 1996). Nachteile einer solchen Methode sind, daß die Festlegung des Schwellenwertes willkürlich erfolgt und daß diese Methode die Signifikanz eines Wertes unberücksichtigt läßt. Aus diesem Grunde werden zunehmend statistische Tests eingesetzt, um zu überprüfen, ob die differenzielle Expression eines Gens signifikant ist. Bei statistischen Test unterscheidet man parametrische und nicht-parametrische (verteilungsfreie) Tests. Bei parametrischen Tests berechnet man Vertrauensbereiche (Konfidenzintervalle), für die anhand einer Stichprobe ermittelten Schätzwerte statistischer Parameter. Parameter sind statistische Masszahlen der Verteilung einer Grundgesamtheit, wie zum Beispiel Mittelwert und Standardabweichung. Bei nicht-parametrischen Tests werden quantitative Werte durch Rangfolgen oder Wahr-Falsch-Zuordnungen ersetzt, aus denen dann eine Prüfstatistik berechnet wird. Nicht-parametrische Tests sind verteilungsunabhängig, das bedeutet, sie setzen nicht das Vorliegen einer bestimmten Verteilung voraus und können auch bei nicht-normalverteilten Grundgesamtheiten eingesetzt werden. Nicht-parametrische Tests, die zur

Identifikation differentiell exprimierter Gene verwendet wurden, sind beispielsweise der Mann-Whitney-Test oder Wilcoxon-Test (Wu, 2001).

Alle statistischen Tests haben eine gewisse Irrtumswahrscheinlichkeit. Ein sogenannter Fehler erster Art (α-Fehler) resultiert in der Ablehnung der Nullhypothese, obwohl sie in Wirklichkeit richtig ist. Im Falle des Testens differenzieller Genexpression würde man ein Gen als differenziell exprimiert einordnen, obwohl es nicht der Fall ist (falsch-positives Ergebnis). Fehler zweiter Art (β-Fehler) führen zu einer Annahme der Nullhypothese, obwohl sie in Wirklichkeit falsch ist. Ein tatsächlich differenzielles Gen bliebe dadurch unentdeckt (falsch-negatives Ergebnis). Je nach der Rate der falsch-positiven und falsch-negativen Ergebnisse eines statistischen Testverfahrens kann ein Testverfahren hohe Spezifizität, d.h. einen niedrigen α-Fehler, oder eine hohe Sensitivität, d.h., einen niedrigen β-Fehler besitzen. Der Anteil des β-Fehlers entscheidet über die Prüfstärke eines statistischen Tests.

Während durch Eingrenzung des Fehlers erster Art die Wahrscheinlichkeit für eine irrtümliche Ablehnung der Nullhypothese nicht größer als α werden kann, wird der Wahrscheinlichkeit β durch die Testkonstruktion keine Grenzen gesetzt, d.h., die Fehlerwahrscheinlichkeit β bleibt bei einem normalen Testverfahren unbekannt. Parametrische Tests haben prinzipiell eine höhere Prüfstärke als parameterfreie Testverfahren. Aus diesem Grund sind die meisten Tests, die zur Überprüfung differenzieller Genexpression vorgeschlagen worden sind, Varianten eines *t*-Tests, und setzten die Normalverteilung der Differenzwerte voraus.

Bei allen statistischen Tests wird zuvor das Signifikanzniveau α festgelegt, das die Irrtumswahrscheinlichkeit des Tests angibt. Ein typisches Signifikanzniveau ist α=0,95, d.h., die Irrtumswahrscheinlichkeit beträgt 5%. Da man bei einem Microarrayexperiment häufig mehrere tausend Gene gleichzeitig testet, ist man mit einem speziellen Problem multipler Testsituationen konfrontiert: Beim Testen von nur 1.000 Genen auf einem Signifikanzniveau von α=0,95 erhält man 50 falsch-positive Testergebnisse. Besonders in Fällen, in denen nur wenige Gene Änderungen der Transkriptionsrate zeigen, wird durch diese Fehlerrate die Stichhaltigkeit der Ergebnisse stark beeinträchtigt. Aus diesem Grund sollte das gewählte Signifikanzniveau an multiple Tessituationen angepaßt werden. Eine einfache Anpassung ist durch die Bonferroni-Korrektur möglich, bei der das Signifikanzniveau mit der Anzahl der durchgeführten Tests multipliziert wird. Obigem Beispiel folgend erhielte man dadurch ein Signifikanzniveau von α=0,00095. Die Anzahl falsch-positiver Ergebnisse würde durch die Anwendung der Bonferroni-Korrektur in diesem Beispiel 0,05 betragen, wäre also vernachlässigbar. Diese Methode geht zu Lasten der Sensitivität, d.h., das viele differenzielle

Gene unentdeckt blieben. Als Alternative Konzepte zur Bonferroni-Korrektur sind zum Beispiel das Konzept der sogenannten *false discovery rate,* FDR (Benjamini und Hochberg, 1995) entwickelt worden. Die FDR ist ein Schätzwert des prozentualen Anteils von Genen, deren Signifikanz zufällig ist. Eine Erweiterung des Konzepts der FDR ist der *q*-Wert (*q-value*) (Storey und Tibshirani, 2003). Der *q-Wert* ist die niedrigste FDR, bei der ein Gen als signifikant eingestuft wird.

Signifikanz-Analyse für Microarrays

In dieser Arbeit ist die von Tusher et al. entwickelte Signifikanzanalyse (*Significance Analysis for Microarrays*, SAM) (Tusher *et al.*, 2001) zur Identifikation differentiell exprimierter Gene der Blutbahnform und der prozyklischen Form von *T. brucei brucei* angewendet worden. Die Signifikanz-Analyse ordnet jedem Gen einen sogenannten *Score* zu. Die Berechnung des *Score* basiert auf der Änderung der differenziellen Expression (Differenzwert) im Verhältnis zur Standardabweichung des für jedes Gen ermittelten Datensatzes an Differenzwerten. Mit diesen Datensätzen werden bei der Signifikanz-Analyse Permutationen vorgenommen. Aus den permutierten Datensätzen wird ein Schätzwert für die FDR ermittelt.

1.2.4. Identifikation stadienspezifisch exprimerter Gene von *T. brucei*

Für die Erforschung von *T. brucei* spielt die Charakterisierung der verschiedenen Lebensstadien eine zentrale Rolle. Die Veränderungen in der Genexpression während des Lebenszyklus können dabei über die genetischen Grundlagen der stadienspezifischen Anpassungen Aufschluss geben. Anders als bei höheren Eukaryonten wird die Genexpression bei Trypanosomen nicht auf der Ebene der Transkriptionsinitiation reguliert, sondern durch posttranskriptionale Mechanismen, wie zum Beispiel der Regulation durch die Degradation von mRNA. Über das Ausmaß der Regulation der Genexpression auf der mRNA-Ebene ist bislang wenig bekannt. Bei essentiellen Genen wie z.B. den Oberflächen-Antigenen findet man jedoch eine starke Transkriptionskontrolle. Man kann deshalb vermuten, daß Gene, deren Expression große Unterschiede in verschieden Stadien zeigen, für das Überleben der Parasiten besonders wichtig sind.

Die Forschung an *T. brucei* wird zu einem großen Teil an den *in vitro* kultivierbaren, replikativen Formen des Parasiten vorgenommen, der *long slender* Blutbahnform und der prozyklischen Form. Die Identifizierung der für diese Stadien spezifischen Gene ermöglicht einen interessanten Einblick in das Ausmaß der Transkriptionsregulation in Kinetoplastiden

und schafft eine wichtige Grundlage für weitergehende experimentelle Arbeiten an diesen Stadien.

Durch die funktionelle Genomanalyse können differentiell exprimierte Gene identifiziert werden. In der vorliegenden Arbeit wurden Methoden der funktionellen Genomanalyse zur Identifikation stadienspezifischer Gene in der Blutbahnform und der prozyklischen Form von *T. brucei* angewendet. Zur Identifkation differentiell exprimierter Gene von Organismen deren genomische Sequenz nicht vollständig bekannt ist, bietet sich das Verfahren der RDA an. Auch Microarrays sind bei zahlreichen anderen Organismen mit bekannter genomischer Sequenz zur Identifikation differentiell exprimerter Gene angewendet worden. Da die komplette genomische Sequenz von *T. brucei* bislang nicht bekannt ist, wurde für die Genexpressionsanalyse ein Microarray mit genomischen Fragmenten zufälliger Sequenzen hergestellt. Mithilfe dieses Verfahrens konnten zahlreiche stadienspezifisch exprimierte Gene identifiziert werden.