Supporting Information for

Effects of Mutations and Post-Translational Modifications on α-Synuclein In Vitro Aggregation

Samantha X. Pancoe, Yanxin J. Wang, Marie Shimogawa, Ryann M. Perez, Sam Giannakoulias, and E. James Petersson*

Department of Chemistry; University of Pennsylvania; 231 South 34th Street; Philadelphia, Pennsylvania 19104, USA

*Email: ejpetersson@sas.upenn.edu

Analysis of Mutational Effects

α-Synuclein (αS) mutations were encoded into bit vectors of length 19, where the first three bits correspond to the position of the mutation within the sequence of αS (N-terminus, NAC, and C-terminus, defined below). The following sixteen bits correspond to the type of mutation. The exact details of these encodings can be found in Tables S1-S2. Using the assigned class labels associated with each mutation, simple correlation analysis was performed by averaging these labels to determine the effect on each type of mutation (Table S3). These analyses were further segmented by the three domains of αS to elucidate rough positional preferences for the varying types of mutations studied (Tables S4-S6).

To further group and visualize the effects of these mutations, the αS mutation bit vectors were then clustered using the unsupervised machine learning KMeans partitioning algorithm. Clustering was performed into k=3 clusters to broadly represent the possible phenotypes of no effect on aggregation, reduction in aggregation rate, and acceleration of aggregation rate. Once cluster labels were assigned, the bit vectors were projected into two dimensions using principal component analysis (PCA) for visualization (Figure 11, main text). Both KMeans clustering and PCA were performed using the scikit-learn python library while visualization utilized matplotlib [1,2]. While it is evident that these simple metrics demonstrate some preferential segregation of αS aggregation phenotypes, more sophisticated machine learning approaches will be required for effective prediction.

[1] Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12: 2825-30.

[2] Hunter, JD Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering.* 2007;9:90-95.

Table S1. Amino Acid Classification

Type	Members
Positive Charge (+)	K, R, H
Negative Charge (–)	D, E
Nonpolar (nP)	A, V, L, I, M, W, F
Polar (P)	C, S, Y, T, N, Q, K, R, H, D, E, Y
Small (Sm)	G, A, S, C
Large (Lg)	R, N, D, E, Q, H, I, L, K, M, F, T, W, Y, V
Bulky (B)	V, T, Y, F, W, I, L, H
Non-bulky (nB)	A, R, N, D, C, E, Q, G, K, M, S
β-Branched (β)	T, V, I
Non- β-Branched (nβ)	A, R, N, D, C, E, Q, H, L, K, M, F, S, W, Y
Proline (Pro)	P
Glycine (Gly)	G
Proline/Glycine (PG)	P, G
All others (X)	All except noted amino acid

Table S2. Mutation Encoding

Bit	Feature
0	N-Terminus: Position 1-60
1	NAC: Position 61-95
2	C-Terminus: Position 96-140
3	+ → –
4	– → +
5	nP → P
6	P → nP
7	Sm → Lg
8	Lg → Sm
9	nB → B
10	B → nB
11	nβ → β
12	β → nβ
13	Pro → X
14	X → Pro
15	Gly → X
16	X → Gly
17	PG → X
18	X → PG

Table S3. Aggregation Scores Averaged Over Entire Protein for Mutation Types

Type	Average	SD	Support
+ → –	-0.500	0.866	4
– → +	1.000	0.000	6
nP → P	-0.300	0.900	20
P → nP	-0.368	0.666	19
Sm → Lg	-0.125	0.857	16
Lg → Sm	-0.429	0.660	21
nB → B	0.000	0.913	12
B → nB	-0.633	0.706	30
nβ → β	0.429	0.904	7
β → nβ	-0.588	0.600	17
Pro → X	1.000	0.000	5
X → Pro	-0.879	0.326	33
Gly → X	-0.286	0.700	7
X → Gly	-0.750	0.433	4
PG → X	0.250	0.829	12
X → PG	-0.853	0.354	34
+ Cntrl	-0.500	0.500	2
– Cntrl			0
nP Cntrl	-0.333	0.667	9
P Cntrl	-0.107	0.900	28
Sm Cntrl	0.600	0.490	5
Lg Cntrl	-0.356	0.821	45
nB Cntrl	0.000	0.771	37
B Cntrl	-0.625	0.484	8
nβ Cntrl	-0.216	0.824	51
β Cntrl	-1.000	0.000	1

Table S4. Aggregation Scores Averaged Over N-Terminus for Mutation Types

Type	Average	SD	Support
+ → –	-0.500	0.866	4
– → +	1.000	0.000	4
nP → P	0.111	0.994	9
P → nP	0.000	1.000	2
Sm → Lg	0.200	0.872	10
Lg → Sm	0.000	1.000	2
nB → B	0.667	0.745	6
B → nB	-0.143	0.990	7
nβ → β	1.000	0.000	5
β → nβ	-0.667	0.471	3
Pro → X			0
X → Pro	-0.813	0.390	16
Gly → X	-0.333	0.471	3
X → Gly			0
PG → X	-0.333	0.471	3
X → PG	-0.813	0.390	16
+ Cntrl	-0.500	0.500	2
– Cntrl			0
nP Cntrl	0.000	0.816	3
P Cntrl	0.118	0.900	17
Sm Cntrl	1.000	0.000	1
Lg Cntrl	-0.048	0.898	21
nB Cntrl	-0.050	0.865	20
B Cntrl	0.000	0.000	1
nβ Cntrl	0.000	0.933	23
β Cntrl			0

Table S5. Aggregation Scores Averaged Over NAC for Mutation Types

Type	Average	SD	Support
+ → –			0
– → +	1.000	0.000	2
nP → P	-0.800	0.400	10
P → nP	0.000	1.000	2
Sm → Lg	-1.000	0.000	4
Lg → Sm	-0.571	0.728	7
nB → B	-1.000	0.000	2
B → nB	-0.688	0.583	16
nβ → β	-1.000	0.000	2
β → nβ	-0.571	0.623	14
Pro → X			0
X → Pro	-0.941	0.235	17
Gly → X	-0.250	0.829	4
X → Gly	-0.750	0.433	4
PG → X	-0.250	0.829	4
X → PG	-0.889	0.314	18
+ Cntrl			0
– Cntrl			0
nP Cntrl	-0.500	0.500	2
P Cntrl	-0.167	0.898	6
Sm Cntrl	0.333	0.471	3
Lg Cntrl	-0.500	0.732	14
nB Cntrl	0.000	0.866	8
B Cntrl	-0.500	0.500	2
nβ Cntrl	0.333	0.943	3
β Cntrl	-1.000	0.000	1

Table S6. Aggregation Scores Averaged Over C-Terminus for Mutation Types

Type	Average	SD	Support
+ → –			0
– → +			0
nP → P	1.000	0.000	1
P → nP	-0.467	0.499	15
Sm → Lg	0.000	0.000	2
Lg → Sm	-0.417	0.493	12
nB → B	-0.500	0.500	4
B → nB	-1.000	0.000	7
nβ → β			0
β → nβ			0
Pro → X	1.000	0.000	5
X → Pro			0
Gly → X			0
X → Gly			0
PG → X	1.000	0.000	5
X → PG			0
+ Cntrl			0
– Cntrl			0
nP Cntrl	-0.500	0.500	4
P Cntrl	-0.800	0.400	5
Sm Cntrl	1.000	0.000	1
Lg Cntrl	-0.800	0.400	10
nB Cntrl	0.111	0.314	9
B Cntrl	-0.800	0.400	5
nβ Cntrl	-0.480	0.574	25
β Cntrl			0

Analysis of Aggregation Conditions

It is widely known that aggregation conditions can significantly influence fibril formation kinetics and the assortment of fibril cryo-EM structures published in recent years make it clear that this also influences fibril structure (see Figure 8 and accompanying discussion). Given the scope of our data set, we took the opportunity to analyze the effects of buffer, salt, shaking speed, and αS concentration on aggregation rate. We did this only among the familial mutations, where there are multiple experiments to compare for the same mutation, performed in different laboratories. Specifically, we categorized the conditions in the following ways:

Buffer – Tris, PBS, or Other

Salt – High, Low, or None (Other not included)

Shaking Speed – Fast, Medium, Slow, or None (Other not included)

αS Concentration – High, Normal, or Low (N/A not included)

We removed the Other and N/A categories as noted since these could not easily be interpreted. For each familial mutant, then aggregation scores were averaged within these categories using a simple NumPy program [3]. Those data are shown in Tables S7, S8, S9, and S10. To examine the effects of switching from one condition to another for a given mutant and between two mutants, we created the heat maps (Figures S1, S2, S3, and S4) that accompany each table in seaborn [4]. The heat map color code indicates whether the change indicated in a given box (from the condition on the abscissa to the condition on the ordinate) results in an increase (red) or decrease (blue) in the aggregation score. The diagonal boxes have been colored black since these are identity operations.

Since the mutational effects are already normalized to the corresponding WT aggregation, our analysis shows that a mutational effect can be exaggerated or decreased in a different buffer. For example, while $A_{30}P$ has a score of -0.75 in PBS, it has a score of -0.31 in Tris, so the slowing of aggregation is muted somewhat in Tris buffer, with a net increase of +0.44. However, if one attempts to generalize this by averaging across all PBS-to-Tris changes, one obtains an average value of -0.12 with a standard deviation of 0.31; the variation is larger than the effect size. This shows that there is no generalizable PBS-to-Tris effect. More generally, when one examines changes to aggregation conditions for a given mutant (within boxed regions near the diagonal on the heat maps) one can see that the effects are typically smaller (colors less intense) than changes between mutants. Collectively, we interpret these data to show that while one must be careful to consider the effects of buffer and have a matched WT experiment to use in normalization, one can safely compare normalized aggregation scores across a set of mutants.

[3] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. Array programming with NumPy. *Nature.* 2020;585:357-62.

[4] seaborn: https://doi.org/10.5281/zenodo.883859

Table S7. Aggregation Scores Averaged Over Each Familial Mutant and Buffer Condition

Type	Average	Support
A30P PBS	-0.75	4
A30P Tris	-0.31	13
A30P Other	-0.67	6
E46K PBS	1.00	3
E46K Tris	0.75	8
E46K Other	0.60	5
H50Q PBS	1.00	1
H50Q Tris	0.83	6
H50Q Other	0.80	5
G51D_PBS	0.00	0
G51D Tris	-0.20	5
G51D Other	-1.00	4
A53E PBS	0.00	0
A53E Tris	-0.50	2
A53E Other	-1.00	3
A53T_PBS	1.00	5
A53T Tris	0.94	16
A53T Other	0.86	7

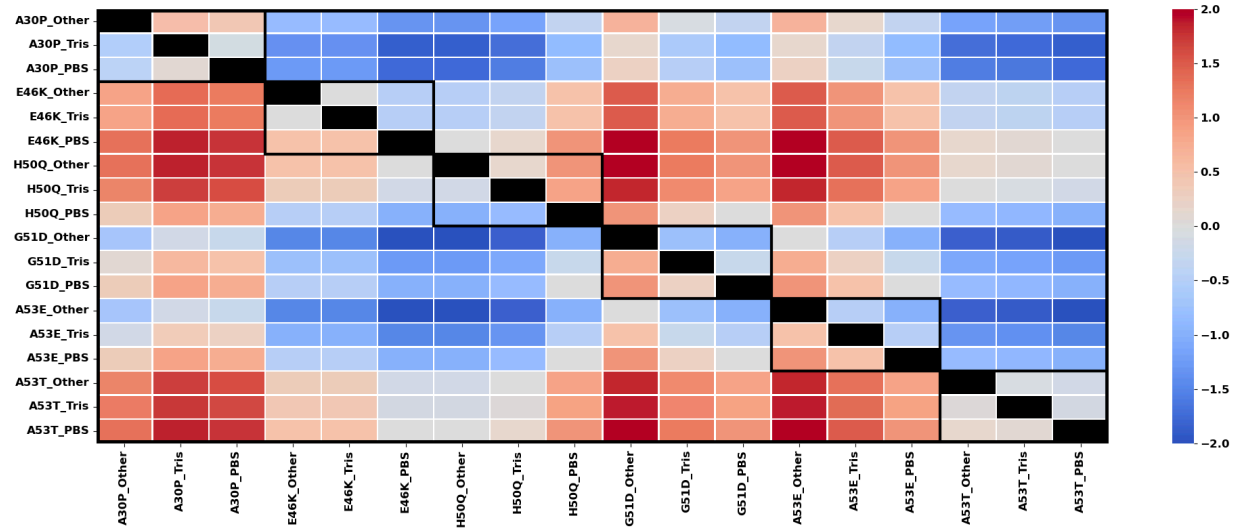

Figure S1. Effects of Buffer Conditions on Aggregation Score

Table S8. Aggregation Scores Averaged Over Each Familial Mutant and Salt Condition

Type	Average	Support
A30P High	-0.58	12
A30P Low	0.67	3
A30P None	-0.75	8
E46K High	1.00	4
E46K Low	1.00	3
E46K None	0.56	9
H50Q High	0.80	5
H50Q Low	0.00	0
H50Q None	1.00	4
G51D High	-0.33	3
G51D Low	0.00	0
G51D None	-0.33	3
A53E High	0.00	0
A53E Low	0.00	0
A53E None	-0.80	5
A53T High	0.85	13
A53T Low	1.00	4
A53T None	1.00	11

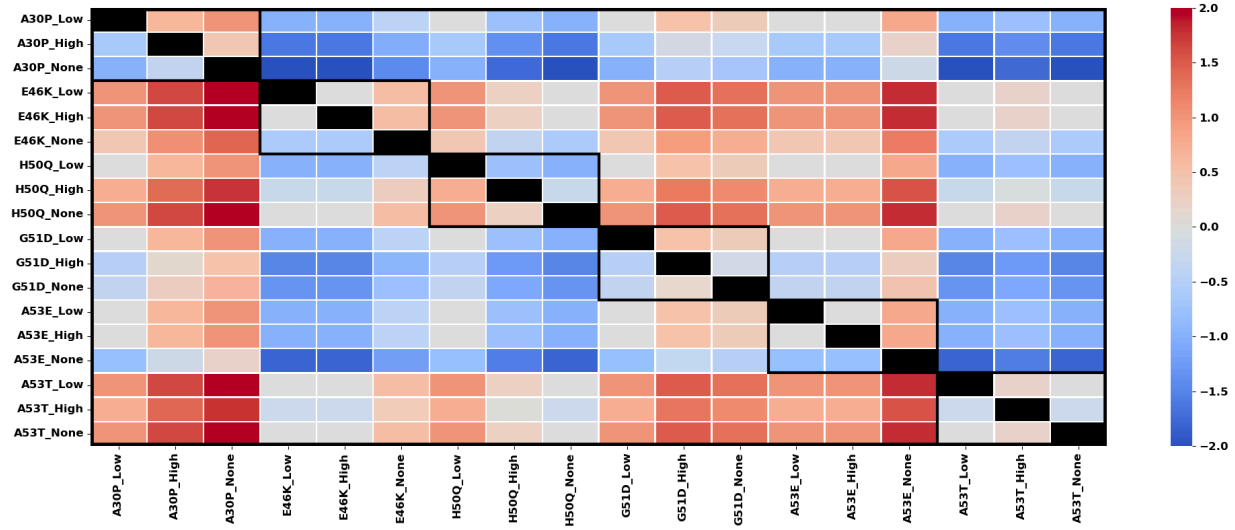

Figure S2. Effects of Salt Conditions on Aggregation Score

Table S9. Aggregation Scores Averaged Over Each Familial Mutant and Shaking Speed

Type	Average	Support
A30P Fast	-0.50	2
A30P Medium	-0.50	2
A30P Slow	-0.75	8
A30P None	-1.00	2
E46K Fast	0.00	0
E46K Medium	0.00	4
E46K Slow	1.00	7
E46K None	0.00	0
H50Q Fast	0.71	7
H50Q Medium	1.00	1
H50Q Slow	1.00	3
H50Q None	0.00	0
G51D Fast	-1.00	4
G51D Medium	0.50	2
G51D Slow	-1.00	2
G51D None	0.00	0
A53E Fast	-1.00	1
A53E Medium	0.00	1
A53E Slow	-1.00	3
A53E None	0.00	0
A53T Fast	1.00	1
A53T Medium	1.00	4
A53T Slow	0.83	12
A53T None	1.00	2

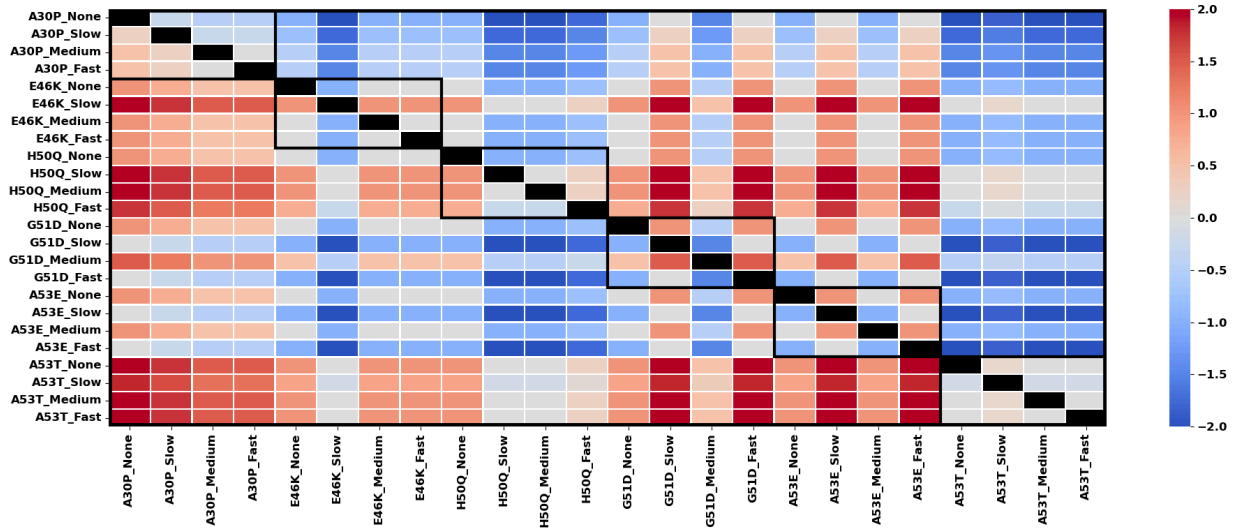

Figure S3. Effects of Shaking Speed on Aggregation Score

Table S10. Aggregation Scores Averaged Over Each Familial Mutant and αS Concentration

Type	Average	Support
A30P High	-0.86	7
A30P Normal	-0.40	5
A30P Low	-1.00	2
E46K High	1.00	5
E46K Normal	0.20	5
E46K Low	1.00	1
H50Q High	1.00	1
H50Q Normal	1.00	2
H50Q Low	0.80	5
G51D High	-1.00	1
G51D Normal	0.50	2
G51D Low	-1.00	2
A53E Normal	-0.50	2
A53E Low	-1.00	1
A53E High	-1.00	2
A53T Normal	0.75	8
A53T Low	1.00	2

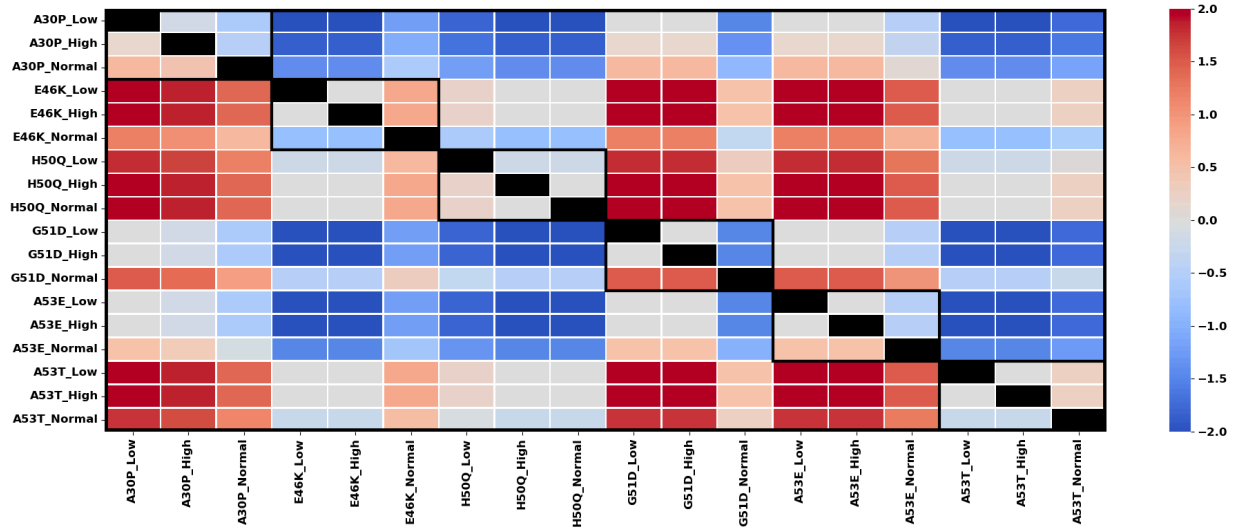

Figure S4. Effects of αS Concentration on Aggregation Score