Chemical chaperones increase the cellular activity of N370S β-glucosidase: a therapeutic strategy for Gaucher disease. *Proc Natl Acad Sci U S A* 99: 15428-15433.

Schiffmann, R., and Brady, R.O. 2002. New prospects for the treatment of lysosomal storage diseases. *Drugs* 62: 733-742.

Schmidt, D.D., Frommer, W., Junge, B., Muller, L., Wingender, W., Truscheit, E., and Schafer, D. 1977. α-Glucosidase inhibitors. New complex oligosaccharides of microbial origin. *Naturwissenschaften* 64: 535-536.

Suami, T., Ogawa, S., and Toyokuni, T. 1983. Sweet tasting pseudo-sugars. *Chem Lett* 611-612.

Suami, T., and Ogawa, S. 1990. Chemistry of carba-sugars (pseudo-sugars) and their derivatives. *Adv Carbohydr Chem Biochem* 48: 21-90.

Suzuki, Y., Crocker, A.C., and Suzuki, K. 1971. G_{M1}-gangliosidosis. Correlation of clinical and biochemical data. *Arch Neurol* 24: 58-64.

Suzuki, Y., Nakamura, N., and Fukuoka, K. 1978. G_{M1}-gangliosidosis: accumulation of ganglioside G_{M1} in cultured skin fibroblasts and correlation with clinical types. *Hum Genet* 43: 127-131.

Suzuki, Y., Furukawa, T., Hoogeveen, A., Verheijen, F., and Galjaard, H. 1979. Adult type G_{M1}-gangliosidosis: a complementation study on somatic cell hybrids. *Brain Dev* 1: 83-86.

Suzuki, Y. 2006. β-galactosidase deficiency: an approach to chaperone therapy. *J Inherit Metab Dis* 29: 471-476.

Suzuki, Y., Ichinomiya, S., Kurosawa, M., Ohkubo, M., Watanabe, H., Iwasaki, H., Matsuda, J., Noguchi, Y., Takimoto, K., Itoh, M., Tabe, M., Iida, M., Kubo, T., Ogawa, S., Nanba, E., Higaki, K., Ohno, K., and Brady, R.O. 2007. Chemical chaperone therapy: clinical effect in murine G_{M1}-gangliosidosis. *Ann Neurol* 62: 671-675.

Suzuki, Y., Nanba, E., Matsuda, J., Higaki, K., and Oshima, A. 2008. β-Galactosidase deficiency (β-galactosidosis): G_{M1}-Gangliosidosis and Morquio B disease. In *The Online*

Metabolic and Molecular Bases of Inherited Disease Chapter 151, D. Valle, A.L. Beaudet, B. Vogelstein, K.W. Kinzler, S.F. Antonarakis, and A. Ballabio, eds. New York, McGraw-Hill, pp. 1-101, <http://www.ommbid.com/>

Tominaga, L., Ogawa, Y., Taniguchi, M., Ohno, K., Matsuda, J., Oshima, A., Suzuki, Y., and Nanba, E. 2001. Galactonojirimycin derivatives restore mutant human beta-galactosidase activities expressed in fibroblasts from enzyme-deficient knockout mouse. *Brain Dev* 23: 284-287.

Tropak, M.B., Reid, S.P., Guiral, M., Withers, S.G., and Mahuran, D. 2004. Pharmacological enhancement of β-hexosaminidase activity in fibroblasts from adult Tay-Sachs and Sandhoff Patients. *J Biol Chem* 279: 13478-13487.

Tsunoda, H., Inokuchi, J., Yamagishi, K., and Ogawa, S. 1995b. Synthesis of glycosylceramide analogs composed of imino-linked unsaturated 5a-carbaglycosyl residues: potent and specific gluco- and galactocerebrosidase inhibitors. *Liebigs Annal Chem* 279-284.

Tsunoda, H., and Ogawa, S. 1994. Synthesis of some 5a-carbaglycosylamides, glycolipid analogs of biological interests. *Liebigs Annal Chem* 103-107.

Tsunoda, H., and Ogawa, S. 1995a. Synthesis of 5a-carba-β-D-glycosylceramide analogs linked by imino, ether and sulfide bridges. *Liebigs Annal Chem*267-277.

Yoshida, K., Oshima, A., Shimmoto, M., Fukuhara, Y., Sakuraba, H., Yanagisawa, N., and Suzuki, Y. 1991. Human β-galactosidase gene mutations in G_{M1}-gangliosidosis: a common mutation among Japanese adult/chronic cases. *Am J Hum Genet* 49: 435-442.

Table 1

(A) Inhibitory activity [K_i (μM)] of some *N*-alkyl-4-epi-β-valienamines against three glycosidases

Compound		n	β-Galactosidase[a]	α-Galactosidase[b]	β-Glucosidase[c]
β-Galacto type	**1**	7	0.87	3.1	3.1
	13a	5	2.3	2.7	1.2
	13b	9	0.13	1.9	2.5
	13c	11	0.01	4.4	0.87

a: Bovine liver, b: Green coffee beans, c: Almonds

(B) Inhibitory activity [K_i (μM)] of some *N*-alkyl-β-valienamines against glucocerebrosidase

Compound		n	Glucocerebrosidase[d]
β-Gluco type	**2**	7	0.03
	11a	5	0.3
	11b	9	0.07
	11c	3	0.12
	11d	7	0.3

d: Mouse liver

Table 2. Neurological examination of genetically engineered G_{M1}-gangliosidosis model mice.

Each test is performed with semi-quantitative time, space, and movement parameters. See Ichinomiya et al (2007) for details.

1. Gait: (hip, knee, spine, and shivering)

Score 0: Normal.

Score 1: Slight gait disturbance.

Score 2: Marked gait disturbance.

Score 3: Marked staggering and shaking; gait impossible.

2. Posture: forelimb (paralysis, deformity)

Score 0: Normal.

Score 1: Starting gait difficult and clumsy.

Score 2: Dragging limbs; inversion of dorsum pedis.

Score 3: Complete paralysis; no spontaneous movement.

3. Posture: hind limb (abduction, extention, posture)

Score 0: Normal; smooth joint flexion and extension.

Score 1: Slight hip abduction, external rotation, and knee extension; wide-based.

Score 2: Severe hip abduction, external rotation, and knee extension; wide-based.

Score 3: No spontaneous movement.

4. Trunk (deformity)

Score 0: Normal.

Score 1: Slight back hump.

Score 2: Moderate back hump.

Score 3: Severe back hump.

5. Tail (posture, stiffness)

Score 0: Normal

Score 1: Slight stiffness and elevation.

Score 2: Severe stiffness and elevation.

Score 3: Severe stiffness and elevation with persistent deformity.

6. Avoiding response (pinching tail root with forceps for one second)

Score 0: Strong rejection, avoidance, and squeaking.

Score 1: Slight decrease of response.

Score 2: Trunk torsion; hind limb extension.

Score 3: No response.

7. Rolling over (turning the tail root three times to left and right)

Score 0: Extending four limbs, resisting passive rolling.

Score 1: Slow passive rolling; prompt recovery.

Score 2: Markedly slow passive rolling; delayed recovery.

Score 3: Posture change impossible; slow body movement.

8. Body righting acting on head (response to vertical hanging, head down by holding tail tip, and quick upward movements)

Score 0: Strong upward righting reaction of the head.

Score 1: Slight decrease in response.

Score 2: Marked decrease in response.

Score 3: No response; trunk rotation only.

9. Parachute reflex (response to vertical hanging, head down by holding tail tip, and quick downward movement, three times, within 30 sec)

Score 0: Extension and abduction of hind limbs; continuous knee extension.

Score 1: Slight decrease in response; intermittent knee extension.

Score 2: Marked decrease in response; flexion and adduction of hind limbs; slow movements.

Score 3: No response; continuous flexion and adduction of hind limbs.

10. Horizontal wire netting (stepping through interstice during walking on horizontal wire netting)

Score 0: No stepping into interstice.

Score 1: 21-30 sec before stepping into interstice.

Score 2: 11-20 sec before stepping into interstice.

Score 3: 0-10 sec before stepping into interstice.

11. Vertical wire netting (clinging and holding body on vertical wire netting)

Score 0: Stay for 30 sec.

Score 1: Stay for 21-30 sec before falling.

Score 2: Stay for 11-20 sec before falling.

Score 3: Stay for 0-10 sec before falling.

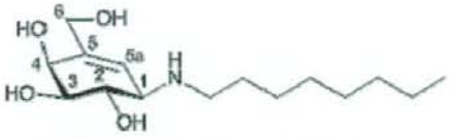
Table 3. Effect of NOEV on G_{M1}-gangliosidosis Tg mice.

Experimental mice were orally fed with water (0 mM NOEV) or NOEV solution (1 mM) for 6 months. Total assessment scores were calculated for each group. Value = mean ± SEM (n); ns=statistically not significant. For details see Suzuki et al (2007)

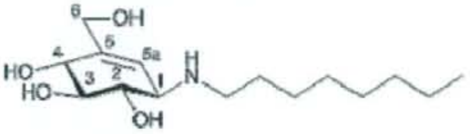
NOEV	0 mM	1 mM	t test
2 months	1.72 ± 0.19 (32)	1.53 ± 0.17 (17)	ns
3 months	2.18 ± 0.38 (11)	1.77 ± 0.24 (17)	ns
4 months	2.53 ± 0.29 (19)	2.06 ± 0.23 (16)	ns
5 months	3.35 ± 0.33 (17)	2.40 ± 0.32 (15)	p<0.05
6 months	3.90 ± 0.31 (30)	2.81 ± 0.25 (16)	p<0.05
7 months	4.88 ± 0.57 (17)	3.43 ± 0.20 (14)	p<0.05

Figure 1

1A

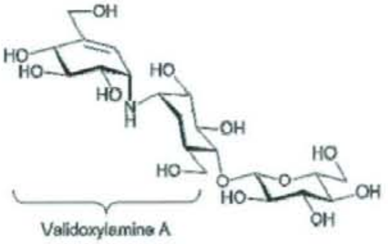


1 NOEV (*N*-Octyl-4-epi-β-valienamine)

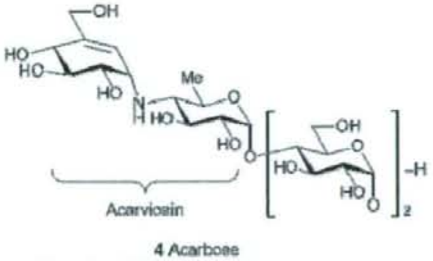


2 NOV (*N*-Octyl-β-valienamine)

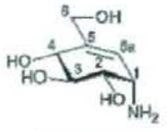
1B

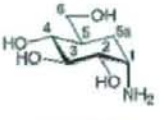


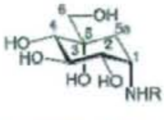
3 Validamycin A



4 Acarbose





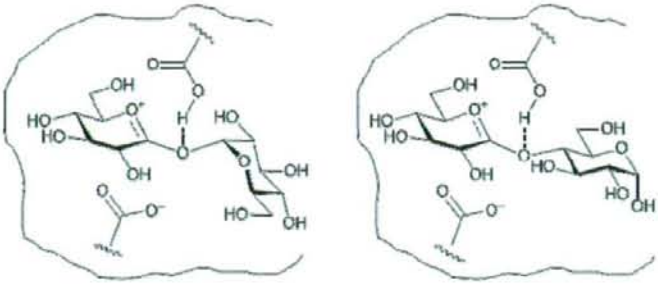


5 Valienamine

6 Validamine

7 Valiolamine (R = H)
8 Voglibose [R = $CH(CH_2OH)_2$]

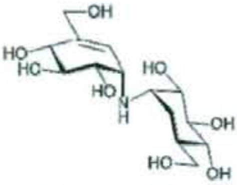
1C



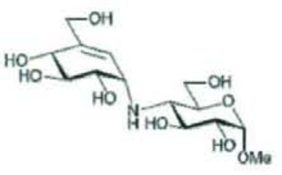
Trehalose

Maltose

Transition-state structures postulated for enzymatic hydrolysis of disaccharides

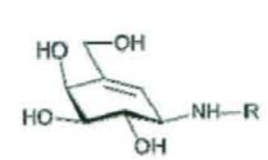


9 Validoxylamine A
Trehalase inhibitor



10 Methyl acarviosin
α-Glucosidase inhibitor

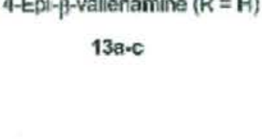
1D



β-Glc type
11 β-Valienamine (R = H)

11a-d

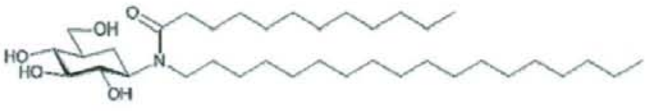
α-Man type
12 2-Epivalienamine (R = H)



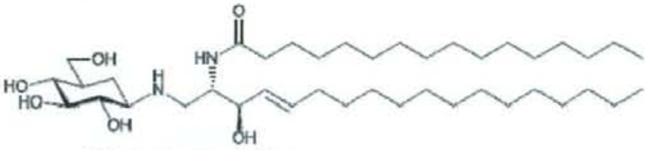
β-Gal type
13 4-Epi-β-valienamine (R = H)

13a-c

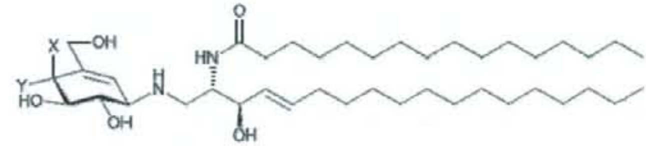
1E



14

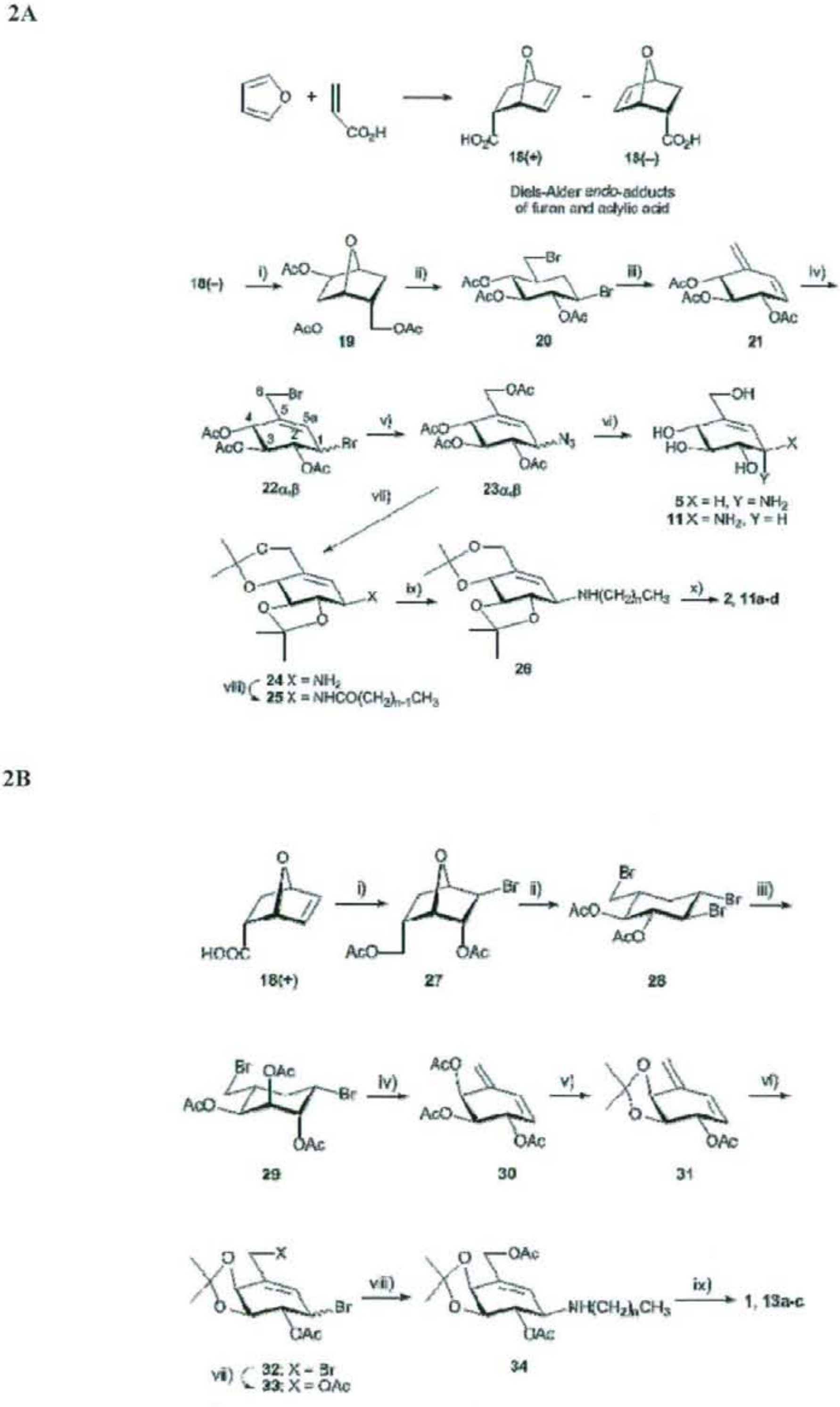


15 Carbaglucosylceramide



16 Unsaturated carbaglucosylceramide: X = H, Y = OH
17 Unsaturated carbagalactosylceramide: X = OH, Y = H

Figure 2


2A
CO2H
HO2C
18(+)
18(–)
Diels-Alder endo-adducts
of furan and aclylic acid
18(–)
i)
AcO
OAc
19
ii)
Br
20
iii)
21
iv)
6
5
5a
4
3
2
1
22α,β
v)
N3
23α,β
vi)
OH
HO
X
Y
5 X = H, Y = NH2
11 X = NH2, Y = H
vii)
24 X = NH2
25 X = NHCO(CH2)n-1CH3
viii)
ix)
NH(CH2)nCH3
26
x)
2, 11a-d
2B
HOOC
18(+)
i)
27
ii)
28
iii)
29
iv)
30
v)
31
vi)
32: X = Br
33: X = OAc
vii)
34
ix)
1, 13a-c

2C

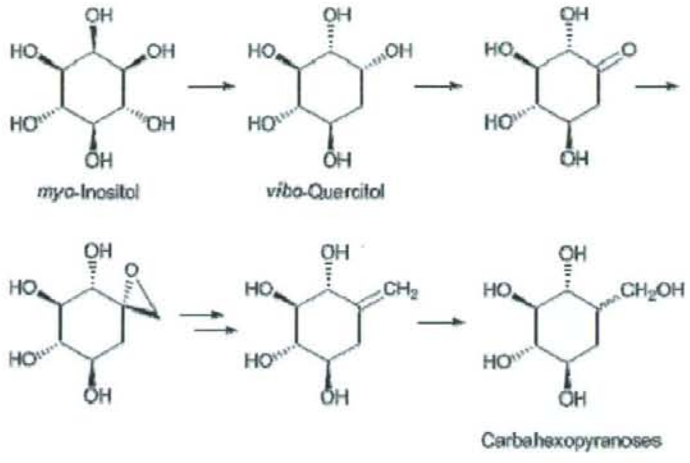


Figure 3

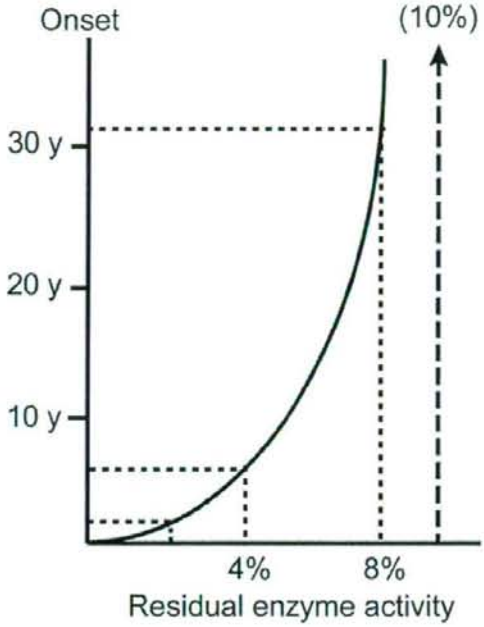


Figure 4 修正版

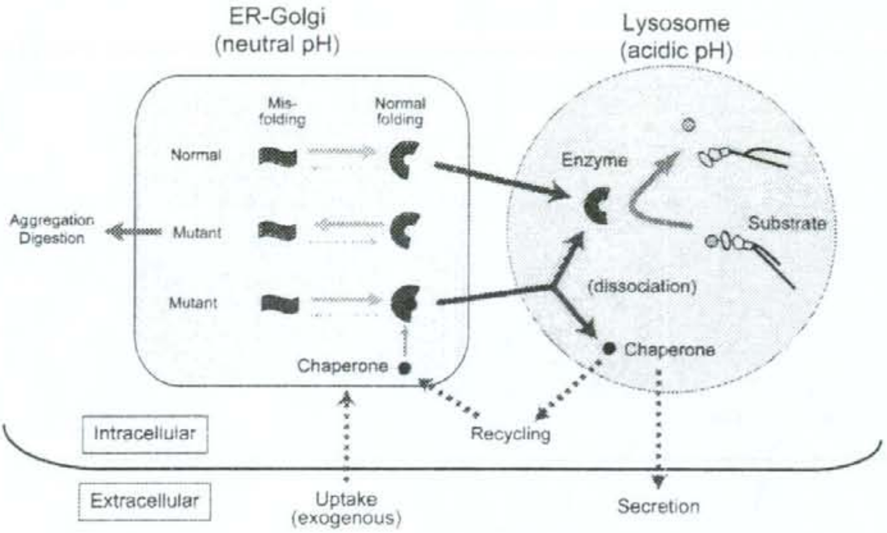


Fig. 4. Postulated molecular events between mutant enzyme molecules and chaperone compounds.

Mutant enzyme protein is unstable in the ER-Golgi compartment at neutral pH, and rapidly degraded or aggregated possibly to cause ER stress. An appropriate substrate analogue inhibitor binds to misfolded mutant protein as chemical chaperone at the endoplasmic reticulum/Golgi compartment in somatic cells, resulting in normal folding and formation of a stable complex at neutral pH. The protein-chaperone complex is safely transported to the lysosome. The complex is dissociated under the acidic condition and in the presence of excessive storage of the substrate. The mutant enzyme remains stabilized, and express catalytic function. The released chaperone is either secreted from the cell or recycled to interact with another mutant protein. These molecular events have been partially clarified by analytical and morphological analyses, and computer-assisted prediction of molecular interactions.

Figure 4

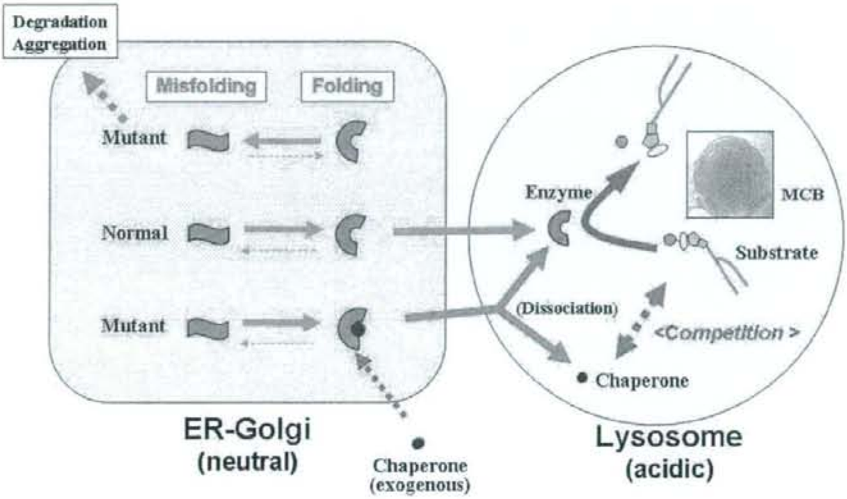
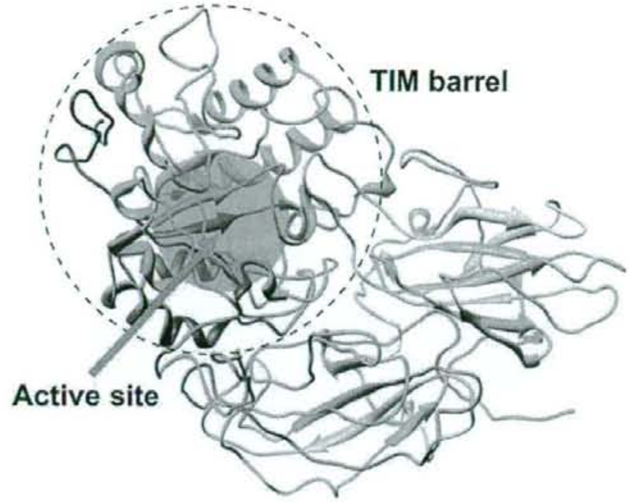

Degradation
Aggregation
Misfolding
Folding
Mutant
Normal
Mutant
Enzyme
MCB
Substrate
(Dissociation)
<Competition >
Chaperone
ER-Golgi
(neutral)
Chaperone
(exogenous)
Lysosome
(acidic)

Figure 5

5A



5B

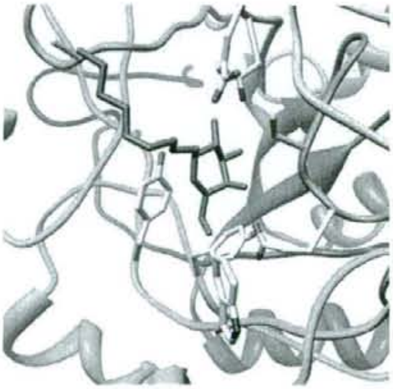


Figure Legends

Fig 1. Structures of valienamines and related compounds.

1A. *N*-Octyl-4-epi-β-valienamine (NOEV) and *N*-octyl-β-valienamine (NOV). **1B.** Antibiotic validamycin A and α-amylase inhibitor acarbose. Carbaglycosylamine-type α-glucosidase inhibitors: valienamine, validamine, and valiolamine. **1C.** Validoxylamine A and methyl acarviosin are mimicking the postulated transition-state structures for hydrolysis of trehalose and maltose, respectively. **1D.** Some biologically interesting valienamine analogues and *N*-alkyl derivatives [R = $(CH_2)_nCH_3$]. **1E.** Biologically active carbaglucosylamide and carbaglucosylceramide, and chemically modified unsaturated derivatives.

Fig. 2. Synthetic pathways of valienamines and related compounds.

2A. Synthesis of valienamines and *N*-alkyl derivatives. (i) H_2O_2, HCOOH; LAH/THF; Ac_2O/Pyrd; (ii) HBr/AcOH; (iii) DBU/toluene; (iv) Br_2/CCl_4; (v) AcONa/$MeO(CH_2)_2OH$; NaN_3/DMF; (vi) MeONa; Ph_3P/MeOH; (vii) MeONa; $(MeO)_2CMe_2$, TsOH/DMF; Ph_3P/MeOH; (viii) $CH_3(CH_2)_{n-1}COCl$/Pyrd; (ix) LAH/THF; (x) aq. AcOH; acidic resin, aq. NH_3.

2B. Synthesis of *N*-alkyl-4-epi-β-valienamines. (i) Br_2, Na_2CO_3/H_2O; LAH/THF; Ac_2O/Pyrd; (ii) HBr/AcOH; (iii) MeONa; aq. H_2SO_4; Ac_2O/Pyrd; (iv) DBU/toluene; (v) MeONa/MeOH; $(MeO)_2CMe_2$, TsOH/DMF; Ac_2O/Pyrd; (vi) Br_2/CCl_4; (vii) AcONa/$MeO(CH_2)_2OH$; (viii) $CH_3(CH_2)_nNH_2$/DMF; (ix) MeONa/MeOH; aq. AcOH; acidic resin, aq. NH_3.

2C. Facile transformation of *vibo*-quercitol into carbahexopyranoses.

Fig. 3. Correlation between residual β-galactosidase activity and clinical onset.

The amount of residual enzyme activity shows positive parabolic correlation with the age of onset in various phenotypic forms of β-galactosidase deficiency disorders. The enzyme activity is generally less than 3% of the control mean in infantile G_{M1}-gangliosidosis, 3-6% in juvenile G_{M1}-gangliosidosis, and more than 6% in late onset (adult/chronic) G_{M1}-gangliosidosis and Morquio B disease. At least 10% of normal enzyme activity is necessary for washout of the storage substrate. The age of onset in patients expressing enzyme activity above this level is theoretically beyond the human life span. This figure is based on the enzyme assay results using cultured skin fibroblasts and a synthetic fluorogenic substrate 4-methylumbelliferyl β-galactopyranoside. In this calculation, for technical reasons, substrate specificity is not taken into account, although mutant enzymes show different spectrum in G_{M1}-gangliosidosis and Morquio B disease.

Fig. 4. Postulated molecular events between mutant enzyme molecules and chaperone compounds.

Mutant enzyme protein is unstable in the ER-Golgi compartment at neutral pH, and rapidly degraded or aggregated possibly to cause ER stress. An appropriate substrate analogue inhibitor binds to misfolded mutant protein as chemical chaperone at the endoplasmic reticulum/Golgi compartment in somatic cells, resulting in normal folding and formation of a stable complex at neutral pH. The protein-chaperone complex is safely transported to the lysosome. The complex is dissociated under the acidic condition and in the presence of excessive storage of the substrate. The mutant enzyme remains stabilized, and express catalytic function. These molecular events have been partially clarified by analytical and morphological analyses, and computer-assisted prediction of molecular interactions.

Fig. 5. Computationally predicted structure of β-galactosidase and its conformation of β-galactosidase and NOEV complex.

5A. Sequence identity in the front part was enough to reconstruct its structure and formed a typical TIM barrel domain that is generally found in glycoside hydrolases. In alignment of this part, active residues of both human and Penicillium sp. β-galactosidase molecules were well matched.

5B. Docking of β-galactosidase and NOEV was performed. In the complex of β-galactosidase and NOEV in pH7, the ring part of NOEV was settled in the active pocket. Oxygen of a glutamic acid in β-galactosidase and hydroxyl of amido in NOEV interacted via hydrogen bonding.

Regular Paper

Support vector machine prediction of N- and O-glycosylation sites using whole sequence information and subcellular localization

Kenta Sasaki,[†1] Nobuyoshi Nagamine[†1]
and Yasubumi Sakakibara[†1]

Background Glycans, or sugar chains, are one of the three types of chain (DNA, protein and glycan) that constitute living organisms; they are often called "the third chain of the living organism". About half of all proteins are estimated to be glycosylated based on the SWISS-PROT database. Glycosylation is one of the most important post-translational modifications, affecting many critical functions of proteins, including cellular communication, and their tertiary structure. In order to computationally predict N-glycosylation and O-glycosylation sites, we developed three kinds of support vector machine (SVM) model, which utilize local information, general protein information and/or subcellular localization in consideration of the binding specificity of glycosyltransferases and the characteristic subcellular localization of glycoproteins.

Results In our computational experiment, the model integrating three kinds of information achieved about 90% accuracy in predictions of both N-glycosylation and O-glycosylation sites. Moreover, our model was applied to a protein whose glycosylation sites had not been previously identified and we succeeded in showing that the glycosylation sites predicted by our model were structurally reasonable.

Conclusions In the present study, we developed a comprehensive and effective computational method that detects glycosylation sites. We conclude that our method is a comprehensive and effective computational prediction method that is applicable at a genome-wide level.

1. Introduction

Glycans, or sugar chains, are one of the three kinds of chain (DNA, protein and glycan) that constitute living organisms: they are often called "the third chain of the living organism". Within an organism, glycans mainly exist as glycolipid or glycoprotein. Efficient chemical synthesis of sugar chains has been well studied in combinatorial chemistry[1)–3)]. Recently, glycosyltransferases that catalyze the transfer of monosaccharides to specific residues in proteins have been well studied in biology and pathology[4)–6)]. In some glycoproteins, glycosylation or attachment of carbohydrate polymers to an amino acid residue has been studied in detail[7)–10)]. However there have been no general approaches that can comprehensively detect glycosylation sites and identify protein-bound glycan structures in living cells. Hence, though there exist several databases on glycans including KEGG GLYCAN[11)] and Glycan Database (`http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp`), there are currently no comprehensive and useful databases on glycosylation.

In the present study, we focused on glycosylation. Glycosylation is one of the most important post-translational modifications, affecting many critical functions of proteins, including cellular communication, and their tertiary structure[12)]. About half of all proteins are estimated to be glycosylated based on the Swiss-Prot database[13)]. There are four different types of glycosylation, namely, via N-glycosylation, O-glycosylation, C-mannosylation and glycophosphatidlyinositol (GPI) anchor attachments. In this study, we developed a method that predicts N-glycosylation, or glycosylation of Asn (N) residues, and O-glycosylation, or glycosylation of Ser (S) and Thr (T) residues, sites, in proteins.

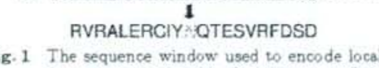


Fig. 1 The sequence window used to encode local information of proteins. k upstream and downstream residues of the target residue (N in italic) were extracted (k=10, in this figure). To encode one residue in the sequence window, we utilized BLOSUM62 profile encoding (the corresponding row in the BLOSUM62 matrix).

[†1] Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan

Several computational approaches to predict O-glycosylation sites in proteins have been developed in recent years[14)–19)]. Statistical learning methods, such as artificial neural network (ANN) and support vector machine (SVM), have been widely utilized for this purpose. In these studies, each amino acid residue was represented by a feature vector in which only local information, or a window of fixed length surrounding the residue (**Fig. 1**), was considered. However, glycosyltransferases attach sugar chains to amino acid residues specifically by recognizing the structure of the whole protein, rather than the individual residue only[20)–22)]. Thus, in predicting glycosylation sites, general protein information, or whole-sequence information should be considered. Moreover, the subcellular localization of glycoproteins is characteristic[15),23),24)]. For example, most membrane proteins have glycans outside the cell membranes and can be regarded as glycoproteins. Hence, we need to utilize not only local information, but also general information and subcellular localization, to predict glycosylation sites.

In this study, we constructed four kinds of SVM model to predict glycosylation sites. The *window* model was based on only local information. The *whole-sequence* and *localization* model utilized, in addition to local information, general information about the proteins and subcellular localization respectively. The *integral* model integrated local information, general protein information and subcellular localization. In our computational experiments, the *whole sequence*, *localization* and *integral* models showed better prediction performances than the *window* model. Moreover, we validated the effectiveness of our model by predicting glycosylation sites that were structurally reasonable in a protein whose glycosylation sites were unknown.

2. Results

2.1 Prediction performance of the proposed method

Table 1 shows the prediction performances when our proposed four kinds of SVM model were applied to the N-glycosylation and O-glycosylation site datasets. Using only local information, the accuracy (described later in Methods) was 0.767 when the model was applied to the N-glycosylation site dataset and 0.784 when applied to the O-glycosylation site dataset. When utilizing all available information, the accuracy was 0.896 when the model was applied to the N-glycosylation site dataset and 0.897 when applied to the O-glycosylation site dataset. The prediction performances with several kernels were shown in Supplementary Material 1.

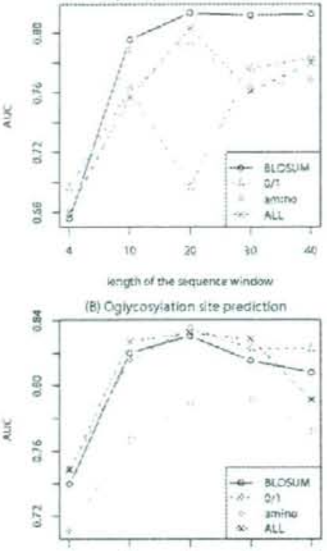


Fig. 2 Comparison of encoding systems. The transition of prediction performances in N-glycosylation sites (A) and O-glycosylation sites (B) were shown. The lengths of sequence window were 4, 10, 20, 30 and 40. Four encoding systems, BLOSUM62 profile encoding system, 0/1 encoding system, amino acid physicochemical properties encoding system and integrated encoding system which was combined by three encoding systems, were applied.

The *whole-sequence* model (N2 and O2 in Table1), using local information and general information, showed significantly better prediction performances than the *window* model (N1 and O1 in Table1). However, the *localization* model, integrating local information and subcellular localization, (N3 and O3 in Table1) showed smaller improvement in performance than the *whole-sequence* model (N2 and O2 in Table1). These results can be elucidated by biological properties represented by both whole-sequence information and subcellular localization infor-