Supporting Information

Alpha Synuclein Fibrils Contain Multiple Binding Sites for Small Molecules

Chia-Ju Hsieh,[1†] John J. Ferrie,[2†] Kuiying Xu,[1] Iljung Lee,[1] Thomas J. A. Graham,[1] Zhude Tu,[3] Jennifer Yu,[4] Dhruva Dhavale,[4] Paul Kotzbauer,[4] E. James Petersson,[2] and Robert H. Mach[1*]

[1]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, ; [2]Chemistry Department, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; [3]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO, 63110; [4]Department of Neurology, Washington University School of Medicine, St. Louis, MO, 63110.

Molecular Docking

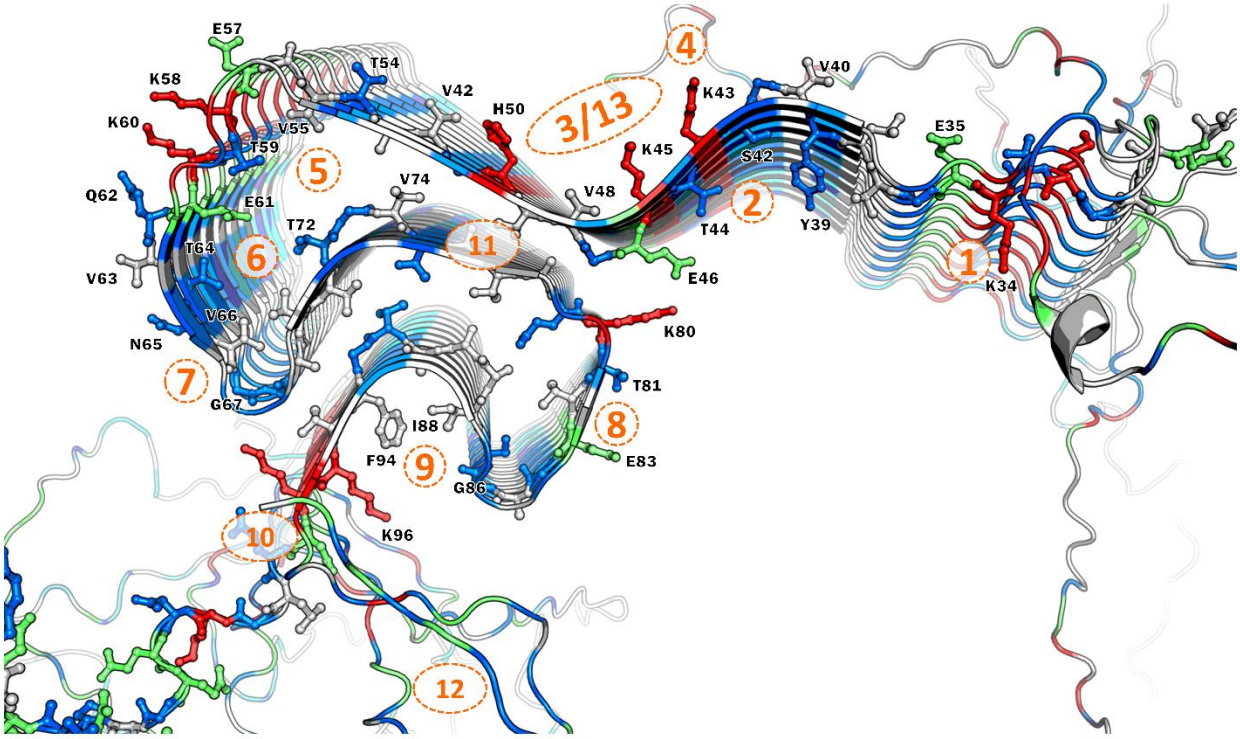


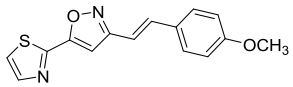
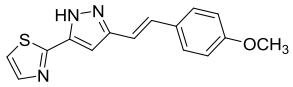
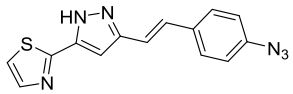
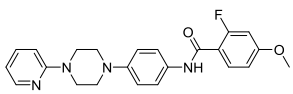
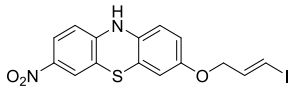
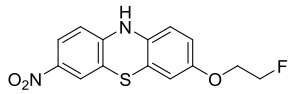
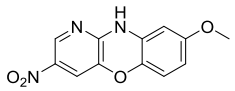
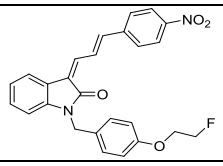
Figure S1. Thirteen putative binding sites in solid-state NMR Asyn structure revealed by molecular blind docking study. Site 1 is located on the two sides of K34. Site 2 is formed by three polar residues Y39-S42-T44. Site 3 is composed of two positive charge and one hydrophobic side chains, L45-V48-H50. Site 4 is located in the N-terminal region of chain C that folding back toward to the NAC region, made a small binding pocket between the N-terminal region and V40 to K45. Site 5 and site 6 are located in the tunnel of Greek key structure, Site 5 is formed by V55-T59-T72-G73, and Site 6 is formed by E61-T64-V69-T72. Site 7 is composed of two polar residue N65-G67 and the backbone of V66. Site 8 is located on the two sides of negative charge residue E83. Site 9 is formed by a polar, two hydrophobic, and one positive charge side chains, which is G86-I88-F94-K96. Site 10 is located at disordered C-terminal area, which formed in between chains at residues of Q99-N103. Site 11 is on the edge of the fibril. Site 12 is located at the highly disordered C-terminal region that formed by Y133 to E139 of chain D and G106 to E110 of chain F. Site 13 mostly is formed by positive charge side chains, located at two sides of K45, which is contains K43-K45 or K43-K45-V48-H50. Since site 3 is contained within site 13, these two sites can collectively be termed site 3/13. (Red, positive charge side chains; Green, negative charge side chains; blue, polar side chains; Gray, hydrophobic side chains)

Table S1. Percentage probability of thirteen putative binding sites of 8 test compounds[a]

Compound	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12	Site 13
1	**29.10%**	**39.70%**	4.70%	0.70%	-	-	0.20%	-	**17.70%**	0.20%	-	-	7.70%
2	**13.80%**	**28.30%**	5.90%	0.10%	2.80%	**22.20%**	3.00%	0.60%	**18.50%**	3.80%	-	1.00%	-
3	**14.00%**	**22.90%**	8.30%	0.60%	3.50%	**25.80%**	2.40%	0.50%	**16.50%**	5.30%	-	0.20%	-
4	9.80%	**22.80%**	4.50%	-	7.20%	**20.50%**	4.70%	0.30%	**30.20%**	-	-	-	-
5	9.40%	-	**25.20%**	0.50%	-	-	0.10%	0.20%	2.80%	-	-	-	**61.80%**
6	5.00%	-	6.40%	8.50%	-	-	0.10%	-	5.80%	-	0.60%	-	**73.60%**
7	-	-	-	0.60%	-	-	-	-	0.80%	-	0.40%	-	**98.20%**
8	4.70%	1.70%	0.80%	7.70%	-	-	0.80%	-	6.00%	0.40%	0.50%	-	**77.40%**

[a] Thirteen putative binding sites in Asyn fibril were revealed by molecular blind docking studies of 8 test compounds. Seven of thirteen putative binding sites (site 4, 5, 7, 8, 10, 11, and 12) that had % probability lower than 10% of all compounds were excluded. Then, the putative binding sites were reduced to six (site 1, 2, 3, 6, 9, and 13). Compounds in the same cluster revealed a similar distribution pattern of % probability of interaction in thirteen putative binding. The preference binding sites for stryrene-based compounds (**1**-**3**) were site 1, 2, 6 and 9. The piperazine analog (**4**) showed a preference for site 2, 6, and 9. The tricyclic analogs (**5**-**7**) and the indolinone-diene compound (**8**) showed a highly preference for site 3 and 13.

Table S2. Probability and predictive binding energy of 4 putative binding sites

		Binding Site	2	3	9	13
#	Compounds	Interaction residue	S42 - T44	K43 - H50	G86 - I88	K43 - K45/ K43 - K45- H50
1		Probability (%)	**39.70%**	4.70%	17.70%	7.70%
		Average binding energy	-6.54±0.14	-6.04±0.09	-6.49±0.13	-6.09±0.09
		Best binding energy	**-6.80**	-6.17	-6.68	-6.18
2		Probability (%)	**28.30%**	5.90%	18.50%	
		Average binding energy	-6.28±0.21	-5.74±0.16	-6.08±0.21	
		Best binding energy	**-6.78**	-6.02	-6.41	
3		Probability (%)	**22.90%**	8.30%	16.50%	
		Average binding energy	-6.49±0.23	-5.98±0.22	-6.44±0.19	
		Best binding energy	**-7.17**	-6.58	-6.84	
4		Probability (%)	22.80%	4.50%	**30.20%**	
		Average binding energy	-7.88±0.25	-7.29±0.29	-8.42±0.20	
		Best binding energy	-8.42	-7.84	**-8.82**	
5		Probability (%)		25.20%	2.80%	**61.80%**
		Average binding energy		-7.73±0.33	-7.34±0.19	-7.74±0.35
		Best binding energy		-8.20	-7.62	**-8.94**
6		Probability (%)		6.40%	5.80%	**73.60%**
		Average binding energy		-6.72±0.14	-7.44±0.18	-7.39±0.24
		Best binding energy		-6.89	**-7.77**	-7.71
7		Probability (%)			0.80%	**98.20%**
		Average binding energy			-6.59±0.17	-7.54±0.11
		Best binding energy			-6.74	**-7.72**
8		Probability (%)	1.70%	0.80%	6.00%	**77.40%**
		Average binding energy	-6.63±0.14	-6.54±0.33	-7.57±0.49	-7.71±0.72
		Best binding energy	-6.85	-7.07	-8.48	**-9.57**

Molecular Dynamics Simulation

The flexibility and stability of Asyn fibril were evaluated by root mean square fluctuation (RMSF) of each side chain (Figure S2). N- and C-terminals showed a dramatic flexibility (high order of RMSF) in the protein, indicating the variability of the potential binding pockets in these regions. The residues of site 2 (RMSF = 4.2 – 5.4 Å), 3/13 (RMSF = 4.1 – 5.1 Å), and 9 (RMSF = 5.0 – 6.0 Å) revealed lower RMSFs than other side chains in the protein, implying that the binding pockets in those areas were relative stable, and better suited for small molecular binding. Although residues V55 to G73 were ordered in the structure that imported from protein data bank, a modest flexibility with a RMSF = 5.4-7.7 Å was observed in the simulation, which is consistent with the article of the Asyn solid-state NMR structure reported these residues are disordered[1].

The solvent accessible surface area (SASA) and water molecular density of site 2, 3/13, 6, and 9 were shown in Table S3. Volumes to evaluate water molecular density of site 2, 3/13 and 9 were calculated by the side chain area that exposed to the solvent extension of 5 Å, and site 6 was calculated from the cavity of the inner code that surrounding by the interactive side chains. Site 2, 3/13, and 9 showed similar level of SASA, and relative higher degree of hydration as compared to site 6, since the SASA of these three sites were more exposed to the solvent. Site 6 had highest level of SASA, but because this site is located in the inner core of the fibril with less exposed to the solvent, showed low degree of hydration indicating relatively inaccessible for small molecular binding.

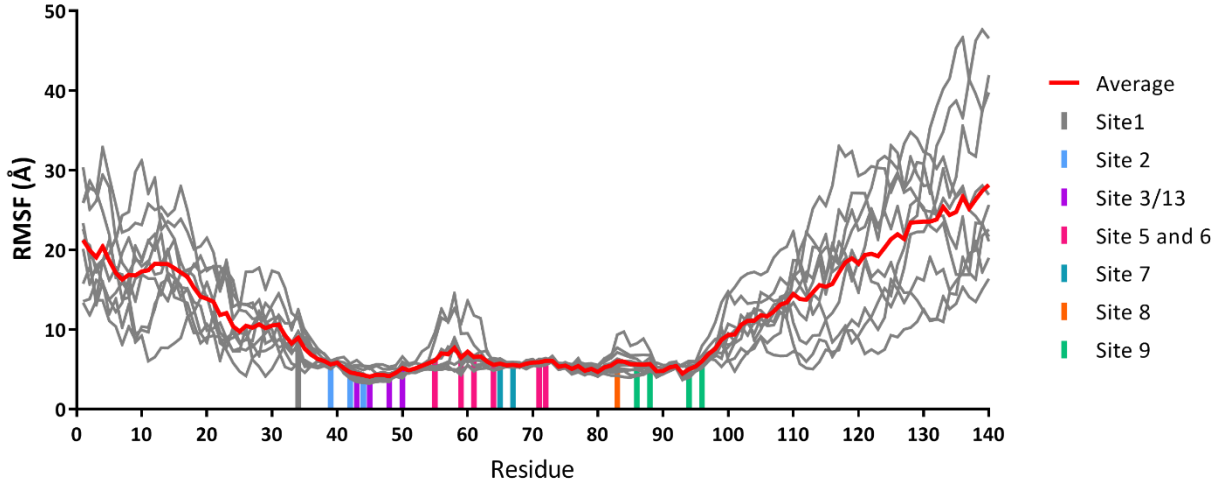


Figure S2. Root mean square fluctuation (RMSF) of each residue of Asyn over 10 ns by molecular dynamics simulation. The gray curves are RMSFs of 140 residue of 10 proteins (chain A, B, C, D, E, F, G, H, I, and J). The interaction residues were color labeled with the corresponding putative binding site. Gray: site 1; blue: site 2; purple: site 3/13; pink: site 5 and 6; cyan: site 7; orange: site 8; green: site 9.

Table S3. Solvent accessible surface area (SASA) and water molecular density of Site 2, 3/13, 6, and 9

	Site2	Site3/13	Site6	Site9
Solvent accessible surface area (SASA) ($Å^2$)	3553.65 ± 69.30	5310.87 ± 135.66	6374.90 ± 160.70	5031.30 ± 115.29
Water molecular density (#/nm^3)	4.92 ± 0.51	6.46 ± 0.58	2.64 ±0.48	3.13 ± 0.51

Photocrosslinking

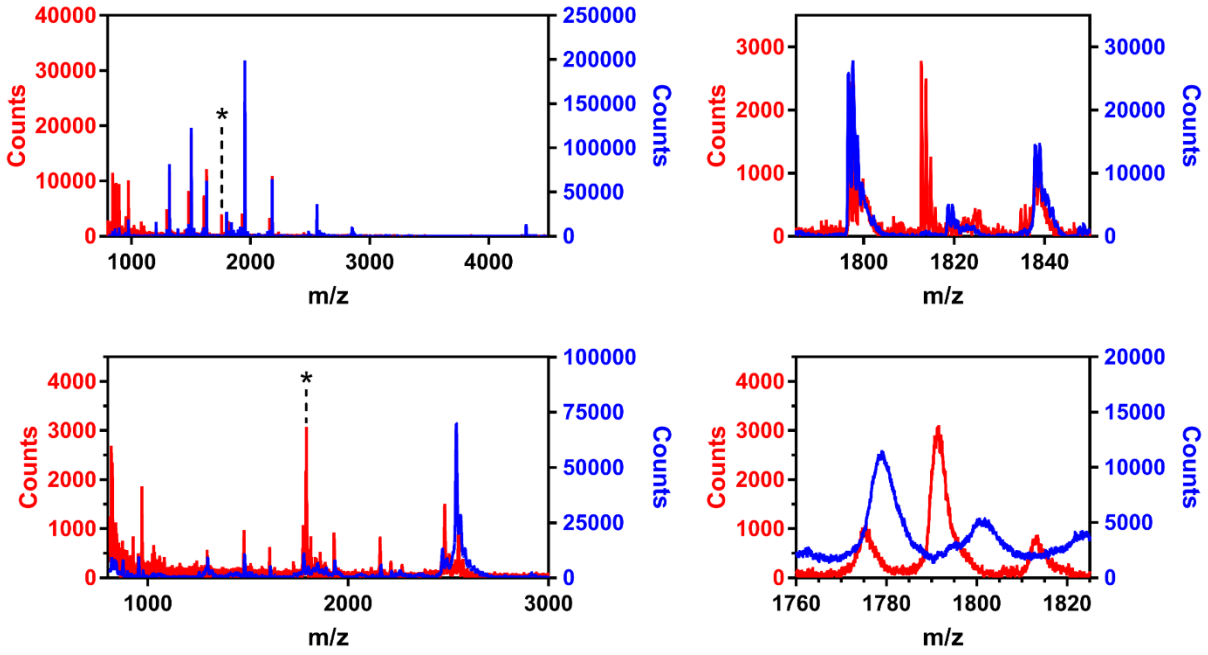


Figure S3. MALDI-MS from photocrosslinking of compound **3** to Asyn fibrils. Figures show comparison of MALDI-MS spectra of trypsin digests of Asyn when fibrils were irradiated in the presence (red) and absence (blue) of Compound **3**. Spectra were acquired using two method: RP (top) and LN (bottom). On each full spectrum (left) a dashed line marked with an asterisk indicates the peak of interest corresponding to Compound **3** covalently crosslinked to Asyn. Zoomed in spectra (right) demonstrate that the indicated peaks are unique to the sample containing Compound **3**.

Table S4. Comparison of measured and predicted values from RP MALDI-MS from photocrosslinking of compound **3** to Asyn fibrils

	Expected	- Compound 3 Observed	+ Compound 3 Observed
$Asyn^{24-32}$ + Na	852.9	-	853.0
$Asyn^{13-21}$ + Na	896.0	896.0	896.1
$Asyn^{35-43}$	952.1	952.0	952.2
$Asyn^{35-43}$ + Na	974.1	974.1	974.2
$Asyn^{24-34}$	1082.2	1082.1	-
$Asyn^{46-58}$	1296.5	1296.3	1296.5
$Asyn^{46-58}$ + Na	1318.5	1318.3	1318.5
$Asyn^{81-96}$	1479.7	1479.5	1479.7
$Asyn^{81-96}$ + Na	1501.7	1501.5	1501.7
$Asyn^{44-58 \text{ or } 46-60}$ + Na	1547.8	1547.6	-
$Asyn^{87-97}$	1607.9	1607.6	1607.9
$Asyn^{81-97}$ + Na	1629.9	1629.7	1629.9
$Asyn^{44-60}$ + Na	1755.0	-	1755.5
$Asyn^{44-58 \text{ or } 46-60}$ + Compound 3 - N_2 + Na	1813.7	-	1812.9
$Asyn^{61-80}$	1929.2	1928.9	1929.2
$Asyn^{61-80}$ + Na	1951.2	1951.0	1951.2
$Asyn^{59-80}$	2158.5	2158.2	2158.4
$Asyn^{59-80}$ + Na	2180.5	2180.3	2180.5
$Asyn^{103-140}$ + Na	4311.5	4311.5	-

Table S5. Comparison of measured and predicted values from LN MALDI-MS from photocrosslinking of compound **3** to Asyn fibrils

	Expected	- Compound 3 Observed	+ Compound 3 Observed
$Asyn^{13-21}$	874.0	876.0	-
$Asyn^{35-43}$	952.1	954.6	-
$Asyn^{46-58}$	1296.5	1299.5	-
$Asyn^{81-96}$	1479.6	1483.4	1480.0
$Asyn^{81-97}$	1607.9	1611.6	1608.5
$Asyn^{44-58 \text{ or } 46-60}$ + Compound 3 - N_2	1791.8	-	1791.4
$Asyn^{61-80}$	1929.2	1933.8	1929.8
$Asyn^{59-80}$	2158.5	2163.8	2159.5

Binding Assay

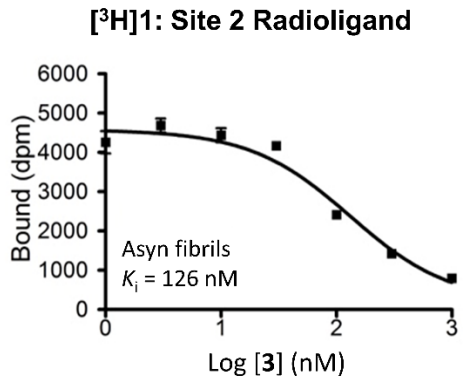


Figure S4. *In vitro* competition binding studies of compound **3** using [^{3}H]**1** as the radioligand.

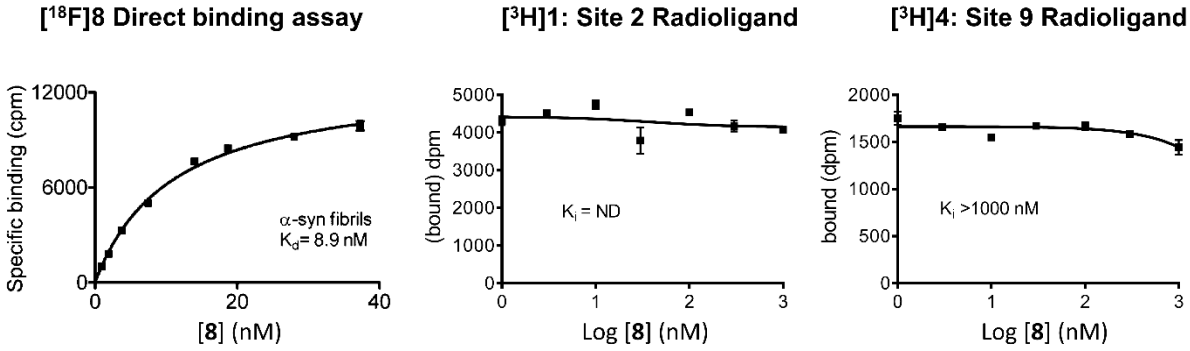


Figure S5. *In vitro* direct binding studies of compound **8** and competition binding studies using either [^{3}H]**1** or [^{3}H]**4** as the radioligand.

References

1. Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A., Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nature Structural and Molecular Biology* **2016,** *23* (5), 409.