

Informatik und Biologie – Eine Symbiose ermöglicht neue Entdeckungen

Ivo F. Sbalzarini (Zürich)

Zusammenfassung

In zunehmendem Masse finden Methoden aus der Informatik ihren Weg in die biologischen Wissenschaften. Handelt es sich dabei nur um eine kurzlebige Modeströmung, die es ermöglicht, neue Forschungsgelder einzufordern? Oder ist es eine einseitige Serviceleistung der Informatik an die Biologie? Oder eine Symbiose, bei der beide Seiten profitieren? Was ist dran an der Behauptung, dass die Informatik für die Biologie sein wird, was die Mathematik heute für die Physik ist? Anhand von konkreten Beispielen aus unserer Forschung gehen wir diesen Fragen nach. Ich werde versuchen aufzuzeigen, dass beide Seiten profitieren können. Die Biologie benötigt die Informatik, um ihre rasch wachsenden Datensätze analysieren zu können und um testbare Modelle der Systeme zu bilden. Die betrachteten biologischen Systeme sind dabei von einer Komplexität, die vielfach neue Methoden erfordert und so wiederum die Forschung in der Informatik stimuliert. Wir betrachten den gesamten Zyklus von Datenanalyse, Modellbildung und Simulation. Die entsprechenden Beispiele stammen aus den Gebieten der Bildverarbeitung, Virus-Infektionen und Transportprozessen im Innern von lebenden Zellen. Die Grenze zwischen Informatik und Biologie verschwimmt dabei in zunehmendem Masse.

Computer Science and Biology – a symbiosis enables new discoveries

An increasing number of methods from computer science are making their way into biology. Is this just a short-lived fashion, allowing the raising of additional research funds? Or is it a one-sided service from computer science to biology? Or is it a symbiosis where both sides equally profit? What is the true core of the claim that computer science will become for biology what mathematics is for today's physics? Using concrete examples from our research we will try to approach these questions. I will show how both sides can profit. Biology needs computer science in order to cope with the vast amounts of experimental data, to analyze them, and to build testable models of the considered systems. The complexity of these systems, however, frequently requires the development of novel computational methods, thus stimulating research in computer science. We consider the complete cycle of data analysis, modeling, and simulation. All examples are taken from the areas of image processing, virus infections, and transport processes in cell organelles. The border between computer science and biology is hereby increasingly blurred.

Schlagwörter: Bildverarbeitung – Bioinformatik – Diffusion – endoplasmatisches Retikulum – Hochleistungsrechnen – Mustererkennung – Simulation – Virus

1 EINLEITUNG

Dieser Artikel ist eine Zusammenfassung meiner Einführungsvorlesung, welche ich am 18. Dezember 2006 an der ETH Zürich gab. Das Gebiet der rechnergestützten Biologie (Computational Biology) ist ein recht junges und es existieren mehrere, zum Teil konkurrierende Definitionen, Ansichten und Illusionen dazu. Ich möchte versuchen, Ihnen dieses faszinierende Forschungsgebiet etwas näher zu

bringen, anhand von konkreten Beispielen unsere Arbeit zu erläutern und Ihnen mein persönlich gefärbtes Verständnis des Gebiets zu vermitteln. Die erste Version des Titels der Vorlesung, «Informatik und Biologie – Symbiose oder Zweckehe?», veranschaulicht das Spannungsfeld, in dem wir uns bewegen. Einerseits wird vermehrt behauptet, dass die moderne Biologie nicht mehr ohne Datenanalyse, Modellierung und Computersimulation auskommt, andererseits besteht die konstante Gefahr, dass diese Kombination

einseitig ausfällt oder zweckentfremdet wird. Ich möchte meine Einführung in das Thema daher mit einem allgemeinen Überblick über die (echten) Computeranwendungen in der Biologie beginnen.

Die Softwarefirma Microsoft hat letztes Jahr ein Komitee von 41 international anerkannten Experten aus Naturwissenschaft und Technik versammelt und sie über die Gestalt der wissenschaftlichen Landschaft im Jahre 2020, sowie über die Rolle und Situation der Informatik darin, nachdenken lassen (MICROSOFT RESEARCH, 2006). Frei ins Deutsche übersetzt lautet ein Fazit dieser Arbeitsgruppe, dass die Informatik im Jahre 2020 für die Biologie sein wird, was die Mathematik heute für die Physik ist. Dies ist eine sehr starke Aussage, kommt doch praktisch kein Physiker ohne mathematische Berechnungen aus. Die Aussage impliziert auch, dass die Biologen von morgen ihre Computermodelle und Simulationen selbst durchführen werden (müssen). Ein Physiker läuft schliesslich auch nicht ins Mathematik-Institut um Hilfe zu holen, wenn er eine Gleichung lösen muss. Die Beherrschung von Informatikmethoden und Programmieretechniken wird in der Biologie zum wettbewerbsentscheidenden Faktor werden, und die biologischen Studiengänge werden sich entsprechend anpassen. Genau wie die Mathematik dies für die Physik tut, wird die Informatik dazu die notwendigen Grundlagen, Algorithmen und Technologien liefern, deren Entwicklung durchaus auch durch die Bedürfnisse der Biologie beeinflusst wird.

Als Wissenschaft beschäftigt sich die Biologie mit dem Studium lebender «Systeme». Im Vergleich zu technischen Systemen sind diese hoch organisiert, alle Prozesse sind reguliert, die auftretenden geometrischen Formen sind komplex, vieles befindet sich im thermodynamischen Ungleichgewicht, und die zahllosen Wechselwirkungen zwischen den Bausteinen eines Lebewesens, sowie zwischen individuellen Lebewesen, sind oft nichtlinear und gekoppelt. Der Fokus in der Biologie verschob sich dabei in der Vergangenheit zunehmend von einer Beschreibung und Klassifikation dieser komplexen Systeme hin zur Mission, sie in ihrer Funktion zu verstehen. Dies erstreckt sich über ein weites Spektrum von Zeit- und Längenskalen. Von einzelnen Molekülen, über Organellen, zu Zellen, Organen und Organismen, bis hin zu Ökosystemen. Der Komplexität des resultierenden Informationsverarbeitungssystems ist dann nicht mehr mit Papier und Bleistift beizukommen.

Die Anwendung von Computern ist zum Systemverständnis oft vorteilhaft oder notwendig. Dabei lassen sich grob die folgenden Indikationen unterscheiden:

1. Im Falle grosser Datenmengen. Sind z. B. Zellkolonien in zehntausenden oder gar Millionen von Fotos auszuzählen, so würde eine manuelle Verarbeitung zu lange dauern.
 2. Zur Sicherstellung von Wiederholbarkeit. Wird derselbe Datensatz (z. B. Positionsbestimmung von fluoreszierenden Markern in Zellen) von zwei verschiedenen Personen oder auch von derselben Person zu zwei Zeitpunkten ausgewertet, so ergeben sich oft zwei verschiedene Resultate. Eine Auswertung durch den Computer hingegen garantiert Reproduzierbarkeit, da alle Arbeitsschritte in Form eines Computerprogramms eindeutig festgehalten und dokumentiert sind.
 3. Zur Verarbeitung von Komplexität. Das Verhalten eines biologischen Systems ist, aus oben genannten Gründen, vielfach nicht direkt aus seiner Beschreibung ersichtlich. Ein im Computer simuliertes Modell ist oft zum Verständnis nötig.
 4. Zur Überbrückung von Zeit- und Längenskalen. Dinge, welche zu gross, zu klein, zu schnell oder zu langsam sind für eine experimentelle Messung, können in «virtuellen Experimenten» im Computer studiert werden.
 5. Aus ethischen Gründen. In virtuellen Experimenten und Computersimulationen sind keine Lebewesen betroffen.
 6. Zum Erreichen von Kontrollierbarkeit oder Beobachtbarkeit. In einer Computersimulation sind alle Variablen kontrollierbar, d. h. wir können ihnen bekannte und vordefinierte Werte zuweisen. Ebenso ist im Computer alles beobachtbar, auch wenn die entsprechende Grösse im Experiment nicht messbar ist.
- Diese Punkte lassen sich in zwei Anwendungsbereiche zusammenfassen. Zum einen werden Computer in der *Datenanalyse* benötigt, zum anderen zur *Modellierung und Simulation* von Systemen.

2 RECHNERGESTÜTZTE BIOPHYSIK – EINE PERSÖNLICHE SICHT

Seit einigen Jahren ist die molekulare Struktur – und für eine rasch wachsende Zahl von Organismen auch die Sequenz – des Erbmaterials bekannt. Andererseits beobachten wir das physikalische Erscheinungsbild und das Verhalten der Lebewesen. Wie hängen jedoch dieser Phänotyp und die Gensequenz, der Genotyp, zusammen? Was und wie viel des Aussehens und Verhaltens sind genetisch bestimmt? Um dies zu erklären, muss, so glaube ich, eine weitere Schicht berücksichtigt werden. Diese umfasst

zwingend die räumliche Organisation und Kompartimentalisierung der Lebewesen und ihrer Bestandteile. Dasselbe Molekül kann z. B. in verschiedenen Kompartimenten einer Zelle unterschiedliche Funktionen haben. *Zytochrom C* ist wohl das bekannteste Beispiel. Ein lebenswichtiger Elektronenträger für die Zellatmung in den Mitochondrien, wird es zum Auslöser des programmierten Zelltodes, wenn es ins Zytoplasma gelangt. Modelle, welche die räumliche Organisation der Systeme ausser Acht lassen, sind daher von begrenzter Vorhersagekraft. Ebenfalls zu berücksichtigen sind die zeitliche Plastizität und Dynamik der Systeme. So werden Gene in verschiedenen Stadien des Zellzyklus unterschiedlich stark ausgedrückt oder haben gar unterschiedliche Funktion. Des Weiteren sind die Einflüsse der Umwelt auf den Phänotyp zu berücksichtigen, und die Physik der Wechselwirkungen muss korrekt abgebildet werden, um herauszufinden, welcher Teil des Phänotyps durch Selbstorganisation zustande kommt. Schliesslich sind auch die zahllosen Regulations- und Kontrollmechanismen zu betrachten, denn das Verhalten eines regulierten Systems ist nicht einfach aus dem Verhalten des entsprechenden unregulierten Systems erklärbar.

Die rechnergestützte Biophysik versucht obiges zu erreichen, indem die Physik als Basis des Funktionierens biologischer Systeme postuliert wird. Wir nehmen also an, dass die belebte Materie denselben physikalischen Gesetzen gehorcht wie die unbelebte. Vor diesem Hintergrund möchten wir dann das Funktionieren lebender Systeme unter minimalen Annahmen verstehen. «Verstehen» heisst hierbei, dass wir das Aussehen und Verhalten des Systems in einer neuen Situation korrekt vorhersagen können. Die Forderung der minimalen Annahmen bezieht sich darauf, dass diese Vorhersage im Idealfall lediglich auf physikalischen Grundgesetzen und einer Beschreibung des Systems basieren sollte. Zur Datenanalyse, Modellbildung und Simulation ist der Computer unser Werkzeug; also: rechnergestützte Bio-Physik.

Der typische Arbeitsablauf beginnt mit einem biologischen Experiment. Die Daten aus dem Experiment (in zunehmendem Masse Bilder und Videos) werden im Computer vollautomatisch analysiert, um diejenigen Informationen zu extrahieren, welche zur Formulierung des physikalischen Modells benötigt werden. Das Verhalten dieses Modells wird dann in Simulationen studiert, und die resultierenden Vorhersagen werden wiederum mit Experimenten verglichen. Das Modell wird dann ggf. angepasst, bis die Realität korrekt abgebildet wird. Danach sind virtuelle Experimente möglich und alles am System

ist beobachtbar und kontrollierbar. Damit die Arbeit echt interdisziplinär ist, müssen jedoch auf beiden Seiten – Informatik und Biologie – neues Wissen oder neue Technologien entstehen. Dies möchte ich in den folgenden zwei Beispielen verdeutlichen.

3 BEISPIEL 1: DATENANALYSE

In einem ersten Beispiel aus unserer aktuellen Arbeit möchte ich den Einsatz des Computers in der Datenanalyse etwas beleuchten. Das Projekt wurde in Zusammenarbeit mit den Gruppen von Prof. Helenius (Biochemie, ETHZ) und Prof. Greber (Zoologie, UniZH) während meiner Zeit in der Gruppe von Prof. Koumoutsakos (Informatik, ETHZ) durchgeführt. Es geht darum, die frühen Stadien einer Virusinfektion zu analysieren, im Besonderen die Zeit zwischen der Bindung des Virus an den Rezeptor auf der Aussenseite der Zelle und dem Eindringen in die Zelle. Durch Beobachtung und Analyse der Bewegung des Virus-Rezeptor-Komplexes auf der Aussenseite der Zellmembran möchten wir etwas über den Aufbau der Zellmembran lernen, sowie darüber, welche Wege das Virus ausnutzt, um ins Innere der Zelle zu gelangen. Dazu wurden die Viren mit einem fluoreszierenden Farbstoff markiert, und die Bewegungen dieser leuchtenden Punkte wurden unter dem Mikroskop beobachtet und auf Video aufgezeichnet (EWERS et al., 2005).

Die ursprüngliche Aufgabe lautete, ein Bildverarbeitungsverfahren und ein Computerprogramm zu entwickeln, mit dem die Bewegungen der hellen Punkte in den Videos automatisch verfolgt und die Bahnkurven der Viren extrahiert werden können. Das entwickelte Computerprogramm verarbeitet über 1000 Punktdetektionen pro Sekunde, was die Auswertung von tausenden von Videos erlaubt (siehe Beispiel in Abb. 1). Eine manuelle Analyse bringt zum Vergleich nicht mehr als 2 Punktdetektionen pro Sekunde, und auch dies nur über eine begrenzte Zeitdauer. Zusätzlich wird die Position jedes Virus vom Computerprogramm reproduzierbar auf eine Genauigkeit von $1/10$ Pixel ermittelt (SBALZARINI und KOUMOUTSAKOS, 2005), während die Genauigkeit bei manueller Auswertung höchstens 1 Pixel beträgt. Durch Analyse dieser Bahnkurven unter verschiedenen Drogenbehandlungen der Zellen konnte schliesslich ein Mechanismus zur Signalübermittlung durch die Zellmembran identifiziert werden (EWERS et al., 2005).

Eine Inspektion der ermittelten Bahnkurven zeigte auch, dass sich die Viren keineswegs immer gleich verhalten. Ihre Bewegung ist vielmehr eine Sequenz von charakteristischen

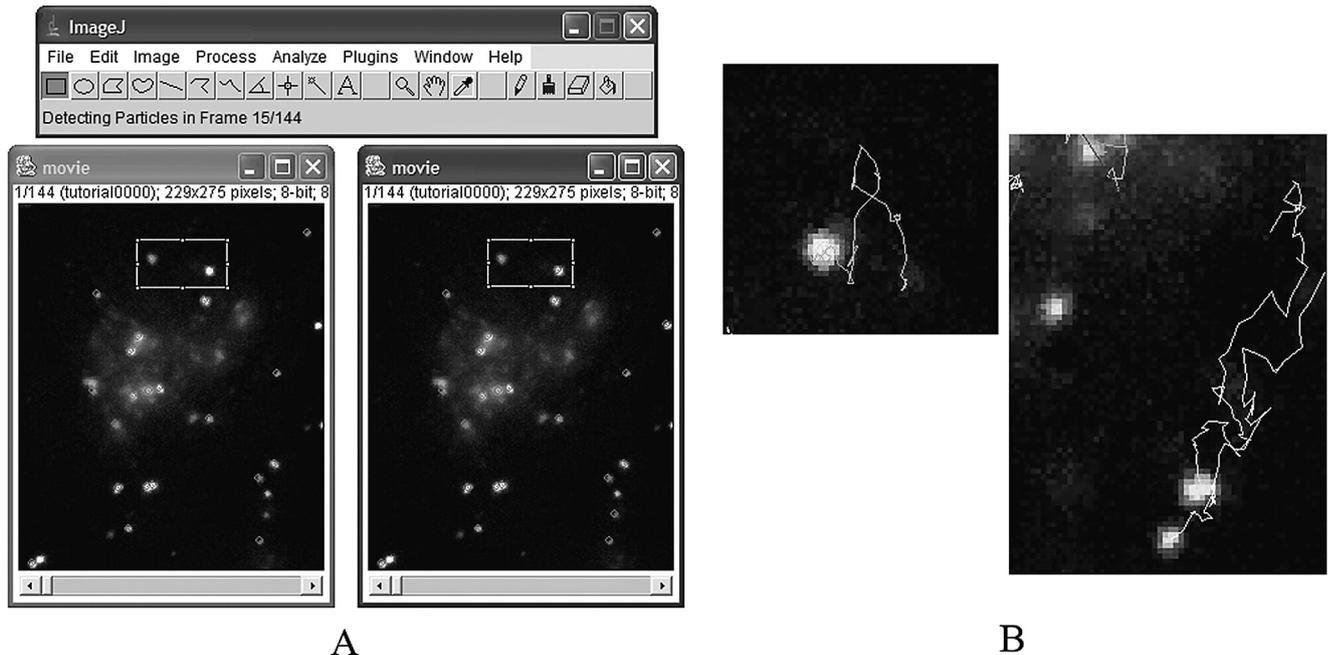


Abb. 1. (A) Benutzeroberfläche der entwickelten Bildverarbeitungssoftware zum Extrahieren der Bahnkurven von Viren aus Mikroskopie-Videos. (B) Zwei Beispiele von verarbeiteten Filmen. Gezeigt ist das letzte Bild des Videos mit den extrahierten Bahnkurven überlagert. Die einzelnen Viren sind als helle Punkte erkennbar. (Daten: menschliche Adenoviren vom Typ 2, Christoph Burckhardt, Greber-Gruppe, Universität Zürich)

Fig. 1. (A) User interface of the developed image processing software for single particle tracking from microscopy videos. (B) Two examples of processed movies. The last frame of the video is shown with the extracted trajectories overlaid. Individual viruses are visible as bright spots. (Data: human Adenovirus of type 2, Christoph Burckhardt, Greber group, University of Zurich)

Mustern aus Festsitzen, Drift und Zufallsbewegung, wie in Abb. 2 veranschaulicht. In einem weiteren Analyseschritt sollte also ein Verfahren entwickelt werden, welches diese Bewegungsmuster in den Bahnkurven automatisch erkennen und extrahieren kann. Wir verwendeten dazu ein neuronales Netzwerk. Dabei handelt es sich um einen lernenden

Algorithmus (Machine Learning), welcher die Funktionsweise des Gehirns nachbildet. Biologisch inspiriert besteht der Algorithmus aus einem Netzwerk miteinander verbundener virtueller Neuronen mit bestimmtem Übertragungsverhalten. Dieser Algorithmus kann dann anhand bekannter Beispiele auf das Erkennen der gesuchten Bewegungsmuster

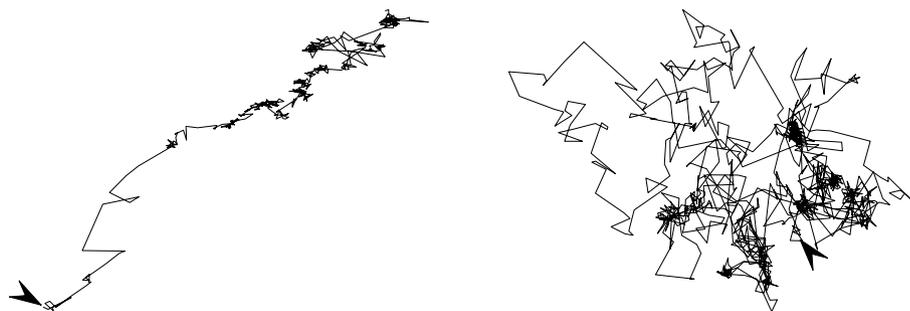


Abb. 2. Zwei Beispiele von automatisch extrahierten Bahnkurven von Polyoma-Viren (Daten: Helge Ewers, Helenius-Gruppe, ETH Zürich). Die Pfeile markieren den Beginn der Kurven. Die Bewegungen der Viren bestehen aus einer Abfolge charakteristischer Muster wie Festsitzen, Drift oder Zufallsbewegung.

Fig. 2. Two examples of automatically extracted trajectories of Polyoma virus (data: Helge Ewers, Helenius group, ETH Zurich). Arrows mark trajectory beginnings. The virus motion consists of a sequence of characteristic patterns such as confinement, drift, or random motion.

trainiert werden. Anschliessend ist er in der Lage, diese Aufgabe auch für neue, bisher ungeschene Bahnkurven durchzuführen. Die Genauigkeit lag dabei in unserem Beispiel bei über 90%, d. h. über 90% der Bewegungsschritte in den Bahnkurven wurden dem korrekten Bewegungsmuster zugeordnet (ermittelt auf synthetischen Daten mit bekannter Zuordnung). Dies ist vergleichbar mit oder besser als eine manuelle Auswertung und bringt zusätzlich wiederum die Vorteile von Geschwindigkeit und Reproduzierbarkeit mit sich.

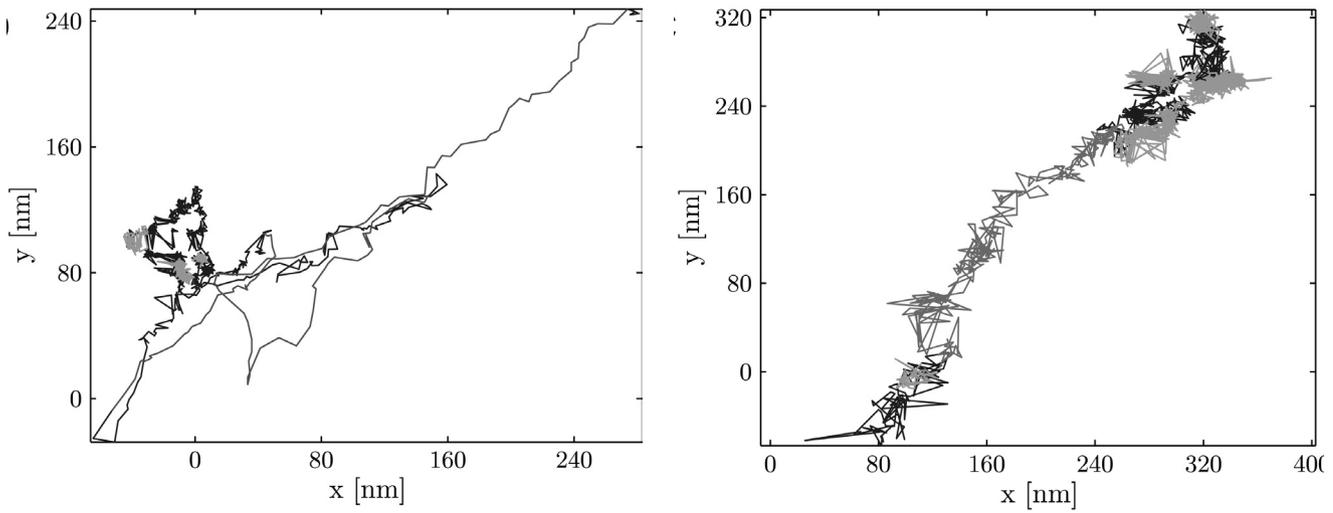


Abb. 3. Vom neu entwickelten Mustererkennungsverfahren (neurales Netz) automatisch erkannte Bewegungsmuster in zwei Bahnkurven menschlicher Adenoviren (Daten: Christoph Burckhardt, Greber-Gruppe, Universität Zürich). Die einzelnen Bewegungsmuster sind: Festsitzen, Zufallsbewegung, Drift und gerichtete Bewegung (in verschiedenen Graustufen dargestellt).

Fig. 3. Automatically detected motion patterns in two trajectories of human Adenovirus (data: Christoph Burckhardt, Greber group, University of Zurich). The following patterns were identified and extracted by the newly developed pattern detection algorithm (neural network): confinement, random motion, drift, and directed motion (depicted in different gray levels).

Das Zerschneiden der Bahnkurven in diese Bewegungsmuster (siehe Abb. 3) ist für die Formulierung des physikalischen Modells essentiell, denn es stellt die Verbindung zwischen den Daten (Bahnkurven) und bekannten physikalischen Prozessen (Diffusion, Festsitzen usw.) her. Nur wenn bekannt ist, welche Teile einer Bahnkurve auch tatsächlich einer Diffusion entsprechen, kann die korrekte Diffusionskonstante bestimmt werden. Und nur wenn bekannt ist, wo und wie lange das Virus festsitzt, kann man die Grössen dieser Zonen des Festsitzens bestimmen. Diese quantitativen Parameter bilden die Grundlage für ein vorhersagekräftiges physikalisches Modell. Der neu entwickelte Algorithmus zur Mustererkennung in Bahnkurven legt somit die Basis, um Veränderungen im Bewegungsverhalten der Viren unter bestimmten Behandlungen zu klassifizieren und neue anti-virale Medikamente mit Zielen im frühen Infektionsprozess zu identifizieren.

4 BEISPIEL 2: MODELLIERUNG UND SIMULATION

Im zweiten Beispiel betrachten wir die Rolle der Informatik in der Modellierung und Simulation komplexer biologischer Systeme. In Zusammenarbeit mit Prof. Koumoutsakos (Informatik, ETHZ) und Prof. Helenius (Biochemie, ETHZ) ging es darum, den Stofftransport in einer Zellorganelle mit komplexer geometrischer Gestalt zu untersuchen. Im

vorliegenden Beispiel war dies das Endoplasmatische Retikulum (ER), ein verästeltes Netzwerk aus Röhren und Lamellen. Diese Organelle ist in allen eukaryotischen Zellen vorhanden, und sie spielt eine wichtige Rolle in der Synthese vieler Proteine und Lipide. Strukturell handelt es sich um eine zusammenhängende Membran, welche den Raum in ein Inneres (Lumen) und ein Äusseres trennt.

Die Standardmethode zur experimentellen Untersuchung intrazellulärer Transportprozesse ist Fluorescence Recovery After Photobleaching (FRAP) (WHITE und STELZER, 1999). Das interessierende Protein wird dabei mit einer fluoreszierenden Domäne versehen und im ER exprimiert. Vom nun fluoreszierenden ER wird ein Teil gebleicht, indem dort mittels eines Lasers der Fluoreszenzfarbstoff irreversibel zerstört wird. Man misst dann den Wiederaufbau der Fluoreszenzintensität in der gebleichten Region durch Einströmen von ungebleichten Molekülen von aussen (siehe Abb. 4). Aus der Dynamik dieses Wiederaufbaus, d. h. aus der Form der Fluoreszenzerholungskurve, kann die Transportgeschwindigkeit der Moleküle ermittelt werden. In unserem Falle handelt es sich beim Transportprozess um Diffusion, wie unabhängige Kontroll-experimente zeigen. Wir möchten also aus Fluoreszenzerholungskurven die Diffusionskonstante von Proteinen im ER lebender Zellen bestimmen.

Die Schwierigkeit hierbei ist, dass nicht nur die Diffusionskonstante des Proteins die Erholungskurve beein-

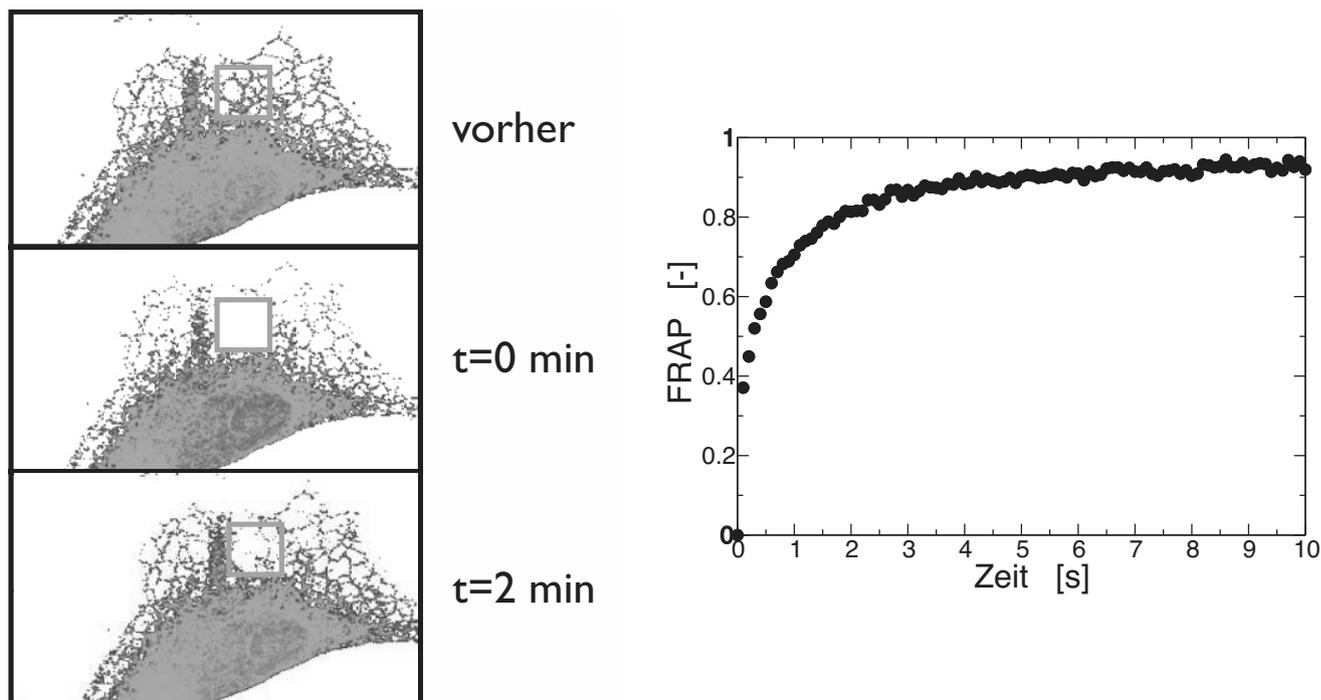


Abb. 4. Die Fluorescence Recovery After Photobleaching (FRAP) Methode. Ein Teil des mit fluoreszierendem Protein gefüllten endoplasmatischen Retikulums wird zur Zeit $t=0$ min mit einem Laser gebleicht (linke Bildsequenz). Aus der Dynamik des Wiederaufbaus der Fluoreszenzintensität in der gebleichten Region (rechtes Diagramm) kann die Transportgeschwindigkeit des fluoreszierenden Proteins bestimmt werden. (Daten: Anna Mezzacasa, Helenius-Gruppe, ETH Zürich)

Fig. 4. The Fluorescence Recovery After Photobleaching (FRAP) method. The endoplasmic Reticulum of a cell is filled with fluorescent protein. At time $t=0$ min, a part of it is bleached with a laser (left image sequence). From the dynamics of fluorescence recovery in the bleached region (diagram on the right) one can determine the transport speed of the fluorescent protein. (Data: Anna Mezzacasa, Helenius group, ETH Zurich)

flusst, sondern auch die lokale Geometrie des ER. Weist das ER lokal nämlich mehr oder dickere Röhren auf, so erfolgt das Rückströmen auch bei gleicher Diffusionskonstante schneller. Das Problem der FRAP-Methode ist, dass die gebleichte Region grösser ist als die einzelnen Strukturen im ER und wir lediglich die totale Fluoreszenz in dieser Region messen, nicht aber die räumliche Verteilung. Zudem beobachten wir im Mikroskop einen zweidimensionalen Schnitt oder eine Projektion eines in Realität dreidimensionalen Objekts. Diese Artefakte der Methode müssen korrigiert werden, sind aber im Experiment nicht kontrollierbar.

Eine mögliche Lösung besteht darin, das System und den Prozess möglichst originalgetreu im Computer nachzubilden, wo die Artefakte dann problemlos beobachtet und die entsprechenden Variablen kontrolliert werden können. Unser Vorgehen sieht also so aus, dass anschliessend an ein FRAP-Experiment in einer lebenden Zelle planparallele Schnittbilder des ER aufgenommen werden. Diese werden dann dazu verwendet, die dreidimensionale Geometrie des

ER dieser spezifischen Zelle im Computer zu rekonstruieren (SBALZARINI et al., 2005a). In dieser originalgetreuen ER-Geometrie (siehe Abb. 5 für ein Beispiel) wird dann der Prozess des Eindiffundierens in die gebleichte Region numerisch simuliert (siehe Abb. 6), und durch Vergleich der simulierten Erholungskurve mit der experimentell gemessenen kann die korrekte Diffusionskonstante ermittelt werden (SBALZARINI et al., 2005a). Ausserdem erlaubt uns dieses Computermodell, den Einfluss der Geometrie und denjenigen der molekularen Diffusionskonstanten zu trennen. Obwohl die Diffusionskonstante im Experiment nicht kontrollierbar ist, ist sie es im Computermodell. Es ist daher ein Leichtes, Simulationen in verschiedenen rekonstruierten ER durchzuführen und dabei in allen Simulationen denselben Wert für die Diffusionskonstante zu verwenden. Die Streuung in den resultierenden Erholungskurven stammt dann ausschliesslich von den Variationen in der Geometrie, denn man kann mathematisch zeigen, dass keine weiteren Einflüsse existieren. Wir fanden so heraus, dass FRAP-Messungen, welche die spezifische Geometrie

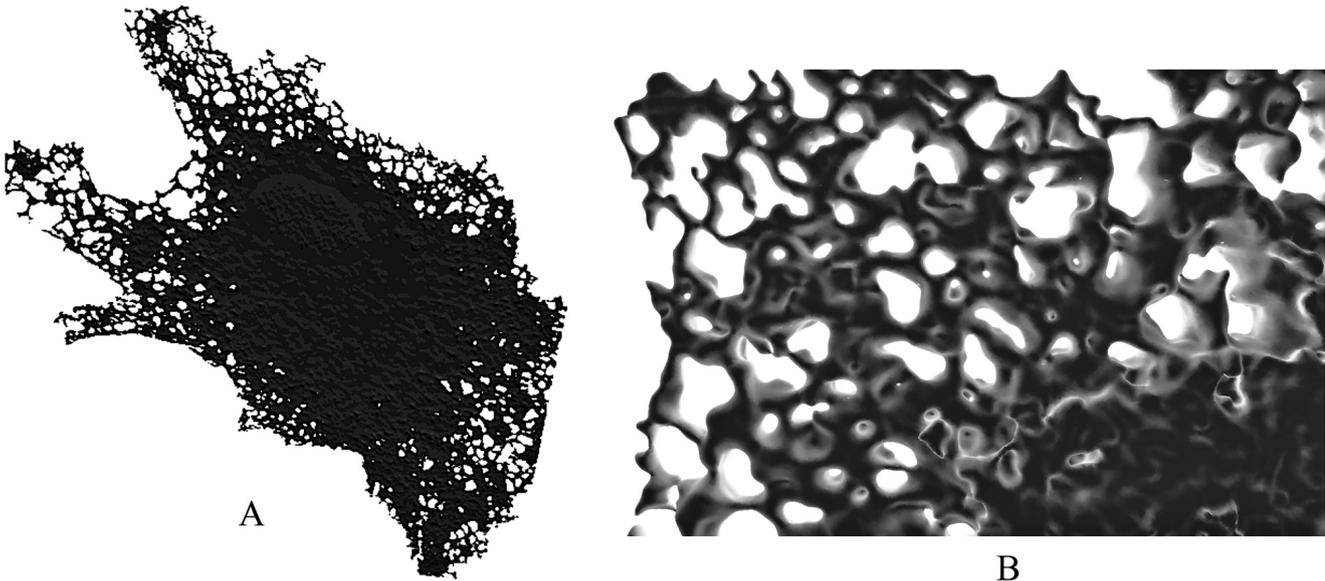


Abb. 5. (A) Beispiel eines im Computer dreidimensional rekonstruierten Endoplasmatischen Retikulums (ER) einer lebenden VERO-Zelle (Mikroskopie: Anna Mezzacasa, Helenius-Gruppe, ETH Zürich). (B) Vergrößerter Ausschnitt eines rekonstruierten ER, wie es für die folgenden Computersimulationen verwendet wurde.

Fig. 5. (A) Example of a three-dimensional computer reconstruction of the Endoplasmic Reticulum (ER) of a live VERO cell (microscopy: Anna Mezzacasa, Helenius group, ETH Zurich). (B) Magnified part of a reconstructed ER as it was used in the subsequent computer simulations.

der individuellen Organelle nicht explizit berücksichtigen, eine Messunsicherheit von mindestens 250% aufweisen, entsprechend dem Variationsanteil der vernachlässigten Geometrie.

In einem weiteren Schritt wurde das numerische Simulationsverfahren vom Lumen auch auf die ER-Membran ausgeweitet (SBALZARINI et al., 2006). Dies bedingte die Entwicklung eines neuen Simulationsverfahrens, da es bis

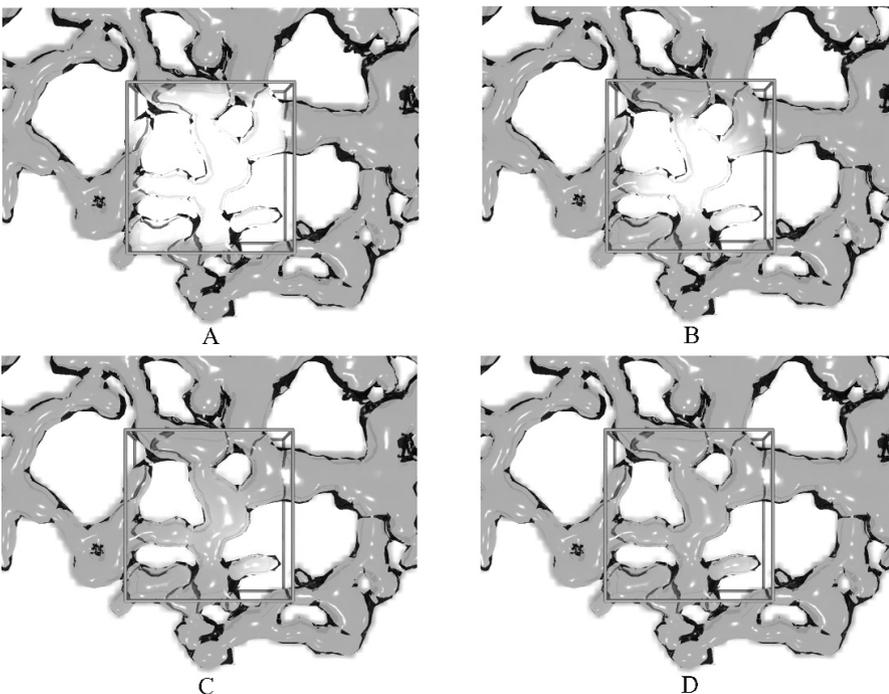


Abb. 6. Ein virtuelles FRAP-Experiment (vgl. Abb. 4): im Computer simulierter Fluoreszenz-Wiederaufbau in einem dreidimensional rekonstruierten Endoplasmatischen Retikulum (ER, siehe Abb. 5). Die lokale Fluoreszenzintensität im Innenraum des ER ist grau dargestellt, der anfänglich gebleichte Bereich ist als Gitterwürfel repräsentiert. Die einzelnen Bilder zeigen die Fluoreszenzverteilung direkt nach dem Bleichen (A), sowie zu den Simulationszeiten 0.25 (B), 1.50 (C) und 3.00 (D).

Fig. 6. A virtual FRAP experiment (cf. Fig. 4): computer simulation of the fluorescence recovery in a three-dimensionally reconstructed Endoplasmic Reticulum (ER, see Fig. 5). The local fluorescence intensity in the ER lumen is depicted in gray; the initially bleached volume is highlighted by the outlines of a cube. Individual images show the fluorescence intensity distribution directly after bleaching (A), as well as at simulation times 0.25 (B), 1.50 (C), and 3.00 (D).

dahin nicht möglich war, Diffusionsprozesse entlang gekrümmter Oberflächen von komplexer Gestalt genau genug zu simulieren.

Dieses Projekt ist ein gutes Beispiel dafür, wie beide Seiten symbiotisch profitieren können. Wurde es in der Biologie erstmals möglich, Diffusionskonstanten mittels FRAP in lebenden Zellen zu messen und die Geometrieinflüsse in Organellen zu quantifizieren, so hat das Projekt auch auf der Informatikseite zahlreiche Verbindungen. Zum einen verbindet die neuentwickelte Methode zur Simulation von Diffusionsprozessen auf gekrümmten Oberflächen direkt zum Gebiet des wissenschaftlichen Rechnens. Andererseits war zur Durchführung der Simulationen ein Supercomputer mit hunderten von parallelen Prozessoren nötig. Dies erforderte die Entwicklung einer neuen Software (SBALZARINI et al., 2006a), wodurch auch die Bereiche Hochleistungsrechnen und Software-Entwicklung profitierten.

5 FAZIT

Zusammenfassend kann man sagen, dass es die Informatik der Biologie nicht nur erlaubt, grosse Datenmengen zu verarbeiten, sondern auch Modelle von bisher unerreichter Komplexität und Wirklichkeitstreue aufzustellen und zu untersuchen. Durch die grossen Datenmengen wird statistische Signifikanz erreicht, und der Bias (systematische Verzerrung) durch den Experimentator wird eliminiert – man erreicht echte Reproduzierbarkeit. Die Computermodelle erlauben es, Variablen zu kontrollieren und zu beobachten, die es im Experiment nicht sind. Durch die Komplexität der von ihr betrachteten Systeme inspiriert die Biologie dabei auch stets neue Entwicklungen in der Informatik. Neue Algorithmen, Programmier Techniken und numerische Verfahren sind vielfach nötig, mit direkten Auswirkungen auf diverse Gebiete der Kerninformatik. Somit werden die Grenzen des Wissens auf beiden Seiten verschoben, eine essentielle Voraussetzung für interdisziplinäre Arbeit. Letztlich haben Biologie und Informatik als gemeinsames

Ziel das Verständnis komplexer informationsverarbeitender Systeme zu fördern – naturgegeben oder menschengemacht.

6 VERDANKUNGEN

Mein herzlichster Dank geht an meinen Doktorvater, Prof. Dr. Petros Koumoutsakos, sowie an alle Mitarbeiter meiner Forschungsgruppe: Jo Helmuth, Guy Levy, Christian Müller und Birte Schrader. Ebenfalls danke ich den Partnergruppen in der Biologie, allen voran den Gruppen von Prof. Helenius und Prof. Greber, ohne die unsere Arbeit nicht möglich wäre.

7 LITERATUR

- MICROSOFT RESEARCH. 2006. Towards 2020 Science. Report.
- EWERS, H., SMITH, A. E., SBALZARINI, I. F., LILIE, H., KOU-MOUTSAKOS, P. & HELENIUS, A. 2005. Single-particle tracking of murine polyoma virus-like particles on live cells and artificial membranes. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15110–15115.
- SBALZARINI, I. F. & KOU-MOUTSAKOS, P. 2005. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology* 151, 182–195.
- SBALZARINI, I. F., MEZZACASA, A., HELENIUS, A. & KOU-MOUTSAKOS, P. 2005a. Effects of organelle shape on fluorescence recovery after photobleaching. *Biophysical Journal* 89, 1482–1492.
- SBALZARINI, I. F., HAYER, A., HELENIUS, A. & KOU-MOUTSAKOS, P. 2006. Simulations of (an)isotropic diffusion on curved biological surfaces. *Biophysical Journal* 90, 878–885.
- SBALZARINI, I. F., WALTHER, J. H., BERGDORF, M., HIEBER, S. E., KOTSALIS, E. M. & KOU-MOUTSAKOS, P. 2006a. PPM – a highly efficient parallel particle-mesh library for the simulation of continuum systems. *Journal of Computational Physics* 215, 566–588.
- WHITE, J. & STELZER, E. 1999. Photobleaching GFP reveals protein dynamics inside live cells. *Trends in Cell Biology* 9, 61–65.

Prof. Dr. Ivo F. Sbalzarini, Computational Biophysics Laboratory, Institute of Computational Science, ETH Zürich CAB H68, Universitätstrasse 6, CH-8092 Zürich, ivos@ethz.ch, www.cbl.ethz.ch