

Koç Üniversitesi'nde  
Bilişimsel Biyoloji ve  
Biyoinformatik  
Merkezi (Center of  
Computational Biology  
and Bioinformatics) adı  
altında yeni bir merkez  
kuruldu.

<http://ccbb.ku.edu.tr>

# KÜ'de Biyoinformatik ve Bilişimsel Biyoloji

Özlem Keskin , Attila Gürsoy ve Metin Türkay

21. yüzyılda biyoloji sadece laboratu-  
varda yapılan bir bilim olmaktan  
çıkışmış, aynı zamanda bilgi teknolojisine  
de dayanan bir bilim dalı haline gelmiştir.  
Bilgisayarların ve bilgisayar yöntemlerinin  
biyolojinin her alanında kullanılmaya  
başlanması günümüzde gitgide kaçınılmaz  
olmuş, biyoinformatik olarak adlandırılan  
yeni bir bilim dalı ortaya çıkmıştır.  
2001 yılının Şubat ayında Science ve  
Nature dergilerinde yayınlanan iki farklı  
makalede insan gen haritasının ilk taslağı  
kamuyuna açıklanmıştır (Human Genome  
Project). Yayınlanan sonuçlar, modern  
biyoloji için yeni bir başlangıcın işaretini  
de vermiştir: Bundan sonra biyoloji ve tıpla  
ilgili araştırmalar protein ve DNA dizilerine  
bağlı olarak yürütülecektir. Bu çalışmalar,  
aslında bir asırlık birikimin sonucudur. İlk  
önce Mendel kanunlarının 20. yüzyılın  
başında tekrar keşfedilmesi ile bu  
araştırmalar başlamış, yüzyılın ilk

çeyreğinde biyologlar hücre içindeki  
bilgilerin nesilden nesile kromozomlar  
tarafından taşındığını bulmuşlardır. İkinci  
çeyrekte ise, DNA'nın üç boyutlu yapısı  
Watson ve Crick tarafından çözülmüştür;  
bu da sonraki çalışmalarda genetik  
özelliklerin taşınma mekanizmasının  
anlaşılmasını sağlamıştır. Bu gelişmelere  
paralel olarak laboratuvar teknikleri ve  
deneysel yöntemlerde ilerlemekteydi. Bilim  
insanları bu teknikleri kullanarak  
laboratuvar ortamında da genetik deneyler  
yapabilmişlerdir.

İnsan Gen Projesi'nin ilk taslağının  
tamamlanması ile DNA moleküllerinin  
nükleotidlerinin (A,T,G,C) sırasının tayini  
hemen hemen bitmiş olup (genbank.org),  
bundan sonraki aşama DNA'mızdaki  
genlerin yerlerinin belirlenmesidir.  
Genlerdeki bilgiler kullanılarak  
sentezlenecek ve vücudumuzdaki  
proteinlerin işlevlerini nasıl yaptıklarını

anlamak mümkün olacaktır. Bu bilgiler  
proteinlerin hücre içinde birbirleriyle, DNA  
ile, diğer küçük molekül ve ilaçlarla  
etkileşimlerini analiz ve kontrol etmemize  
yardımcı olacaktır. Proteinlerin bu  
marifetlerini anlamak için proteinlerin  
amino asit dizileri ile üç boyutlu yapıları,  
yapıları ile işlevleri arasındaki bağlantıları  
bulmak günümüzde yapılan araştırmaların  
önemli bir kısmını oluşturmaktadır. Şu  
anda farklı organizmalara ait yüzbinlerce  
proteinin amino asit dizisi ve 20,000  
proteinin üç boyutlu yapısı bilinmektedir.  
Bu yapılar protein veri bankalarında  
toplanmaktadır ([www.rcsb.pdb.org](http://www.rcsb.pdb.org)). Bu  
bilgiler birçok biyolojik olayların daha iyi  
anlaşılmasını sağlayacak; dolayısı ile  
genetik hastalıkların belirlenmesine,  
teşhisine ve tedavisine faydalı olacaktır.

Bütün bu bilgilerin süratli bir şekilde  
ortaya çıkması, istatistiksel kuramlar ve  
bilgisayar yöntemleri ile düzenlenebilecek  
milyarlarca nükleotit ve amino asit dizisi  
ve bu dizilerden anlamlı bilgi üretilmesi  
gerekmesini ortaya çıkarmıştır. Bu  
dizilerden anlamlı bilgi çıkarabilmek için  
her bir amino asit dizisinin hangi moleküler  
yapıya sahip olduğu, ayrıca her bir moleküler  
yapının hangi dinamik özelliklere sahip  
olduğu, dolayısıyla fonksiyonun ne  
olduğunun araştırılması gerekmektedir.  
Bütün bu çalışmaları gerçekleştirebilmek  
için de yeni nesil biyologlara ihtiyaç  
duyulmaktadır. Bu bilim insanları biyolojinin  
yanı sıra bilgisayar ve mühendislik  
bilimlerinin bilgilerini ve kurallarını da  
bilmek durumundadırlar. Bu ihtiyacı  
karşılama için dünyanın çeşitli ülkelerinde  
Biyoinformatik bölümleri ve enstitüleri  
kurulmaya başlanmıştır. En önde gelen  
merkezler yandaki tabloda verilmiştir.

## DNA Yaşam molekülü

**Trilyonlarca hücre**  
Her hücrede:

- 46 kromozom
- 2 m. DNA sarmalı
- 3 milyar DNA alt ünitesi (A, T, C, G)
- 30.000 gen kodu

Bunlar insanın hayati fonksiyonlarını yerine getiren proteinlerdir.

## ÖNEMLİ BİYİNFORMATİK MERKEZLERİ

National Center for Biotechnology Information	<a href="http://www.ncbi.nlm.nih.gov">http://www.ncbi.nlm.nih.gov</a>
The Genome Sequencing Center (GSC)	<a href="http://genome.wustl.edu/gsc/index.shtml">http://genome.wustl.edu/gsc/index.shtml</a>
The Sanger Center	<a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>
The Institute for Genomic Research	<a href="http://www.tigr.org">http://www.tigr.org</a>
The Protein Information Resources (PIR)	<a href="http://pir.georgetown.edu">http://pir.georgetown.edu</a>
European Bioinformatics Institute (EBI)	<a href="http://www.ebi.ac.uk">http://www.ebi.ac.uk</a>

### KÜ Bilişimsel Biyoloji ve Biyoinformatik Merkezi

Koç Üniversitesi'nde de Bilişimsel Biyoloji ve Biyoinformatik Merkezi (Center of Computational Biology and Bioinformatics) adı altında yeni bir merkez kurulmuştur (<http://ccbb.ku.edu.tr>). Bu merkez Fen-İnsani Bilimler ve Edebiyat Fakültesi ile Mühendislik Fakültesinin öğretim üyeleri Alper Demir, Burak Erman, Attila Gürsoy, Özlem Keskin, Serdar Taşırın, Metin Türkay ve Deniz Yüret'den oluşmaktadır.

Merkezde yapılan araştırmaların bir kısmı organizmalardaki tüm işlevleri tek başlarına veya diğer proteinlerle etkileşimde bulunarak yerine getirmekten sorumlu olan proteinlerin yapısının belirlenmesine yöneliktir. Proteinlerin işlevleriyle üç boyutlu yapıları arasında doğrudan bir bağlantı vardır; dolayısıyla bir organizmadaki tüm proteinlerin üç boyutlu yapısı belirlendiğinde, o organizmanın biyolojik ve kimyasal aktivitelerin açıklanmasında önemli bir aşama kaydedilmiş olur.

Proteinlerin yapısını belirlemede kullanılan deneysel yöntemlerin çok pahalı olması ve deneylerin uzun sürmesi sebebiyle istatistiksel ve bilişimsel yöntemlerin kullanılması zorunluluk haline gelmiştir. Protein yapısı her biri değişik seviyede detay içeren dört kategoride belirlenmektedir: birincil yapı proteini oluşturan amino asit dizinlerini, ikincil yapı protein yapısının temel taşları olan sarmal ve levhaların yerlerini, üçüncül yapı proteini oluşturan atomların üç boyutlu uzayda koordinatlarını, dördüncül yapı ise birden fazla proteinin birlikte işlev gördüğü durumlarda birbirlerine

göre konumlarını içermektedir. Merkezimizde proteinlerin ikincil ve üçüncül yapılarının tahminine yönelik araştırmalar yapılmaktadır. Ayrıca protein veri bankasındaki bilgileri kullanarak, protein-protein ara yüzeylerinin veri setini oluşturmak ve daha sonra bu veri setini kullanarak ara yüzeylerin istatistiksel olarak incelenmesi de araştırma konularımızın içindedir. Bu veriler proteinlerin birbirleriyle etkileşim mekanizmalarının tayininde çok önemli bir rol olacaktır.

### Veri madenciliğinin biyoinformatikte kullanılması

Bilgisayar bilimlerinde son zamanlarda gelişmeye başlayan alanlardan biri olan veri madenciliğinin biyoinformatikte kullanılması gün geçtikçe artmaktadır. Biyolojik veriler hem içerik olarak çok farklılık göstermekte (gen, protein dizileri, evrimsel ilişkiler, üç boyutlu yapılar v.b.), hem de farklı ortamlarda ve formatlarda saklanmaktadır (hatta çoğu bilgiler literatürde verilmiş olmasına rağmen bilişimsel yöntemlerde kullanılabilir halde değildir). Son yıllarda bu verileri belli standartlarda, ulaşımı ve sorgulaması kolay, aralarındaki ilintiler doğru şekilde kurulmuş ve bir bütünlük içinde araştırmacılara sunulması konusunda çalışmalar hız kazanmıştır.

DNA mikroçip teknolojisi binlerce genin aynı zamanda deneylerinin yapılmasını sağlar. Bu deneylerden elde edilen veriler gen ekspresyonu, gen fonksiyonlarının kontrolü ve hücrede gelişen olayların anlaşılması için çok önemli veri oluşturur. Bu deneylerde elde edilen verilerin biyoinformatik araçları ile incelenmesi bir

çok hastalığa sebep olan genlerin teşhis edilmesine olanak sağlamaya başlamıştır. Merkezimizde yapılan bir başka çalışmada da 8,799 fare geni incelenmiştir (veriler University of Pittsburgh, Medical School'dan alınmıştır). Fare beyininin 5 farklı bölgesinde bulunan 8,799 genin

ekspresyon seviyeleri ölçülmüş, daha sonra aynı deney belli süre kokaine maruz bırakılmış fare beyin hücrelerinde tekrar edilmiştir. İki data seti istatistiksel yöntemlerle filtre edilip en çok farklılık gösteren genler matematiksel ve istatistiksel yöntemlerle (kendini örgütleyen haritalar ve tekil değer ayrışımı) sınıflandırılmıştır. Bu çalışma ile kokain bağımlılığına sebep olan genler bulunup, bu genlerin sentezledikleri proteinler incelenmektedir.

Bu doğrultuda, Biyoinformatik'te, veri madenciliği, yapay zeka, internete dayalı bilgi alışverişi, ve programlama gibi bilgi teknolojileri ile istatistiksel, matematiksel modelleme ve analiz yöntemlerinin kullanımı artmaktadır. Bütün bu gelişmelerin doğal bir sonucu olarak disiplinlerarası ortak çalışmalar önem kazanmaktadır.

**Dünyanın çeşitli ülkelerinde Biyoinformatik bölümleri ve enstitüleri kuruluyor. Biyoinformatik ve Bilişimsel Biyoloji alanında yeni adımlar atmaya hazırlanan Koç Üniversitesi'nde çalışmalar bütün hızıyla sürüyor...**

