QSAR Modelling of LRRK2 Inhibitors for the Treatment of Parkinson's Disease

Anirudh Rao

BE21B004

Introduction

Parkinson's disease is one of the most common neurodegenerative disorders in the world [1]. It primarily affects motor function, leading to slow movement, tremors, rigidity, and imbalance. It can also lead to non-motor symptoms like cognitive impairment, dementia, and sleep disorders. There is no cure for Parkinson's and current treatments only offer symptomatic relief [2].

Figure 1: Symptoms of Parkinson's Disease
(Source: Stanford Medicine)

Studies have shown that there is a correlation between certain genes and the occurrence of Parkinson's. Of these genes, *lrrk2* (leucine-rich repeat kinase 2) has been hypothesized to be behind a significant fraction of Parkinson's cases [3]. The protein encoded by this gene, LRRK2, is a 286 kDa protein with multiple domains. Its most important domains include a kinase domain and a GTPase domain.

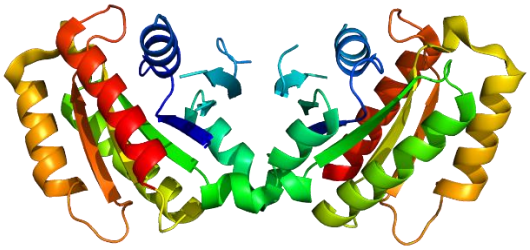


Figure 2: LRRK2, the target protein
(Source: Wikimedia Commons)

While no studies have confirmed the complete function of this protein, it has been shown to be involved in neuron plasticity, vesicle trafficking, mitochondrial activity, autophagy, and apoptosis [4]. Hyperactivation of LRRK2 may be involved in the onset and progression of Parkinson's disease. Abnormal activity in its kinase domain, due to hyperactivation, can lead to the accumulation of α-synuclein, which is a hallmark of Parkinson's [5]. Thus, targeting LRRK2 is a good strategy for the treatment of this disease.

Using previously identified inhibitors of this protein, we develop a quantitative structure-activity relationship (QSAR) model. This will help us predict the activity of any newly discovered inhibitors of LRRK2.

Methodology

Identifying LRRK2 inhibitors from literature

13 small molecule drugs that have been shown to inhibit LRRK2 were found from existing literature [6, 7]. These drugs target the kinase domain of LRRK2. Their details are summarised below:

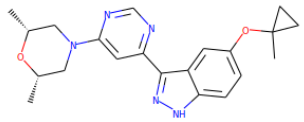
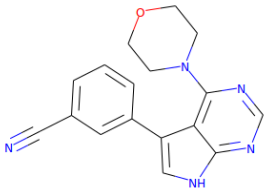
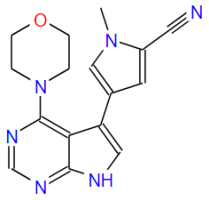
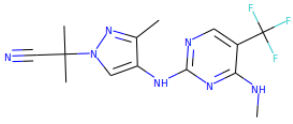
Table 1: Small molecule LRRK2 inhibitors

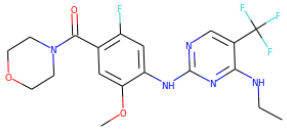
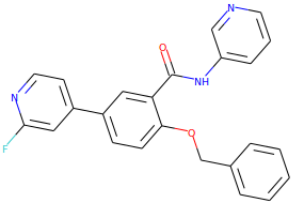
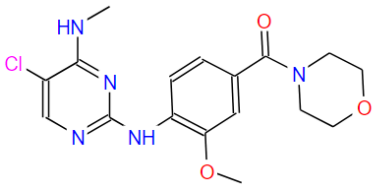
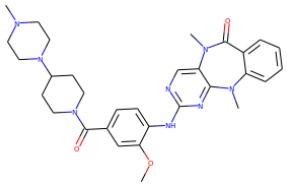
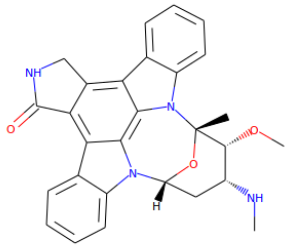
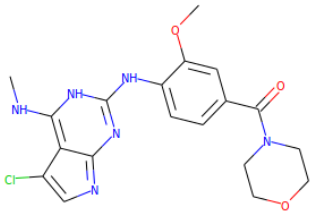
Sl.	Drug Name	ZINC ID	IC50 (nM)	Reference
1	MLi-2	ZINC220966210	0.8	[6]
2	PF-06447475	ZINC210747484	3.0	[6]
3	PF-06685360	Not available	2.3	[6]
4	GNE-0877	ZINC103260600	3.0	[6]
5	GNE-7915	ZINC95578650	9.0	[6]
6	GSK2578215A	ZINC68267183	10.9	[6]
7	HG-10-102-1	Not available	20.3	[6]
8	LRRK2-IN-1	ZINC84605049	13.0	[6]
9	Staurosporine	ZINC3814434	1.0	[7]

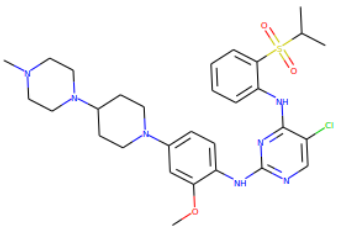
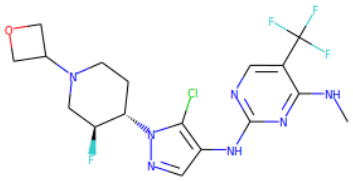
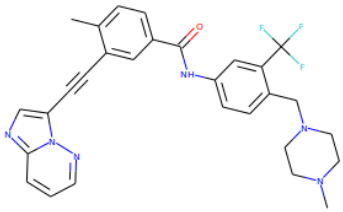
10	JH-II-127	ZINC253387881	6.0	[7]
11	TAE684	ZINC55760827	6.1	[7]
12	GNE-9605	ZINC103260612	19	[7]
13	Ponatinib	ZINC36701290	31	[7]

Using ZINC15, the SMILES notations and structures for these compounds were obtained.

Table 2: SMILES notations and structures

Sl.	Drug Name	SMILES	Structure
1	MLi-2	C[C@H]1CN(c2cc(-c3n[nH]c4ccc(OC5(C)CC5)cc34)ncn2)C[C@@H](C)O1	
2	PF-06447475	N#Cc1cccc(-c2c[nH]c3ncnc(N4CCOCC4)c23)c1	
3	PF-06685360	Cn4cc(c2c[nH]c3ncnc(N1CCOCC1)c23)cc4C#N	
4	GNE-0877	CNc1nc(Nc2cn(C(C)(C)C#N)nc2C)ncc1C(F)(F)F	

5	GNE-7915	CCNc1nc(Nc2cc(F)c(C(=O)N3CCOCC3)cc2OC)ncc1C(F)(F)F	
6	GSK2578215A	O=C(Nc1cccnc1)c1cc(-c2ccnc(F)c2)ccc1OCc1ccccc1	
7	HG-10-102-1	CNc3nc(Nc2ccc(C(=O)N1CCOCC1)cc2OC)ncc3Cl	
8	LRRK2-IN-1	COc1cc(C(=O)N2CCC(N3CCN(C)CC3)CC2)ccc1Nc1ncc2c(n1)N(C)c1ccccc1C(=O)N2C	
9	Staurosporine	CN[C@@H]1C[C@H]2O[C@@](C)([C@@H]1OC)n1c3ccccc3c3c4c(c5c6ccccc6n2c5c31)C(=O)NC4	
10	JH-II-127	CNc1[nH]c(Nc2ccc(C(=O)N3CCOCC3)cc2OC)nc2ncc(Cl)c1-2	

11	TAE684	COc1cc(N2CCC(N3CCN(C)CC3)CC2)ccc1Nc1ncc(Cl)c(Nc2ccccc2S(=O)(=O)C(C)C)n1	
12	GNE-9605	CNc1nc(Nc2cnn([C@H]3CCN(C4COC4)C[C@@H]3F)c2Cl)ncc1C(F)(F)F	
13	Ponatinib	Cc1ccc(C(=O)Nc2ccc(CN3CCN(C)CC3)c(C(F)(F)F)c2)cc1C#Cc1cnc2cccnn12	

Obtaining descriptors for the inhibitors

The SMILES notations for all 13 drugs were written in a single .smi file. This was given as input to the ChemDes webserver [8].The webserver was used to compute 1538 PaDEL descriptors. These are all 1D and 2D descriptors. The output of the webserver was downloaded as an .xlsx file for further analysis.

Developing and evaluating a one-parameter QSAR model

To develop the QSAR model, we divided the data into a training set and a validation set. We took 3 data points to be part of our validation set. The remaining 10 points were used to develop the model.

The data was divided into all ${}^{13}C_3 = 286$ possible training-validation splits. For each split, the training set was used to identify the descriptor with the best performance in predicting the IC50. Based on this analysis, the best training-validation split and best descriptor were chosen.

A one-parameter QSAR model was then built using the training data and the chosen descriptor. Its performance was then assessed on both the training data as well as the validation data. The statistics of the model were computed using the Analysis ToolPak available in Microsoft Excel.

Developing and evaluating a two-parameter QSAR model

To develop a two-parameter QSAR model, the top 2 descriptors for all training-validation splits were found. The split with the best performance was chosen for model building.

A two-parameter QSAR model was built using the training data and the two chosen descriptors. Its performance was then assessed on both the training data as well as the validation data. The statistics of the model were computed using the Analysis ToolPak available in Microsoft Excel.

Computing the ADME properties of the identified inhibitors

The SwissADME webserver [9] was used to compute the ADME properties of the 13 LRRK2 inhibitors. The SMILES notation for the drugs was given as input to the server. The computed ADME properties were downloaded in .csv format.

Results

1538 PaDEL descriptors for each of the drugs were obtained using the ChemDes webserver.

To build a one-parameter QSAR model, it was found that using 'GATS5i' as the descriptor and drugs #2, #5, and #10 as the validation set gave the best performance. The linear regression fit for the training data was plotted.

The one-parameter QSAR model obtained using this was given by:

$$\boldsymbol{y = -76.51x + 86.771}$$

Here, y = IC50 and x = GATS5i.

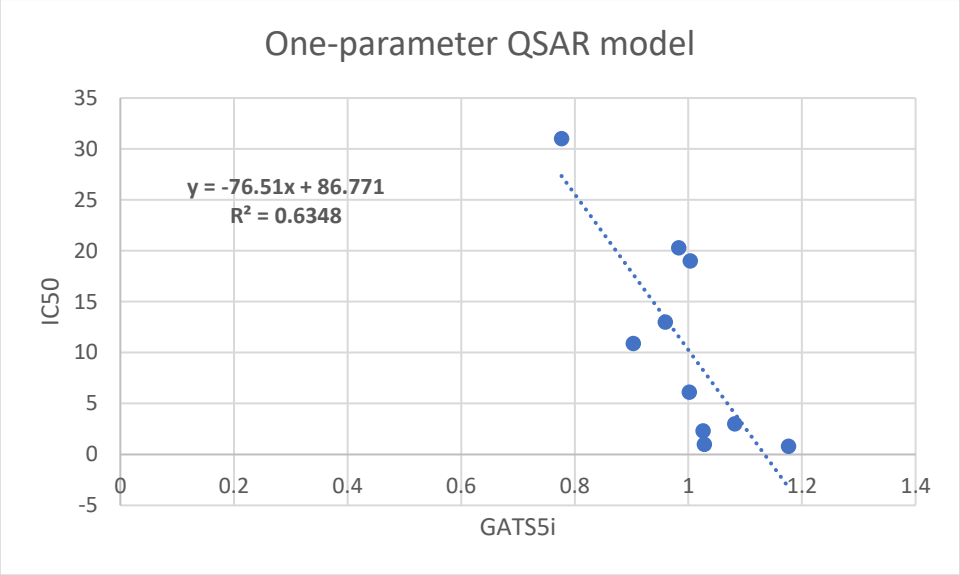


Figure 3: The one-parameter QSAR model

The statistics of the model are shown below:

Table 3a: Regression statistics of the one-parameter QSAR model

Regression Statistics	
Multiple R	0.796731834
R Square	0.634781615
Adjusted R Square	0.589129317
Standard Error	6.486255815
Observations	10

Table 3b: ANOVA statistics of the one-parameter QSAR model

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	584.991884	584.991884	13.90470229	0.005796877
Residual	8	336.572116	42.0715145		
Total	9	921.564			

Table 3c: Coefficient statistics of the one-parameter QSAR model

	Coefficients	Standard Error	t Stat	P-value
Intercept	86.77059316	20.49245448	4.234270386	0.002859683
GATS5i	-76.51031282	20.51819393	-3.72890095	0.005796877

To build a two-parameter QSAR model, it was found that using 'GATS5i' and 'MATS5c' as the descriptors and drugs #1, #2, and #8 as the validation set gave the best performance. The linear regression fit for the training data was plotted for each of the chosen descriptors.

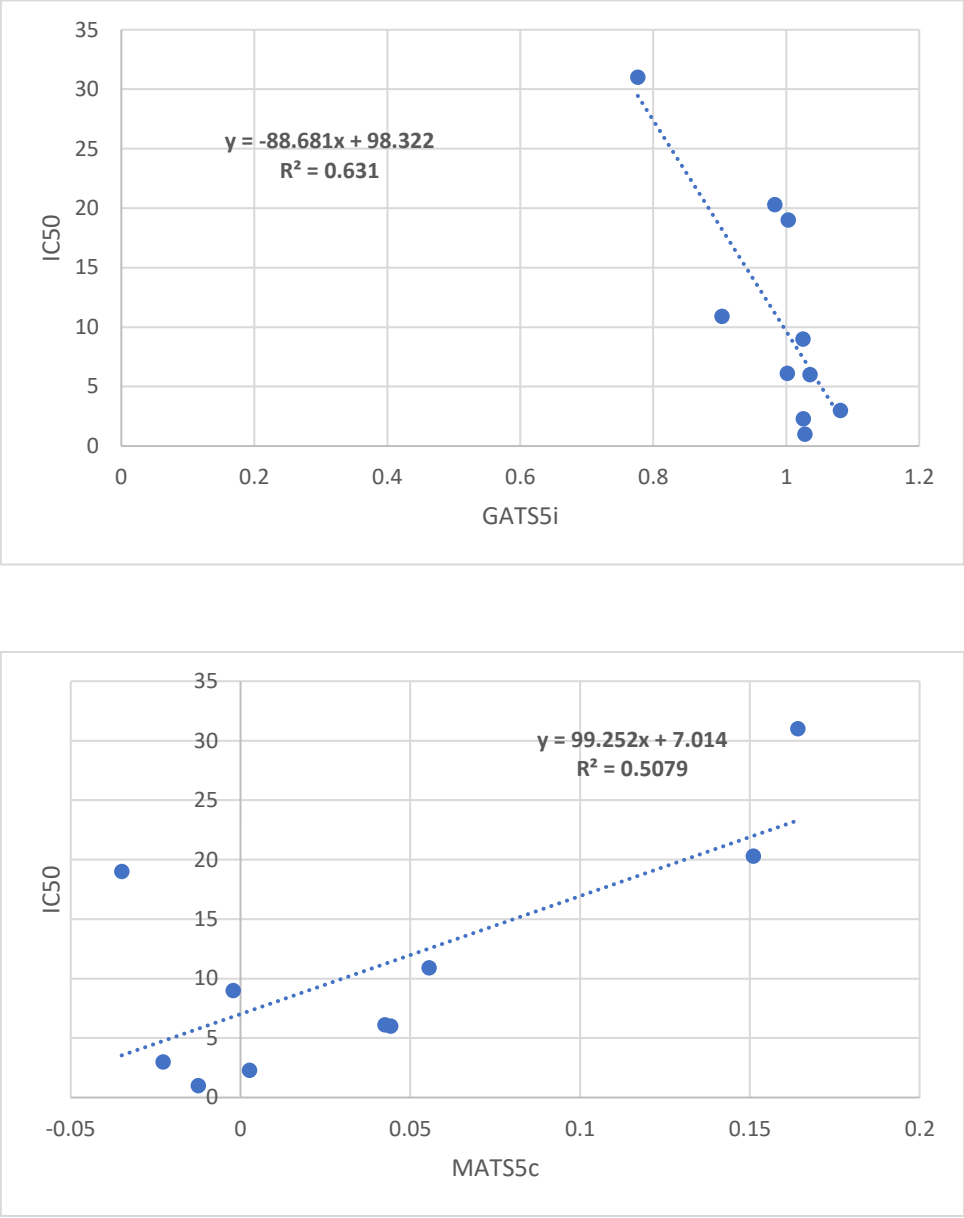


Figure 4: The two-parameter QSAR model

The two-parameter QSAR model obtained using these descriptors was given by:

$$\boldsymbol{y = -65.768x_1 + 38.778x_2 + 74.222}$$

Here, y = IC50, x_1 = GATS5i, and x_2 = MATS5c.

The statistics of the model are shown below:

Table 4a: Regression statistics of the two-parameter QSAR model

Regression Statistics	
Multiple R	0.816339458
R Square	0.666410111
Adjusted R Square	0.571098714
Standard Error	6.338291914
Observations	10

Table 4b: ANOVA statistics of the two-parameter QSAR model

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	561.7863893	280.8931946	6.991924713	0.021441002
Residual	7	281.2176107	40.17394439		
Total	9	843.004			

Table 4c: Coefficient statistics of the two-parameter QSAR model

	Coefficients	Standard Error	t Stat	P-value
Intercept	74.22217313	36.92487909	2.010085746	0.084355095
GATS5i	-65.76847412	36.06219438	-1.823751307	0.110956947
MATS5c	38.77786224	44.98684436	0.861982271	0.417245665

The ADME properties of the drugs obtained from SwissADME are tabulated below:

Table 5: ADME properties of the selected drugs

Sl. No.	1	2	3	4	5	6	7
Log P	2.96	2.07	1.36	2.36	3.1	3.97	2.02
Solubility class	Moderate	Soluble	Soluble	Soluble	Moderate	Moderate	Soluble
GI absorption	High	High	High	High	High	High	High
BBB permeant	Yes	No	No	No	No	Yes	No
Pgp substrate	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CYP1A2 inhibitor	Yes	Yes	Yes	Yes	Yes	Yes	No
CYP2C19 inhibitor	Yes	No	No	No	Yes	Yes	No
CYP2C9 inhibitor	Yes	Yes	No	No	Yes	Yes	Yes

CYP2D6 inhibitor	Yes	Yes	Yes	No	Yes	Yes	Yes
CYP3A4 inhibitor	Yes	Yes	Yes	No	Yes	Yes	Yes
log Kp (cm/s)	-6.35	-6.68	-7.36	-6.62	-7.02	-5.8	-7.13
Lipinski #violations	0	0	0	0	0	0	0
Bioavailability	0.55	0.55	0.55	0.55	0.55	0.55	0.55

Sl. No.	8	9	10	11	12	13
Log P	2.67	3.09	2.27	4.29	2.72	4.3
Solubility class	Moderate	Moderate	Soluble	Poor	Moderate	Moderate
GI absorption	High	High	High	High	High	High
BBB permeant	No	Yes	No	No	No	Yes
Pgp substrate	Yes	Yes	Yes	Yes	Yes	Yes
CYP1A2 inhibitor	No	No	No	No	No	No
CYP2C19 inhibitor	Yes	Yes	No	Yes	No	Yes
CYP2C9 inhibitor	Yes	No	Yes	No	No	Yes
CYP2D6 inhibitor	Yes	Yes	Yes	Yes	Yes	Yes
CYP3A4 inhibitor	Yes	Yes	Yes	No	Yes	No
log Kp (cm/s)	-7.64	-6.85	-7.2	-6.02	-7.06	-6.63
Lipinski #violations	2	0	0	1	0	1
Bioavailability	0.17	0.55	0.55	0.55	0.55	0.55

Most of the drugs have good absorption properties, demonstrating good solubility and GI absorption. Except drug #8, all the drugs have good distribution properties in terms of bioavailability. All the drugs show Pgp efflux. The drugs have varying metabolism and excretion properties since they inhibit cytochromes differently.

Only 4 of the drugs can permeate through the blood-brain barrier, which is a crucial property for a LRRK2 inhibitor. Thus, preference should be given to those drugs that can pass through the blood-brain barrier.

Discussion

The one-parameter QSAR model achieves a decently high R^2 of 0.635. This model achieves a root mean square error (RMSE) of 5.801484 on the training set and 1.082830 on the validation set. Its F-statistic is sufficiently larger than that required for significance.

The two-parameter QSAR model achieves a decently high R^2 of 0.666, which is slightly better than then one-parameter model. This model achieves a root mean square error (RMSE) of 5.302995 on the training set and 2.42394 on the validation set. Its F-statistic is larger than that required for significance.

The descriptors found to have the best ability to predict the IC50 of the LRRK2 inhibitors were GATS5i and MATS5c. GATS5i is the Geary autocorrelation of lag 5 weighted by ionization potential and MATS5c is Moran coefficient of lag 5 weighted by gasteiger charge [10].

GATS5i calculates the autocorrelation of ionization potential along the molecular graph with a lag of 5, i.e., it is considering atoms that are 5 bonds apart. Since this descriptor is based on ionization potential, it suggests that electronic properties at specific parts of the LRRK2 inhibitors (particularly atoms 5 bonds apart) may play a key role in their binding or activity.

Similarly, MATS5c uses Moran's index to measure how the atomic charge distribution varies along the molecular graph. The "lag 5" again refers to interactions between atoms that are 5 bonds apart. Charge distribution influences how the drug might engage in electrostatic interactions with LRRK2.

The selected drugs have varying ADME properties. Only a few of the chosen drugs are able to penetrate the blood-brain barrier, which is a crucial property for a LRRK2 inhibitor to have in order to be an effective Parkinson's treatment.

Conclusion

Using PaDEL descriptors, one-parameter and two-parameter QSAR models were built to predict the IC50 of selected LRRK2 inhibitors. Both models achieved an $R^2 > 0.6$, indicating good performance.

Both the chosen descriptors (GATS5i and MATS5c) point to the importance of electronic and charge distribution over a molecular scale of 5 bonds apart for optimal LRRK2 inhibitor activity.

The ADME properties of the drugs revealed good absorption properties. However, not all the drugs were found to penetrate the blood-brain barrier as required.

References

[1] Mhyre, T. R., Boyd, J. T., Hamill, R. W., & Maguire-Zeiss, K. A. (2012). Parkinson's Disease. Sub-cellular Biochemistry/Subcellular Biochemistry, 389–455. https://doi.org/10.1007/978-94-007-5416-4_16

[2] World Health Organization (WHO). (2023). Parkinson disease. https://www.who.int/news-room/fact-sheets/detail/parkinson-disease
[3] Monfrini, E., & Di Fonzo, A. (2017). Leucine-Rich Repeat kinase (LRRK2) Genetics and Parkinson's Disease. Advances in Neurobiology, 3–30. https://doi.org/10.1007/978-3-319-49969-7_1
[4] Cookson, M. R. (2016). Cellular functions of LRRK2 implicate vesicular trafficking pathways in Parkinson's disease. Biochemical Society Transactions, 44(6), 1603–1610. https://doi.org/10.1042/bst20160228
[5] Di Maio, R., Hoffman, E. K., Rocha, E. M., Keeney, M. T., Sanders, L. H., De Miranda, B. R., Zharikov, A., Van Laar, A., Stepan, A. F., Lanz, T. A., Kofler, J. K., Burton, E. A., Alessi, D. R., Hastings, T. G., & Greenamyre, J. T. (2018). LRRK2 activation in idiopathic Parkinson's disease. Science Translational Medicine, 10(451). https://doi.org/10.1126/scitranslmed.aar5429
[6] Wojewska, D. N., & Kortholt, A. (2021). LRRK2 Targeting Strategies as potential treatment of Parkinson's Disease. Biomolecules, 11(8), 1101. https://doi.org/10.3390/biom11081101
[7] Hu, J., Zhang, D., Tian, K., Ren, C., Li, H., Lin, C., Huang, X., Liu, J., Mao, W., & Zhang, J. (2023). Small-molecule LRRK2 inhibitors for PD therapy: Current achievements and future perspectives. European Journal of Medicinal Chemistry, 256, 115475. https://doi.org/10.1016/j.ejmech.2023.115475
[8] CBDD Group. ChemDes – Molecular descriptors computing platform. http://www.scbdd.com/chemdes/
[9] SwissADME. http://www.swissadme.ch/
[10] Descriptor List — mordred 1.2.1a1 documentation. https://mordred-descriptor.github.io/documentation/master/descriptors.html

Supplementary Files

All the data generated during this project can be found at this link:

https://docs.google.com/spreadsheets/d/1Ufg2wF9caJRtRsLgZoqcyS9DimRyTICO/edit?usp=sharing&ouid=117158162975771603067&rtpof=true&sd=true