

Mai 2000

life **Technologien** science

Ein gemeinsames Supplement von LaborPraxis, PROCESS und UmweltMagazin

Labor

Proteomics – was kommt
nach den Genen?

Pharma/Bioinformatik

Virtuelle Suche nach
neuen Wirkstoffen

Agrar

Maßgeschneiderte
nachwachsende Rohstoffe

Umwelt

Bakterien als Ölfresser

Messen

BioEngineering & Products

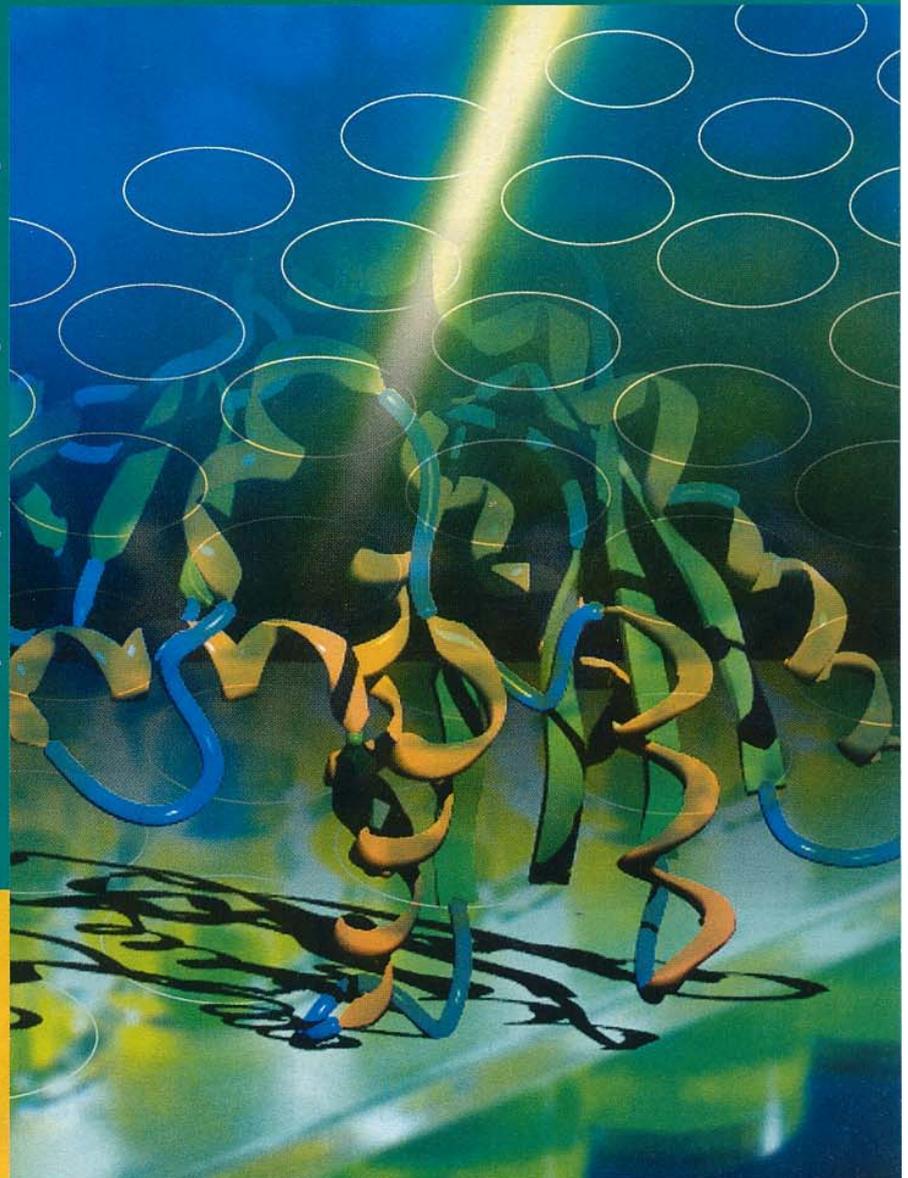
Bio-Gen-Tec-Forum NRW

Synthesis
Screening
Sequenzing

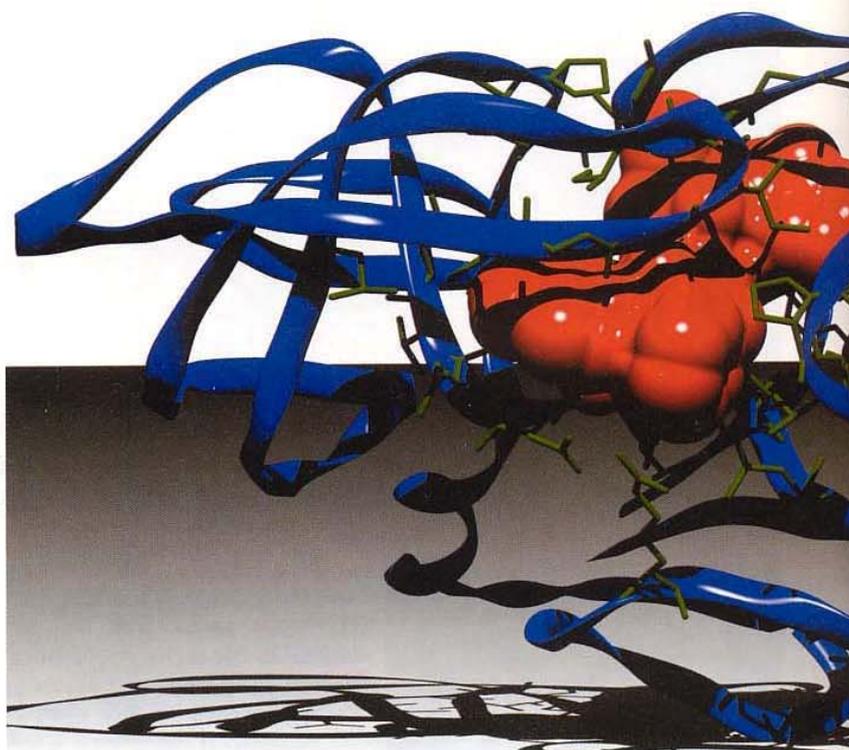
ACHEMA 2000



Life Science
Technologien
finden Sie in
Halle 9, Stand A10



Eine Schlüsselrolle für die pharmazeutische Forschung spielen Zielproteine, an denen Wirkstoffe ansetzen und eine Krankheit verhindern bzw. lindern können. Die aufwendige und kostspielige Jagd nach den geeigneten Zielproteinen und Wirkstoffen lässt sich mittlerweile mit der entsprechenden Software beschleunigen. Einen besonderen Kick erhält die Suche, wenn man spezielle Softwarelösungen auf Parallelrechner portiert. So geschehen mit Hilfe eines Technologietransfer-Projektes „Dechema“, das mit Hilfe des europäischen Förderprogrammes „Esprit“ unterstützt wurde.



Virtuelle Suche nach neuen Zielproteinen und Wirkstoffen in Moleküldatenbanken

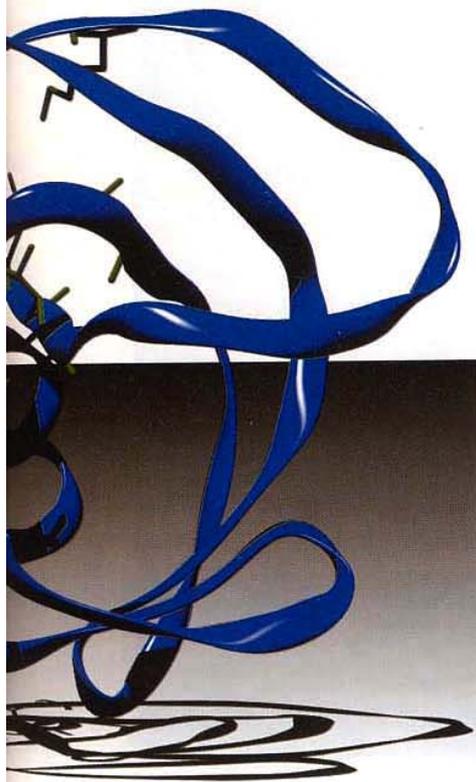
Schneller Check beim Screening

STEPHAN SPRINGSTUBBE, MATTHIAS RAREY UND RALF ZIMMER*

Im Gegensatz zu vielen Problemen aus dem ingenieurwissenschaftlichen Umfeld, z.B. dem Entwurf komplexer Anlagen, Maschinen oder Microchips, sind die zugrunde liegenden Modelle für Fragestellungen im biochemischen Umfeld noch sehr ungenau. Beispielsweise lassen sich zentrale Größen wie die Bindungsenergie von Protein-Ligand-Komplexen nur mit hohen Fehlerraten vorher sagen. Bei der Entwicklung von Medikamenten kommt es zudem nicht nur auf die Wechselwirkung des Wirkstoffs mit dem Zielprotein an. Vielmehr ist der Einfluss des Wirkstoffs auf den gesamten Organismus zu unter-

suchen. Schon wegen dieser Komplexität wird der Computer nicht das Allheilmittel für alle Fragestellungen in der pharmazeutischen Forschung sein. Trotzdem nimmt die Bedeutung des Rechners in der Wirkstoffforschung zu. Der Grund hierfür liegt in der sehr großen Menge potenzieller Moleküle, die als Wirkstoff für eine spezielle Applikation in Frage kommen. Mit dem Computer lässt sich eine sinnvolle Vorauswahl treffen und Molekülbibliotheken können auf bestimmte Eigenschaften optimiert werden, bevor sie synthetisiert werden. Damit lassen sich auch die Folgeexperimente besser planen.

* ST. SPRINGSTUBBE, DR. M. RAREY, DR. R. ZIMMER, SCAI, GMD – FORSCHUNGSZENTRUM INFORMATIONSTECHNIK GMBH.



1) Die Abbildung zeigt das HIV-Protease Protein (blau) im Komplex mit einem Inhibitor (rot). HIV-Protease ist Teil des Replikationsmechanismus des HIV-Virus. Durch die Bindung des Inhibitors an das aktive Zentrum der Protease wird diese in ihrer Funktion blockiert. Obwohl HIV-Protease-Inhibitoren nicht die Infektion heilen, verringern sie jedoch drastisch die Viruslast und können somit deutlich die Lebenserwartung eines HIV-Infizierten erhöhen.

Mit Esprit auf die Suche

Eine Möglichkeit, die Suche zu beschleunigen, sind spezielle Programme für die Vorhersage von Proteinstrukturen und von ihren Wechselwirkungen mit Wirkstoffen. Diese Programme sind jedoch trotz effizienter Algorithmen sehr rechenintensiv, insbesondere beim Durchsuchen großer Datenbanken. Eine spezielle Softwaretechnologie – das Bearbeiten mit Parallelrechnern – könnte die Rechenzeiten erheblich verkürzen. Allerdings sind diese Rechenkünstler nicht gerade kostengünstig, daher scheuen gerade kleinere Unternehmen die hohe Investition. Die Europäische Kommis-

sion wollte dieser Technologie eine Chance geben und hat innerhalb des ESPRIT-Forschungsrahmens ein spezielles Programm initiiert: Preparatory, Support and Transfer Activities oder kurz PST-Programm. Das Ziel des Programmes ist die Erhöhung der Wettbewerbsfähigkeit der europäischen Industrie und die Verbesserung der Akzeptanz von parallel rechnenden Computersystemen. Derzeit gibt es innerhalb des PST-Programmes 175 Technologietransfer-Projekte mit industriellen Partnern aus fast allen Wirtschaftszweigen.

Eines dieser Projekte entstand aus der Kooperation mit der DECHEMA – Deutsche Gesellschaft für Chemisches Apparatewesen, Chemische Technik und Biotechnologie e.V. und der Forschungsgruppe von Prof. Lengauer am Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI) im GMD – Forschungszentrum Informationstechnik GmbH. Es beinhaltet drei Erfahrungsstudien mit industriellen Anwendern in der Pharmaindustrie. In diesen Studien werden neue Softwarelösungen in der Pharmaforschung eingesetzt, aus denen Vorhersagen zur Struktur und Wechselwirkung von Wirkstoffen für die Arzneimittel-Entwicklung gewonnen werden.

Medikamente in drei Phasen

Um die entwickelten Programme zu verstehen, muss man die Vorgehensweise bei der Suche nach neuen Wirkstoffen begreifen. Die Entwicklung eines neuen Medikaments lässt sich in drei Phasen einteilen. In der ersten Phase, der Suche nach Zielproteinen, muss die Krankheit auf molekularbiologischer Ebene soweit verstanden werden, dass einzelne Proteine und deren Bedeutung für das Krankheitsgeschehen bekannt sind. Ist ein Zusammenhang zwischen der Unterdrückung (oder Verstärkung) einer Proteinfunktion und dem Krankheitsbild erkannt, wird dieses Protein zum Zielprotein.

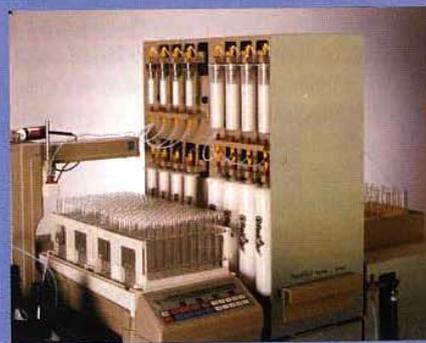
In der zweiten Phase wird der eigentliche Wirkstoff entwickelt. Das Molekül soll gut an das Zielprotein binden (molecular docking), und es dadurch in seiner Funktion hemmen. Gleichzeitig muss es gut verträglich sein und sich in hoher Konzentration am Wirkort anreichern.

Man beginnt mit der Suche nach einer Leitstruktur – also einem Molekül,

AISS



Sq 16 x



Automatische sequentielle Flash-Reinigung grosser Probenzahlen

- binäre Gradienten-Elution
- automatische Probenaufgabe
- detektorgesteuerte Fraktionsnahme
- mit PC-Steuerung

Fragen Sie auch nach unseren Einsäulen- bzw. Parallel-Flash-Reinigungssystemen

Axel Semrau GmbH & Co
 Tel.: 0 23 39 / 1209-0
 Fax: 0 23 39 / 60 30
 Email: info@axelsemrau.de
 Internet: www.axelsemrau.de

▲ Weitere Informationen über Kennziffer 166

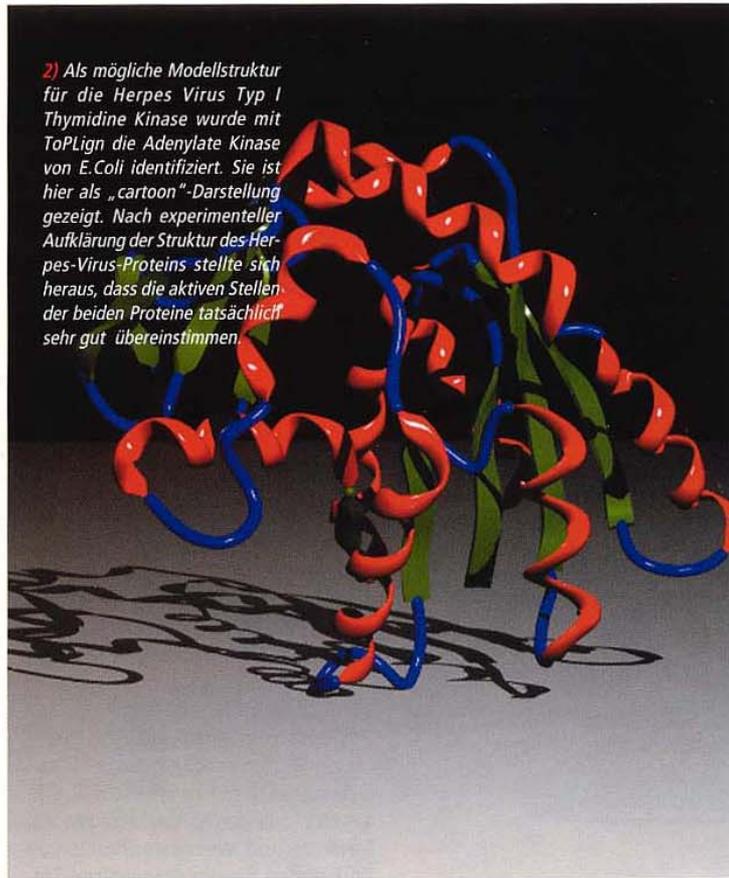
das als erster Vorschlag für das Medikament dient. Idealerweise kann die Leitstruktur anschließend so modifiziert werden, dass sich das resultierende Molekül zum Wirkstoff eignet. In der dritten Phase wird der Wirkstoff in mehreren Schritten auf seine Verträglichkeit und Wirksamkeit als Medikament getestet.

Steht ein Strukturmodell des Proteins zur Verfügung, kann dieses direkt zur Suche nach passenden Wirkstoffen verwendet werden. Steht jedoch kein solches Modell zur Verfügung, wird die Ähnlichkeit zu Molekülen, die an das Zielprotein binden, als Maß für die Eignung zum Wirkstoff verwendet. Das Spektrum der hierbei zu durchsuchenden Datenmenge reicht von wenigen hundert bereits vorselektierten Molekülen über große Datenbanken mit mehreren Millionen Molekülen bis zu virtuellen kombinatorischen Molekülbibliotheken, aus denen sich theoretisch Billionen verschiedene Moleküle synthetisieren lassen.

Für eine Reihe von Zielproteinen lässt sich deren dreidimensionale Struktur mit experimentellen Methoden oder durch Modellierung bestimmen. Auf der Basis der Proteinstruktur kann man nun direkt nach Molekülen suchen, die sowohl geometrisch als auch chemisch in das aktive Zentrum des Proteins passen (strukturbasiertes Wirkstoffdesign). Zu diesem Zweck muss man das molekulare Docking-Problem lösen: Wenn ein Protein und ein kleineres Molekül – in diesem Zusammenhang Ligand genannt – gegeben ist, muss man herausfinden, ob der Ligand einen Komplex mit dem Protein bildet, wie stark die Bindung ist und wie der Komplex geometrisch aussieht.

Schnellere Vorhersage

Für dieses Problem hat das GMD-Institut SCAI in Kooperation mit der BASF AG, Ludwigshafen und unter Beteiligung der Merck KGaA, Darmstadt das Softwarewerkzeug FlexX entwickelt. Die Schwierigkeit des Docking-Problems ergibt sich zum einen aus der Schätzung der freien Energie eines molekularen Komplexes in wässriger Lösung und zum anderen aus der Flexibilität der beteiligten Moleküle. In FlexX wird die wichtige Flexibilität des Liganden durch eine dreiphasige Platzierungsstrategie in der Berechnung berücksichtigt. Der Ligand wird



2) Als mögliche Modellstruktur für die Herpes Virus Typ 1 Thymidine Kinase wurde mit ToPLign die Adenylate Kinase von *E.Coli* identifiziert. Sie ist hier als „cartoon“-Darstellung gezeigt. Nach experimenteller Aufklärung der Struktur des Herpes-Virus-Proteins stellte sich heraus, dass die aktiven Stellen der beiden Proteine tatsächlich sehr gut übereinstimmen.

hierzu in Fragmente zerlegt. In der ersten Phase werden einzelne Fragmente zur Platzierung ausgewählt, die dann in der zweiten Phase unabhängig vom Rest des Moleküls in die Bindungstasche des Proteins platziert werden. Hierzu wurde ein Verfahren aus der Mustererkennung an das Docking-Problem adaptiert. Schließlich wird in der dritten Phase der Ligand in inkrementellen Schritten in der Bindungstasche des Proteins wieder zusammengebaut. Hierfür werden nacheinander noch nicht platzierte Fragmente an die bereits platzierten Teile angehängt, energetisch bewertet und ausgewählt, bis der Ligand vollständig in der Bindungstasche des Proteins platziert ist.

Die Vorhersage eines Protein-Ligand-Komplexes mit FlexX dauert im Durchschnitt sensationelle 90 Sekunden; damit können ca. 1000 Ligandmoleküle pro Tag auf einem einzelnen Prozessor bearbeitet werden. FlexX ist in ca. 70 Prozent der Testfälle in der Lage, eine der experimentell ermittelten Struktur hinreichend ähnliche vorzuschlagen. In etwa 50 Prozent der Testfälle wird diese Lösung sogar auch en-

ergetisch am günstigsten bewertet. Dr. Matthias Rarey, Entwickler des FlexX-Systems in der GMD, stellt fest: „Mit FlexX haben wir eines der schnellsten heute verfügbaren Docking-Programme entwickelt und damit virtuelles High Throughput Screening zur effizienten Wirkstoffsuche mit dem Computer ermöglicht!“

Einsatz in der Praxis

Die Merck KGaA testete außerdem ein weiteres Programm, das von der GMD entwickelte Softwarepaket ToPLign (Toolbox for Protein Alignment). Mit ToPLign lassen sich Vorhersagen zur dreidimensionalen Struktur und damit häufig auch zur biochemischen Funktion gegebener Proteinsequenzen treffen. Hierbei ist insbesondere die Strukturvorhersage von Bindungsregionen von Interesse. Dr. Ralf Zimmer, Entwickler des ToPLign-Systems im GMD-Institut SCAI, stellt fest: „Unsere Erfahrungen zeigen, dass auf der Basis einer 20 Prozent Übereinstimmung mit der Sequenz eines strukturell bekannten Proteins meist schon sehr gute Vorhersagen zur Faltung ei-

Die Neuheit !!!

**Vollständig automatisierte
Bioreaktoren mit Mikro-
prozessor, Fernsteuerung und
Netzanbindung für**

**Analytik, Biotechnologie,
Pharmazie und Medizintechnik**

- Multireaktorsysteme mit 3 bis 48 Reaktionsgefäßen
- Variable und voll automatisierte Bioreaktoren von 50 ml bis 10 l
- Manuelle und automatische Probenehmer für alle Reaktorsysteme am Markt
- Autoklavierbare pH- und Sauerstoffsonden Ø 6 und 12 mm
- pH- und Sauerstoffregler
- Mass-Flow und Gasmischer und viele weitere Extras



**„Absolute Präzision bei
einfachem Handling“**

**Das Schlauchpumpensystem
für alle Anwendungen**

- Pumpraten im Bereich von 0,1 µl bis 2000 ml/min
- bis zu 8 Rollen für geringste Pulsation
- sekundenschneller und leichter Schlauchwechsel
- hohe Schlauchlebensdauer
- feste Drehzahl bis hin zur Mikroprozessor gesteuerten kalibrierbaren Dosierpumpe
- mit Digitalanzeige und externer Ansteuerung

*Bitte besuchen Sie uns auf der
ACHEMA 2000 in Frankfurt
22.–27. Mai 2000
Halle 7.0 Stand D2–E4*

meredos GmbH

Alte Dorfstraße 37
D-37120 Bovenden

Telefon ++49 (0) 55 03/9 13-24

Fax ++49 (0) 55 03/9 13-25

E-Mail: info@meredos.com

<http://www.meredos.com>

nes unbekanntes Proteins möglich sind!“ Anhand unterschiedlicher Testfälle wird die Leistungsfähigkeit und vor allen Dingen die Genauigkeit der mit ToPLign gewonnenen Strukturvorhersagen im Technologietransfer-Projekt „Dechema“ von Merck geprüft und bewertet. Die Studienobjekte bestehen aus bekannten Testsequenzen, aber auch aus Anwendungen aus der aktuellen Forschung bei Merck, z. B. der Vorhersage aller ORFs (open reading frames), der Genome vollständig sequenzierter Mikroorganismen und der Analyse von großen Datenbanken mit menschlichen ESTs (expressed sequence tags).

Merck erwartet von dem Einsatz dieses Programmsystems neue Erkenntnisse und eine wesentlich effektivere Verfahrensweise bei der Strukturanalyse von Proteinsequenzen. Diese Softwarelösungen sind nach Aussagen von Merck derzeit die einzige Alternative zu sehr zeit- und kostenaufwendigen experimentellen Verfahren. Die ToPLign-Software steht über das Internet kostenfrei für Protein-Strukturvorhersagen zur Verfügung.

Die beiden weiteren Erfahrungsstudien wurden mit der Boehringer Ingelheim Pharma KG und Groupe Lipha S.A. durchgeführt. Bei Boehringer Ingelheim findet das bereits erwähnte Softwarepaket FlexX und bei Lipha die Software GOLD von dem Cambridge Crystallographic Data Centre Anwendung. Beide Programme berechnen und sagen Wechselwirkungen zwischen Proteinen und potenziellen Wirkstoffen vorher. Im Vergleich zu FlexX kann in GOLD auch partielle Proteinflexibilität berücksichtigt werden, was jedoch auch höhere Laufzeiten bedingt.

Gewinn von Zeit und Genauigkeit

Allen drei Projekten gemeinsam ist, dass eine enorme Zeitersparnis gelang. Bei Merck konnte die Proteinstrukturvorhersage für ganze Genome auf einem 24-Prozessor SGI-Parallelrechner um mehr als den Faktor 20 beschleunigt werden. Der parallelisierte FlexX-Code wurde von Boehringer Ingelheim zum Docking von 100 000 Liganden in das aktive Zentrum eines Proteins eingesetzt. Die ursprüngliche Zeit von drei Monaten bei sequentieller Vorgehensweise konnte bei Nutzung von sechs Prozessoren

auf einen halben Monat reduziert werden. Während bei Anwendung der sequentiellen GOLD-Version die komplette Bindungsanalyse von 2502 Stereo-Isomeren in drei existierende Proteinkonformationen mit Standardparametern 240 Tage benötigt, dauerte es mit der parallelen Version und modifizierten Parametern auf einem 4-Prozessor SGI-System nur 12,6 Tage. Bei allen Softwarelösungen zeigte sich, dass eine Zunahme von beteiligten Prozessoren immer auch zu einer proportionalen Verringerung der Rechenzeit führte und somit die Laufzeiten weiter beschleunigt werden können.

Ausblick

Da bei der Wirkstoffsuche manchmal mehr als 250 000 Kandidaten auf ihre Bindung an ein Protein hin untersucht werden, ist eine Vorauswahl von wirklich geeigneten Kandidaten von größtem Interesse. Ähnliches gilt für die Auswahl von Zielstrukturen aus einer großen Zahl von mehreren 10 000 Kandidaten, z. B. aus DNA-chip-Experimenten. Die nun zur Verfügung stehende Softwaretechnologie macht diesen Auswahlprozess und damit ein virtuelles „High Throughput Screening“-Verfahren erstmalig effizient möglich. Ein kurzes Rechenbeispiel zeigt das Potenzial des parallelen Rechnens. Eine Verringerung der zu analysierenden Bindungskandidaten um den Faktor zehn verkürzt die Suche nach einem geeigneten Kandidaten von neun Monaten auf einen Monat. Dies lässt eine Kosteneinsparung von einer Million US-Dollar durchaus realistisch erscheinen.

Weitere Informationen erhalten Sie von:

Dipl.-Inform. Stephan Springstube,
Institut für Algorithmen und
Wissenschaftliches Rechnen –
GMD – Forschungszentrum
Informationstechnik GmbH,
53754 Sankt Augustin,
E-Mail: springstube@gmd.de,
URL: www.gmd.de/SCAI/ttn.

▲ Weitere Informationen über Kennziffer **168**

