Déjà Vu - Réjà Vu : on knowledge-based approaches linking ligand and target information to bioactivity

Westen, G.J.P. van

Samenvatting

De hoofdvraag in dit proefschrift was of het combineren van data uit meerdere disciplines (hier chemie, biologie en bioactiviteit) synergistische effecten zou laten zien boven methoden die slechts data uit één discipline gebruiken. Deze vraag probeerden wij te beantwoorden door middel van een aantal preklinische studies en een studie op een klinische data set.

In **hoofdstuk 1** werden een aantal gebruikelijke concepten uit de wereld van Computational Chemistry geintrodcuceerd (waaronder: chemical space, chemical similarity, target space en target similarity). Tevens werd in dit hoofdstuk kort ingegaan op de manier waarop de '-ínformatics' methoden het chemisch en biologisch onderzoek blijvend hebben veranderd. Daarnaast werd een korte introductie gegeven in Röntgen diffractie kristallografie. Het hoofdstuk werd afgesloten met een opsomming van een aantal tekortkomingen van de huidige methoden waarbij tevens uiteen werd gezet waarom er een behoefte bestaat aan nieuwe methoden zoals proteochemometrics.

Aansluitend werd in **hoofdstuk 2** een review gegeven van proteochemometrics. Het concept werd uitgelegd en het review bevat een nagenoeg compleet overzicht van al het werk uit het veld tot 2010 in een tabel. Tevens bevat deze tabel een overzicht van de data sets waarop PCM werd toegepast, de gebruikte descriptoren en machine learning technieken. Het hoofdstuk werd afgesloten met een opsomming van mogelijke valkuilen en daarnaast nieuwe mogelijkheden en toepassingsgebieden.

In **hoofdstuk 3** werden een 5-tal nieuwe eiwit descriptoren geintroduceerd. Deze en andere eerder gepubliceerde descriptoren (totaal 13) werden aan verschillende analyses onderworpen. Het is in PCM gebruikelijk de overeenkomsten tussen aminozuren te quantificeren door middel van fysisch-chemische eigenschappen. Als gevolg daarvan zullen in een overzicht de aromatische aminzuren een cluster vormen evenals de geladen aminozuren enzovoorts. Om een dergelijk gesimplificeerd beeld te krijgen is het echter noodzakelijk de data te comprimeren. Dientengevolge is de uiteindelijke matrix slechts een benadering van de werkelijkheid en ligt het voor de hand uit te zoeken welke descriptor uiteindelijk het beste werkt in combinatie met PCM.

Wij concludeerden echter dat de descriptoren gemiddeld nagenoeg gelijk presteren desalniettemin kunnen er grote verschillen per eiwit ontstaan. Hierom is het aan te raden voor elk experiment de verschillende descriptoren te testen. Wij observeerden echter ook dat het includeren van meer informatie in de datacompressie niet noodzakelijkerwijs tot betere prestaties leidt.

Het is waarschijnlijk dat dit een direct gevolg is van het feit dat datacompressie methoden het merendeel van de variatie in de eerste factoren verklaren en de hierop volgende factoren steeds minder verklaren.

Hoofdstuk 4 bevatte de eerste experimentele toepassing van PCM, in dit hoofdstuk voerden wij een preklinische studie uit om nieuwe liganden voor de humane adenosine receptoren te identificeren. Om dit doel te bereiken combineerden wij de historische data beschikbaar voor liganden die op humane receptoren getest waren met de data over liganden die op rat receptoren getest waren. Deze combinatie leverde een grotere variatie op in de chemische structuren waartoe ons model toegang had. Onze hypothese was dan ook dat wij met dit model nieuwe liganden konden identificeren in plaats van liganden die minimaal van de bestaande verschillen.

Uiteindelijk kochten wij 54 liganden en valideerden de voorspellingen van ons model experimenteel. Wij identificeerden 6 nieuwe liganden (11 % hit rate) waarvan één een affiniteit had van 7 nM voor de humane A1 receptor. Wij concludeerden dat ons model beter presteert dan de huidige methoden gezien wij een hoge hitrate hadden en nieuwe liganden konden identificeren.

De tweede preklinische studie werd beschreven in **hoofdstuk 5.** Hier werd echter gericht op een latere fase in het medicijn onderzoek (lead optimization). Onze modellen werden getraind op een nagenoeg complete dataset (voor 64 % van de mogelijke ligand – eiwit paren hadden wij een acitiviteitswaarde (pEC_{50})). Door middel van het model konden wij de missende 36 % invullen met een nauwkeirigheid die vergelijkbaar was met die van de experimentele assay.

Vanwege deze hoge nauwkeurigheid konden wij bioactiviteitsspectra voorspellen (waarmee de activiteitsverschillen van liganden op virale mutanten voorspeld kon worden). De belangrijkste ontdekking van dit hoofdstuk is dan ook dat op deze manier het optimale kandidaat medicijn gekozen kan worden zonder dat alle 6,314 experimenten gedaan hoeven te worden. Tot slot voegden we voor dit model een betrouwbaarheidswaarde toe gebaseerd op de aminozuur overeenkomst tussen de mutanten.

Hoofstuk 6 bevatte een studie die gezet was in een klinisch scenario. In dit hoofdstuk werd de grootste dataset waarop ooit PCM uitgevoerd is, gebruikt om persoonlijke behandelplannen voor HIV patienten te voorspellen. De dataset zelf bevatte duizenden unieke HIV mutanten (protease en reverse transcriptase) en alle klinisch beschikbare medicijnen voor de behandeling van HIV.

Onze resultaten demonstreerden dat wij dit inderdaad succesvol kunnen en dat de gebruikte PCM aanpak beter presteert dan de huidige modellen in de kliniek die gebaseerd zijn op alleen de mutanten. Tevens konden wij aantonen dat onze modellen de onderliggende structuur activiteitsrelatie modelleren aangezien de modellen ook succesvol de activiteit van medicijnen op mutanten die niet in de dataset zaten kon voorspellen. Daarnaast konden wij ook de activiteit van medicijnen voorspellen in behandeling van HIV van patienten waarbij niet één enkele mutant dominant was maar meerdere. Tot slot konden wij ook hier een betrouwbaarheid aan de voorspellingen geven, een belangrijk gegeven in een klinische toepassing.

Hoofdstuk 7 is het enige hoofdstuk met een meer structuur georienteerde aanpak. Hier introduceren wij nieuwe technieken die gebruik maken van meerdere kristal structuren van een enkel eiwit om tot nieuwe inzichten te komen, welke met het gebruik van één eiwit niet te breiken zijn. Wij gebruikten HIV reverse transcriptase als model eiwit om tot nieuwe inzichten te komen op het gebied van de interactie tussen dit eiwit en zijn liganden. We konden ook succesvol nieuwe aanknopingspunten in de bindingsplaats identificeren door middel van onze methode ‘consensus structures’ welke nog niet geidentificeerd waren (terwijl er voor dit eiwit sinds 1995 kristal structuren beschikbaar zijn). Het exploiteren van deze aanknopingspunten zal leiden tot betere medicijnen voor het inhiberen van HIV reverse transcriptase.

Tot slot trokken wij in **hoofdstuk 8** een eindconclusie en introduceerden we een aantal toekomst perspectieven. De hoofdstelling van dit proefschrift was dat het samenvoegen van data uit verschillende disciplines (hier chemie, biologie en bioactiviteit) synergistisch is, wat wij ook hebben aangetoond. Echter de methoden die wij hiervoor gebruikten verschaffen een raamwerk, wat afhankelijk is van de bovengenoemde data voor het doen van voorspellingen. Dientengevolge kunnen deze methoden ook in nieuwe gebieden aangewend worden, zoals het voorbeeld van ‘drug target residence time’. Dit concept wordt momenteel uitgebreid onderzocht en deze data kunnen de fundering vormen van modellen die wellicht de ‘residence time’ van nieuwe liganden op een eiwit kunnen voorspellen.

Als laatste stellen wij voor dat elk onderzoeksprogramma uitgebreid moet worden met een preliminaire fase waarin een gedegen literatuuronderzoek alle relevante resultaten uit de literatuur combineert voor een effectieve hypothese vorming. Het verschil met huidige methoden is echter dat men zich hierbij niet moet beperken tot het eigen onderzoeksveld maar ook naar andere disciplines moet kijken. Een voorbeeld is het includeren van liganden voor een eiwit dat slechts weinig gemeen heeft met het eiwit dat onderzocht wordt maar desalniettemin nuttige informatie kan verschaffen. Daarnaast kan het erg lonend zijn om *voor* het maken van een structuur – activiteits model in informatica op zoek te gaan naar nieuwe (mogelijk efficientere) algoritmen. Een uitgebreid onderzoek kan op deze manier een nieuw project een vliegende start verschaffen.