

MEETING ABSTRACT

Open Access

Applications of proteochemometrics - from species extrapolation to cell line sensitivity modelling

Isidro Cortes-Ciriano^{1*}, Gerard JP van Westen², Daniel S Murrell³, Eelke B Lenselink⁴, Andreas Bender³, Therese E Malliavin¹

From Third International Society for Computational Biology (ISCB) European Student Council Symposium 2014

Strasbourg, France. 6 September 2014

Background

Proteochemometrics (PCM) is a predictive bioactivity modelling method which simultaneously models the bioactivity of multiple ligands against multiple targets. PCM permits exploration of the selectivity and promiscuity of ligands on biomolecular systems of different complexity. This includes proteins and even cell-line models [1,2]. The suitability of PCM to predict compound polypharmacology has been validated both retrospectively and in prospective experimental validation [1,2]. In practice, each ligand-target interaction is encoded by the concatenation of ligand and target descriptor vectors used to train a single machine learning model. The inclusion of both chemical and target information enables the extra- and interpolation on the chemical and on the biological space. Therefore, PCM permits to predict compound bioactivities on targets not present in the training phase [3].

Results

In this contribution, we show a methodological advancement in the field [4], namely how Bayesian inference (Gaussian Processes) can be successfully applied in the context of PCM for (i) the prediction of compound bioactivity along with the error estimation of the prediction; (ii) the determination of the applicability domain of a PCM model; and (iii) the inclusion of experimental uncertainty of bioactivity measurements. We illustrate how the application of PCM can be useful in medicinal chemistry to concomitantly optimize compounds

selectivity and potency, in the context of two application scenarios: (a) modelling isoform-selective cyclooxygenase inhibition; and (b) large-scale cancer cell line drug sensitivity prediction, where we benchmark the predictive signal of basal gene expression, gene copy-number variation, exome sequencing, and protein abundance data. We present the R package **Chemically Aware Model Builder** (*camb*) [5], which is able to perform the above mentioned modelling tasks. *camb* is an open source platform for the generation of Structure-Activity and Structure-Property models. The functionalities of *camb* include: (i) standardisation of chemical structure representation, (ii) calculation of 905 one-dimensional descriptors and 14 fingerprints for small molecules, (iii) 8 types of amino acid descriptors, (iv) 13 whole protein sequence descriptors, and (iv) training, validation and visualization of predictive models.

Conclusions

Overall, the application of PCM in these two case scenarios let us conclude that PCM is a suitable technique, on this data, to model the activity of ligands exhibiting diverse bioactivity profiles across a panel of targets, which can range from protein binding sites (a), to cancer cell-lines (b). The *camb* package constitutes a platform encompassing all steps for the generation of predictive models from chemical structures and their associated bioactivities/properties, which will provide reproducibility and simplify the generation of predictive bioactivity/property models.

¹Institut Pasteur, Unité de Bioinformatique Structurale; CNRS UMR 3825; Département de Biologie Structurale et Chimie, 25, rue du Dr Roux, 75015, Paris, France

Full list of author information is available at the end of the article

Authors' details

¹Institut Pasteur, Unité de Bioinformatique Structurale; CNRS UMR 3825; Département de Biologie Structurale et Chimie, 25, rue du Dr Roux, 75015, Paris, France. ²ChEMBL Group, European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, CB10 1SD, Hinxton, Cambridge, UK. ³Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Cambridge, UK. ⁴Division of Medicinal Chemistry, Leiden Academic Center for Drug Research, Leiden, The Netherlands.

Published: 13 February 2015

References

1. van Westen GJP, Wegner JK, Ijzerman AP, van Vlijmen HWT, Bender A: **Proteochemometric Modeling as a Tool to Design Selective Compounds and for Extrapolating to Novel Targets.** *Med Chem Commun* 2011, **2**:16-30.
2. Cortes-Ciriano I, Ain QU, Subramanian V, Lenseink EB, Mendez-Lucio O, Ijzerman AP, Wohlfahrt G, Prusis P, Malliavin TE, van Westen GJP, Bender A: **Polypharmacology Modelling Using Proteochemometrics (PCM): Recent Methodological Developments, Applications to Target Families, and Future Prospects.** *Med Chem Commun* .
3. van Westen GJP, Wegner JK, Gelyukens P, Kwanten L, Vereycken I, Peeters A, Ijzerman AP, van Vlijmen HWT, Bender A: **Which Compound to Select in Lead Optimization? Prospectively Validated Proteochemometric Models Guide Preclinical Development.** *PLoS ONE* 2011, **6**:e27518.
4. Cortes-Ciriano I, van Westen GJP, Lenseink EB, Murrell DS, Bender A, Malliavin TE: **Proteochemometric Modelling in a Bayesian framework.** *J Cheminf* 2014, **6**:35.
5. Murrell DS, Cortes-Ciriano I, van Westen GJP, Stott IP, Bender A, Malliavin TE, Glen RC: **Chemically Aware Model Builder (camb): An R package for property and bioactivity modeling of small molecules.**[<http://www.github.com/cambDI/camb>].

doi:10.1186/1471-2105-16-S3-A4

Cite this article as: Cortes-Ciriano *et al.*: Applications of proteochemometrics - from species extrapolation to cell line sensitivity modelling. *BMC Bioinformatics* 2015 **16**(Suppl 3):A4.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

