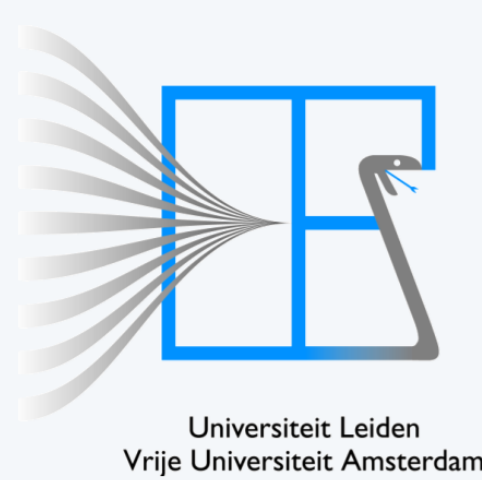




Contact info



Poster reprint

3D-Neighbourhood Protein Descriptors for Proteochemometric Modeling

The search for a perfect protein descriptor for the Adenosine receptors

Remco F. Swier¹, Gerard J.P van Westen¹, Jörg K. Wegner², Adriaan P. IJzerman¹, Herman W.T. van Vlijmen^{1,2}, Andreas Bender^{1,3}

¹ Leiden / Amsterdam Center for Drug Research, Leiden University, Netherlands ; ² Tibotec BVBA, Beerse, Belgium; ³ University of Cambridge, United Kingdom

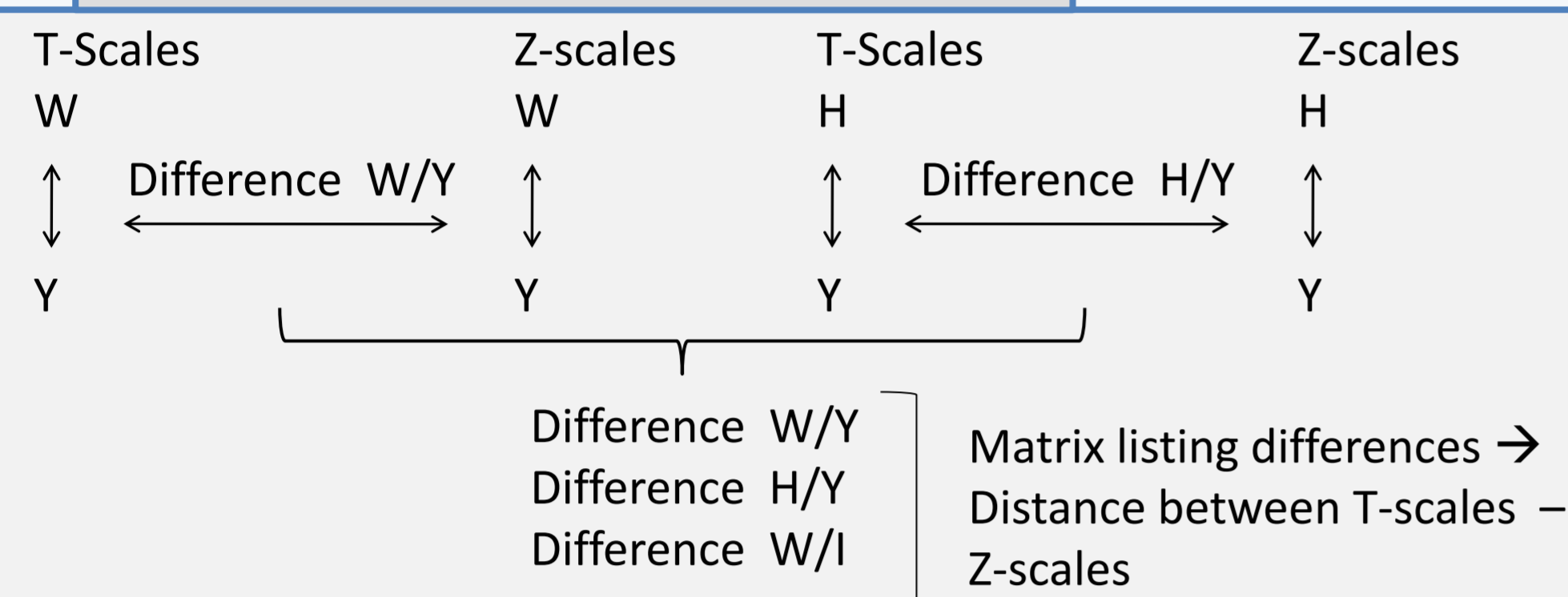
Introduction

- Proteochemometric Modeling (PCM) enables modeling of several targets against several compounds¹
- Based on preliminary results, incorporating more targets in PCM results in better extrapolation towards unseen targets
- Choosing the correct protein descriptor is of vital importance

Aims

1. Map the difference in performance of the protein descriptors
2. Explaining differences in the protein descriptor performance

PCA plot



A similarity matrix (scales between 0 - 1) listing the euclidian distances between the individual amino acids was calculated for each descriptor. From these matrices the difference between amino acids distances was calculated. This matrix served as input for a PCA analysis.

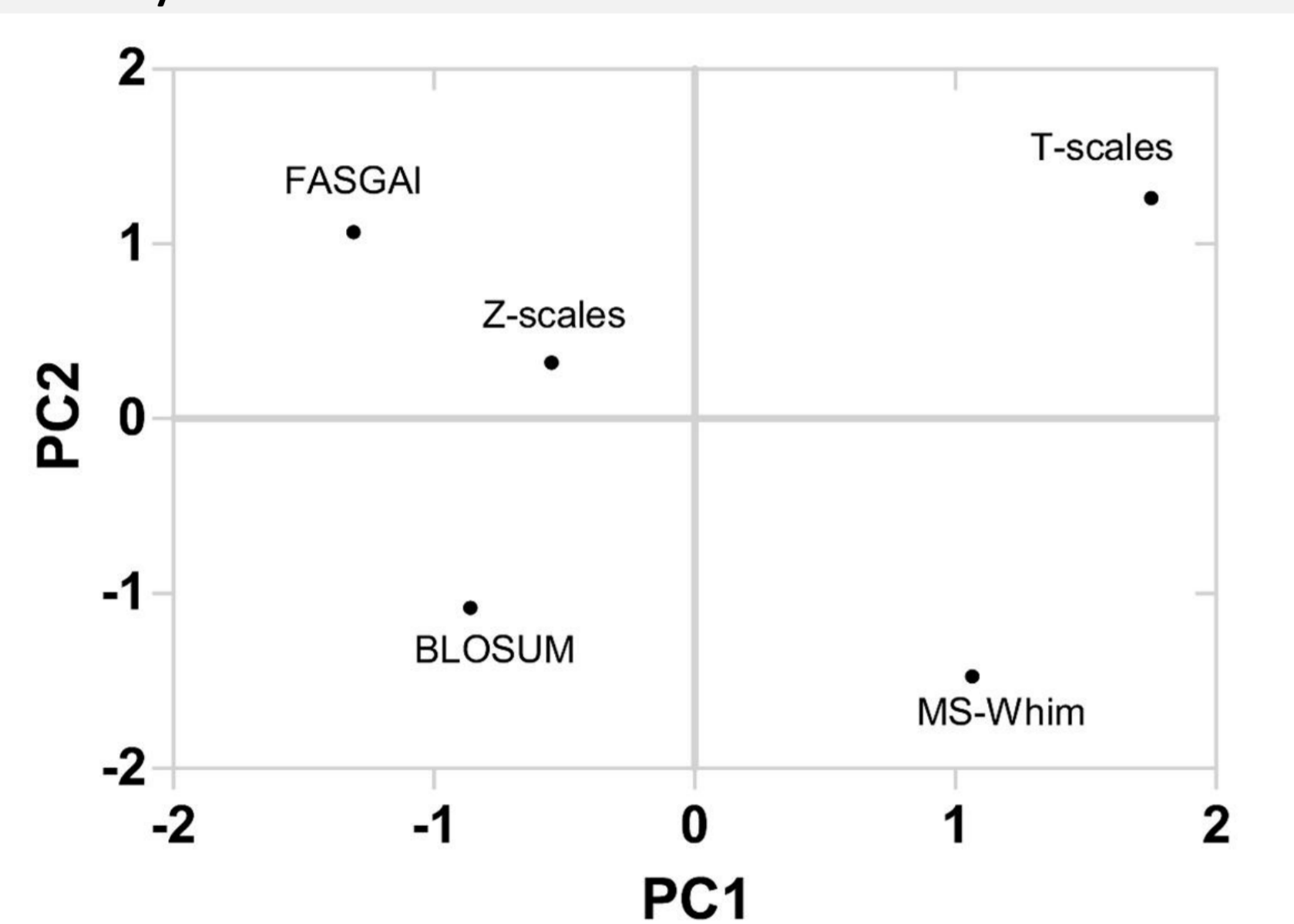


Figure 1. PCA analysis was performed on the average distance between the 20 natural amino acids similarity matrix of all amino acid descriptors. As expected the physicochemical property based descriptors cluster close together and BLOSUM, MS-Whim and T-Scales form different classes.

Methods and Data set

- PCM models were trained on Human and Rat Adenosine receptors (class A GPCRs)
 - 13,625 human and 9,296 rat activity data points from ChEMBL_09 Database
 - ECFP_6 was used as molecular fingerprint
- Bindingsite defined by 5Å around ligand in crystal structures of human Adenosine A_{2A} with agonist and antagonist (PDB codes: 3QAK and 3EML)
- Aligning 7TM domains allowed translation to other Adenosine subtypes (A₁, A_{2B} and A₃ and Rat receptors)
- Models were based on Support Vector Machines (SVM) with 5 fold cross validation
 - Parameters were optimized (cost of 10, γ of 1/nx and ϵ was 0.1)
- External validation was performed by Leave On Sequence Out method
- ProtFP is based on a unique feature per amino acid, hashed from 59 of the amino acid indices by Kawashima *et al.*²
- 3D Z-scales was constructed by selecting physicochemical neighbourhoods of 3.8Å around the amino acid in focus from crystal structures

Descriptor	BLOSUM	MS-Whim	T-scales	FASGAI	Z-scales	3D Z-scales	Prot_FP
Based on	Frequency / Alignment matrix	Molecular Electrostatic Potential	Topological property	Physico-chemical property	Physico-chemical property	Physico-chemical property	Feature based
Derived from	Frequency and occurrence of aligned amino acid pairs	PCA on 36 3D molecular electrostatic potential properties	PCA on 67 topological properties	Factor analysis of 335 physico-chemical properties	PCA on 29 physico-chemical properties	Neighbourhood selection of Z-scales	Hashed feature obtained from 59 amino acids indices

Descriptor Benchmark

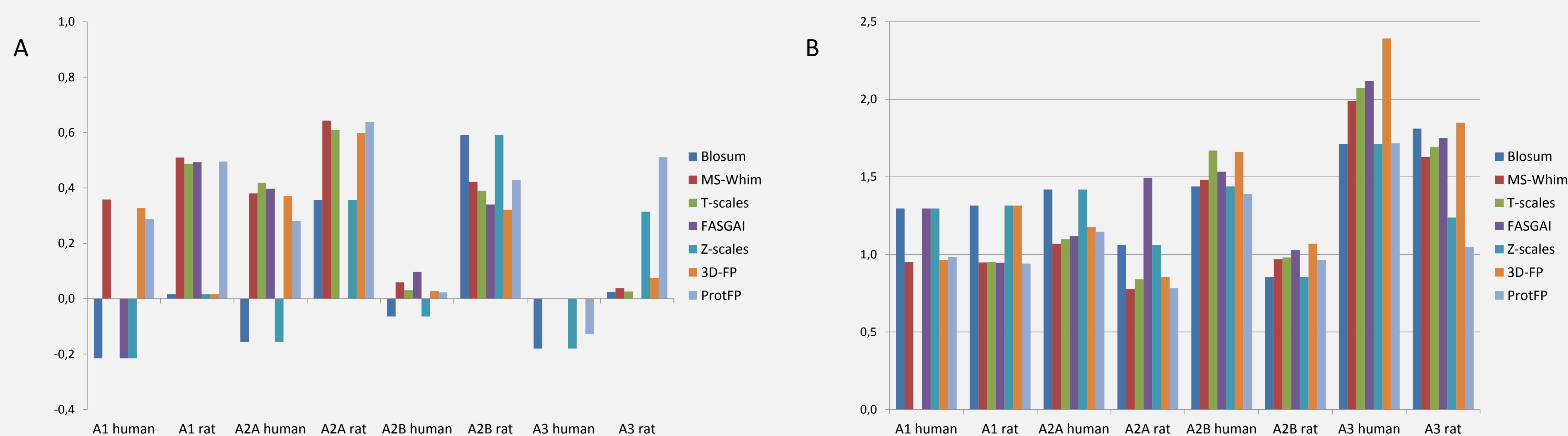


Figure 2. Performance of the protein descriptor benchmarking where (A) represents the R_0^2 and (B) the Root-Mean-Squared Error (RMSE). The Adenosine rat receptors are extrapolated better than the human receptors (higher R_0^2). The Adenosine A₃ receptors perform worse in the extrapolation but this could be due to the dissimilarity of the receptor compared to the other Adenosine receptors. The Adenosine A_{2A} and A_{2B} are most similar which contributes to the better extrapolation to the left out receptor. Overall it is seen that the ProtFP can extrapolate well for almost every left out receptor, specially for Adenosine A₃. However the Adenosine A₃ have higher RMSE scores and the A_{2A} and A_{2B} rat receptors have lower RMSE scores.

Conclusion

- PCA analysis can distinguish differences in protein descriptors
- The Z-scales and ProtFP perform better than the other protein descriptors implying that physicochemical property based protein descriptors are better than other descriptors
- The 3D Z-scales descriptor performs better compared with the normal Z-scales in the A_{2A} receptor from which a crystal structure is available
- Receptors that are too dissimilar to other receptors in the PCM show low performance in extrapolation

References

- 1: van Westen, G.J.P.; Wegner, J.K.; IJzerman, A.P.; van Vlijmen, H.W.T.; Bender, A.; **Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets**. Med. Chem. Commun., **2011**; 2(1): 16-30
- 2: Kawashima, S. *et al.* **AA index: amino acid index database**. Nucleic Acids Res. **2000** 28, 374

