

Genetik Kod Nasıl Çözüldü? Tarihsel Bir Bakış

How Was the Genetic Code Deciphered? A Historical Look: Scientific Letter

Dr. Murat KASAP^a

^aTıbbi Biyoloji AD,
Kocaeli Üniversitesi Tıp Fakültesi,
Kocaeli

Geliş Tarihi/Received: 03.11.2007
Kabul Tarihi/Accepted: 05.02.2008

Yazışma Adresi/Correspondence:
Dr. Murat KASAP
Kocaeli Üniversitesi Tıp Fakültesi,
Tıbbi Biyoloji AD, Kocaeli
TÜRKİYE/TURKEY
mkasap@vt.edu

ÖZET Bu makale, genetik kodun çözülmesinde meydana gelen gelişmeleri tarihsel bir bakış açısı ile incelemektedir. Makalede özellikle iki fizik kökenli bilim adamı George Gamow ve Francis Crick'in düşüncelerine ağırlık verilmiştir. Genetik kod hakkında önceden ne biliniyordu ve bizler ne biliyoruz soruları ışığı altında kronolojik bir sıra takip edilmiştir. Kodun kırılmasında rol alan olaylar 2 ana kısımda incelenmiştir. Birinci kısımda geliştirilen hipotezler, ikinci kısım da ise deneysel yaklaşımlar üzerinde durulmuştur. Hipotezlerden ilki George Gamow'un ortaya koyduğu elmas kod hipotezi ikincisi ise Francis Crick'in ortaya koyduğu virgülden arınmış kod hipotezidir. Deneysel yaklaşımlar içerisinde Marshall Nirenberg'in ortaya koyduğu yaklaşımdan bahsedilmiş ve genetik kodun açığa çıkarılmasında umulanın aksine kodun dejenere olduğu yani birden fazla kodonun aynı amino asidi kodladığı gösterilmiştir. Bu makalede ayrıca model üzerinden hipotez kurularak varılan sonuçların deneysel yollarla elde edilen sonuçlar kadar önemli olabileceği, yanlış olsa bile bu hipotezlerin doğrunun üretilmesi için kaynak teşkil edeceği vurgulanmak istenmiştir. İlginçtir ki doğa bazen bizim olması gerektiğini düşündüğümüz kanunları görmezlikten gelebilmektedir.

Anahtar Kelimeler: Genetik kod; DNA

ABSTRACT This is an article describing the events that led to the breaking of the genetic code. Emphasis was given to the ideas developed by two physicists, George Gamow and Francis Crick. The article took a chronological view by asking the questions "What did they know then?" and "What do we know today?". The code was deciphered in two stages. Stage I included the hypothesis about the genetic code based on the models. Stage II included the experimental aspects and the role of biochemical reasoning in deciphering the code. Unexpectedly, the code was degenerated and many codons stood for the same amino acid. The article also highlights the fact that hypothesis driven by logical thinking and supported by physical modeling can sometimes be as important as experimental data although they may not truly points out the facts. The doors opened by such model-based hypotheses ultimately lead to finding the facts in science. Amazingly, nature sometimes overrules any rules that we think should be there.

Key Words: Genetic code; DNA

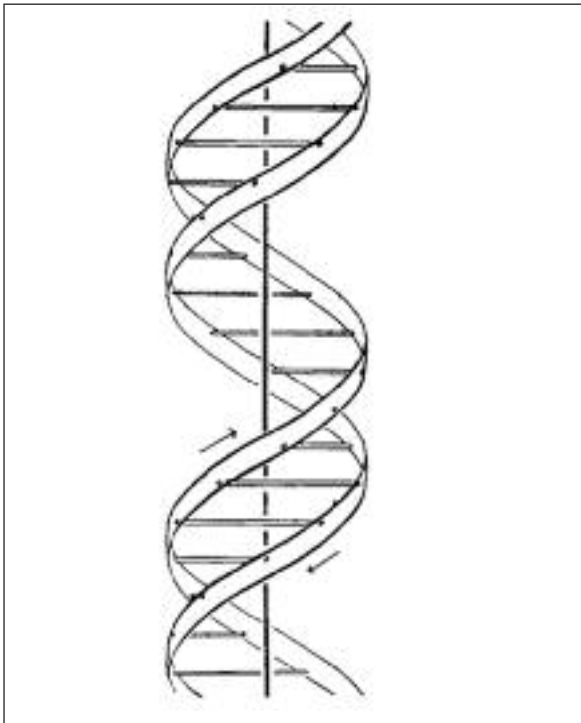
Türkiye Klinikleri J Med Sci 2008;28:721-726

8 Şubat 1953'te Francis Crick ve James Watson İngiltere'de bir barda etraflarındaki bir grup insana hayatın sırrını keşfettiklerini anlatmaya çalışıyorlardı. Aynı yılın 25 Nisan'ında yayınladıkları "A Structure for Deoxyribose Nucleic Acid" adlı bilimsel makale ile bu keşfi tüm dünyaya duyurdular.¹ Bugün hemen herkesin az çok bildiği ama o gün için oldukça yeni olan bu bilgiler DNA'nın çift sarmal yapıda olduğunu ve en-

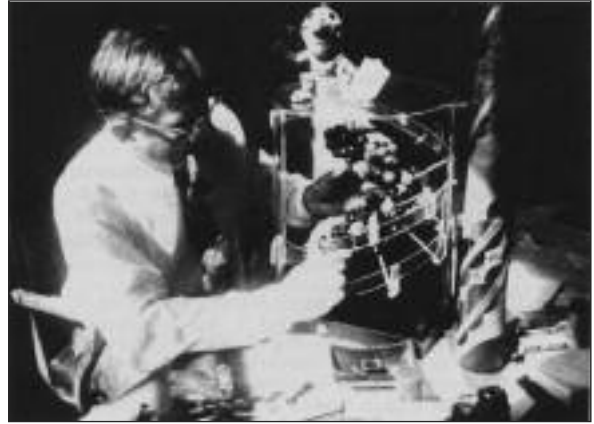
mel bileşenlerinin 5 karbonlu şeker, bir fosfat grubu ve 4 baz (adenin, guanin, sitozin ve timin) olduğunu göstermiştir (Şekil 1). Ancak DNA'nın yapısı açıklanmış olsa dahi, bu yapının canlıları oluşturan bilgiyi nasıl kodlandığına dair hiçbir ipucu yoktu. Başka bir deyişle yaşamı kodlayan alfabe bulunmuş, henüz bu alfabe ile yazılmış kitap okunamamıştı. Aynı yıllarda proteinlerin birincil yapıları (amino asit dizileri) ile ilgili bilgiler daha yeni elde edilmeye başlanmış, Frederick Sanger insülin proteinini oluşturan amino asit dizisini yeni açığa çıkarmıştı.² Bu bilgi doğada 20 adet amino asidin proteinlerin birincil yapısında rol aldığını gösteriyordu. Geriye kalan soru ise hayatın kitabı olan DNA'nın 20 amino asitten oluşmuş proteinleri nasıl kodladığını bulmaktı. Bunun yanında bilim adamları sorulan diğer bazı sorulara da cevap aramak zorundaydılar. Bunlar arasında ön plana çıkanlar:

1) DNA çift sarmal ise iplikçiklerden hangisinde bilgi saklanmakta?

2) Eğer iplikçiklerden sadece birisi mesajı taşıyorsa bu hangi iplikçik olmalı?



ŞEKİL 1: Watson ve Crick'in 1953 yılında yayınladıkları makalelerinde gösterdikleri DNA molekülünün modellenmiş yapısı (Kaynak 1'den alınmıştır).



ŞEKİL 2: George Gamow elmas kodu adını verdiği modeli üzerinde çalışırken.

3) Bilgi hangi yönde okunmalı?

4) DNA'yı oluşturan bazlar nasıl ve ne şekilde sıralanırsa bir anlam ifade etmeli?

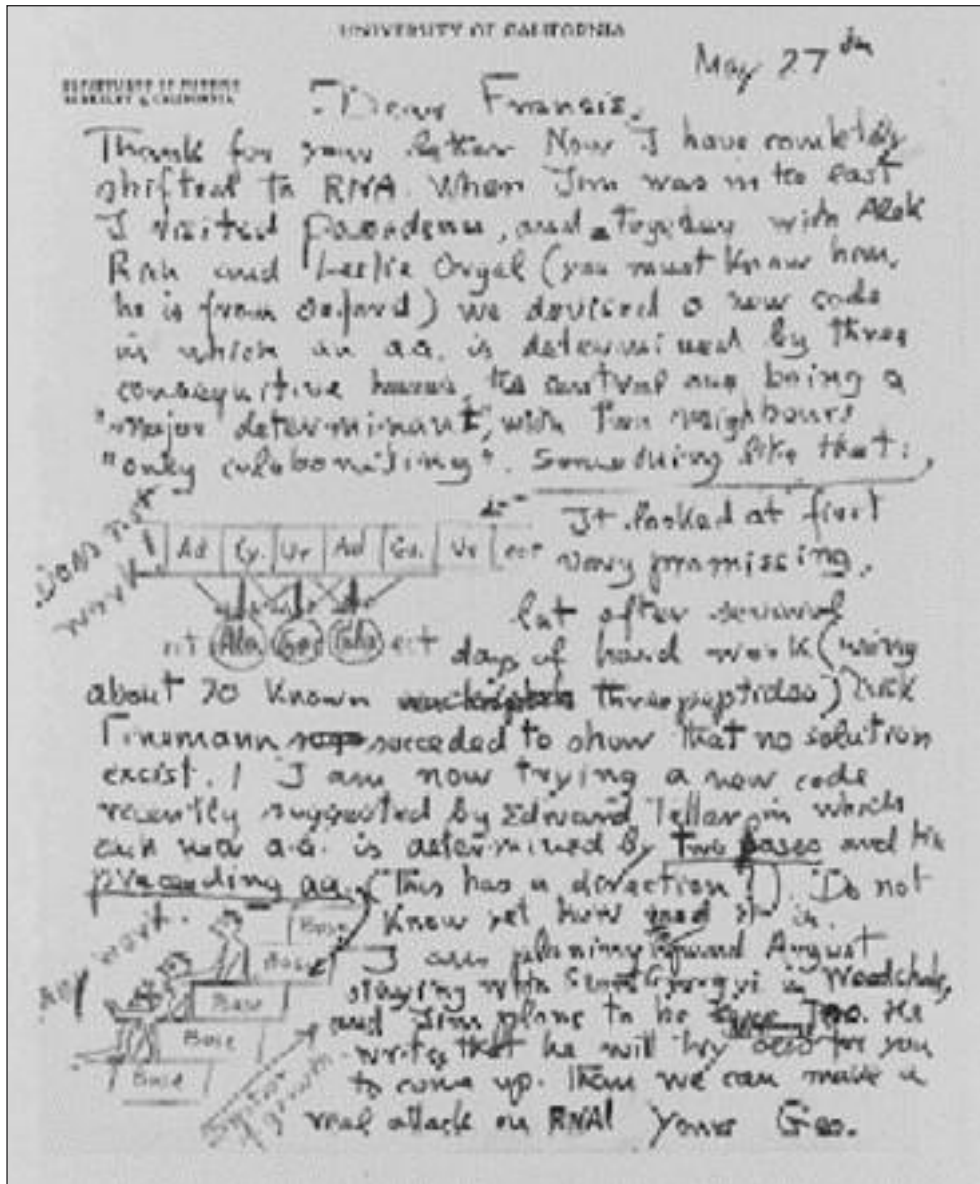
soruları idi.

DNA'nın yapısını açıklayan yayını takip eden ikinci bir yayın ile Watson ve Crick DNA'nın genetik bilgi akışındaki (DNA'nın replikasyonunda) rolünden bahsetmekteydiler.³ Bu yayın o yıllarda bir fizikçi olan ve biyolojik bilimlere de ilgi duyan George Gamow tarafından fark edildi (Şekil 2). Rus asıllı olan Gamow aslında kuantum mekaniğine yaptığı katkılarla bilinen ve özellikle alfa parçacıklarını açıklayıcı teorileri ile tanınan çok başarılı bir bilim insanıydı.⁴ Gamow 1953 ve 1954 yılları arasında Watson ve Crick'e genetik kod hakkındaki düşüncelerini içeren bir dizi mektup yazdı (Bu mektupların bazılarını <http://profiles.nlm.nih.gov/SC/Views/Exhibit/documents/deciphering.html> web adresinde bulmak mümkündür) (Şekil 3). Yazdığı bu mektupların ilkinde Gamow "...Eğer sizin bahsettiğiniz bakış açısı doğru ise (ki eminim en azından detaylar kaçmış olsa bile ana bilgi doğru) her bir organizma uzun uzadıya yazılmış quadrucal (?) sistemlerle karakterize edilmeli...". Gamow'un yazdığı bu ve benzeri mektuplar biyoloji tarihinde daha sonra "genetik kod" adı verilecek olan (Gamow'un quadrucal (?) sistemleri) bir dizi kuralın oluşmasında temel teşkil etmiş ve DNA üzerinde taşınan genetik bilgiyi proteinlerin yapısını oluşturan amino asitler ile ilişkilendirmiştir.

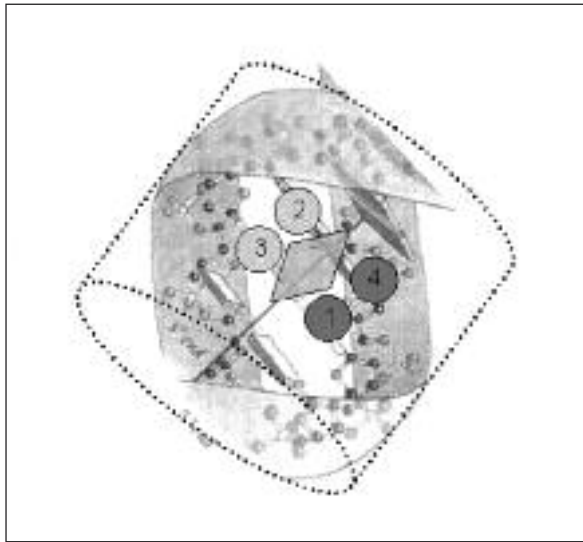
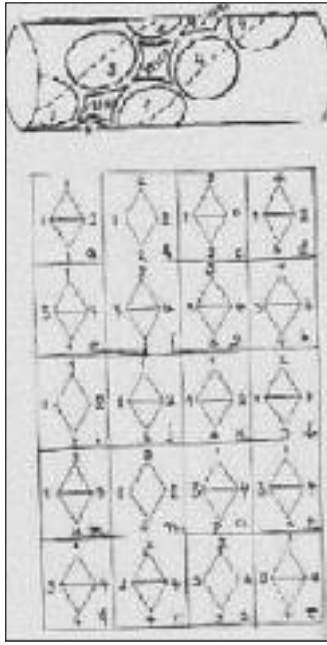
Crick, Gamow ile ilgili daha sonra yazdığı bir yazısında Gamow'un modellerinin önemini, modellerin genetik kodun özeti şeklinde olmaları ve gereksiz kimyasal detaylardan uzak kalmaları olarak özetlemiştir.⁵ Gamow'un modelleri gerçekten de bir öz formatında olup bilim dünyasındaki etkisini uzun süre muhafaza etmiştir.

Proteinleri toplam 20 çeşit amino asidin oluştu-
rması genetik kodun 2 harfli bir kodlama sistemi
ile çalışma ihtimalini ortadan kaldırmaktadır. Basit
bir mantıkla düşünüldüğünde de tek harfli bir kod-

lama maksimum 4 amino asidi ve 2 harfli bir kod-
lama maksimum 16 amino asidi kodlayabilmekte-
dir. O halde kod en az üç harfli olmalıdır. Gamow
elmas kodu adını verdiği hipotezini bu basit bilgi-
den yola çıkarak ileri sürmüştür (Şekil 4). Ancak
kodun üç harfli olması durumunda 4 baz toplamda
64 adet kod üretir ki bu da gerekli olan 20 adet ami-
no asit ihtiyacının 3.2 katıdır. Üçlü kodlamanın
üreteceği bu bolluk uzun süre kafaları karıştırmış
olsa da Gamow, modelinde bu olayı açıklamaya ça-
lışmıştır. Elmas kodu hipotezinde Gamow amino



ŞEKİL 3: George Gamow tarafından Francis Crick'e yazılmış modelini anlatan mektuplardan biri. (Kaynak: <http://profiles.nlm.nih.gov/SC/Views/Exhibit/documents/deciphering.html>)



ŞEKİL 4: George Gamow'un kendi çizimiyle gösterdiği elmas kodlar ve bu kodların DNA üzerine uyarlanmış hali.

asitlerden proteinlerin oluşmasında çift sarmal DNA'nın bir şablon görevi gördüğünü ileri sürmüştür. Çift sarmal DNA yapısı üzerinde bulunan ve sarmalın doğal dönme hareketinden oluşan oluklardan her biri amino asitlerin yan zincirlerinin sığabileceği belirgin yapıda açıklıklar oluşur. Her bir açıklık, kendine has bir amino asidi fiziksel olarak içine alacak biçimde şekillenir. Yan yana açıklıklar içerisine yerleşen amino asitler en-

zimler aracılığı ile polimerize edilerek protein zincirleri haline dönüşürler. Gamow'un ileri sürdüğü açıklıklardan her biri, elmas şeklinde boşluklardan oluşmuş olup bazlar, elmasın dört köşesine gelecek şekilde yer alırlar. Elmasların dört köşesi olmasına ve her bir köşeye bir baz düşmesine rağmen, DNA içerisinde yatay ve çapraz olarak yerleşmiş bulunan bazlardan ikisi Watson-Crick eşleşmesine göre belirleneceği için bazlardan sadece üçü kod olarak kullanılır. Örneğin Şekil 4'te gösterilen DNA elması tarafından belirtilen kodun 324(1) olarak okunması gerekmektedir. Bu şekilde oluşturulan bir kodlama ile toplamda 64 adet kodun olduğu ortaya çıkmaktadır. Gamow, yaptığı modellerde amino asitlerin çoğunun yan zincirlerinin simetrik olduğunu ve bu yüzden elmasın köşelerinin yer değiştirilebileceğini ve bu değişimin kodu etkilemeyeceğini not etmiştir. Örneğin Gamow'un modeline göre CAG ve GAC aynı bilgiyi kodlamalıdır. Tüm bu simetriler göz önüne alındığında Gamow toplam 20 adet belirgin kodun olduğunu görmüştür. Bu rakam Gamow'un aradığı büyülü rakamın da kendisidir. Ayrıca dikkatli bakıldığında Gamow'un ortaya koyduğu kodun çakışan (overlapping) kodlar olduğu gözlemlenebilir. 131(4) ile 324(1) kodlarındaki 3'lerin aynı 3 olması hali gibi. Moleküler biyoloji dili ile bu şu demektir. GATTACA dizisinde 5 adet üst üste gelen kod bulmak mümkündür GAT ATT TTA TAC ve ACA.

Gamow'un bu hipotezi ilk bakışta zihinlerdeki boşlukları dolduran bir hipotezmiş gibi görünmüştür. Bunun en temel nedenleri arasında DNA'daki bazlar arasında oluşan mesafenin proteinlerdeki amino asitler arasındaki mesafe kadar olması, DNA'nın proteinleri amino asitler ile bire bir eşleşerek kodlayabilmesi, protein uzunluğunun istenilen büyüklükte olabilmesi ve son olarak da çakışan kodların tek bir diziden birçok alternatif protein üretimine izin verebilmesini saymak mümkündür. İlk bakışta çekici gelen çakışan kodlar bilgisi Gamow'un hipotezinin aslında sonunu getirmiştir. Brenner çakışan kodların stokiyometrik nedenlerden dolayı imkansız olduğunu ispatlamıştır.⁶ Örnek vermek gerekirse, baz dizisi 12345 olan bir DNA molekülünde birbirini takip eden amino asitler 123, 234 ve 345 bazlarınca kodlanacaktır. O

halde ilk amino asit 123 kodunca belirlendikten sonra ikinci amino asit 23 ile başlayan bir kod ile belirlenmek zorundadır. O tarihte hali hazırda elde bulunan protein dizilerinin incelenmesi sonucu, proteinlerin bu kurala uymadığı gösterilmiştir. İstatistiksel açıdan incelendiğinde de mesela bir dipeptid için 20^2 adet (400 tane) muhtemel peptid oluşturulabilirken 4^4 adet (256 tane) çakışan kod elde edilebilmektedir. Bu hesaba göre 144 adet dipeptidin proteinlerde görülmemesi gerekmektedir. Proteinler Gamow'a ve koyduğu kurala uymak istememiş olmaları ki bu 144 adet dipeptid formuna proteinlerde hem o gün hem de bugün rastlanmaktadır. Gamow'un teorisine, ilk başlarda sorunu çözücü bir yaklaşım olmasından dolayı ilgi ile bakılsa da zamanla teorideki eksikliklerin ortaya konması ile ciddi boyutta eleştiriler almıştır.

1950'li yılların sonlarına doğru mRNA'yı destekleyen fikirler ortaya çıkmaya başlamıştır. Bunun temel nedenleri arasında protein sentezinin sitoplazmada gerçekleşmesi ve DNA'nın hücrenin çekirdeğinde bulunması yatıyordu. Bilim adamları ayrıca DNA üzerindeki tek bir genin hücre için yeterli proteini yapamayacağına ve bu nedenle hücrenin ihtiyacına cevap verebilecek sayıda DNA kopyasının yapılmasının mantıklı olacağına kanaat getirmişlerdi. Aynı yıllarda Francis Crick, adaptör hipotezi adını verdiği bir hipotez geliştirdi.⁷ Hipotezinde Crick amino asitlerin mRNA ile doğrudan değil mRNA üzerindeki kodları tanıyan aracı moleküller vasıtası ile etkileştiklerini söylüyordu. Bugün bu adaptör moleküllerin tRNA molekülleri oldukları ortaya konulmuştur. Crick hipotezinde tRNA'ların mRNA üzerindeki üçlü kodlar üzerine oturduklarını ve tRNA'ların taşıdığı uygun amino asitlerin daha sonra polimerize edilerek proteinlere çevrildiğini ileri sürdü. Bu hipotez Gamow'un çakışan kodlar problemini ortadan kaldırmış ve kodların çakışmayan üçlü gruplar (kodonlar) halinde olduğunu varsaymıştır. Ancak bu hipotezdeki en önemli problem RNA molekülü üzerindeki üçlü kodonları birbirinden ayıran herhangi bir boşluk ya da virgülün olmayışydı. Başka bir deyişle mRNA kodlarının nereden okunmaya başlayacağını bilinmeyişydı. Bu nedendir ki Crick'in bu hipotezine daha sonra "virgülden arın-

mış kodlar (comma-free codes)" adı verilmiştir. Örnekleme gerekirse, mesela mRNA üzerinde bulunan birUGUCGUAAG..... dizisi UGU, CGU, AAG şeklinde okunabileceği gibiU, GUC, GUA, AG.... şeklinde de okunabilir. Crick bu konuya açıklık getirebilmek için hücredeki adaptör moleküllerin sadece 64 kodonu tanıyacak sayıda olduğunu ileri sürmüştür. Örneğin ...CGUAAGT dizisinde eğer CGU ve AAG manalı kodonlarsa ve bunlara karşı adaptör moleküller varsa, GUA ve AGT manasız kodonlar olmak zorunda ve bu kodonların hücrede adaptör molekülleri bulunmamak durumundadır. Buna ilaveten AAA, CCC, GGG ve UUU gibi kodonlara karşı da hücrede doğal olarak adaptör moleküller bulunmamalıdır. Aksi takdirde TAAAAAGTC dizisine benzer bir DNA dizisinde TAA ve AAA manalı kodonlar ise AAA ve AAG'de manalı kodonlar olmak zorundadır. Crick siklik permutasyon kullanarak tekrar eden 4 kodonu çıkardıktan sonra geriye kalan 60 kodonun tamamına 20 amino asidi kodladığını göstermiştir. Örneğin AAC, ACA ve CAA kodonlarından sadece biri amino asidi kodlamalıdır. Aynı şekilde ACG, CGA ve GAC kodonlarından da sadece biri amino asit kodlamalıdır. Bu şekilde yaptığı hesaplarla Crick de, Gamow gibi mucize sayı olan 20'yi yakalamış ve hipotezinin doğruluğuna kendi dahil birçok bilim adamını inandırmıştır. Yaklaşık 5 yıl boyunca Crick'in bu hipotezini destekler ve daha da formüle hale getirir nitelikte bilimsel çalışmalar yapılmıştır.

1961 yılında tüm bu hipotezleri yıkan Marshall Nirenberg'in ünlü çalışması ile genetik kod bilim dünyasının bir kere daha ilgi odağı olmuştur.⁸ Marshall Nirenberg 1957 yılında önce doktora sonrası araştırmacı olarak gittiği ve daha sonra uzman olarak çalışmaya başladığı Amerikan Ulusal Sağlık Enstitüsünde (NIH) yaptığı deneylerle genetik kodun ne Gamow'un tarif ettiği ne de Crick'in ileri sürdüğü kurallara uymadığını göstermiştir. Hatta ve hatta, yaptığı ilk deneylerde Crick'in söyleminin aksine poliurasil'in (UUUUUUUUUUU) fenil alanin amino asidini kodladığını ispatlamıştır. Beş yıllık bir süreç sonunda Nirenberg 20 kişilik bir grup ile tüm genetik şifreyi çözmekle kalmamış aynı zamanda genetik şifrenin türler arasında benzer

olduğunu da göstermiştir. 2004 yılında yayınladığı kişisel anılarını anlatan makalesinde Nirenberg, genetik kodu çözmeye dinlediği bir öğle arası semineri sayesinde başladığını anlatmakta ve 5 yıllık süreç içerisinde diğer bilim adamları ile yaptığı ortak çalışmaların öneminden bahsetmektedir.⁹

Kodonlar ve onlara atfedilen amino asit tabloları ortaya çıktıktan sonra 20 ve 64 rakamlarının aslında hiçbir büyüünün olmadığı da anlaşıldı. Ekstra olan kodonların sadece bolluk ifade ettiği, bazı amino asitlerin 4, bazılarının ise 6 kodonla kodlandığı görüldü. Hayes'in deyimiyile "doğa okuma çerçevesi problemini çözmek için tüm matematiksel becerisini göz ardı etmiş ve çözümü tabiri doğrusu üstünkörü bir yolda bulmuştur".¹⁰ Ribozomların başlayıp duracağı kodonlar hariç genetik kodda aslında hiçbir gizem yoktur. Gizemin daha sonra ribozomların protein sentezine nereden başlayacağını belirleyen ribozom bağlama bölgelerinde olduğu anlaşılmıştır.

Aslında ilk bakışta Gamow ve Crick'in tasarladığı kodlar bize daha zekice ve ustaca gelmekte-

dir. Çünkü bu kodlarda mutasyona ve hataya yer yoktur. Genetik kod üzerinde oluşacak herhangi bir mutasyon Gamow ve Crick'in kodlarında ölümcül olmalıdır. Doğada ise durum tam tersinedir. Genetik koddaki bolluk (redundancy), kodun hatalara tolerans göstermesini sağlamak içindir. Mutasyonların birçoğu sinonim kodonların birbirlerine dönüştüğünü göstermektedir. Herhangi bir amino asit değiştiğinde ise değişen amino asidin özellikleri, değiştiği amino asidin özellikleriyle çok benzer olmaktadır. Haih ve Hurst'un yapmış olduğu bilgisayar simülasyonları genetik kodun neredeyse en uygun olduğunu destekler niteliktedir.¹¹ Gamow ve Crick'in genetik kodun belirlenmesinde aldıkları roller ve ürettikleri hipotezler onları bilim tarihinde hak ettikleri konuma getirmiştir. 1950'li yıllarda ortaya konulan bu hipotezler bugün bizlerin biyoçip teknolojilerine kadar uzanan bilgi birikimine sahip olmamıza ön ayak olmuştur.¹² Günümüzde dahi bazı bilim adamları bu hipotezleri yanlışlıkları ispat edilmiş en dâhice hipotezler olarak görmektedirler.

KAYNAKLAR

1. Watson JD, Crick FH. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 1953;171:737-8.
2. Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. Biochem J 1951;49:463-81.
3. Watson Jd, Crick Fh. Genetical implications of the structure of deoxyribonucleic acid. Nature 1953;171:964-7.
4. Gamow G. On the formation of protogalaxies in the turbulent primordial gas. Proc Natl Acad Sci U S A 1954;40:480-4.
5. Crick FH. The genetic code--yesterday, today, and tomorrow. Cold Spring Harb Symp Quant Biol. 1966;31:1-9.
6. Brenner S. On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. Proc Natl Acad Sci U S A 1957;43:687-94.
7. Ibbá M, Becker HD, Stathopoulos C, Tumbula DL, Söll D. The adaptor hypothesis revisited. Trends Biochem Sci 2000;25:311-6.
8. Nirenberg Mw, MATTHAEI JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polynucleotides. Proc Natl Acad Sci U S A 1961;47:1588-602.
9. Nirenberg M. Historical review: Deciphering the genetic code--a personal account. Trends Biochem Sci 2004;29:46-54.
10. Hayes B. The invention of the genetic code. American Scientist 1998;86:8-14.
11. Haih D, Hurst LD. A quantitative measure of error minimization in the genetic code. J Mol Evol 1991;33:412-7.
12. Gürán Ş, Yakıcıer, C. [New horizon on molecular biology: Biochip technology]. Türkiye Klinikleri J Med Sci 2003; 416-9.