SUPPORTING INFORMATION

Engineering Faster Transglycosidases and their Acceptor Specificity

Linh T. Tran[a,b,†,§], Vincent Blay[c,d,§,*], Sukanya Luang[e], Chatchakorn Eurtivong[f], Sunaree Choknud[a,b], Humbert González-Díaz[g,h], James R. Ketudat Cairns[a,b,g,*]

[a]School of Chemistry, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.
[b]Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.
[c]Fisher College of Business, The Ohio State University, Gerlach Hall, 2108 Neil Ave. Columbus, OH 43210, United States.
[d]Molecular Topology & Drug Design Research Unit, Departamento de Química Física, Universitat de València, Av. V. A. Estellés, s/n, 46100 Burjassot, Spain.
[e]Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand.
[f]Chulabhorn Graduate Institute, Bangkok 10210, Thailand.
[g]Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940, Leioa, Spain.
[h]IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain.
[g]Laboratory of Biochemistry, Chulabhorn Research Institute, Bangkok 10210, Thailand.

*E-mail: blayroger.1@osu.edu (V.B.); cairns@sut.ac.th (J.K.C).

[§]These authors contributed equally.

[†]Current Address: Research Institute for Interdisciplinary Science, Okayama University, Okayama, Japan.

Table S1. Oligonucleotide primers used for the site-directed mutagenesis of Os9BGlu31.

Primer name	Sequence (5'-3')
AK121679_fwd	CACCATGGCGGCGGGGATCACCAG
AK121679_rev	CTCGAGAACCTTGATCACTGGGAGTAGGCTC
I172T_fwd	GCACTGTCAATGAGCCTAACACCGAGCCGATTGGCGGATACG
I172T_rev	CGTATCCGCCAATCGGCTCGGTGT TAGGCTCATTGACAGTGC
L183Q_fwd	CGGATACGATCAAGGAATCCAACCGCCACGGCGATGCTCATTCC
L183Q_rev	GGAATGAGCATCGCCGTGGCGGTTGGATTCCTTGATCGTATCCG
L241D_fwd	GGACAAATTGGGCTCACATTGGACGGTTGGTGGTACGAGCCCG
L241D_rev	CGGGCTCGTACCACCAACCGTCCAATGTGAGCCCAATTTGTCC
W243C_fwd	GGGCTCACATTGCTCGGTTGCTGGTACGAGCCCGGGACG
W243C_rev	CGTCCCGGGCTCGTACCAGCAACCGAGCAATGTGAGCCC
W243E_fwd	GGGCTCACATTGCTCGGTTGAAGGTACGAGCCCGGGACG
W243E_rev	CGTCCCGGGCTCGTACCATTCACCGAGCAATGTGAGCCC
W243G_fwd	GGGCTCACATTGCTCGGTGGGTGGTACGAGCCCGGGACG
W243G_rev	CGTCCCGGGCTCGTACCACCCACCGAGCAATGTGAGCCC
W243H_fwd	GGGCTCACATTGCTCGGTCATTGGTACGAGCCCGGGACG
W243H_rev	CGTCCCGGGCTCGTACCAATGACCGAGCAATGTGAGCCC
W243I_fwd	GGGCTCACATTGCTCGGTATCTGGTACGAGCCCGGGACG
W243I_rev	CGTCCCGGGCTCGTACCAGATACCGAGCAATGTGAGCCC
W243K_fwd	GGGCTCACATTGCTCGGTAAGTGGTACGAGCCCGGGACG
W243K_rev	CGTCCCGGGCTCGTACCAATGACCGAGCAATGTGAGCCC
W243L_fwd	GGGCTCACATTGCTCGGTCTGTGGTACGAGCCCGGGACG
W243L_rev	CGTCCCGGGCTCGTACCACAGACCGAGCAATGTGAGCCC
W243P_fwd	GGGCTCACATTGCTCGGTCCGTGGTACGAGCCCGGGACG
W243P_rev	CGTCCCGGGCTCGTACCACGGACCGAGCAATGTGAGCCC
W243Q_fwd	GGGCTCACATTGCTCGGTCAGTGGTACGAGCCCGGGACG
W243Q_rev	CGTCCCGGGCTCGTACCACTGACCGAGCAATGTGAGCCC
W243R_fwd	GGGCTCACATTGCTCGGTAGATGGTACGAGCCCGGGACG
W243R_rev	CGTCCCGGGCTCGTACCATCTACCGAGCAATGTGAGCCC
W243S_fwd	GGGCTCACATTGCTCGGTAGCTGGTACGAGCCCGGGACG
W243S_rev	CGTCCCGGGCTCGTACCAGCTACCGAGCAATGTGAGCCC
W243T_fwd	GGGCTCACATTGCTCGGTACCTGGTACGAGCCCGGGACG
W243T_rev	CGTCCCGGGCTCGTACCAGGTACCGAGCAATGTGAGCCC
W243V_fwd	GGGCTCACATTGCTCGGTGTGTGGTACGAGCCCGGGACG
W243V_rev	CGTCCCGGGCTCGTACCACACACCGAGCAATGTGAGCCC

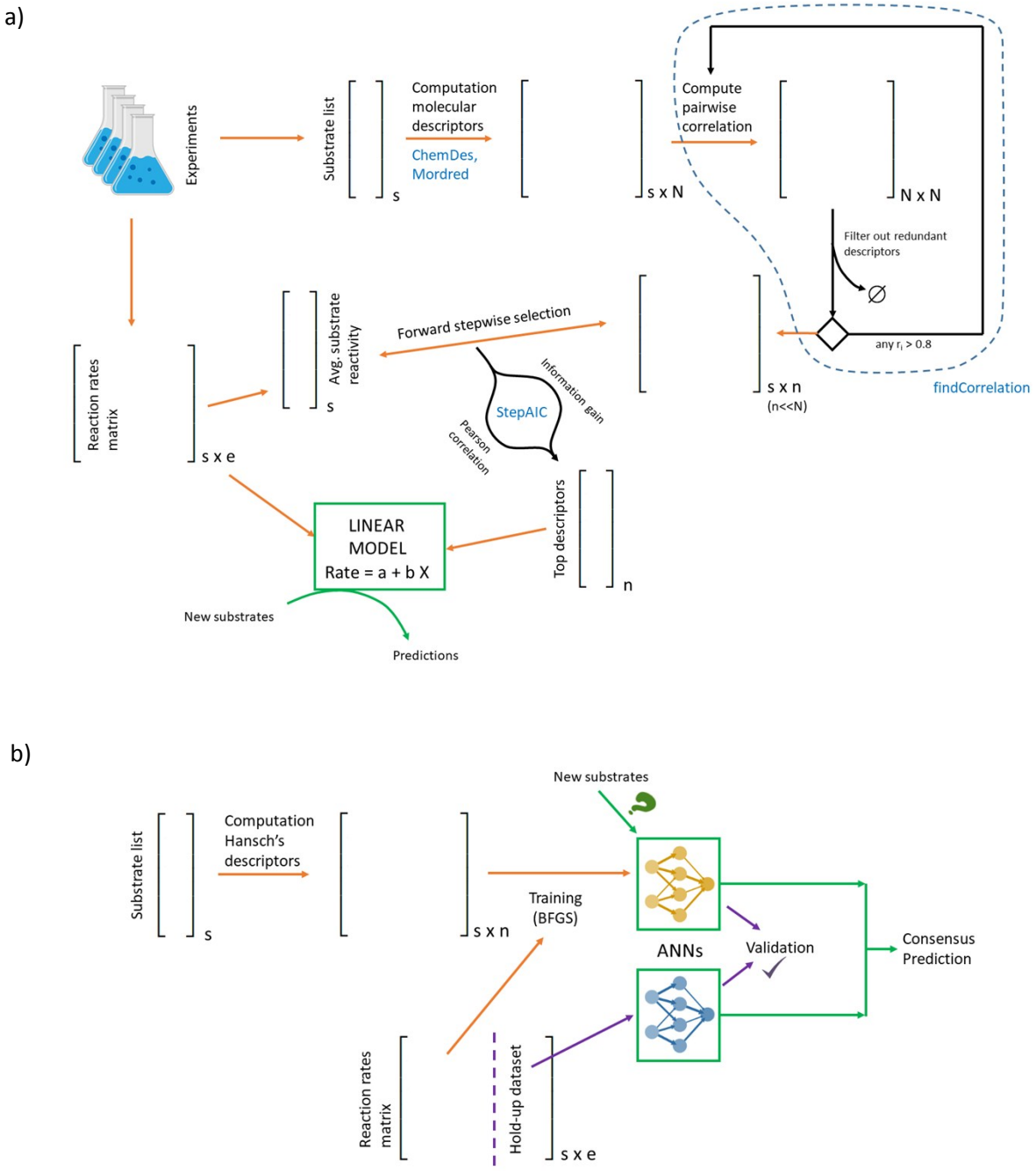

Figure S1. Flowchart of the data analysis used to build the a) linear and b) nonlinear ANN models in the present work.

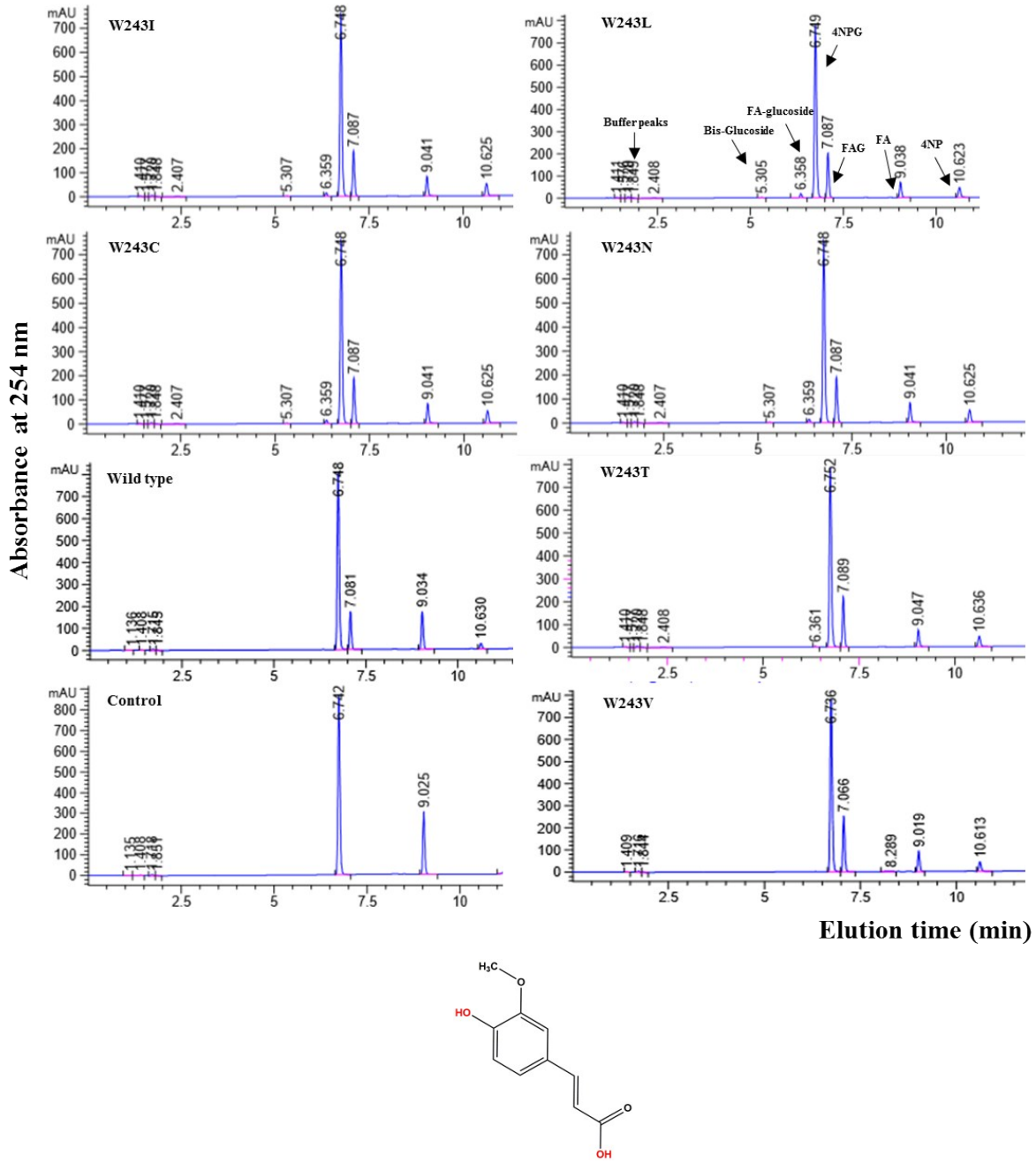

Figure S2. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243 mutants in the transfer of glucose to ferulic acid. Os9BGlu31 mutants can transfer to multiple positions to provide different glucosides. The possible glycosylation positions are marked on the structure. The peaks are labeled based on the elution time of standards for ferulic acid (FA), ferulic acid glucose ester (FAG), 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG), and the relative positions of products FA-β-D-glucoside and FA-bis-β-D-glucoside (Bis-glucoside), which identified by MS[18].

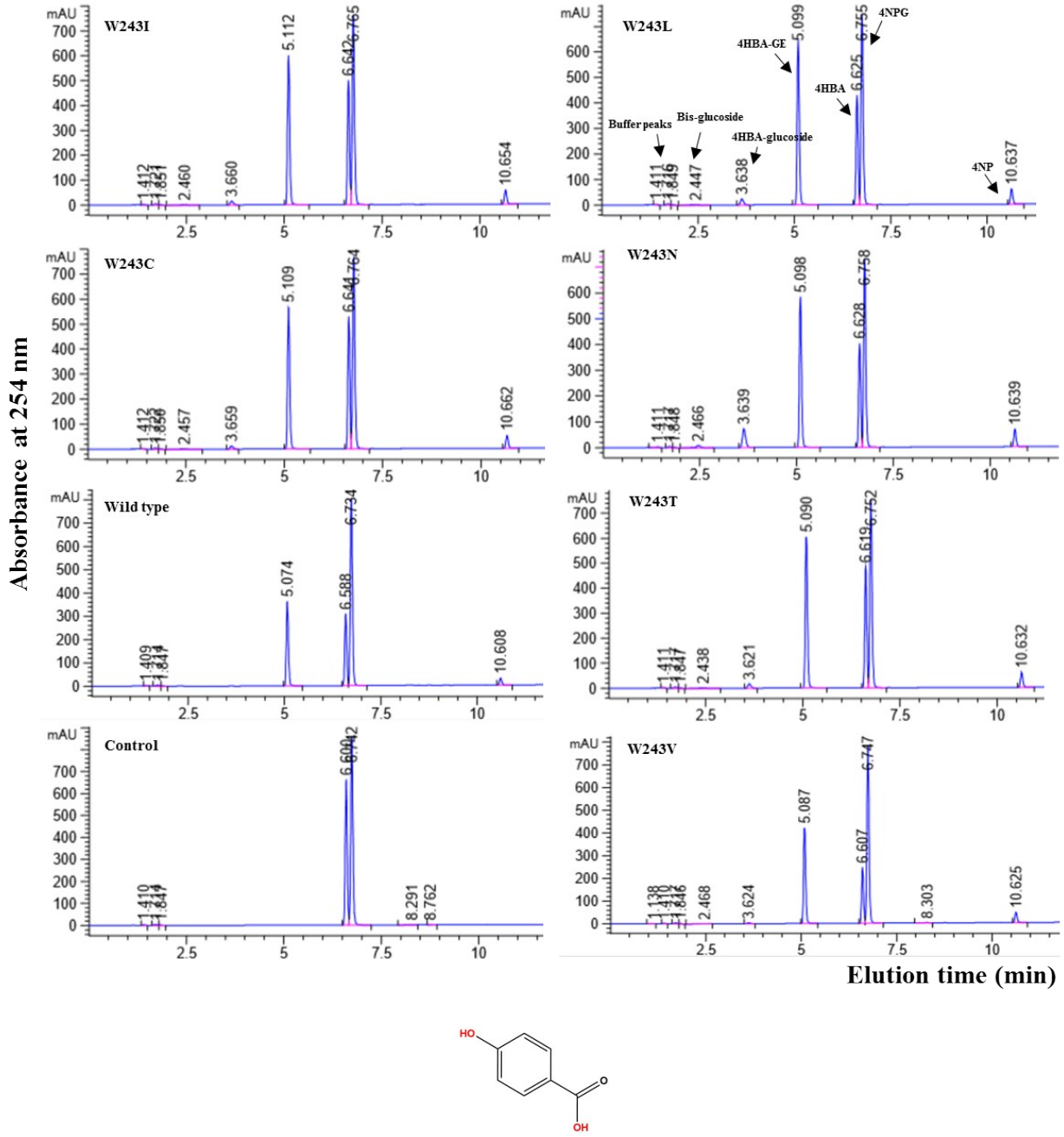

Figure S3. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243 mutants in the transfer of glucose to 4-hydroxybenzoic acid (4HBA). Bis-glucoside, glucoside and glucose ester peaks eluted before the donor 4NPGlc. The peaks are labeled based on the elution time of standards for 4HBA, 4HBA-glucose ester (4HBA-GE, which was previously identified as the product made by WT Os9BGlu31[14]), 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG). The relative positions of products 4HBA-β-D-glucoside (4HBA-glucoside) and 4HBA-bis-β-D-glucoside (Bis-glucoside) were identified by mass spectrometry[18].

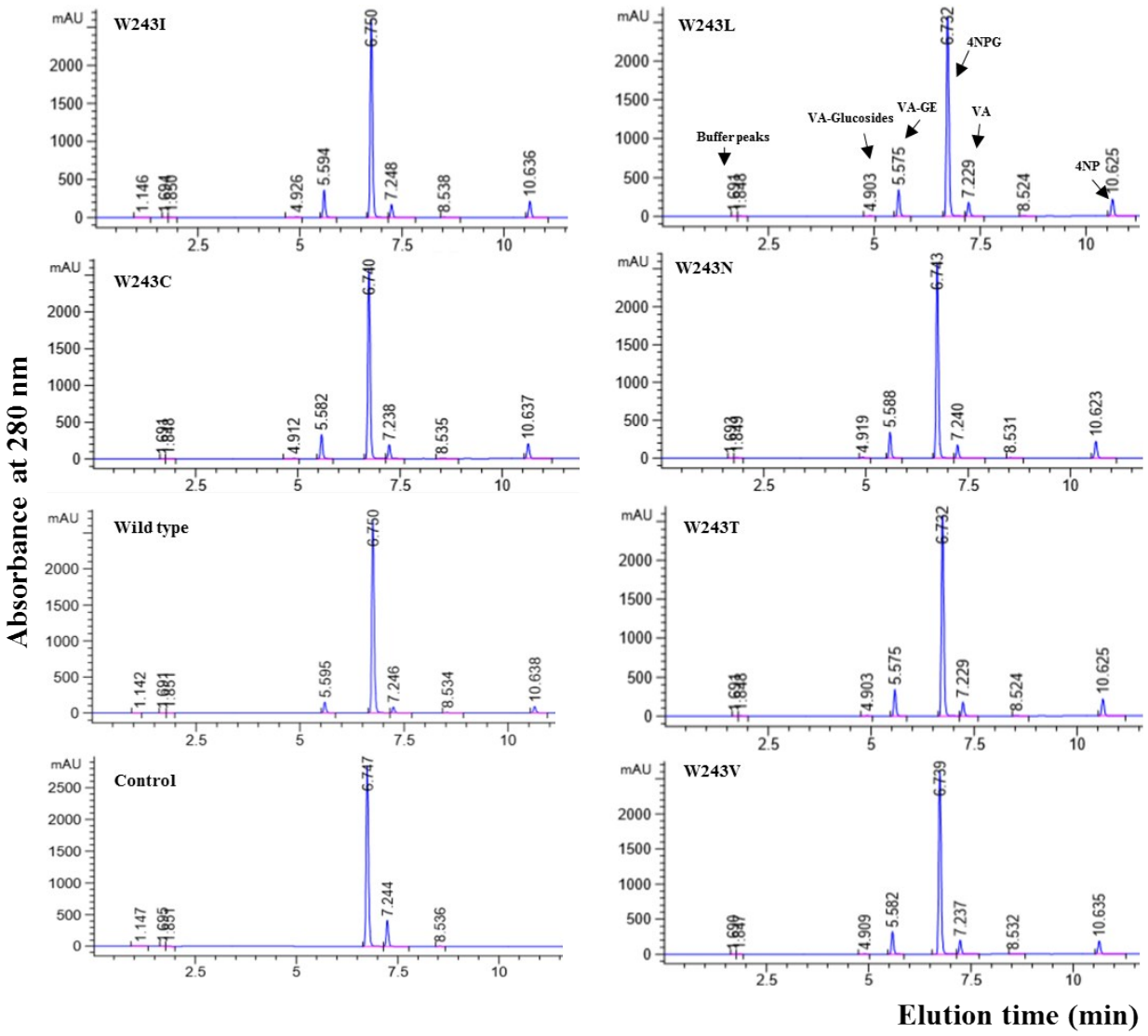

Figure S4. Chromatogram of reaction mixtures of wild type and W243 mutants. Os9BGlu31 mutants in the transfer glucose to both the hydroxyl group and carboxyl groups in vanillic acid (VA) to produce glucoconjugates. The peaks are labeled based on the elution time of standards for VA, 4HBA-glucose ester (4HBA-GE, which was previously identified as the product made by wild type Os9BGlu31[14,18]), 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG). The relative positions of the VA-β-D-glucoside (4HBA-glucoside) was identified by mass spectrometry[18]. Unlike the other phenolic acids, no bis-glucoside was detected for VA.

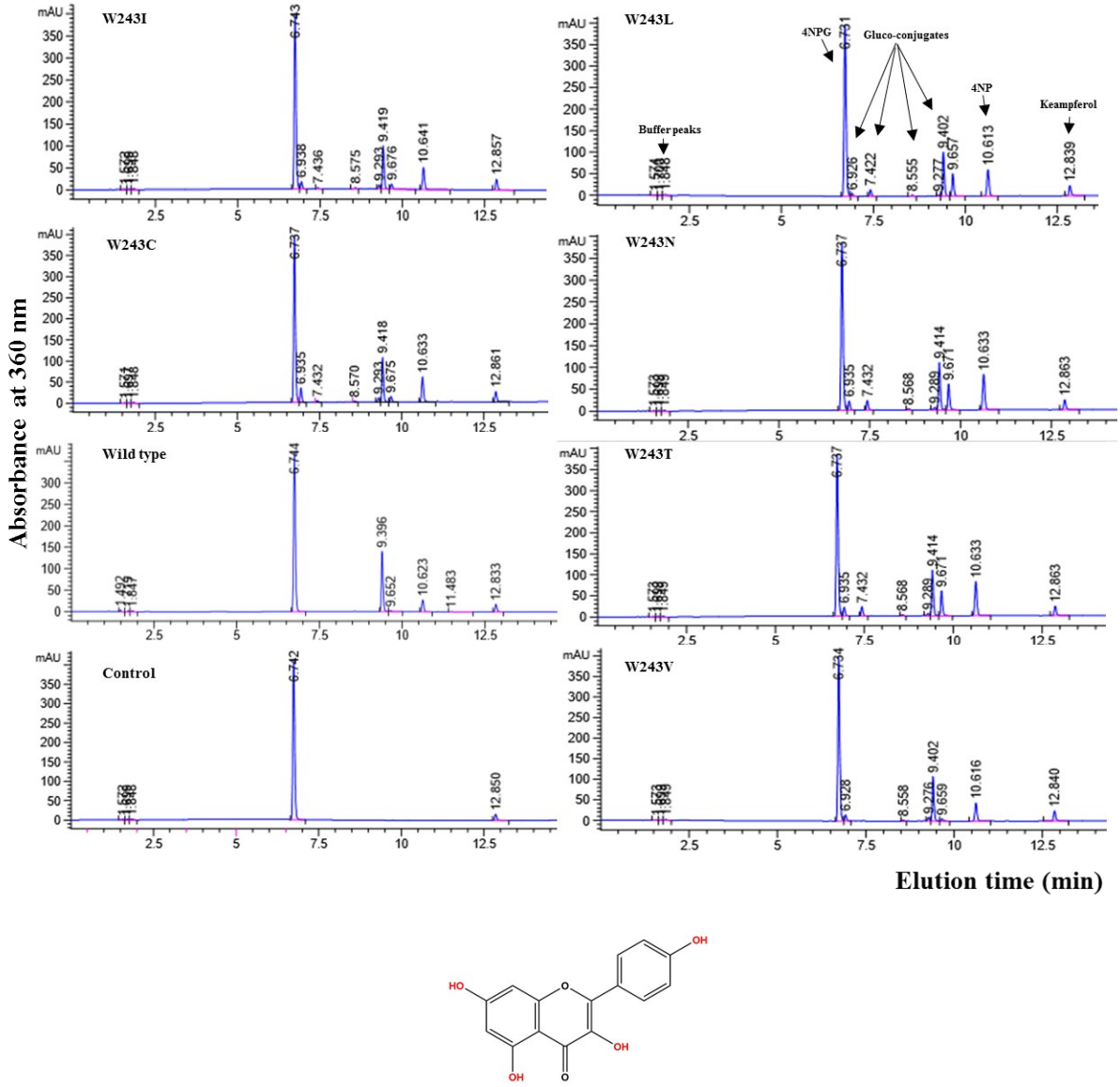

Figure S5. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its mutants in the transfer of glucose to kaempferol. Os9BGlu31 mutants can transfer to multiple positions to provide different glucosides. The peaks are labeled based on the elution time of standards for kaempferol, 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG). According to MS results, the later eluting glucosyl conjugates are mono-glucosides, while the earlier eluting kaempferol glucoconjugates are bis-glucosides[18].

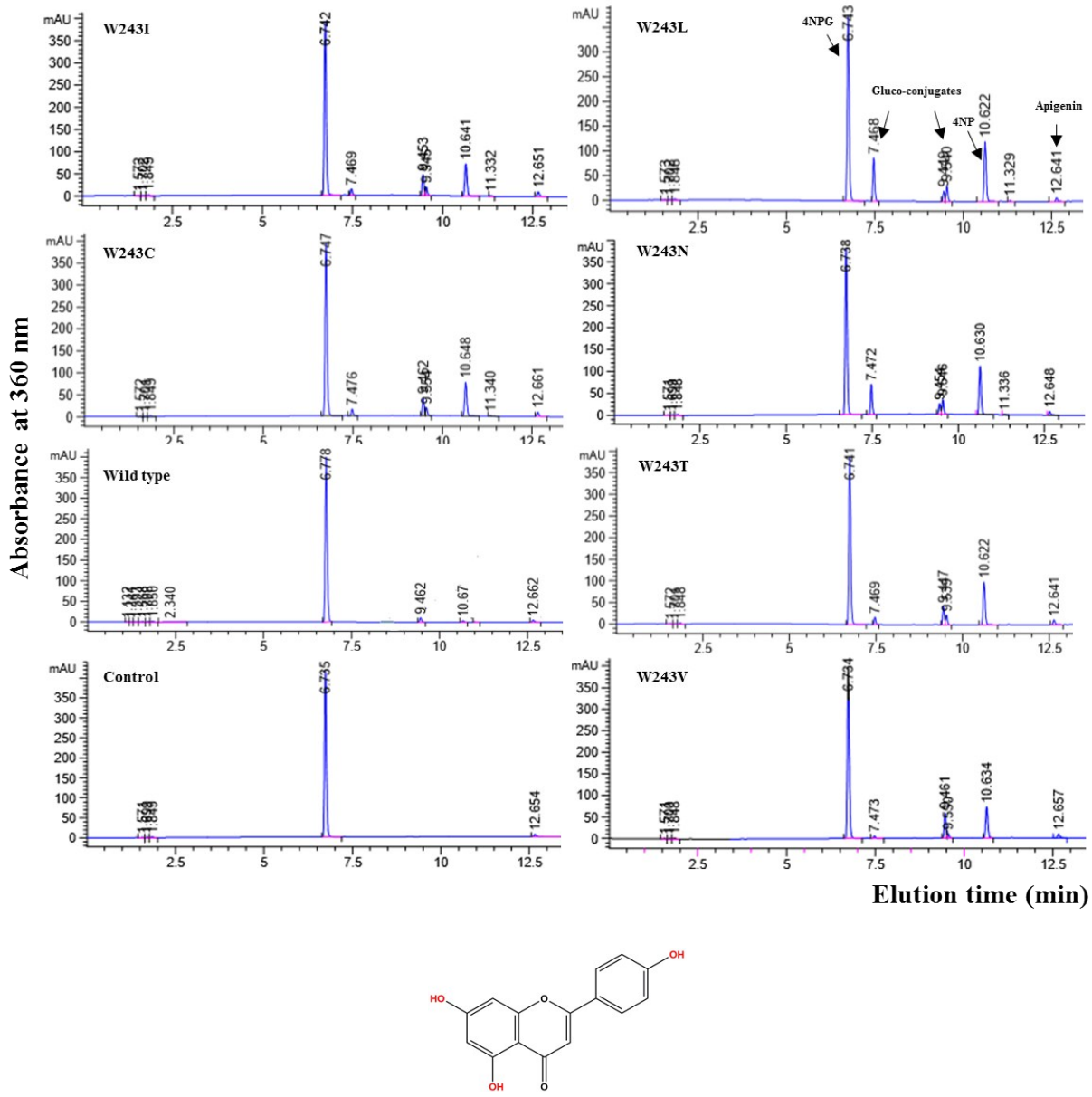

Figure S6. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243 mutants in the transfer of glucose to apigenin. The peaks are labeled based on the elution time of standards for apigenin, 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG). By comparison with the position of kaempferol glucoside identified by MS[18], the closely spaced peaks at 9 min were assigned to apigenin monoglucosides, while the peak at 7.47 min is corresponds to a bis-glucoside.

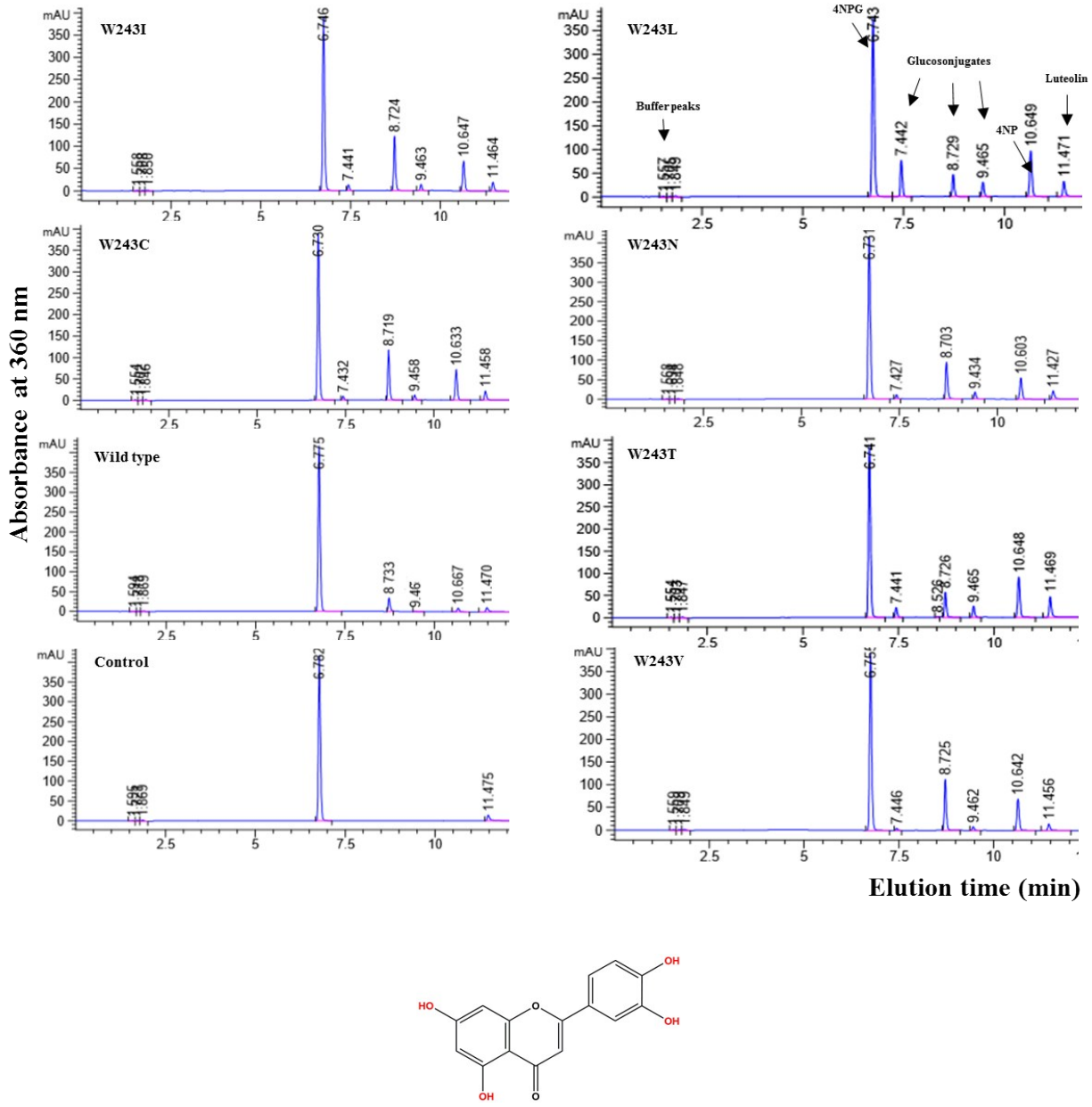

Figure S7. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243 mutants in the transfer of glucose to luteolin. The peaks are labeled in the W243L reaction chromatogram based on the elution time of standards for luteolin, 4-nitrophenol (4NP), and 4-nitrophenyl β-D-glucoside (4NPG). By comparison with the position of kaempferol glucoside identified by MS[18], the peaks at 8.7 and 9.4 min were assigned to luteolin monoglucosides, while the peak at 7.4 min corresponds to a bis-glucoside.

Table S2. Top scores (GoldScore) calculated when docking the acceptor substrates on the wild-type Os9BGlu31 and its W243L mutant.

Acceptor substrate	Wild type	W243L
4-coumaric acid	39.72	43.16
Ferulic acid	43.92	48.68
Caffeic acid	44.71	47.16
Trans-cinnamic acid	39.34	48.58
4-Hydroxybenzoic acid	36.58	37.35
Vanillic acid	41.43	41.25
Syringic acid	40.74	46.34
Sinapic acid	46.03	41.77
Salicylic acid	34.71	35.19
1-Naphthaleneacetic acid	43.65	46.78
1-Naphthol	39.15	41.67
6-Hydroxyflavone	41.61	52.99
Chrysin	43.74	49.02
Apigenin	44.53	54.76
Hesperitin	47.06	51.10
Kaempferol	47.60	49.12
Luteoline	45.20	51.42
Water	10.28	10.29

Table S3. Nomenclature of the molecular descriptors identified by the method in this work.

Variable	Name
AATS7dv	Averaged Moreau-Broto autocorrelation of lag 7 weighted by valence electrons
ALOGP	Ghose-Crippen $\log K_{ow}$
ATSC2pe	Centered moreau-broto autocorrelation of lag 2 weighted by pauling EN
ATSC5c	Centered moreau-broto autocorrelation of lag 5 weighted by gasteiger charge
ATSC5dv	Centered moreau-broto autocorrelation of lag 5 weighted by valence electrons
ATSC8c	Centered moreau-broto autocorrelation of lag 8 weighted by gasteiger charge
BCUTd.1l	First lowest eigenvalue of Burden matrix weighted by sigma electrons
EstateVSA0	Estate VSA Descriptor 0
GATS1Z	Geary coefficient of lag 1 weighted by atomic number
GATSp6	Geary coefficient of lag 1 weighted by atomic number
JGI10	10-ordered mean topological charge
JGI7	6-ordered mean topological charge
largestChain	The number of atoms in the largest chain.
MATS3dv	Moran coefficient of lag 3 weighted by valence electrons
MATSe6	Moran autocorrelation of lag6 weighted by atomic Sanderson electronegativities
MATSv5	Moran autocorrelation of lag5 weighted by atomic van der Waals volumes
MinAbsEStateIndex	Returns a tuple of estate indices for the molecule
Mor06m	3D-morse weighted by mass (distance = 6)
Mor13	3D-morse (distance = 13)
Mor29	3D-morse (distance = 29)
Mor31se	3D-morse weighted by Sanderson electronegativities (distance = 31)
MR	Molar refractivity
RPCS	Relative positive charge surface area
TPSA	Topological polar surface area based on fragments

Table S4. Correlation matrix between selected molecular descriptors. Absolute values are presented for convenience. All colors belong to a single-color scale considering absolute values.

	TPSA	MR	logP	Mor31se	JGI10	AATS7dv	MinAbsEStateIndex	JGI7	BCUTd.1l	EstateVSA0	MATSe6	MATSv5	RPCS	GATS1Z	Mor06m	ATSC2pe	ATSC8c	GATSp6	Mor13	Mor29	ATSC5c	ATSC5dv	MATS3dv	largestChain
largestChain	0.27	0.51	0.57	0.39	0.45	0.08	0.37	0.60	0.60	0.15	0.50	0.39	0.37	0.68	0.57	0.47	0.58	0.70	0.55	0.53	0.52	0.51	0.72	1
MATS3dv	0.27	0.49	0.47	0.26	0.32	0.02	0.22	0.42	0.38	0.03	0.12	0.33	0.31	0.62	0.70	0.53	0.56	0.52	0.37	0.61	0.42	0.75	1	
ATSC5dv	0.58	0.70	0.43	0.27	0.30	0.36	0.39	0.10	0.27	0.21	0.10	0.13	0.42	0.47	0.71	0.68	0.52	0.40	0.59	0.67	0.75	1		
ATSC5c	0.54	0.58	0.40	0.35	0.19	0.37	0.26	0.07	0.24	0.19	0.29	0.14	0.24	0.37	0.35	0.34	0.58	0.35	0.57	0.49	1			
Mor29	0.70	0.35	0.07	0.13	0.17	0.13	0.32	0.48	0.12	0.41	0.43	0.15	0.49	0.19	0.37	0.62	0.46	0.61	0.61	1				
Mor13	0.80	0.72	0.34	0.10	0.35	0.58	0.47	0.01	0.31	0.41	0.06	0.34	0.75	0.28	0.51	0.58	0.50	0.25	1					
GATSp6	0.16	0.04	0.04	0.02	0.23	0.25	0.33	0.76	0.47	0.17	0.70	0.02	0.20	0.23	0.26	0.36	0.26	1						
ATSC8c	0.36	0.53	0.48	0.53	0.48	0.15	0.02	0.30	0.35	0.30	0.22	0.13	0.53	0.40	0.60	0.34	1							
ATSC2pe	0.57	0.66	0.35	0.28	0.49	0.25	0.41	0.23	0.34	0.04	0.10	0.67	0.62	0.41	0.71	1								
Mor06m	0.28	0.69	0.65	0.46	0.56	0.15	0.34	0.15	0.51	0.17	0.00	0.46	0.52	0.53	1									
GATS1Z	0.27	0.72	0.74	0.71	0.68	0.17	0.36	0.11	0.13	0.24	0.19	0.53	0.35	1										
RPCS	0.72	0.68	0.31	0.30	0.66	0.57	0.32	0.04	0.14	0.32	0.04	0.44	1											
MATSv5	0.26	0.51	0.40	0.32	0.49	0.10	0.24	0.12	0.20	0.14	0.13	1												
MATSe6	0.12	0.10	0.13	0.01	0.08	0.39	0.31	0.69	0.08	0.00	1													
EstateVSA0	0.54	0.09	0.32	0.33	0.06	0.24	0.12	0.15	0.36	1														
BCUTd.1l	0.21	0.25	0.57	0.30	0.21	0.08	0.09	0.32	1															
JGI7	0.07	0.23	0.15	0.11	0.01	0.66	0.22	1																
MinAbsEStateIndex	0.53	0.37	0.06	0.06	0.26	0.13	1																	
AATS7dv	0.65	0.68	0.33	0.21	0.36	1																		
JGI10	0.33	0.68	0.63	0.75	1																			
Mor31se	0.00	0.64	0.83	1																				
logP	0.03	0.77	1																					
MR	0.64	1																						
TPSA	1																							
Correlation to other descriptors	0.39	0.51	0.39	0.32	0.38	0.30	0.27	0.26	0.28	0.22	0.21	0.29	0.41	0.41	0.45	0.44	0.40	0.32	0.44	0.40	0.37	0.45	0.41	0.48
Correlation to Rate	0.37	0.46	0.36	0.32	0.39	0.11	0.23	0.28	0.36	0.05	0.26	0.21	0.42	0.42	0.45	0.46	0.46	0.46	0.47	0.48	0.48	0.49	0.49	0.61

Table S5. Predicted transglycosylation rates (in nmol 4NP min^{-1} mg^{-1}) to several putative acceptors by wild-type Os9BGlu31 enzyme and different variants by the ensemble of the 10 neural networks in Table 3 (2.5 vol.% DMSO).

Compound	MR	logP	TPSA	L183QW243N	L241DW243N	W243C	W243I	W243L	W243N	W243T	W243V	WT
3,4-Dihydroxybenzoic acid	37.447	0.796	77.76	151	713	280	424	814	813	510	389	139
3-Indoleacetate	47.901	-0.022	55.92	640	1298	833	1061	1506	1508	1191	1022	639
6-Benzylaminopurine	65.476	1.483	66.49	444	1065	604	785	1195	1226	916	758	428
Ampicillin	92.564	0.318	112.73	915	1556	1073	1328	1792	1872	1546	1319	923
Arbutin	62.61	-1.429	119.61	864	1538	1058	1324	1798	1821	1487	1288	881
cis-Jasmone	52.129	3.022	17.07	222	788	350	482	876	897	584	452	200
Cyanidin-3-glucoside	108.294	0.382	191.6	365	582	425	532	685	750	614	537	368
Digitoxin	194.937	3.247	182.83	756	0	349	210	0	0	374	491	818
Dopamine	42.968	0.599	66.48	316	951	483	666	1086	1086	770	627	307
Hydroquinone	30.488	1.098	40.46	183	764	323	479	884	870	567	440	170
Indole-3-acetic acid	49.843	1.313	53.09	342	984	509	689	1113	1122	800	651	330
Serotonin	52.802	0.893	62.04	429	1090	610	806	1238	1249	925	768	420
Tetracycline	110.218	-0.214	181.62	720	1097	835	1028	1316	1381	1164	1023	744
trans-Zeatin	60.905	0.173	86.72	586	1257	777	996	1435	1455	1133	960	582

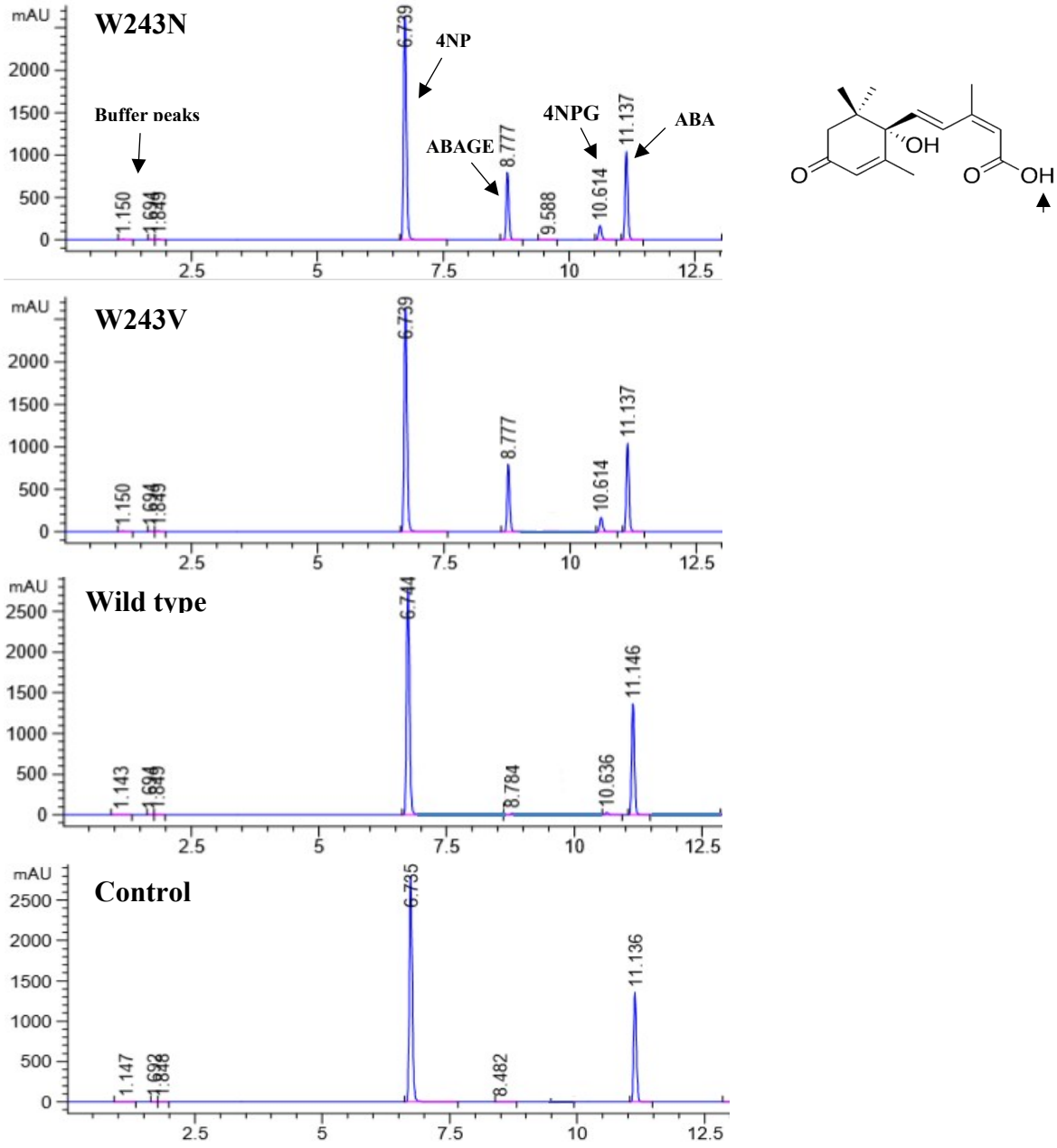

Figure S8. Chromatograms of reaction mixtures of Os9BGlu31 wild type and its W243N and W243V mutants with 4NPGlc and abscisic acid (ABA), a phytohormone acceptor, to produce abscisic acid glucose ester (ABA-GE). Os9BGlu31 wild type and its mutants are shown to catalyze the transfer to a carboxyl group to produce abscisic acid glucose ester. The product identity was confirmed by ^{1}H NMR.