Proteinlerin Yapı Tahmininde Biyoinformatik Yöntemler

Proteinler, amino asit dizi bilgisini taşıyan mRNA'lardan çevrilir. Proteinler, yaşam için önemli olan çeşitli işlevleri yerine getirmelerine izin veren karmaşık moleküllerdir. Protein dizisinin üç boyutlu (3D) katlanma şekli ve kimyasal grubun özellikleri 3D uzayının düzenlemelerine izin veren benzersiz yapılara yol açar. Kimyasal varlığın bu önemli yerleşimi; proteinlerin önemli yapısal, düzenleyici, katalitik ve taşıma fonksiyonları gerçekleştirmesine izin verir.

Proteinlerin Yapı Taşları

Protein dizisi; doğal olarak bulunan 20 çeşit amino asitten meydana gelir. Her amino asit; bir amino grubuna (NH2), değişken (radikal) R grubuna, bir hidrojen atomuna (H) ve bir karboksil grubu (COOH) bağlı bir merkezi alfa karbon (Cα) içerir. Aminoasitler, Cα' (alfa karbon)a bağlı R grubu tarafından temsil edilen yan zincirin varlığı nedeniyle birbirinden farklıdır. R grubundaki çeşitlilikler, proteinin işlevini yöneten rezidünün (kalıntının) spesifik kimyasal özelliklerini verir. Doğal olarak oluşan bu 20 aminoasidin yanı sıra, protein sentezinden sonra enzimatik modifikasyonlar nedeniyle doğal olmayan amino asitler de nadir durumlarda ortaya çıkabilir. Yan zincir R grubundaki çeşitlilik ve su gibi bir çözücü ile temas etme eğilimi, amino asitleri hidrofobik, polar ve yüklü olmak üzere üç sınıfa ayırır. Aromatik veya alifatik gibi ek alt sınıfları da vardır. Hidrofobik amino asitler; lizin, izolösin, alanin, prolin, valin aromatik amino asitler; fenilalanin ve triptofan sülfür içeren amino asitler; metiyonin ve sisteindir. Yüklü amino asitler; pozitif yüklü (+) lizin, arginin ve histidin ve negatif yüklü (-) aspartat ve glutamatı içerir. Polar amino asitler arasında asparagin, glutamin, treonin, serin ve prolin bulunur. Glisin, yan zincirinde sadece tek H atomuna sahip bir istisnadır. Protein moleküllerinin çoğu, membran proteinleri ile çözücü ortamıyla temas halindedir ayrıca yüzeydeki su ve polar amino asitler gibi çözücü tarafından erişilemeyen bir hidrofobik yapıya da sahiptir. Bir polipeptit dizisi, iki veya daha fazla aminoasidin peptit bağı oluşturmasıyla sentezlenir. Amino asitler, komşu amino asitlerin NH2 ve COOH grupları arasındaki peptit bağları (alfa peptit bağları) ile birbirine bağlanır.

Protein Yapıları

Birincil (Primer) Yapı

Birincil yapı basitçe, polipeptidin amino asit dizisidir ve polipeptidi kodlayan gendeki kodon dizisi ile belirlenir.

İkincil (Sekonder) Yapı

Bir proteinin ikincil yapısı, polipeptit zincirinin yerel konformasyonu olarak doğrusal amino asitlerin tekrarlayan uzaysal düzenlenmelerine karşılık gelir. Protein yapısında tanımlanan ikincil yapısal elemanlar; alfa (a) sarmal ve beta (β) tabakalarıdır. Alfa-sarmal ilk olarak 1951'de Linus Pauling ve arkadaşları tarafından ortaya konmuştur. Tipik bir alfa sarmal konformasyonu, polipeptit omurgasının düzenli bir sarmal üretmek üzere kıvrılmasıyla oluşur. Bu sarmal yapıda; polipeptit zincirinin omurgası, molekülün ekseni etrafında yan zincir R grupları sarmal omurgadan dışarı doğru çıkıntı yapacak şekilde sarılır. Tek bir dönüşü tamamlamak için gereken rezidü sayısı α-sarmalında 3.6'dır. Polipeptidin omurgası iki yönde sarılabilir (sağ veya sol). Sol sarmalların oluşumunda sterik engel vardır; bu nedenle proteinlerin çoğunda sağ el sarmalı bulunmaktadır. β-tabakalar, zincirlerdeki bitişik polipeptit omurgaları arasındaki H-bağı ile oluşturulur. Polipeptit zincirlerinin bu bölümleri p-tabakaları olarak bilinir. Alfa-sarmalının aksine, H-bağı β tabakalara neredeyse diktir. B-tabaka iki olası konfigürasyondadır; paralel tabaka (aynı N- ile C-terminal yönü) veya karşıt paralel tabaka (N- ile C-terminal yönünün tersi) olabilir. α-sarmallar ve β-tabakalar proteinlerde ikincil yapısal elemanlar olarak kabul edilmelerine rağmen, ilmek (loop) olarak adlandırılan bir bölgeye de sahip olabilir. Bu bölgeler genellikle diğer biyolojik moleküller ile etkileşim için aracı olarak hareket eden proteinlerin yüzeylerinde bulunurlar. Yan zincirlerinde (R grubu) genellikle; prolin, sistein, serin, aspartat, asparagin ve glisin aminoasitleri bulunur.

Üçüncül (Tersiyer) Yapı

Proteinin üçüncül yapısı, bir polipeptit zincirindeki tüm amino asitlerin 3D uzayda oluşturduğu yapıya denir. Üçüncül yapı ilk olarak 1936'da Alfred Mirsky ve Linus Pauling tarafından benzersiz bir konfigürasyon oluşturmak için katlanmış polipeptitten oluşan molekül olarak tanımlanmıştır. Protein yapısındaki amino asitlerin yan zincirlerinin etkileşimlerindeki H-bağlarının rolünü tahmin etmişlerdir. Üçüncül yapının oluşumu, birincil yapıdaki komşu olmayan amino asit rezidülerini birbirine yakınlaştırır ve protein fonksiyonları için belirleyici bir faktör olan katlanmaların oluşturulmasını sağlar. Üçüncül yapı aynı zamanda ikincil yapının; elektrostatik etkileşimler, van der Waals etkileşimleri, hidrofobik temaslar, disülfit, tuz köprüleri ve H bağı ile etkileştiği kompakt bir molekülü oluşturur.

Dördüncül (Kuarterner) Yapı

Proteinin dördüncül yapısı, kovalent olmayan etkileşimler yoluyla protein içindeki iki veya daha fazla bağımsız olarak katlanmış polipeptidin bir birleşiminden oluşur. Çoğu protein bir monomer (tek alt birim) olarak işlev görmez, daha çok çoklu alt birimli (multimerik veya oligomerik proteinler) olarak işlev görür. Bazı protein türlerinde, dördüncül yapı oluşumu fonksiyonel perspektiften çok önemlidir, çünkü tek bir alt birim protein durumunda mümkün olmayan alt birimlerin arayüzleri arasında bağlanma veya katalitik alanların oluşumuna izin

verirler. Enzimlerin, oligomerik proteinlerde ligand / substrat bağlanması nedeniyle meydana gelen konformasyonel değişiklikler nedeniyle sıklıkla ortaya çıkan allosterik regülasyonda (düzenlenmede) rol oynadığı bilinmektedir. Kuaterner yapılarda, aynı alt birimler homomerik bir protein oluştururken farklı alt birimler heteromerik bir protein oluşturur. Kuaterner yapıdaki stabilizasyon kuvvetleri ve etkileşimleri, ikincil ve üçüncül yapı stabilizasyonunda gözlenen tiplerle aynıdır. Alt birim etkileşimlerinde yer alan monomer biriminin yüzey bölgeleri, polar olmayan yan zincir rezidüleri, H-bağı ve disülfit bağlarını oluşturabilen rezidülerden oluşur.

Protein Yapı Tahminleri

Anfinsen katlanmamış proteinlerin, yalnızca birincil yapıları veya amino asit dizileri ile kendi doğal konformasyonlarına geri katlanabileceğini göstermiştir. Böylece birincil sekanstan üçüncül yapıları tahmin etmek için hesaplama yöntemlerinin temelini attı. Protein yapı tahmininin nihai amacı, X-ışını kristalografisi ve NMR (nükleer manyetik rezonans) kullanarak deneysel olarak elde edilen sonuçlarla karşılaştırılabilir doğrulukla bir yapıyı birincil diziden ortaya çıkarmaktır. Termodinamik olarak kararlı katlanmayı diğer konformasyonlardan daha iyi elde etmek için, protein rezidüleri içindeki ve arasındaki etkileşimleri optimize ederek katlanmayı geliştirmeli ve bir peptit bağının atomları arasındaki tüm uzamsal kısıtlamaları karşılamalıdır. Birçok protein NMR için oldukça büyüktür ve X ışını kırınımı için kırınım kalitesi oluşturma eğiliminden yoksundur. Bu nedenle bu gibi durumlarda yapı tahmini için hesaplama yöntemi tek yaklaşımdır. Birincil yapıdan gelen 3D yapı tahmini, yapısal biyoinformatiğin çok tartışılan bir alanıdır. Son zamanlardaki muazzam algoritmaların geliştirilmesine ve protein yapısı tahmini için hesaplama yöntemlerine rağmen, protein katlanmasının doğru tahmini için kapsamlı bir çözüm hala anlaşılması zor olmaya devam etmektedir. Küresel çapta, deneysel olarak birkaç protein yapısı üzerinde çeşitli yapı tahmin araçlarının/yazılımlarının kullanılarak performansını değerlendirmek için bir mücadele oluşturulmuştur. Bu mücadele 1995 yılında, kapsamlı hesaplama amacı için küresel ölçüt sağlayan Protein Yapı Tahmini Tekniklerinin Eleştirel Değerlendirilmesi (CASP) olarak başlatılmıştır. On yıllar süren araştırmalar; kalıp tabanlı yöntemler (homoloji modelleme), katlanma tanıma yöntemi 8(threading tekniği) ve de novo (ab initio) yöntemleri gibi proteinlerin 3D yapılarını doğru bir şekilde tahmin etmenin çeşitli yollarına dair fikirler ortaya konmuştur. Homoloji modelleme yöntemleri, karşılaştırmalı modelleme olarak da bilinir. Sorgulanan yapının tahminini Protein Veri Bankası (PDB)'nda biriktirilen ve deneysel olarak bilinen yapının yakın homologlarına dayanarak yapar. Katlanma tanıma yöntemi (threading yöntemi), genellikle benzer katlanmaya sahip bir yapı mevcut olduğunda ve homoloji modellemesine yakın bir akraba bulunmadığında kullanılır. Ab initio yöntemindeki yapı tahmini sadece amino asit bilgileri kullanılarak sıfırdan gerçekleştirir.

Homoloji Modellemesi

Karşılaştırmalı modelleme olarak da bilinen homoloji modellemesi; bilinen yapılarla (kalıplar), dizi homolojisine dayalı olarak kullanıcı tanımlı dizinin (hedef) protein yapılarını tahmin etmeye çalışır. Bu ilişki ilk olarak 1986 yılında Chothia ve Lesk tarafından belirlenmiştir. Yıllarca evrim geçirmesine rağmen yapının daha kararlı ve daha az duyarlı olduğu sonucuna varmışlardır. Protein sekanslarından birinin bilinen bir yapısı varsa, yapısal bilgiler yüksek derecede güvenilir olmayan proteinle karşılaştırılabilir. Sorgu dizisi ve bilinen yapılar arasındaki özdeşlik yüzdesi "Güvenli" bölgeye düşerse, iki dizinin pratik olarak özdeş bir katlanması olabilir. Genel bir kural olarak, bu "Güvenli" bölge optimal bir dizi hizalamasında en az %30-50 özdeş amino aside sahip olmalıdır ve sonuçta elde edilen homoloji modeli, başka bir uygulama için yeterince kullanılabilir olmalıdır. Homoloji modeli çıktısı, basit bir şekilde polipeptidin hizalanmış bölgelerinin kalıp/homolog yapıdan kopyalanması, gerektiğinde yan zincirlerin değiştirilmesi ve dizi hizalamaları arasında farklılık gösteren rezidülerin mutasyona uğratılması yoluyla üretilir. Homoloji modellemesi ile oluşturulan son model, yapıya dayalı ilaç keşfi ve bölgeye yönelik mutajenez gibi sonraki deneylerin tasarımı için gerekli ve rezidülerin 3D düzenlemesi hakkında yeterli bilgi içerir. Bir homoloji modeli oluşturmak için çok sayıda araç vardır ve bunların çoğu, aşağıdaki gibi yapı oluşumunda yer alan adımlara sahiptir:

a. Kalıp tanımlama
b. İlk dizi hizalaması
c. Hizalama düzeltmesi
d. Omurga üretimi
e. İlmek (loop) modelleme ve optimizasyonu
f. Yan zincir modelleme ve optimizasyonu
g. Genel model optimizasyonu
h. Model doğrulaması

Kısaca işlem; BLAST, PSIBLAST, katlanma tanıma ve PDB (Protein Data Bank) gibi araçlar kullanılarak hedef dizi ve proteinin bilinen yapısı için dizi benzerliği araştırması yapmakla başlar. Benzerlik yüzdesinin genellikle düşük olduğu bir durumda, bir veya daha fazla olası kalıp seçilir ve daha fazla hizalama düzeltmesi gerçekleştirilir. İlk hizalamadan sonra, düşük yüzdeli bölgelerin konuma özgü puanlama matrislerini üretmek için çoklu dizi hizalaması kullanılarak hizalama düzeltmesi gerçekleştirilir. Bu hizalamanın geliştirmesi ve düzeltmesi için genellikle tercih edilen araçlar MUSCLE ve T-COFFEE'dür. Bu düzeltmeler, en kaliteli modeli öngörmek için oldukça önemlidir. Kalıp ve modeldeki hizalanmış kök koordinatlarının basitçe

ilk omurga modeline kopyalandığı başka omurga üretimi gerçekleştirilir. Deneysel olarak belirlenen yapıların zayıf elektron yoğunluk haritası nedeniyle hatalar içerdiği ve manuel inceleme için PDBREPORT (http: // www. Cmbi.ru.nl/gv/pdbreport/) gibi yapı doğrulama aracı kullanabilir. Swiss Model sunucusu (http://swissmodel.expasy.org/), tek bir kalıp seçiminde hizalanmayan eksik rezidüleri telafi etmek ve omurga oluşturmak için birden fazla kalıp kullanır. Swiss Model, Phyre ve MODELLER serbest homoloji modelleme araçları olarak yaygın olarak kullanılmaktadır. MODELLER, tek bir omurga oluşturmayan en yaygın kullanılan homoloji modelleme aracıdır. Bir sonraki adım, daha önce modellenmemiş hizalamadaki ekleme veya silme işleminin ilmek (loop) modellemesidir. Yaygın olarak kullanılan iki yaklaşım bilgiye dayalıdır. İlk yaklaşımda, bilinen benzer ilmekler için ilmekler arasındaki uç noktaları eşleştirerek ve ilmeklerin konformasyonunu kopyalayarak PDB'de bir arama yapılır. İkinci yaklaşımda; rastgele ilmek konformasyonunun örneklenmesi, maksimum enerji ve minimize sarmal döngü konformasyonunu bulmak için Monte Carlo veya moleküler dinamik kullanılır. Ayrıca yapısal olarak benzer proteinler, yan zincir konformasyonları karşılaştırılırken özdeş bir kıvrılma açısına (psi-açı) sahip olduklarından kalıptan korunmuş rezidülerin kopyalandığı yan zincir modellemesi gerçekleştirilir. İlk model oluşturma sonrasındaki model optimizasyonu; yanlış tahmin edilen rotamerlerin, atom pozisyonunu sınırlandırarak ve hataları düzeltmek için moleküler simülasyonlar kullanır ve birkaç adım enerji minimizasyonu uygulayarak optimize edilir. Son adım modelin önceki adımlar sırasında biriken her türlü hata için değerlendirildiği model doğrulama basamağıdır. Model doğrulamasında yöntem, doğru bağ açılarını ve uzunluklarını kontrol eder ancak bu doğrulama modelin doğru katlanıp katlanmadığına karar veremediğinden dikkatli olunmalıdır.

Ab Initio Modelleme Yöntemi

Ab initio yöntemi, 3D yapıyı sadece birincil diziye dayanarak tahmin etmeye çalışan de novo tahmini olarak bilinir. Genel olarak, bu yöntem yapısal homoloğun eksik olduğunda protein yapısı tahmininde son çare olarak kullanılır. Ab initio yöntemi, en uygun enerji konformasyonlarını hesaplayarak doğal konformasyonları tahmin etmeye çalışır. Çok az hesaplama programı bu yönteme güvenmektedir çünkü büyük hesaplama gücü gerektirir. Ayrıca protein katlanma modelleri hakkında mevcut olan sınırlı bilgi nedeniyle de az güvenilir bir yöntemdir. De novo tahmininde yer alan kilit alanlar; doğru enerji, puanlama fonksiyonları ve verimli örnekleme konformasyon alanlarıdır. Zhang laboratuvarının QUARK ve ROSETTA sunucuları, proteinlerin üçüncül yapısını tahmin etmek için diğer yöntemlerle birleştirilmiş ab initio yöntemlerini kullanan en iyi iki tahmin sunucusudur. Ab initio tahmini sürecinde, “tuzak’’ olarak da adlandırılan birkaç aday yapı kümesi hesaplanır. Ayrıca, doğal benzeri

konformasyonlar daha sonra doğal protein katlamasının en düşük entropi ve serbest enerjiye sahip olduğuna dayanarak bu tuzak grupları arasından seçilir.

Katlanma Tanıma Yöntemi

Katlanma tanıma yöntemi, herhangi bir protein modelinin bilinen benzer bir katlanmaya sahip olduğunu incelemek için geliştirilmiştir. Katlanma tanıma yöntemleri, PDB'de biriken yapılar ve sorgu protein dizisi arasındaki ilişkinin istatistiksel bilgisini kullanarak çalışır. Bu yöntemler, hizalamada "alacakaranlık bölgesi" çevresinde olduğunda ve bilinen 3D yapılara homoloji olmadan sorgu proteininin uygun katlanmayı bulmasını mümkün kılmıştır. Katlanma tanıma alanında yaygın olarak kullanılan iki yöntem; 3D profil tabanlı yöntem ve threading tekniğidir. 3D profil tabanlı yöntem; 3D yapısındaki ortama uyması gereken sorgu proteini rezidülerinin fizikokimyasal özelliklerine dayanarak çalışmaktadır. Threading yaklaşımında dizideki rezidüler, 3D uzayda karşılaştırma yapılarak bilinen yapıların omurga koordinatlarına göre yerleştirilir. "Threading" terimi tahmindeki her kalıbın, kalıp yapısına göre sorgu dizisine yerleştirilip hizalanması ve daha sonra hedefin kalıba ne kadar iyi uyduğunun değerlendirilmesiyle yapılır. Threading yaklaşımı temel bir gerçeğe dayanır; az miktarda farklı olduğu bilinen katlanmaların (yaklaşık 1400) olması ve veri tabanında yeni gönderilen yapıların çoğunun PDB'de depolanmış benzer yapısal katlanmalara sahip olmasıdır. Homoloji modellemesi ve katlanma tanıma yöntemi arasındaki temel fark; homoloji modellemesinin, eşleşen kalıbı bir hizalamada dizi olarak işleyerek tahmin etmek için sadece dizi homolojisini kullanmasıdır. Dizi homolojisinin yokluğunda, threading sadece yapısal bilgiye dayanarak bir tahmin yapar. Threading yönteminin önemli sınırlamalarından biri, doğru tahminin ancak bilinen yapıların kütüphanesinde uygun bir katlanma eşleşmesi mevcut olduğunda mümkün olmasıdır.

Threading yaklaşımının temel adımları aşağıdaki gibidir:

a. Yapı kalıbı veri tabanının oluşturulması: PDB, SCOP ve CATH gibi protein yapı veri tabanlarından farklı protein katlanmaları için bir kütüphane oluşturulur.
b. Puanlama işlevinin tasarımı: Sorgu dizisi ile kalıplar arasındaki uygunluk; puanlama işlevleri, dizi ve yapılar arasındaki bilgiye dayalı ilişkilerle ölçülür.
c. Kütüphaneye en uygun yerleştirme: Daha sonra hedef dizi, ilmek (loop) bölgelerindeki eklemeler ve silmeler dikkate alınarak her bir kütüphane katlanmasına uygun şekilde yerleştirilir.
d. Threading hizalama: Sorgu dizisinin her kalıp yapısıyla hizalanması, puanlama işlevinin optimize edilmesiyle oluşturulur. Her olası uyumun skorlama fonksiyonu, ikili etkileşimlerin ve enerji çözümlemesinin esas olarak iki yaklaşımla toplanmasıyla

hesaplanır: biri etkileşimli partnerlerin kalıp proteinden seçildiği donmuş bir ayrıştırma ortamıyla dinamik programlama ve ardından Monte Carlo yöntemi kullanılır.

e. Son threading tahmini: Katlanmalar kütüphanesinden en uygun threading eşleşmesi, toplam enerjinin artan sırasına ve en düşük enerji kıvrımlarına sahip bir kalıp seçilmesine göre yapılır. Nihai yapı modeli, her bir rezidünün omurga atomlarını, seçilen yapısal kalıbın hizalanmış omurga uzamsal koordinatlarına sorgu dizisine yerleştirerek oluşturulur.

REFERANS

Shaik, N. Ahmad, Hakeem, K. Rehman, Banaganapalli, B., & Elango, R. 2019. Essentials of Bioinformatics, Volume I: Understanding Bioinformatics: Genes to Proteins. 1st ed.

Choudhuri, S. (2014). Bioinformatics for Beginners: Genes, genomes, molecular evolution, databases and analytical tools.

Ibrahim, K. S., Gurusubramanian, G., Yadav, R. P., Kumar, N. S., Pandian, S. K., Borah, P., & Mohan, S. (2017). Bioinformatics - A Student's Companion. Springer Singapore.