G Protein-Coupled Receptor Kinases Motifs, Hydrophobicity and Weakly Connected Components in Neighborhoods of Nitrogen Spatial Distributions

Jeff Cromwell, PhD

1 Abstract

G proteins, guanine nucleotide-binding proteins act as molecular switches inside cells and involved in signal transmission from diverse stimuli outside to the interior of the cell. Their activity is regulated by factors that control their ability to (a) bind to and (b) hydrolyze guanosine triphosphate (GTP) to (c) guanosine diphosphate (GDP). Here the emphasis is placed on a collection of G proteins crystals and their sequences examined with molecular properties such as Stability Index Binding Potential ALiphatic f.1 CpH5 CpH7 CpH9 [1001] Based on the similarities of the proteins, it is expected that the variations would be at the boundaries of the thresholds for the A and not A categorification. This evidence for a change in behavior at that aggregate boundaries is presented.

Keywords: G Protein Crystals 4WBO 4L9I 6N4B 7D7M 1XVS 4WNK 2BCJ 6PJX 4TND 4TNB 3NYO 3NYN 3CAP 1OMW 3D7M 3JBQ 1YM7 3V5W 4GBR 4GPO 7L1U 7L1V 1TNY 3HYT 3HYR 4YHJ 7K7L 7K7Z 3Q1C 3UZT 3UZS 1XHM 2JX0 Molecular Properties, Stability, Binding Potential, Community Detection, Guanine Nucleotide-Binding Proteins

2 Introduction

The classification of a graph into weakly or strongly connected is based on The weakly connected components are found by a simple breadth-first search. The strongly connected components are implemented by two consecutive depth-first searches. components finds the maximal (weakly or strongly) connected components of a graph. The count of components does almost the same as components but returns only the number of clusters found instead of returning the actual clusters. The component distribution creates a histogram for the maximal connected component sizes. [101] Community detection for the subgraphs of (a) 10 nodes in the vertices 1:30 and (b) 10 nodes in the vertices of 10:20 was also performed.

3 G Protein Crystals

G proteins, guanine nucleotide-binding proteins a part of GTPases enzymes are molecular switches inside cells and transmit signals from a variety of stimuli outside a cell to its interior. Control factors include the ability to bind to and hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP). When bound to GTP, the state is A or 'on', and, when bound to GDP, the state is not A or 'off'. G proteins are. [20] Table 1 has a collection of G Protein Crystals descriptions. [1]

	Pdb ID	Description
1	4WBO	Bovine G Protein Coupled Receptor Kinase 1 in Complex with Amlexanox
2	4L9I	Bovine G Protein Coupled Receptor Kinase 1 in Complex with Paroxetine
3	6N4B	Cannabinoid Receptor 1-G Protein Complex
4	7D7M	Cryo-EM Structure of the Prostaglandin E Receptor EP4 Coupled to G Protein
5	1XVS	Crystal structure of apaG Protein from Vibrio cholerae
6	4WNK	Crystal Structure of Bovine G Protein Coupled-Receptor Kinase 5 in Complex with CCG215022
7	2BCJ	Crystal Structure of G Protein-Coupled Receptor Kinase 2 in Complex with Galpha-q and Gbetagamma Subunits
8	6PJX	Crystal Structure of G Protein-Coupled Receptor Kinase 5 (GRK5) in Complex with Calmodulin (CaM)
9	4TND	Crystal Structure of G Protein-Coupled Receptor Kinase 5 in Complex with AMP-PNP
10	4TNB	Crystal Structure of G Protein-Coupled Receptor Kinase 5 in Complex with Sangivamycin
11	3NYO	Crystal Structure of G Protein-Coupled Receptor Kinase 6 in Complex with AMP
12	3NYN	Crystal Structure of G Protein-Coupled Receptor Kinase 6 in Complex with Sangivamycin
13	3CAP	Crystal Structure of Native Opsin: the G Protein-Coupled Receptor Rhodopsin in its Ligand-free State
14	1OMW	Crystal Structure of the complex between G Protein-Coupled Receptor Kinase 2 and Heterotrimeric G Protein beta 1 and gamma 2 subunits
15	3D7M	Crystal Structure of the G Protein Fast-Exchange Double Mutant I56C/Q333C
16	3JBQ	Domain Organization and Conformational Plasticity of the G Protein Effector, PDE6
17	1YM7	G Protein-Coupled Receptor Kinase 2 (GRK2)
18	3V5W	Human G Protein-Coupled Receptor Kinase 2 in Complex with Soluble Gbetagamma Subunits and Paroxetine
19	4GBR	N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor
20	4GPO	Oligomeic Turkey Beta1-Adrenergic G Protein-Coupled Receptor
21	7L1U	Orexin Receptor 2 (OX2R) in Complex with G Protein and Natural Peptide-Agonist Orexin B (OxB)
22	7L1V	Orexin Receptor 2 (OX2R) in Complex with G Protein and Small-Molecule Agonist Compound 1
23	1TNY	Rat Protein Geranylgeranyltransferase Type-I Complexed with a GGPP analog and a FREKKFFCAIL Peptide Derived from the Heterotrimeric G Protein Gamma-2 Subunit
24	3HYT	Structural Basis of GDP Release and Gating in G Protein Coupled Fe2+ Transport
25	3HYR	Structural Insight into G Protein Coupling and Regulation of Fe2+ Membrane Transport
26	4YHJ	Structure and Function of the Hypertension Variant A486V of G Protein-coupled Receptor Kinase 4 (GRK4)
27	7K7L	Structure of a hit for G Protein Coupled Receptor Kinase 2 (GRK2) Inhibitor for the Potential Treatment of Heart Failure
28	7K7Z	Structure of a hit for G Protein Coupled Receptor Kinase 2 (GRK2) Inhibitor for the Potential Treatment of Heart Failure
29	3Q1C	Structure of EspG Protein
30	3UZT	Structure of the C13.18 RNA Aptamer in Complex with G Protein-Coupled Receptor Kinase 2
31	3UZS	Structure of the C13.28 RNA Aptamer Bound to the G Protein-Coupled Receptor Kinase 2-Heterotrimeric G Protein Beta 1 and Gamma 2 Subunit Complex
32	1XHM	The Crystal Structure of a Biologically Active Peptide (SIGK) Bound to a G Protein Beta:Gamma Heterodimer
33	2JX0	The paxillin-binding domain (PBD) of G Protein Coupled Receptor (GPCR)-kinase (GRK) interacting protein 1 (GIT1)

4 Molecular Properties

Table 1 has A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. A protein have high binding potential if the index value is higher

than 2.48. The relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) is considered to be a positive factor for the increase of thermostability of globular proteins. net charge of a protein sequence based on the Henderson-Hasselbalch equation for pH levels of 5 7 and 9. f.1 is the crosscovariance based on one of the sequence S at lag 1 is given by f.1 with property1 = Hydrophobicity based on the KyteDoolittle Scale and property2 =Hydrophobicity based on the Eisenberg scale. [1002]

	Protein	Stability Index	Binding Potential	ALiphatic	f.1	CpH5	CpH7	CpH9
1	4WBO	40.0	1.66510	82.5	-0.3432	46.566	-3.232	-45.01
2	4L9I	39.2	1.65866	82.2	-0.3353	24.375	-0.628	-21.91
3	6N4B	37.7	1.25281	90.9	0.0392	40.894	11.216	-24.97
4	7D7M	38.8	1.53155	88.1	-0.1159	37.549	14.382	-17.54
5	1XVS	50.5	1.61202	87.9	-0.2637	-8.757	-13.579	-17.24
6	4WNK	52.4	2.05826	76.9	-0.5617	21.019	4.457	-12.37
7	2BCJ	45.4	1.91278	82.3	-0.3956	41.942	-2.250	-40.56
8	6PJX	46.6	1.97087	77.4	-0.5188	8.277	-9.821	-27.95
9	4TND	49.9	2.04450	75.9	-0.5715	19.532	3.160	-14.88
10	4TNB	49.9	2.04450	75.9	-0.5715	19.532	3.160	-14.88
11	3NYO	44.0	1.87472	80.6	-0.4633	49.989	18.217	-19.50
12	3NYN	44.0	1.87472	80.6	-0.4633	49.989	18.217	-19.50
13	3CAP	34.2	-0.00377	92.4	0.5445	13.460	0.257	-20.84
14	1OMW	48.0	1.88378	81.0	-0.3890	38.754	3.338	-30.48
15	3D7M	37.9	1.74131	85.4	-0.2675	16.818	-1.860	-12.66
16	3JBQ	37.4	1.52043	86.1	-0.2473	39.977	-25.982	-85.60
17	1YM7	54.1	1.92583	81.6	-0.4768	114.410	18.356	-55.35
18	3V5W	48.5	1.93166	80.4	-0.4118	36.014	0.403	-31.93
19	4GBR	35.9	0.77422	107.1	0.3123	18.300	10.101	-1.69
20	4GPO	34.2	0.08490	124.5	0.7859	16.946	10.342	-7.30
21	7L1U	40.1	1.36656	88.9	-0.0256	35.658	7.531	-27.45
22	7L1V	39.8	1.35472	87.7	-0.0345	39.287	10.470	-26.63
23	1TNY	46.8	1.95191	88.7	-0.4026	81.852	-48.842	-166.89
24	3HYT	40.8	1.43593	114.6	-0.0416	5.613	-13.815	-31.26
25	3HYR	37.3	1.38403	113.4	-0.0292	2.022	-17.810	-35.26
26	4YHJ	55.3	2.12553	75.1	-0.5648	34.288	5.255	-31.82
27	7K7L	48.8	1.92954	80.7	-0.4124	37.866	2.401	-29.93
28	7K7Z	48.5	1.92305	80.5	-0.4134	37.866	2.401	-29.93
29	3Q1C	48.7	1.87467	83.4	-0.3359	0.776	-10.769	-19.05
30	3UZT	54.2	1.91085	82.3	-0.4650	27.655	5.147	-13.37
31	3UZS	47.6	1.87318	80.7	-0.3895	33.853	-0.142	-33.45
32	1XHM	33.7	1.74505	82.0	-0.2151	6.713	-4.289	-19.09
33	2JX0	66.6	1.79664	91.6	-0.3702	3.609	0.379	-3.29

Table 1: A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. A protein have high binding potential if the index value is higher than 2.48. The relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) is a positive factor for the increase of thermostability of globular proteins. net charge of a protein sequence based on the Henderson-Hasselbalch equation for pH levels of 5 7 and 9. f.1 is the crosscovariance based on one of the sequence S at lag 1 is given by f.1 with property1 = Hydrophobicity based on the KyteDoolittle Scale and property2 =Hydrophobicity based on the Eisenberg scale.

5 Results

Figure 1 has the matrix with the indices of points for the set of the k nearest neighbours of each other C1, k=1 and Figure 2 has the matrix with the indices of points for the set of the k nearest neighbours of each other C2, k=3. Here a separate graph for each component of a graph is created where the mode is a type of the components, wither weak for weakly connected components or strong for strongly connected components. The max components is the maximum number of components to return and the minmum vertices are the minimum number of vertices a component should contain in order to place it in the result list. [101]
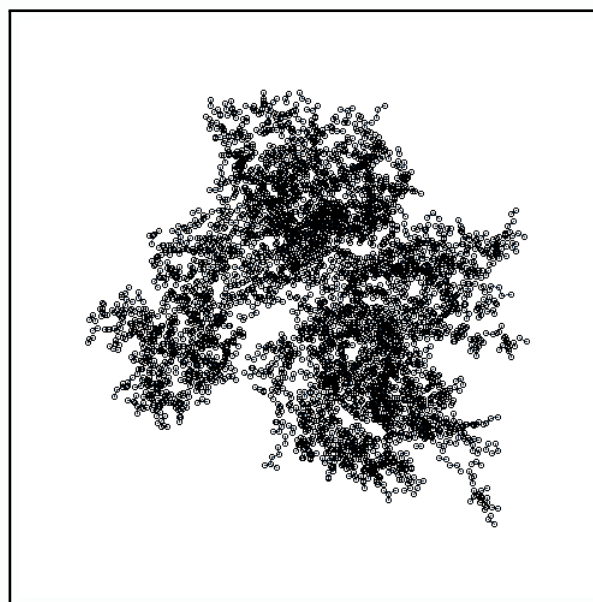

Figure 1: Matrix with the indices of points for the set of the k nearest neighbours of each other in coordinate system C1, k=1
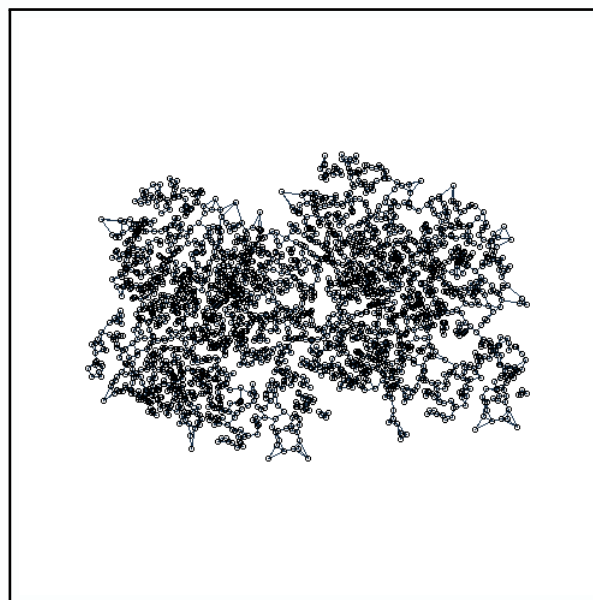

Figure 2: Matrix with the indices of points for the set of the k nearest neighbours of each other coordinate system C2, k=3

For the entire G Protein the components with a mimimum of 10 vertices has diameter for each component as 4 4 7 4 4 5 4 and 0.0000000000 0.2789354758 0.4737007281 0.2162942506 0.0305674115 0.0005021341. In the filter for just Nitrogen, the components with a minimum vertices of 2 has diameter 69 1 1 6 3 2 9 6 8 1. The degree distribution is given by 0.0000000000 0.0000000000 0.0000000000 0.0205203371 0.0982044705 0.2304873580 0.3089043606 0.2257237083 0.0894100403 0.0234518139 0.0029314767 0.0003664346. [101]
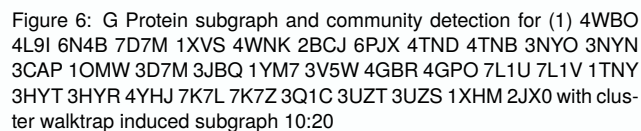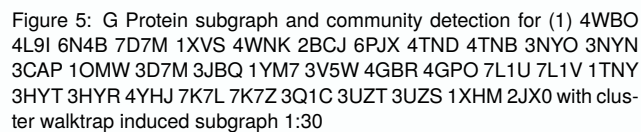
Figure 4 has minimum of 10 vertices for all elements in the G Protein and Figure 5 has minimum of 2 vertices of just Nitrogen in the G Protein

2

Figure 3: minimum of 10 vertices for all elements in the G Protein

Figure 4: minimum of 2 vertices of just Nitrogen in the G Protein
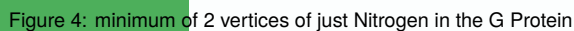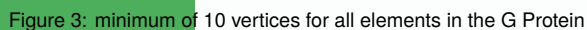
The closeness centrality of a vertex is defined by the inverse of the average length of the shortest paths to/from all the other vertices in the graph such that $\frac{1}{\sum(d(v,i),i!=v)}$ If there is no (directed) path between vertex v and i then the total number of vertices is used in the formula instead of the path length. [101]

The mean distance of the graph G1 is 17.91795 while the mean distance of G is 1.448668. The authority scores of the vertices are defined as the principal eigenvector of t(A)*A, where A is the adjacency matrix of the graph.The authority score of the graph G is 4 and the authority score of G1 is 15.17024. [101]

Figures 5 and 6 have the community detection for the subgraphs of (a) 10 nodes in the vertices 1:30 and (b) 10 nodes in the vertices of 10:20. [101]

Figure 5: G Protein subgraph and community detection for (1) 4WBO 4L9I 6N4B 7D7M 1XVS 4WNK 2BCJ 6PJX 4TND 4TNB 3NYO 3NYN 3CAP 1OMW 3D7M 3JBQ 1YM7 3V5W 4GBR 4GPO 7L1U 7L1V 1TNY 3HYT 3HYR 4YHJ 7K7L 7K7Z 3Q1C 3UZT 3UZS 1XHM 2JX0 with cluster walktrap induced subgraph 1:30

Figure 6: G Protein subgraph and community detection for (1) 4WBO 4L9I 6N4B 7D7M 1XVS 4WNK 2BCJ 6PJX 4TND 4TNB 3NYO 3NYN 3CAP 1OMW 3D7M 3JBQ 1YM7 3V5W 4GBR 4GPO 7L1U 7L1V 1TNY 3HYT 3HYR 4YHJ 7K7L 7K7Z 3Q1C 3UZT 3UZS 1XHM 2JX0 with cluster walktrap induced subgraph 10:20

6 Conclusion

In this short discussion on G proteins, the emphasis was placed on a collection of G proteins crystals with the molecular properties Stability

Index Binding Potential ALiphatic f.1 CpH5 CpH7 CpH9 [1001] Based on the similarities of the proteins, it is expected that the variations would be at the boundaries of the thresholds for the A and not A distinction. This evident for several proteins with a change in behavior at that aggregate boundaries. Community detection for the subgraphs of (a) 10 nodes in the vertices 1:30 and (b) 10 nodes in the vertices of 10:20 was also performed.

7 References

[1] Wikipedia contributors. "WD40 repeat." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 6 Dec. 2020. Web. 11 Apr. 2021.

[2] Wikipedia contributors. "Gaussian network model." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 12 Dec. 2020. Web. 27 Mar. 2021.

[3] Wikipedia contributors. "Lp space." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 2 Mar. 2021. Web. 27 Mar. 2021.

[4] Wikipedia contributors. "Minkowski distance." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 30 Dec. 2020. Web. 27 Mar. 2021.

[10] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199–D205 (2014).

[20] Wikipedia contributors. "G protein." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Jul. 2021. Web. 14 Aug. 2021.

[101] Wikipedia contributors. "Graph theory." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Mar. 2021. Web. 24 Mar. 2021.

[102] Wikipedia contributors. "Spectral graph theory." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Mar. 2021. Web. 24 Mar. 2021.

[400] Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353-D361 (2017).

[401] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457-D462 (2016).

[402] Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

[420] GeneCards Version 3: the human gene integrator

[430] DrugBank 5.0: a major update to the DrugBank database for 2018.

[440] COSMIC: the Catalogue Of Somatic Mutations In Cancer.

[450] Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders

[460] The ClinicalTrials.gov Results Database — Update and Key Issues

[470] PubChem Substance and Compound databases

[480] The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible

[490] MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search

[803] Cromwell, J. "Mathematical Learning Space Research Portfolio" Mathematical Learning Space Research Portfolio, http://mathlearningspace.weebly.com/ 8 3 2021. Web. 3 Aug. 2021.

[1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[1001] Osorio, D., Rondon-Villarreal, P. and Torres, R. Peptides: A package for data mining of antimicrobial peptides. The R Journal. 7(1), 4-14 (2015).

[1002] Spedicato Giorgio Alfredo Discrete Time Markov Chains with R. The R Journal 7 (2017).

[1003] Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S: New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998, 41:2481-2491.

[1004] Andrew O. Finley, Sudipto Banerjee, Bradley P. Carlin (2007). spBayes: An R Package for Univariate and Multivariate Hierarchical Point-Referenced Spatial Models. Journal of Statistical Software, 19(4), 1-24. URL http://www.jstatsoft.org/v19/i04/.

[1005] Andrew O. Finley, Sudipto Banerjee, Alan E.Gelfand (2015). spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models. Journal of Statistical Software, 63(13), 1-28. URL http://www.jstatsoft.org/v63/i13/.

[1006] Roger Bivand, Gianfranco Piras (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software, 63(18), 1-36. URL http://www.jstatsoft.org/v63/i18/.

[1007] Bivand, R. S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis, 45(2), 150-179.