



Protein-Ressourcen

Aus den großen Sequenzierungsprojekten ergibt sich eine Flut genomischer Informationen, allen voran aus dem Human-Genom-Projekt. Die „Protein Information Resource“ (PIR) ist eine der Institutionen, die diese Flut kanalisieren und für die Forschung nutzbar machen. Kernstück der PIR ist die Protein Sequence Database (PSD), eine wachsende Datenbank von heute mehr als einer viertel Million Proteinsequenzen, auf die Benutzer frei über das Web zugreifen können. Diese Sequenzen sind annotiert, d.h. sie sind versehen mit zusätzlicher biologisch oder medizinisch relevanter Information in Form kurzer Texte oder von Verweisen auf Originalliteratur oder Hyperlinks auf andere Datenbanken.

Die PIR wird von einem Team aus etwa einem Dutzend Spezialisten der National Biomedical Research Foundation gepflegt, die an der Georgetown-Universität in Washington angesiedelt

ist. Man darf dieses Team beglückwünschen, denn der Pflögel ist wohl geraten. Das sieht man bereits bei der funktionellen und ergonomischen Eingangsseite, die auf einen Blick eine Übersicht der wichtigsten Teile und Funktionen der PIR gibt. Schon hier kann eine einfache Suche in der PSD gestartet werden. Zu Eingangsseiten verschiedener Datenbanken und zu bioinformatischen Werkzeugen gelangt man mit nur einem Mausklick. Die Funktionen sind im Allgemeinen gut dokumentiert und ihre Benutzung leicht erlernbar.

Ein Anwendungsbeispiel: Wir suchen nach Informationen über das Enzym Tyrosin-Hydroxylase. Dazu schreiben wir auf der Eingangsseite „tyrosine hydroxylase“ in das Suchfenster links oben, drücken die Enter-Taste oder klicken auf „Go!“ und nach wenigen Sekunden erhalten wir eine Liste von Titeln einschlägiger PSD-Einträge, in denen unser Suchbegriff vorkommt. Wir können per Mausklick nun einzelne Einträge aus der Liste auf den Bildschirm rufen oder die Suche modifizieren, zum Beispiel wenn sich herausstellt, dass der Suchbegriff zu grob war und damit unsere Liste länger als nötig geraten ist. Zur Modifikation von Suchen stellt die PIR eine elegante Funktion bereit, die man bei verwandten Datenbanken vermisst (siehe Abbildung 1 unten). In unserem Beispiel könnten wir die Suche nach Tyrosin-Hydroxylasen auf die Spezies Mensch einschränken. Dazu stellen wir neben dem zweiten Suchfenster die Suchkategorie auf „Species“, tragen „homo sapiens“ ein und klicken auf den „Submit“-Knopf unten. Nun schrumpft unsere Liste von 27 Einträgen auf nur noch zwei, wovon nur einer tatsächlich die menschliche Tyrosin-Hydroxylase beschreibt. Abbildung 2 zeigt einen Ausschnitt aus diesem Eintrag. Neben der Sequenz der Aminosäuren enthält der Eintrag unter anderem Verweise auf Originalarbeiten und Hyperlinks auf Ab-

FUNCTION	
#description	catalyzes the 3'-hydroxylation of tyrosine to 3', 4'-dihydroxyphenylalanine by tetrahydrobiopterin and oxygen
#pathway	catecholamine biosynthesis
#note	this is the rate-limiting step in catecholamine biosynthesis
CLASSIFICATION	
#superfamily	phenylalanine 4-monooxygenase
KEYWORDS	
	alternative splicing; biotin; catecholamine biosynthesis; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein
FEATURE	
1-528	#product tyrosine 3-monooxygenase, splice form 4 #status predicted #label MAT4\
1-34,62-528	#product tyrosine 3-monooxygenase, splice form 2 #status predicted #label MAT2\
1-30,35-528	#product tyrosine 3-monooxygenase, splice form 3 #status predicted #label MAT3\
1-30,62-528	#product tyrosine 3-monooxygenase, splice form 1 #status predicted #label MAT1\
8	#binding site phosphate (Thr) (covalent) (by unidentified kinase) #status predicted\

Abbildung 2. Ergebnis der Suche nach Tyrosinhydroxylase (Ausschnitt).

stracts in PubMed, Angaben zur Genetik und Hyperlinks auf Einträge der OMIM-Datenbank der Erbkrankheiten, Information zur Funktion, zu Modifikationen und Wechselwirkungen, Hyperlinks auf räumliche Strukturen in der PDB. Außerdem lässt sich die gefundene Sequenz direkt an einige Bioinformatik-Programme senden, wie zum Beispiel an Psi-Blast oder Fasta. So lassen sich schnell einfache Sequenzanalysen und -vergleiche durchführen, aus deren Resultaten man häufig weitergehende Schlüsse auf die räumliche Struktur und Funktion des jeweiligen Proteins ziehen kann. Suchresultate wie auch ganze Datenbanken lassen sich in gängigen Formaten herunterladen.

Das obige Anwendungsbeispiel zeigt nur einen kleinen Teil der Möglichkeiten der PIR. Sie bietet mehr Datenbanken als nur die PSD und auch einiges an Bioinformatik zur Analyse, wenn auch gerade die gebotene Bioinformatik etwas hinter heutigen Möglichkeiten zurückbleibt, zum Beispiel was die Vorhersage räumlicher Strukturen angeht. Hier könnte man sich eine direkte Anbindung (ähnlich wie z.B. bei Psi-Blast) von Threading- oder anderen ähnlichkeitsbasierten Methoden vorstellen, die bereits vorhandene Strukturinformation für die Vorhersage nutzen.

Daniel Hoffmann
CAESAR, Bonn

PIR-PSD Query Results

Modify Query

2 entries were found

WBIHUP phenylalanine 4-monooxygenase (BC 1.14.16.1) - human
WBIHUY4 tyrosine 3-monooxygenase (BC 1.14.16.2), splice form 4 - human

Modify Query

Field	Query String
1. Create the list of all entries containing:	All fields <input type="text" value="tyrosine hydroxylase"/>
2. refine, add to, remove from the list with entries containing:	Species <input type="text" value="homo sapiens"/>
3. refine, add to, remove from the list with entries containing:	All fields <input type="text" value="1"/>

Submit

Abbildung 1. Suchergebnis und Maske zur Verfeinerung der Suche in der Protein Information Resource.

Für weitere Informationen besuchen Sie:
<http://pir.georgetown.edu/>
 oder nehmen Sie Kontakt auf mit
pirmail@nbrf.georgetown.edu