

Proteomske baze in orodja za študijo posttranslacijskih modifikacij

Uvod

Proteom je proti genomu bistveno bolj kompleksen. To je posledica večih procesov. Prvi teče na ravni RNA in mu rečemo alternativno izrezovanje (splicing). Iz enega gena tako lahko nastane več različnih mRNA in posledično več izooblik proteinov. Drugi pa poteka na ravni proteinov, gre za posttranslacijske modifikacije. Oba procesa prispevata k dejstvu, da je proteinov za tri velikostne razrede več kot genov. Osredotočili se bomo na slednjega.

PTM: poznamo več kot 300 različnih tipov. So biološko in kemijsko dinamični procesi, gre za kovalentno adicijo ali cepitev. Ključni celični kontrolni mehanizmi, odločajo o usodi proteinov, kar se odraža v zvitju, strukturi, aktivnosti in posledično funkciji. Večina proteinov bi bila brez PTM nefunkcionalnih. Reverzibilne-fosforilacija (dodajajo kinaze, odcepljajo fosfataze); Ireverzibilne-proteolitična cepitev signalnih peptidov. Ne potekajo na a.k. ostankih Leu, Ile, Val, Ala, Phe.

Ker so postranslacijske modifikacije pomembni regulatorni mehanizmi, njihove okvare lahko vodijo v celično smrt ali pa v razvoj bolezni. Z okvarami PTM mehanizmov je povezan nastanek raka, kjer gre za napake v modifikacijah jedrnih receptorjev. Tak primer sta abnormalna fosforilacija in acetilacija androgenih receptorjev, ki vodita v nastanek raka prostate. Diabetes tipa 2 se lahko pojavi zaradi odsotnosti PTM na modularnih domenah proteinov vpletenih v metabolno signalizacijo. Alzheimerjeva bolezen se razvije zaradi hiperfosforilacije tau proteinov in njihove agregacije v možganih.

Za boljše razumevanje in lažjo analizo so zbrali PTM v baze in razvili programe in orodja. Z različnimi metodami lahko pridemo do podatkov o PTM, sledi analiza podatkov, nato skladiščenje. Različni strežniki nam omogočajo dostop do podatkov preko spleta. Podatki se uporabijo tudi za učenje algoritmov, ki na podlagi simulacij napovedujejo možne PTM za vhodno sekvenco. Računalniška analiza nam prihrani veliko časa in sredstev. Procesi na shemi tečejo hkrati.

Sledenje PTM:

Gre za relativno majhno spremembo v masi.

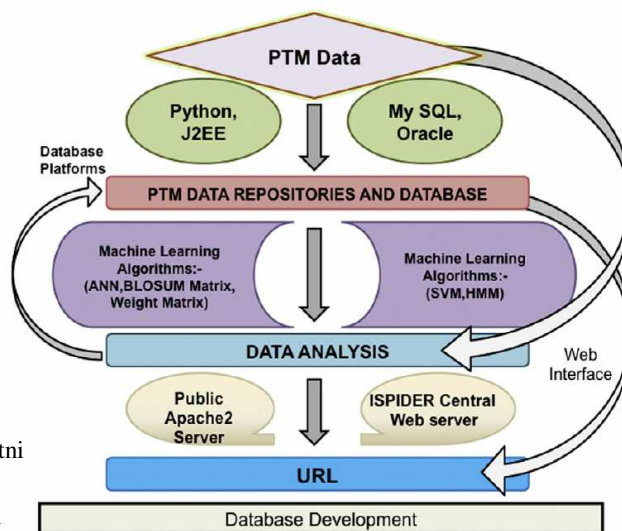
Za študijo PTM-jev so se uveljavile 3 tehnike **2DE**- ločevanje proteinov na podlagi razlik v izoelektrični točki in molekularni masi.

Po ločitvi sledi PTM specifično barvanje.

MS - najbolj uporabljena, v zadnjem desetletju izjemen napredek. Še posebej veliko se uporablja shotgun MS.

Mikromreže - zasledujemo lahko veliko št. proteinov vzporedno. Če pride do reakcije med proteinom in sondo, to opazimo kot fluorescentni signal. Metoda je hitra, avtomatizirana, ekonomična in zelo občutljiva, poraba majhnih količin vzorca. Konkretno se za študijo PTM uporablja samo reverzno fazna proteinska mikromreža.

Kljub velikemu napredku na področju analize, ostaja študija kratkoživih in kemično labilnih PTM velik izziv zaradi omejitev v obsegu in meji detekcije.



Klasifikacija podatkovnih baz

Podatkovne baze post translacijskih modifikacij (PTM) hitro in neprekinjeno rastejo. To je predvsem posledica razvitih tehnologij in tehnik določevanja. Baze vsebujejo podatke od virusov do človeka. Imamo specifične baze, ki se osredotočajo na eno vrsto modifikacije (PhosphoBase) in baze s podatki o več PTM-jev (Swiss-prot). Baze

so nastajale na predvsem na podlagi eksperimentov in nato še s pomočjo verjetnostne analize iz že znanih modifikacij.

Podatkovne baze fosforilacij

Fosforilacija je ena najbolj raziskanih PTM. To je posledica zelo pomembne biološke vloge, med njimi so regulacija in kontrola mehanizmov, celične rasti, diferenciacija, apoptoza, dna popravila itd. Predvidevajo, da je 500000 fosforilacijskih mest v človeškem genomu.

PhosphoBase je bila prva podatkovna baza za fosforilacije. Imela je 156 fosfoproteinov z 398 fosforiliranimi mesti. Potem obstaja tudi Phospho.ELM baza, ki vsebuje 42575 fosforiliranih mest in informacije 310 različnih kinazah. Te informacije vsebujejo sekvenco, povezane kinaze, evolucijske podobnosti, 3D struktura, motivi vezave in molekularno interakcijsko mrežo.

Naslednja znan je Phosida. Ta je izpeljana iz masnih spektrov 2244 proteinov in s tem 6600 fosforiliranih mest. PhosphoSitePlus je ena največjih kolekcij PTM informacij, od in vivo do in vitro, od strunarjev in nestrunarjev itd.

Obstajajo še Phospho3D, PlantsP, PhosphoPep, LymPHOS.

Podatkovne baze glikozilacije

Poznamo štiri vrste glikozilacije: N-, O-glikozilacija, C-manozilacija in glikofosfatidinositolatija. Sodeluje pri prepoznavanju antigenov, razpolovni dobi, proteinsko zvijanje, celica-celica interakcije in proteinska stabilnost. GlycosuiteDB 8.0 je podatkovna baza, ki vsebuje 3238 unikatnih mest glikozilacije in 1851 karakteriziranih. Ta baza se povezuje z več drugimi splošnimi bazami. O-Glycbase in BOLD vsebujeta informacije o O-glikozilacijah. BOLD vsebuje informacije o glikanski strukturi, biološkem viru, reference o glikanu in metode identifikacije in karakterizacije glikanov.

N-glikozilacija je bolj predvidljiva, ker ima značilno prepoznavno meto in sicer Asn-Xaa-Ser/Thr (Xaa ni Pro).

Podatkovne baze ostalih PTM

Obstajajo tudi podatkovne baze za ostale post translacijske modifikacije, in sicer acetilacija, lipidna modifikacija, ubikvinacija (Ubiprot), disulfidne vezi (DSDBASE) itd.

Orodja za določevanje posttranslacijskih modifikacij

Vzrok za razvoj orodij:

Dandanes se v znanosti zelo veliko uporabljajo takoimenovane HT tehnike (High-throughput screening). Gre za prilagojene avtomatizirane teste, ki vključujejo predvsem tandemsko masno spektroskopijo in nam producirajo ogromne količine informacij. Vendar pa je treba te informacije, v tem primeru lokacije PTM še vseeno potrditi, kar pa je eksperimentalno in tudi časovno precej zahtevno, zato znanstveniki poskušajo razviti orodja, ki bi lahko potrdila oziroma z neko gotovo verjetnostjo napovedala lokacije možnih tarč za PTM.

Delovanje:

Orodja za preprosto prileganje sekvenc kot je na primer Blast se ne da uporabiti za napovedovanje mest posttranslacijskih modifikacij, ker je zanesljivost napovedi veliko prenizka. Zato se uporablja boljše algoritme kot so utežne matrike in »machine learning« (tehnike strojnega učenja), ki v bistvu poskušajo posnemati sam encim, ki izvaja posttranslacijske modifikacije.

Proces učenja v osnovi sestavljata dva aspekta:

Učenje programa in nato testiranje. Za učenje je potrebno pripraviti posebno optimizirano bazo podatkov, ki vsebuje zelo značilna mesta kjer se in kjer se ne pojavljajo posttranslacijske modifikacije. Ko je učenje končano se program testira na neodvisni bazi podatkov, da se preveri v kakšni meri pravilno določi posttranslacijske

modifikacije, nato pa je potrebna še optimizacija in sicer tako, da se variera razmerje med občutljivostjo in specifičnostjo programa in se določi neko optimalno razmerje.

Učne funkcije umetnointeligentnih programov:

Nevronske mreže

Metoda podpornih vektorjev

Skriti model Markova

Eden prvih programov ki so jih razvili, je bil za napovedovanje mest fosforilacij. (NetPhosk) Zanesljivost napovedi je bila precej nižja od napovedi sodobnih programov, ker je mesta napovedoval samo na podlagi sekvenčnih motivov. Nato pa so delovanje programov izboljšali z dodatnimi informacijami kot so evolucijsko konzervirana mesta, podatki o sekundarnih in terciarnih strukturah proteinov, hidrofobnosti, dostopnosti mest encimu in informacijami specifičnimi za vsak posamezen encim.