New trends and strategies in the chemistry of advanced materials with relevance in biological systems, technique and environmental protection
7-8 October, 2021

Fingerprinting G-protein coupled receptors superfamily

Liliana HALIP, Ramona CURPAN, Ana Borota, Alina BORA, Sorin AVRAM

"Coriolan Dragulescu" - Institute of Chemistry, 24 Mihai Viteazu, 300223-Timisoara, Romania
e-mail: lili.ostopovici@acad-icht.tm.edu.ro

SUMMARY

G-protein coupled receptors (GPCR) represent a family of proteins with significant pharmacological importance, they being involved in the completion of most physiological processes in the human body. Although it is one of the most studied families of proteins - about a third of the drugs on the market target a member of this receptor family - there are still understudied GPCR receptors for which no active or selective ligands have been identified. The present work aims to explore the relationship degrees among GPCR members and to classify all human GPCR receptors based on new approach that allow associations of known GPCR with unstudied members. Thus, innovative binary fingerprints were generated using the 3D protein structures, active ligands and activation/coupling information of GPCR family. Next, GPCR superfamily were topographical mapped in Kohonen networks and results were analyzed. A set of 29 understudied GPCR have been found of interest. The results suggest the applicability and potential of this approach to classify understudied GPCRs.

METHODS

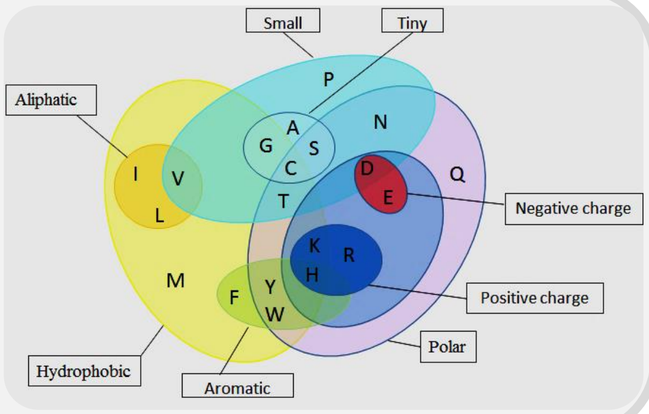
Amino acid encoding. **For each essential amino acid a unique 11-bits length binary fingerprint was generated. The fingerprints were defined based on Venn diagram classification [1] and other physico-chemical descriptors: Cbeta-branching, Calpha and OH side-chain containing, etc.**

Sequence Alingment. **The GPCRs sequences were downloaded in FASTA format from IUPHAR database [2] and aligned with T-Coffee bioinformatics package [3] using a multistep alignment procedure. Amino acids were numbered using the numbering system proposed by Ballesteros and Weinstein in 1995 [4] where the most highly conserved residue in each helix was assigned a value of 50 preceded by the number of the transmembrane helix.**

Binding site definition. **Based on the Gloriam's binding site definition [5], 52 aIDs representing the relevant positions from GPCR primary binding site were labeled in the sequence alignment. Amino acids which reside in the aIDs and corresponding neighbors were extracted from the sequence alignment and labeled accordingly their type (binding site residue, neighbor). By replacing the amino acid type with the unique fingerprint, each receptor was described by a 572-biths length binary finger print. The influence of the binding site environment for each receptor was estimated according to equation (1):**

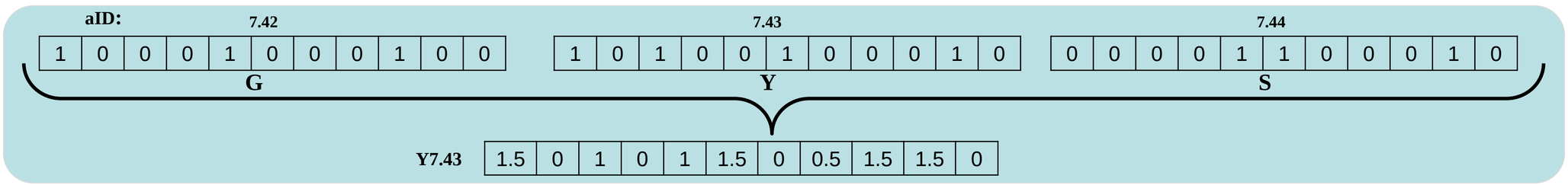
$$F(aID_i) = 0.5 * F(aID_{i-1}) + F(aID_i) + F(aID_{i+1}) \quad (1)$$

Self-organizing map classification. **Self-organizing maps were generated with *kohonen* package [6] in RStudio [7,8]. Each unit has assigned a codebook or weight vector which has the same dimension as the input space and describes the variables pattern of the enclosed samples (compounds). For a relevant classification of the receptors, we aim for at least 1 receptor per unit. SOM grid of 5 × 5 units with a hexagonal topology showed the best chemical space coverage and was selected to analyze the results.**



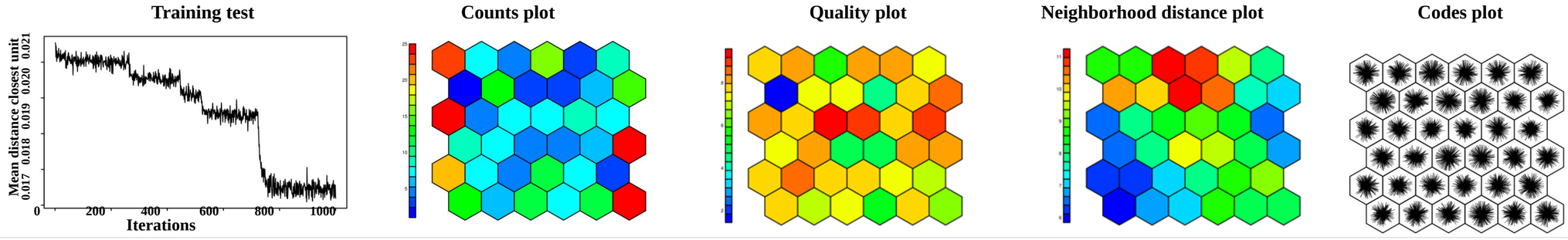
RESULTS AND DISCUSSIONS

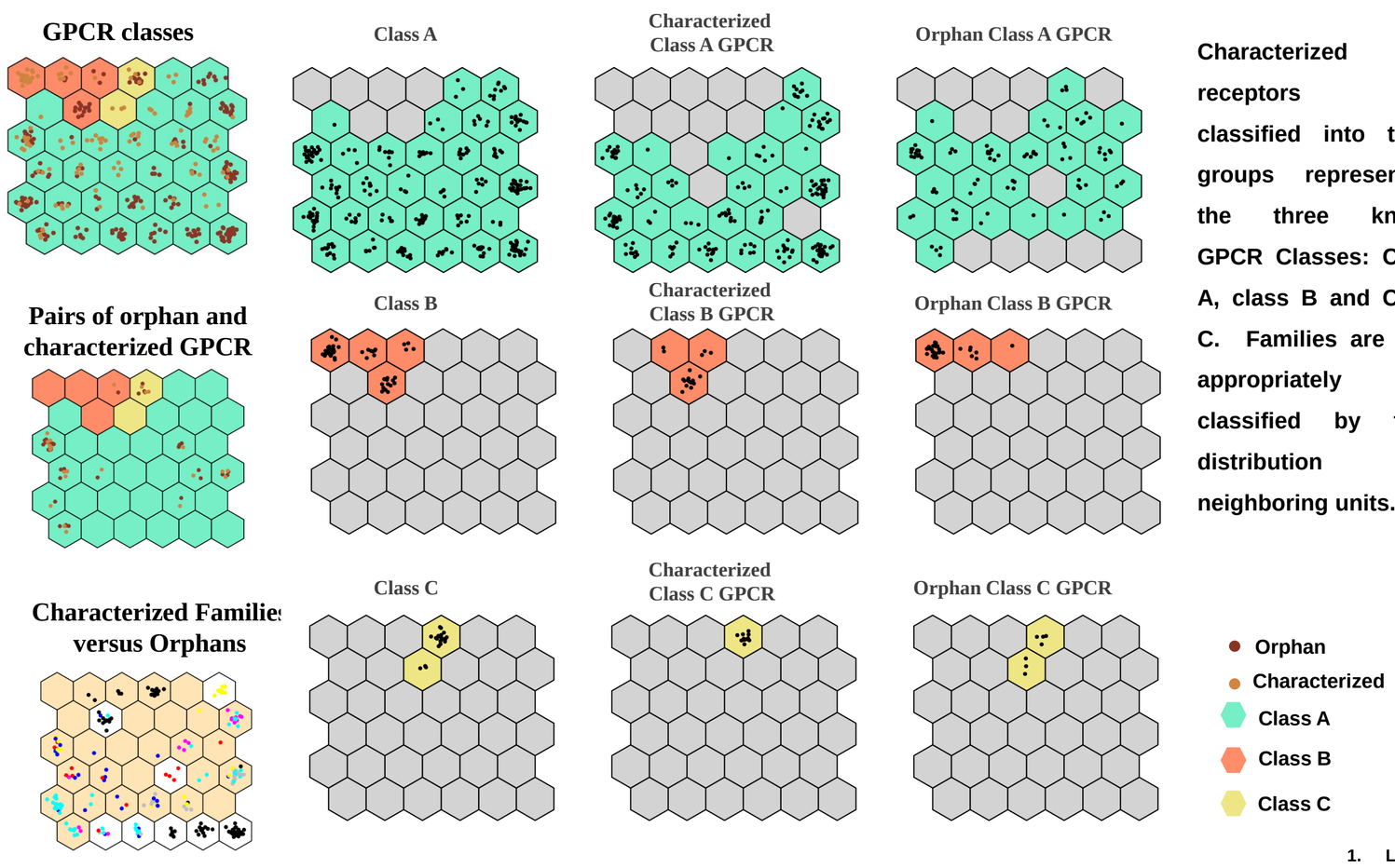
The binding site environment, specific for each receptor, was estimated as exemplified in scheme 1, for an aID filled with tyrosine whose neighbors are glycine and serine (GYS motif).



Scheme 1 – Estimation of the binding site environment

SOM evaluation





Characterized receptors are classified into three groups representing the three known GPCR Classes: Class A, class B and Class C. Families are also appropriately classified by their distribution to neighboring units.

Table I – Orphan-Characterized GPCR pairs selected for study

oGPCR	cGPCR	oGPCR	cGPCR	oGPCR	cGPCR
BB3	GRPR	GPR87	P2Y14	P2RY8	PAR3
GPR42	FFA3	GP132	PAR3	PSYR	P2Y12
GPR1	AGT1	GP151	BKRB1	SUCR1	P2Y1
GPR4	LPA6	GP171	P2Y12	LGR4	FSHR
GPR31	HCA2	GP174	LPA6	LGR5	FSHR
GPR33	AGT1	GP183	PTAFR	LGR6	FSHR
GPR34	P2Y14	CCRL2	CCR3	GPR157	hBAI1
GPR35	CLTR2	CML1	BKRB1	OXGR1	P2Y4
GPR39	NT1	MTR1L	MT1a	P2Y10	LPA6
GPR45	CCKB	OGR1	PTAFR	GPR83	TAC1

CONCLUSIONS

By using the information about the GPCR primary binding site, weighted on the Gloriam's set, human characterized GPCR were classified with high accuracy into conventional classes and families. Orphans were also accurately classified according to their class. A set of 30 pairs orphan-characterized GPCR have been found of interest. The results suggest the applicability and potential of this approach to classify orphan GPCRs based on the GPCR primary binding site.

REFERENCES

1. Livingstone CD, Barton GJ. CABIOS, 9, Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation.(1993), 9(6):745-756.
2. Harding SD, Sharman JL, Faccenda E, et al. Nucl. Acids Res. The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. 46 (2018) (Issue D1): D1091-D1106.
3. Notredame C, Higgins DG, Heringa J. J Mol Biol, T-Coffee: A novel method for multiple sequence alignments. (2000), 302:205-217.
4. Ballesteros JA, Weinstein H. Methods Neurosci, Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. (1995), 25:366-428.
5. Gloriam DE, Foord SM, BlaneyFE, Garland SL. Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. J Med Chem (2009) 52(14):4429-4442.
6. Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics. 1982;43(1):59-69.
7. Team R. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/. 2015.
8. Wehrens R, Buydens LMC. Self- and Super-organizing Maps in R: The kohonen Package. Journal of Statistical Software. 2007;21(5):1-19.

This work was supported by the Theme 1.2.1 of the Project No. 1.2 of the Institute of Chemistry Timisoara of Romanian Academy.