

Funktionelle Annotation regulatorischer Genombereiche

Zusammenfassender Abschlußbericht

Teilnehmer und Teilprojekte:

Handhabung und Aufbereitung genregulatorischer Daten

Dr. Edgar Wingender (Koordinator)

GBF, AG Bioinformatik

Mascheroder Weg 1

D-38124 Braunschweig

Systematische Erkennung potentieller Promotor-Regionen auf der Grundlage von Transkriptionsfaktor-Bindungsstellen

Dr. Thomas Werner

GSF

AG BIODV/Institut für Säugetiergenetik

Ingolstädter Landstraße 1

D-85764 Oberschleißheim

Die molekulare Evolution regulativer Einheiten

Prof. Dr. Andreas Dress

Forschungsschwerpunkt Mathematisierung - Strukturbildungsprozesse

Universität Bielefeld

Postfach 10 01 31

D-33501 Bielefeld

Strukturelle Beschreibung und Analyse genregulatorischer DNA-Sequenzen

Dr. Heinz Sklenar

Max-Delbrück-Centrum für Molekulare Medizin (MDC)

Robert-Rössle-Str.10

D-13122 Berlin

Ziele

Unmittelbare Projektziele:

- Identifizierung, systematische Analyse und Klassifizierung regulatorischer Regionen in genomischen DNA-Sequenzen.
- Systematische Erfassung der relevanten experimentellen Daten
- Weiterentwicklung bereits existierender Programme für die Charakterisierung und Identifizierung regulatorischer Regionen
- Neuentwicklung und Implementation von Algorithmen
- Anwendung der verschiedenen Tools auf die Annotation und Klassifizierung genomischer Sequenzen
- Erstellung einer Ergebnisdatenbank

Mittelfristige Ziele, die soweit wie möglich erreicht werden sollten:

Ableitung einer systematischen funktionellen Klassifizierung regulatorischer Regionen aus der Fülle annotierter genomischer Sequenzen

Langfristige Ziele, die über den unmittelbaren Projektrahmen hinausreichen:

Signifikanter Beitrag zu einem umfassenden Verständnis der "Sprache des Genoms", sowohl in Hinblick auf die Semantik wie auf die Grammatik.

Organisatorische Ziele:

- Behauptung der weltweiten Spitzenstellung im Bereich der Interpretation regulatorischer Potentiale in genomischen Sequenzen
- Überführung der Methodik in eine systematische Anwendung
- Etablierung einer sich selbst tragenden Service-Einrichtung bzw. Übergabe an Ausgründungen

Ergebnisse

Wie in den Berichten der Teilprojekte ausführlich dargestellt ist, konnten sowohl die Teilprojektziele als auch das Gesamtziel im Wesentlichen erreicht und in einzelnen Gebieten sogar übertroffen werden. So wurden völlig neue und neuartige Wissensbasen über genomische Merkmale erstellt, die maßgeblich an der Genregulation beteiligt sind, und der wissenschaftlichen Öffentlichkeit zur Verfügung gestellt (Braunschweig).

Es wurden bereits in das Projekt eingebrachte Softwaretools für die Erkennung regulatorischer Bereiche signifikant verbessert (z. B. MatInspector; Braunschweig/München), insbesondere aber Algorithmen für die Erkennung kontextabhängiger Merkmale entwickelt und als öffentlich verfügbare Programme implementiert. Vor allem konnte erstmals eine Methode für die Erkennung eukaryoter Promotoren entwickelt werden, die verifizierbare Ergebnisse in genomischem Maßstab zu liefern vermag (München).

Bei der Berechnung der dreidimensionalen Struktur von DNA-Sequenzen wurden signifikante Fortschritte erzielt (Berlin). Die Korrelation struktureller mit funktionellen Merkmalen, d. h. mit der Fähigkeit eines Sequenzelementes, als Transkriptionsfaktorbindungsstelle zu wirken, konnte jedoch noch nicht eindeutig gezeigt werden (Berlin/Braunschweig/München), wobei sich Hinweise auf zusätzlich zu berücksichtigende Parameter ergaben. In dieser Richtung hat gegen Projektende ein neuer Ansatz, der über die Berechnung der *a priori*-Struktur hinausgehend die Deformationsenergie für die im DNA-Proteinkomplex vorliegende Struktur als *scoring function* nutzt, bereits erste vielversprechende Ergebnisse erbracht. Dieses Verfahren wird im Rahmen eines anderen Projektes weiterverfolgt werden.

Um einer umfassenden funktionellen Klassifizierung regulatorischer Genombereiche näher zu kommen, wurde ein Formalismus zur Abbildung von "Genexpressionsquellen" (Organen/Geweben, Zelltypen, physiologischen Systemen und deren Entwicklungsstadien) entworfen (Bielefeld/Braunschweig). Auf dieser Grundlage werden zukünftig Genexpressionsmuster dargestellt werden können.

Insgesamt flossen Projektergebnisse in 19 (5+6+?+8) wissenschaftliche Publikationen ein.

Über die Projektlaufzeit fanden 4 Projekttreffen aller Beteiligter (08./09.12.1997 in Bielefeld; 13./14.10.1998 in Braunschweig; 22.02.1999, Statusseminar in Braunschweig; 08./09.05.2000 Abschlußtreffen in München) sowie verschiedene bilaterale Arbeitstreffen statt (z. B. 8.05.1998 BI-BS; 17.-20.05.1998 M-B; 18.06.1999 BI-BS; 12.05.2000 BS-M). Der Austausch von Zwischenergebnissen sowie die Feinabstimmung der Arbeiten erfolgte kontinuierlich und reibungslos über das Internet.

Schließlich ist es gelungen, die öffentliche Förderung der genannten Beiträge zu einer internationalen Bioinformatik-Infrastruktur nachhaltig wirksam zu machen, indem sie an zwei Ausgründungen in München (Genomatix Software GmbH) und Braunschweig (BIOBASE GmbH) übertragen wurden. Entsprechende Abkommen über den Technologietransfer wurden zwischen GSF und Genomatix sowie zwischen GBF, MDC und BIOBASE abgeschlossen. Beide Firmen sind inzwischen erfolgreich international tätig und haben 10 (Genomatix) bzw. 38 (BIOBASE) Arbeitsplätze geschaffen (Stand November 2000). Beide Firmen haben damit auch die Verpflichtung zur Pflege und Weiterentwicklung der jeweiligen Entwicklungen übernommen.

Bewertung

Die maßgeblichen Fortschritte insbesondere bei der Erarbeitung integrativer Tools (Datenressourcen und Programmentwicklung) wurden in der engen Kooperation zwischen den Projektbeteiligten in Braunschweig und München erzielt, wo auch die teilprojektspezifischen Ziele vollständig erreicht wurden. Auch das Berliner Teilprojekt hat, trotz Schwierigkeiten bei den Stellenbesetzungen, beachtenswerte Leistungen erzielt; Es wurden auch Anstrengungen aller Beteiligter unternommen, die Berliner Tools mit den Braunschweiger Datenbanken und der Münchner Software zu verknüpfen, allerdings stellte sich heraus, daß die dreidimensionale *a priori*-Struktur der DNA nicht hinreichend indikativ für regulatorische Merkmale zu sein scheint. Gravierende personelle Probleme haben leider dazu geführt, daß das Bielefelder Teilprojekt nicht im vorgesehenen Umfang zum Gesamtergebnis beitragen konnte.

Durch die "Übererfüllung" der Vorgaben in anderen Bereichen kann der Verbund insgesamt jedoch als ein großer Erfolg angesehen werden, der auch die methodischen Grundlagen für die Erlangung seiner mittelfristigen Ziele in vollem Umfang legen konnte. Der Beitrag zur Erreichung der langfristigen, über die Projektlaufzeit hinausreichenden Ziele ist kaum zu überschätzen, wie sich bereits jetzt am wissenschaftlichen Erfolg der beteiligten Arbeitsgruppen und am sich abzeichnenden wirtschaftlichen Erfolg der Ausgründungen Genomatix und BIOBASE zeigt.