Modeling of Human Neuraminidase-1 and Its Validation by LERE-Correlation Analysis

Seiji Hitaoka1,*, Yuto Shibata1, Hiroshi Matoba1, Akihiro Kawano1, Masataka Harada1, M Motiur Rahman1, Daisuke Tsuji1, Takatsugu Hirokawa2, Kohji Itoh1, Tatsusada Yoshida1, Hiroshi Chuman1

1Institute of Health Biosciences, The University of Tokushima Graduate School, 1-78 Shomachi, Tokushima 770-8505, Japan
2Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan
*E-mail: c401041004@tokushima-u.ac.jp (medicine6025007@gmail.com)

(Received November 21, 2012; accepted January 5, 2013; published online January 31, 2013)

Abstract

Four human neuraminidases (hNEUs1–4) have been identified. Among them, hNEU1 has been studied extensively as a target for sialidosis. It has been desired to understand the biological functions of hNEU1 at the molecular and atomic levels. The three-dimensional structure of hNEU1 is not known at present. In the present work, we constructed a three-dimensional structure of hNEU1 by homology modeling, and then performed correlation analyses between observed and calculated free-energy changes (quantitative structure−activity relationship (QSAR) analyses), coupled with LERE (linear expression by representative energy terms) procedure using the modeled three-dimensional structure in order to confirm the validity of the modeled structure. The atomic coordinates of all atoms in the verified model of hNEU1 are available. The proposed structure of hNEU1 will be useful and helpful for further studies concerning the biological and chemical functions of hNEU1.

The present article is one of continuous works derived from the one that won the CBI Award for the best presentation in the CBI/JSBi 2011 Joint Conference (presented by Seiji Hitaoka).

Key Words: Anti-influenza drugs, Homology modeling, Human neuraminidase-1, LERE, QSAR, Sialic acid analogues

Area of Interest: Molecular Recognition
1. Introduction

Human neuraminidases (hNEUs) have been identified, all of which hydrolyze the terminal sialic acid residues of sialylated glycoconjugates, and are involved in various cellular functions such as cell proliferation, differentiation, and apoptosis [1]. There are four types [2], lysosomal neuraminidase (hNEU1) [3], cytosolic neuraminidase (hNEU2) [4], plasma membrane-associated neuraminidase (hNEU3) [5], and lysosomal/mitochondrial membrane-associated neuraminidase (hNEU4) [6], according to several characteristics, such as their subcellular distributions, enzymatic properties, and substrate specificities. Among the four hNEUs1–4, hNEU1 has been studied extensively as a target for sialidosis, and was proved to associate with lysosomal protective protein/cathepsin A (PPCA) and β-galactosidase (GLB) to form a multienzymic complex in lysosomes [7][8][9][10]. In particular, the association of hNEU1 with PPCA is essential for expression of sialidase activity [7][11][12]. Sialidosis is an autosomal recessive metabolic inborn error caused by germ-line mutation of NEU1 [3][13], which is associated with excessive accumulation in tissues of patients and urinary excretion of sialyloligosaccharides [14][15]. Another hNEU1 deficiency, galactosialidosis, is caused by a primary defect of PPCA with simultaneous decreases in PPCA, hNEU1, and GLB activities, clinical manifestations similar to those of sialidosis due to a hNEU1 deficiency being developed [7][16]. It is difficult to carry out crystallization and structural analyses of hNEU1, because of its extremely unstable properties [9].

The four hNEUs1–4 have the unique amino acid sequences common to the neuraminidase family, i.e., the arginine triad [17], Asp box [18][19][20], and RIP motif [21]. Although the amino acid sequence identity score among hNEUs2–4 (average) is 34 (±5.2)%, that of hNEU1 with hNEUs2–4 is 23 (±0.4)%. This possibly suggests that hNEU1 can take on a three-dimensional structure somewhat different from those of hNEUs2–4. At present, the crystallographic structure of hNEU2 [22] is the only one available. For homology modeling of the three-dimensional structure of hNEU1, hNEU2 and bacterial (*Micromonospora viridifaciens*, *Salmonella typhimurium*, and *Vibrio cholerae*) neuraminidases (identity score with hNEU1 (average) = 23 (±5.1)% ) have been taken as the reference proteins [12][23][24][25][26][27][28], of which the three-dimensional structures are known. Several interesting studies involving homology modeling have been reported concerning the binding structure of hNEU1 with sialic acid analogues [23], the association structure with PPCA [12], the effect of amino acid mutations on the association [24][25][26][27], and possible binding sites for glycosidation [28]. However, the important question remains whether or not a reliable structural model of hNEU1 has been prepared for discussion of the structure−activity (function) relationship, because the identity score of hNEU1 with the reference proteins is considerably lower than the general criterion value (> 50%) required for constructing a reliable structure of a protein [29][30][31].

Although DANA (2-deoxy-2,3-didehydro-N-acetyl-neuraminic acid (Neu5Ac2en)) [32], a well-known and experimentally well-used inhibitor, is not selective for influenza virus neuraminidases, DANA shows moderate inhibitory activity toward a variety of neuraminidases including hNEUs1–4. Commercially available anti-influenza drugs, as typified by zanamivir (Relenza) [33] and oseltamivir (Tamiflu) [34], exhibit remarkably potent inhibitory activity toward influenza virus neuraminidases.

In this work, we aimed to obtain a reliable model of hNEU1 by homology modeling, and to confirm the validity of the modeled structure by correlation analyses (quantitative structure–activity relationship (QSAR) analyses) coupled with LERE (linear expression by representative energy terms) procedure, according to our previous reports [35][36][37][38], i.e., (1) modeling of the
hNEU1 structure by homology modeling, (2) construction of complexes of hNEU1 with several sialic acid analogues (including DANA, zanamivir, and oseltamivir), and (3) LERE-QSAR analysis of complexes of hNEU1 with the analogues in quantitative comparison with those of hNEU2 and influenza virus neuraminidase-1 (N1-NA). The proposed structure of hNEU1 obtained through the above procedures will be helpful for understanding the binding mechanism of hNEU1 with a ligand, and for discovering drugs for disorders associated with hNEU1, because successful crystallographic structural determination of hNEU1 does not look promising at present [12].

2. Methods

2.1 Homology modeling

We obtained the full amino acid sequence of hNEU1 from the UniProt database [39] (accession no. Q99519), and removed the N-terminal 52 amino acids including a 47 signal peptide region [40] from the full sequence prior to the homology modeling. Among neuraminidases of which the crystallographic structures are available, we selected two reference proteins for the homology modeling of hNEU1, i.e., human and bacterial neuraminidases, hNEU2 (complex with the inhibitor DANA (PDB code: 1VCU, chain B)) [22] and one isolated from Micromonospora viridifaciens (1EUR, chain A) [41], respectively. Sequence similarity analysis involving PSI-BLAST [42] showed that the alignment scores of hNEU1 with 1VCU (hNEU2) and 1EUR (M. viridifaciens) are relatively small (the closer the alignment score is to zero, the more similar the two structures are), $10^{-9}$ and $10^{-21}$ (E-value) [43], and that the sequence identity scores are considerably low, 22.7 and 29.8% [44], respectively. The sequence identity score with the two reference proteins (“templates”) does not reach the general criterion value (> 50%) for the homology modeling procedure in order to construct a reliable three-dimensional model. Therefore, we used double templates (1VCU and 1EUR) and manual alignment matching of sequences of hNEU1 with the templates in order to improve the quality of the homology model [45]. The VAST program [46] (implemented in the MOE package [47]) suggested that the arginine triad, Asp boxes, and the RIP motif in hNEU1 can occupy corresponding positions in the templates. The insertions and deletions in the amino acid sequence of hNEU1 were further processed manually so that the arginine triad, Asp boxes, and RIP motif nicely matched, as shown in Figure 1. A number of possible three-dimensional structures (~100 conformers) of hNEU1, which satisfy the alignment shown in Figure 1, were generated exhaustively using MODELLER [48], and then the structure having the highest Verify3D [49] score was selected as the most probable model of hNEU1.

2.2 Calculation of binding free-energy change

Table 1 shows the chemical structures of the sialic acid analogues including anti-influenza drugs along with their inhibitory potencies IC$_{50}$ toward hNEU1, hNEU2, and N1-NA. The compounds are classified into Set I (compounds 1–3) and II (3–6) compounds: Set I compounds differ in fragments A, B, and C, and Set II only in fragment A. Pockets A, B, and C in neuraminidases comprise amino acid residues surrounding (in close contact with) fragments A, B, and C, respectively. Compounds 1–3 and 3–6 were separately subjected to LERE-QSAR analyses. The IC$_{50}$ values of Set I compounds were taken from the reports by Hata et al. [50] and Woods et al. [51], and those of Set II compounds from the report by Magesh et al. [52] (note that compound 3 is common in the two reports [50][52]).

The generation of initial complex structures is summarized concisely in Figure 2: 1–6 represent
the compound numbers in Table 1. a, b, and c denote the ligand-free proteins, hNEU1, hNEU2, and N1-NA, respectively. The crystallographic structures of complexes 2-c (compound 2-c (N1-NA), PDB code: 2HU4) [53] and 3-b (compound 3-b (hNEU2), 1VCU) [22] were used as the initial structures for the subsequent calculations. The initial complex structures other than the above two complexes were constructed in the manner illustrated in Figure 2.

\[\Delta G_{\text{obs}} = a (\Delta E_{\text{bind}} + \Delta G_{\text{sol}}) + \text{const} \]  

Figure 1. Manually modified multiple sequence alignment of hNEU1 with 1VCU (hNEU2) and 1EUR (M. viridifaciens).

The conserved Asp boxes and RIP motif are represented by blue and magenta dots, respectively, and the arginine triad is shown in bold with red. The identical residues are shown in green background.

Then, energy minimization for each complex was performed by molecular mechanics calculations, according to the previously reported procedure [37][54]. The optimized complex structures were used in the following LERE-QSAR analysis.

The LERE-QSAR procedure [35][36][37][38] is based upon the portioning of the overall free-energy change into additive contributions from several representative free-energy terms, which are typically the intrinsic binding and solvation energies associated with complex formation, $\Delta E_{\text{bind}}$ and $\Delta G_{\text{sol}}$, respectively.

$$\Delta G_{\text{obs}} = a (\Delta E_{\text{bind}} + \Delta G_{\text{sol}}) + \text{const}$$  

Coefficient $a$ on the right-hand side of eq. 1 is a constant, which is determined by the entropic contribution and a balance of the representative energy terms with other energy terms that are not considered explicitly in eq. 1. $\Delta G_{\text{obs}}$ on the left-hand side is the overall free-energy change obtained from the observed inhibitory potency $IC_{50}$, assuming $\Delta G_{\text{obs}} = RT \ln IC_{50} (T = 310 \text{ K})$. 
Table 1. Chemical structure and inhibitory potency of compounds 1–6\(^a\) toward hNEU1, hNEU2, and N1-NA

| no. | set | R\(_1\) (fragment A) | R\(_2\) (fragment C) | IC\(_{50}\) \(^b\) (µM) | hNEU1 | hNEU2 | N1-NA |
|-----|-----|----------------------|----------------------|------------------------|--------|--------|--------|
| 1   | I   | CH(OH)CH(OH)CH\(_2\)(OH) | NHC(=NH\(_2\))NH\(_2\) | 2700\(^c\) | 16\(^c\) | 0.0016\(^c\) |
| 2   | I   | OCH(CH\(_2\)CH\(_3\))\(_2\) | NH\(_3\)^+ | \(~10000\(^c\) | \(~6000\(^c\) | 0.0071\(^c\) |
| 3   | I   | CH(OH)CH(OH)CH\(_2\)(OH) | OH | 170\(^c\) | 46\(^c\) | 3.5\(^d\) |
| 3   | II  | CH(OH)CH(OH)CH\(_2\)(OH) | OH | 140\(^e\) | – | – |
| 4   | II  | CH(OH)CH(OH)CH\(_2\)NHCO-CH\(_3\) | OH | 58\(^e\) | – | – |
| 5   | II  | CH(OH)CH(OH)CH\(_2\)NHCO-C\(_3\)H\(_7\) | OH | 32\(^e\) | – | – |
| 6   | II  | CH(OH)CH(OH)CH\(_2\)NHCO-n-C\(_4\)H\(_9\) | OH | 10\(^e\) | – | – |

\(^{a}\) All ionizable groups were treated as ionized forms.

\(^{b}\) In µM.

\(^{c}\) Taken from Ref. [50].

\(^{d}\) Taken from Ref. [51].

\(^{e}\) Taken from Ref. [52].

\(\Delta E_{\text{bind}}\) and \(\Delta G_{\text{sol}}\) in eq. 1 are defined as \(E(\text{complex}) - [E(\text{protein}) + E(\text{ligand})]\) and \(G_{\text{sol}}(\text{complex}) - [G_{\text{sol}}(\text{protein}) + G_{\text{sol}}(\text{ligand})]\), respectively. Eq. 1 implicitly contains the entropic contribution of the overall free-energy change (\(-T\Delta S_{\text{overall}}\)) via the coefficient and intercept terms. Table 2 concisely summarizes generally usable estimation methods of \(\Delta E_{\text{bind}}, \Delta G_{\text{sol}},\) and \(-T\Delta S_{\text{overall}}\) in the LERE procedure. In the current study, \(\Delta E_{\text{bind}}\) and \(\Delta G_{\text{sol}}\) on the right-hand side of eq. 1 were calculated by molecular mechanics and generalized Born/surface area (GB/SA) calculations, respectively, and

Table 2. Quantitative estimation of representative energy terms

| Energy term          | Symbol | Method                                      |
|----------------------|--------|---------------------------------------------|
| Binding energy       | \(\Delta E_{\text{bind}}\) | MM\(^a\), MFCC\(^b\), QM/MM\(^c\), FMO\(^d\) |
| Solvation energy     | \(\Delta G_{\text{sol}}\) | GB/PB\(^e\), MO/Continuum media\(^f\), RISM\(^g\) |
| Entropic energy      | \(-T\Delta S_{\text{overall}}\) | Normal mode calculation\(^h\), Quasi-harmonic approximation\(^i\), \(\Delta H - \Delta S\) compensation\(^j\) (empirical) |

\(^{a}\) Ref. [55].

\(^{b}\) Ref. [56].

\(^{c}\) Ref. [57].

\(^{d}\) Ref. [58].

\(^{e}\) Ref. [59].

\(^{f}\) Ref. [60].

\(^{g}\) Ref. [61].

\(^{h}\) Ref. [62].

\(^{i}\) Ref. [63].

\(^{j}\) Ref. [64].
Figure 2. Generation of initial complex structures.

1: 1–6 represent compound numbers in Table 1. a, b, and c denote ligand-free proteins, hNEU1, hNEU2, and N1-NA, respectively. For instance, 2-c represents the complex of compound 2 with N1-NA. 2: Since PDB 2F0Z contains 24 missing residues, complex 1-b is reconstructed by replacement of compound 3 (DANA) in the referred complex (PDB code: 1VCU) with compound 1 (zanamivir) in the compared protein (2F0Z). 3: Superimposition of a Cα atoms in the compared protein (the terminal point of the arrow) on the corresponding Cα atoms in the referred protein (the initial point), and replacement of a ligand in the compared complex with that in the referred complex. The Cα atoms used for superimposition are taken from amino acid residues in the secondary structural regions matched between the referred and compared proteins. 4: Replacement of fragment A in compound 3 (DANA) with each one in compounds 4–6. (a) and (b): one and two water molecules are placed in the active site of protein, respectively [37][54].

$-T\Delta S_{overall}$ was estimated by introducing the empirical entropy–enthalpy compensation rule (i.e., $-T\Delta S_{overall}$ is assumed to be linear with $\Delta H_{overall}$) [35][36][37][38]. All the calculations were carried out using the AMBER10 package [65] with the parm99 [66] parameter for the three proteins and with the general AMBER force field (GAFF) [67] parameter for compounds 1–6. $\Delta E_{bind}$ consists of the electrostatic and van der Waals interaction energies, $\Delta E_{bind}^{\text{ele}}$ and $\Delta E_{bind}^{\text{vdW}}$ ($\Delta E_{bind} = \Delta E_{bind}^{\text{ele}} + \Delta E_{bind}^{\text{vdW}}$), respectively. Although $\Delta G_{\text{sol}}$ further consists of the polar and nonpolar components, $\Delta G_{\text{sol}}^{\text{polar}}$ and $\Delta G_{\text{sol}}^{\text{nonpolar}}$ ($\Delta G_{\text{sol}} = \Delta G_{\text{sol}}^{\text{polar}} + \Delta G_{\text{sol}}^{\text{nonpolar}}$), respectively, the contribution of $\Delta G_{\text{sol}}^{\text{nonpolar}}$ was negligible in the current case. Short-time molecular dynamics calculations with a 150 ps production run at 300 K were performed to average $G_{\text{sol}}^{\text{polar}}$ sampled at 3.0 ps intervals, because $G_{\text{sol}}^{\text{polar}}$ was sensitive to subtle conformational changes in the current case.

3. Results and discussion

3.1 Validation of chain folding in the modeled structure of hNEU1

Figure 3 shows the three-dimensional structure of the modeled hNEU1 along with the
crystallographic ones of 1VCU (hNEU2) and 1EUR (M. viridifaciens), all three structures having on a six-bladed β-propeller fold, and having Asp boxes and the RIP motif at corresponding positions. The final model of hNEU1 passes structural verification tests, i.e., Verify3D (score = 0.77), PROCHECK [68], and DSSP [69] ones, as shown in Tables A1 and A2 (Appendix).

Figure 3. Three-dimensional modeled structure of (a) hNEU1, (b) 1VCU (hNEU2), and (c) 1EUR (M. viridifaciens). The α-helixes and β-strands are colored in red and yellow, respectively. The conserved Asp boxes and RIP motif are shown in blue and magenta, respectively.

3.2 Validation of the hNEU1 structure by the correlation analysis

3.2.1 LERE-QSAR of Set I compounds
The observed overall free-energy change $\Delta G_{\text{obs}}$ for complex formation of hNEU1, hNEU2, and N1-NA with Set I compounds 1–3 is excellently reproduced with the LERE-QSAR procedure. Figure 4 shows excellent agreement between $\Delta G_{\text{obs}}$ and $\Delta G_{\text{calc}}$ ($r = 0.985$). Both $\Delta G_{\text{obs}}$ and $\Delta G_{\text{calc}}$ show that compounds 1–3 exhibit more potent inhibitory activity toward influenza virus neuraminidase-1 (N1-NA) than human ones (hNEU1 and hNEU2), and that compound 2 (oseltamivir) exhibits very weak activity toward hNEU1 and hNEU2.

![Figure 4](image)

**Figure 4.** Plots between $\Delta G_{\text{obs}}$ and $\Delta G_{\text{calc}}$.

Circle, triangle, and square represent complexes of three neuraminidases (hNEU1 (a, white), hNEU2 (b, gray), and N1-NA (c, black)) with compounds 1 (zanamivir), 2 (oseltamivir), and 3 (DANA), respectively. $\Delta G_{\text{calc}}$ is calculated by: $\Delta G_{\text{obs}} = a(\Delta E_{\text{bind}} + <\Delta G_{\text{sol polar}}>) + \text{const}$ ((a): $r = 0.905, a = 0.0442, \text{const} = -1.05 (n = 3)$, (b): $r = 0.978, a = 0.0694, \text{const} = -0.254 (n = 3)$, and (c): $r = 0.917, a = 0.345, \text{const} = 8.79 (n = 3)$).

### 3.2.2 LERE-QSAR of Set II compounds

Next, LERE-QSAR analysis was performed on complexes of hNEU1 with DANA and its analogues (Set II compounds 3–6), in order to further confirm the validity of the modeled structure of hNEU1. Compounds 3–6 vary only in fragment A. We previously reported [37][54] that the overall free-energy change $\Delta G_{\text{obs}}$ is governed mostly by the electrostatic and dispersion interaction energies of fragment A in ligands with pocket A in hNEU2 and N1-NA. Table 3 lists $\Delta G_{\text{obs}}, \Delta G_{\text{calc}}, \Delta E_{\text{bind ele}}, <\Delta G_{\text{sol polar}}>, \text{and } \Delta E_{\text{bind vdW}}$. The correlation equation 2 shows that $\Delta G_{\text{calc}}$ excellently reproduces $\Delta G_{\text{obs}}$.

$$
\Delta G_{\text{obs}} = 0.0739 (\Delta E_{\text{bind ele}} + <\Delta G_{\text{sol polar}}> + \Delta E_{\text{bind vdW}}) + 1.32
$$

$n = 4, r = 0.926, s = 0.318, F = 12.0$
The variances of $\Delta E_{\text{bind}}^{\text{ele}}$ and $<\Delta G_{\text{sol polar}}>$ are considerably large (430 and 474 kcal/mol$^2$, respectively). There is an excellent anti-correlation between $\Delta E_{\text{bind}}^{\text{ele}}$ and $<\Delta G_{\text{sol polar}}>$ ($<\Delta G_{\text{sol polar}}> = -1.04 \, \Delta E_{\text{bind}}^{\text{ele}} - 91.4, n = 4, r = -0.995$). Consequently, the variance of electrostatic contributions ($\Delta E_{\text{bind}}^{\text{ele}} + <\Delta G_{\text{sol polar}}>$) becomes considerably small (5.3 kcal$^2$/mol$^2$), because of the nearly perfect compensation between $\Delta E_{\text{bind}}^{\text{ele}}$ and $<\Delta G_{\text{sol polar}}>$. Therefore, only $\Delta E_{\text{bind}}^{\text{vdW}}$ appears to survive. In fact, the variance of $\Delta E_{\text{bind}}^{\text{vdW}}$ (53 kcal$^2$/mol$^2$) is considerably larger than that of ($\Delta E_{\text{bind}}^{\text{ele}} + <\Delta G_{\text{sol polar}}>$). $\Delta E_{\text{bind}}^{\text{vdW}}$ alone shows an excellent correlation with $\Delta G_{\text{obs}}$ ($r = 0.996$). The above results clearly suggest that the variation of $\Delta E_{\text{bind}}^{\text{vdW}}$ is a dominant contributor to the variation of $\Delta G_{\text{obs}}$ although the complex formation is strongly driven by the electrostatic contributions, i.e., $\Delta E_{\text{bind}}^{\text{ele}}$ and $<\Delta G_{\text{sol polar}}>$.

### Table 3. Overall free-energy change $\Delta G$ and representative energy terms\(^a\) (hNEU1)

| compound no. | $\Delta G_{\text{obs}}$ | $\Delta G_{\text{calc}}$ | $\Delta E_{\text{bind}}^{\text{ele}}$ | $<\Delta G_{\text{sol polar}}>$ | electrostatic energy $\Delta E_{\text{bind}}^{\text{ele}} + <\Delta G_{\text{sol polar}}>$ | van der Waals energy $\Delta E_{\text{bind}}^{\text{vdW}}$ |
|--------------|----------------|----------------|----------------|----------------|---------------------------------|----------------|
| 3            | −5.45          | −5.75          | −250.37        | 168.82 (9.51)  | −81.55                          | −14.08         |
| 4            | −6.01          | −5.69          | −284.88        | 209.70 (13.87) | −75.18                          | −19.63         |
| 5            | −6.38          | −6.47          | −307.20        | 227.71 (10.28) | −79.49                          | −25.83         |
| 6            | −7.09          | −7.03          | −291.29        | 212.03 (14.27) | −79.26                          | −33.64         |
| variance\(^e\) |                 |                |                |                | 430.46                           | 473.87         |

\(^a\) In kcal/mol.  
\(^b\) $\Delta G_{\text{obs}} = RT \ln \text{IC}_{50}$ ($T = 310$ K).  
\(^c\) Calculated from eq. 2.  
\(^d\) Average value and standard deviation (in parenthesis).  
\(^e\) In kcal$^2$/mol$^2$.

It should again be noted that $\Delta G_{\text{calc}}$ in Sections 3.2.1 and 3.2.2 was calculated using the modeled structure of hNEU1. These two QSAR results demonstrated that the modeled structure of hNEU1 consistently accounts for the two observed structure–activity datasets, the observed activity of Set I compounds toward hNEU1 as well as hNEU2 and N1-NA, and that of Set II ones toward hNEU1.

### 3.3 Interaction of Set II compounds with hNEU1

$\Delta E_{\text{bind}}^{\text{vdW}}$ shows the largest variance among interaction energy terms, as can be seen in Table 3. The variance of $\Delta E_{\text{bind}}^{\text{vdW}}$ comes mostly from differences in fragment A, obviously because Set II compounds 3–6 differ only in fragment A. Figure 5 shows the binding structure of compound 6 (having the largest size substituent in fragment A) with amino acid residues in pocket A of hNEU1. All of the Set II compounds have two hydroxyl groups at the 7- and 8-positions (as shown in Table 1) in fragment A. These common groups are involved in hydrogen-bonding and electrostatic interactions with the side chains of Lys161, Asp263, and Glu264 in pocket A. The 9-OH group attached to fragment A, existing only in compound 3, shows an additional hydrogen-bonding interaction with Asp263 (not shown in Figure 5). However, these electrostatic binding interactions do not make a significant contribution to the variance of the overall interaction energy, because the electrostatic binding and solvation contributions compensate for each other. The 9-OH group in
fragment A of compound 3 is replaced by the NHCO-n-alkyl group in compounds 4–6. This replacement causes extensive stabilization due to further van der Waals interactions of the NHCO-n-alkyl group with His220, Gly221, Asp226, Pro262, and Asp263 in pocket A. As a result, the overall free-energy changes of compounds 4–6 are more stabilized than that of compound 3.

An above-mentioned interpretation of the QSAR results in Section 3.2.2 probably reflects a general interaction profile of hNEU1 with the side chain (fragment A) in sialic acid analogues.

Figure 5. Interaction between fragment A of compound 6 and pocket A of hNEU1.

4. Summary and conclusion

Since the three-dimensional structure of hNEU1 is not available at present, the interesting and important biological functions of hNEU1 are not well understood at the molecular and atomic levels. In the current study, we constructed a homology model of hNEU1 from the crystallographic structures of two other neuraminidases. Then, we performed LERE-QSAR analyses using the modeled structure to determine whether or not the observed structure–activity datasets can be explained reasonably. The QSAR results confirmed that the proposed model of hNEU1 is nicely consistent with the observed structure–activity datasets, which show the selectivity and specificity of hNEU1 toward a ligand.

The proposed structure of hNEU1 will be of significance for understanding the detailed binding mechanism of hNEU1 with various ligands including PPCA [12][70], and for the discovery of drugs for disorders associated with hNEU1. The atomic coordinates of all atoms in the verified model of hNEU1 are available from the corresponding author.

The combined approach of homology modeling and correlation analysis (LERE-QSAR) will be a powerful one in related fields, as demonstrated in the current work.
Acknowledgements

This work was supported by the Ministry of Agriculture, Forestry and Fisheries (Agri-Health Translational Research Project) and Grants-in-Aid for Scientific Research (Nos. 24590051 and 23710272) from the Ministry of Education, Culture, Sports, Science and Technology.

Appendix

Table A1. Summary of Ramachandran plot for the modeled structure of hNEU1 and crystallographic ones of 1VCU (hNEU2) and 1EUR (M. viridifaciens)

| structure          | core<sup>a</sup> | allowed<sup>b</sup> | generous<sup>c</sup> | disallowed<sup>d</sup> | number of Gly, Pro, and end-residues | Total |
|--------------------|------------------|---------------------|----------------------|------------------------|-------------------------------------|-------|
| hNEU1              | 262              | 85.9                | 37                   | 12.1                   | 4                                  | 2     | 0.7     | 58     | 363     |
| 1VCU (hNEU2)       | 257              | 80.8                | 61                   | 19.2                   | 0                                  | 0     | 0       | 55     | 373     |
| 1EUR (M. viridifaciens) | 258          | 88.1                | 31                   | 10.6                   | 4                                  | 1.4   | 0       | 68     | 361     |

<sup>a</sup> Residues in most favored regions.
<sup>b</sup> Residues in additional allowed regions.
<sup>c</sup> Residues in generously allowed regions.
<sup>d</sup> Residues in disallowed regions.

Table A2. Secondary structure distribution of the modeled structure of hNEU1 and crystallographic ones of 1VCU (hNEU2) and 1EUR (M. viridifaciens)

| structure          | α-helix | β-strand | turn | random coil |
|--------------------|---------|----------|------|-------------|
| hNEU1              | 14      | 3.9      | 153  | 42.1        | 36 | 9.9 | 160 | 44.1 |
| 1VCU (hNEU2)       | 24      | 6.4      | 187  | 50.1        | 44 | 11.8 | 118 | 31.6 |
| 1EUR (M. viridifaciens) | 11      | 3.0      | 156  | 43.2        | 47 | 13.0 | 147 | 40.7 |

References

[1] Miyagi, T. Aberrant expression of sialidase and cancer progression. *Proc. Jpn. Acad., Ser. B* 2008, 84, 407–418.
[2] Miyagi, T.; Kato, K.; Ueno, S.; Wada, T. Aberrant expression of sialidase in cancer. *Trends Glycosci. Glycotechnol.* 2004, 16, 371–381.
[3] Pshezhetsky, A. V.; Richard, C.; Michaud, L.; Igdoura, S.; Wang, S.; Elsliger, M.-A.; Qu, J.; Leclerc, D.; Gravel, R.; Dallaire, L.; Potier, M. Cloning, expression and chromosome mapping of human lysosomal sialidase and characterization of mutations in sialidosis. *Nat. Genet.* 1997, 15, 316–320.
[4] Monti, E.; Preti, A.; Rossi, E.; Ballabio, A.; Borsani, G. Cloning and characterization of NEU2, a human gene homologous to rodent soluble sialidases. *Genomics* 1999, 57, 137–143.
[5] Wada, T.; Yoshikawa, Y.; Tokuyama, S.; Kuwabara, M.; Akita, H.; Miyagi, T. Cloning,
expression, and chromosomal mapping of a human ganglioside sialidase. *Biochem. Biophys. Res. Commun.* **1999**, *261*, 21–27.

[6] Monti, E.; Bassi, M. T.; Bresciani, R.; Civini, S.; Croci, G. L.; Papini, N.; Riboni, M.; Zanchetti, G.; Ballabio, A.; Preti, A.; Tettamanti, G.; Venerando, B.; Borsani, G. Molecular cloning and characterization of *NEU4*, the fourth member of the human sialidase gene family. *Genomics* **2004**, *83*, 445–453.

[7] d’Azzo, A.; Hoogeveen, A.; Reuser, A. J.; Robinson, D.; Galjaard, H. Molecular defect in combined β-galactosidase and neuraminidase deficiency in man. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4535–4539.

[8] Verheijen, F. W.; Palmeri, S.; Hoogeveen, A. T.; Galjaard, H. Human placental neuraminidase. Activation, stabilization and association with β-galactosidase and its ‘protective’ protein. *Eur. J. Biochem.* **1985**, *149*, 315–321.

[9] Galjart, N. J.; Gillemans, N.; Harris, A.; van der Horst, G. T.; Verheijen, F. W.; Galjaard, H.; d’Azzo, A. Expression of cDNA encoding the human “protective protein” associated with lysosomal beta-galactosidase and neuraminidase: homology to yeast proteases. *Cell* **1988**, *54*, 755–764.

[10] van der Horst, G. T.; Galjart, N. J.; d’Azzo, A.; Galjaard, H.; Verheijen, F. W. Identification and *in vitro* reconstitution of lysosomal neuraminidase from human placenta. *J. Biol. Chem.* **1989**, *264*, 1317–1322.

[11] van der Spoel, A.; Bonten, E.; d’Azzo, A. Transport of human lysosomal neuraminidase to mature lysosomes requires protective protein/cathepsin A. *EMBO J.* **1998**, *17*, 1588–1597.

[12] Bonten, E. J.; Campos, Y.; Zaitsev, V.; Nourse, A.; Waddell, B.; Lewis, W.; Taylor, G.; d’Azzo, A. Heterodimerization of the sialidase NEU1 with the chaperone protective protein/cathepsin A prevents its premature oligomerization. *J. Biol. Chem.* **2009**, *284*, 28430–28441.

[13] Carrillo, M. B.; Milner, C. M.; Ball, S. T.; Snoek, M.; Campbell, R. D. Cloning and characterization of a sialidase from the murine histocompatibility-2 complex: low levels of mRNA and a single amino acid mutation are responsible for reduced sialidase activity in mice carrying the Neu1a allele. *Glycobiology* **1997**, *7*, 975–986.

[14] Lowden, J. A.; O’Brien, J. S. Sialidosis: a review of human neuraminidase deficiency. *Am. J. Hum. Genet.* **1979**, *31*, 1–18.

[15] Thomas, G. H. Disorders of glycoprotein degradation: α-mannosidosis, β-mannosidosis, fucosidosis, and sialidosis. In *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed.; Scriver, C. R.; Beaudet, A. L.; Sly, W. S.; Valle, D., Eds.; McGraw-Hill: New York, 2001; pp 3507–3534.

[16] d’Azzo, A.; Andria, G.; Strisciuglio, G.; Galjaard, H. Galactosialidosis. In *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed.; Scriver, C. R.; Beaudet, A. L.; Sly, W. S.; Valle, D., Eds.; McGraw-Hill: New York, 2001; pp 3811–3826.

[17] Taylor, G. Sialidases: structures, biological significance and therapeutic potential. *Curr. Opin. Struct. Biol.* **1996**, *6*, 830–837.

[18] Copley, R. R.; Russell, R. B.; Ponting, C. P. Sialidase-like Asp-boxes: sequence-similar structures within different protein folds. *Protein Sci.* **2001**, *10*, 285–292.

[19] Chaudhuri, I.; Söding, J.; Lupas, A. N. Evolution of the β-propeller fold. *Proteins* **2008**, *71*, 795–803.

[20] Quistgaard, E. M.; Thirup, S. S. Sequence and structural analysis of the Asp-box motif and Asp-box beta-propellers; a widespread propeller-type characteristic of the Vps10 domain family and several glycoside hydrolase families. *BMC Struct. Biol.* **2009**, *9*, 46.

[21] Roggentin, P.; Schauer, R.; Hoyer, L. L.; Vimar, E. R. The sialidase superfamily and its spread by horizontal gene transfer. *Mol. Microbiol.* **1993**, *9*, 915–921.
[22] Chavas, L. M. G.; Tringali, C.; Fusi, P.; Venerando, B.; Tettamanti, G.; Kato, R.; Monti, E.; Wakatsuki, S. Crystal structure of the human cytosolic sialidase Neu2. Evidence for the dynamic nature of substrate recognition. *J. Biol. Chem.* **2005**, *280*, 469–475.

[23] Magesh, S.; Suzuki, T.; Miyagi, T.; Ishida, H.; Kiso, M. Homology modeling of human sialidase enzymes NEU1, NEU3 and NEU4 based on the crystal structure of NEU2: hints for the design of selective NEU3 inhibitors. *J. Mol. Graph. Model.* **2006**, *25*, 196–207.

[24] Lukong, K. E.; Elsliger, M.-A.; Chang, Y.; Richard, C.; Thomas, G.; Carey, W.; Tylki-Szymanska, A.; Czartoryska, B.; Buchholz, T.; Criado, G. R.; Palmeri, S.; Pshezhetsky, A. V. Characterization of the sialidase molecular defects in sialidosis patients suggests the structural organization of the lysosomal multienzyme complex. *Hum. Mol. Genet.* **2000**, *9*, 1075–1085.

[25] Naganawa, Y.; Itoh, K.; Shimamoto, M.; Takiguchi, K.; Doi, H.; Nishizawa, Y.; Kobayashi, T.; Kamei, S.; Lukong, K. E.; Pshezhetsky, A. V.; Sakuraba, H. Molecular and structural studies of Japanese patients with sialidosis type 1. *J. Hum. Genet.* **2000**, *45*, 241–249.

[26] Lukong, K. E.; Landry, K.; Elsliger, M.-A.; Chang, Y.; Lefrancois, S.; Morales, C. R.; Pshezhetsky, A. V. Mutations in sialidosis impair sialidase binding to the lysosomal multienzyme complex. *J. Biol. Chem.* **2001**, *276*, 17286–17290.

[27] Itoh, K.; Naganawa, Y.; Matsuzawa, F.; Aikawa, S.; Doi, H.; Sasagasaki, N.; Yamada, T.; Kira, J.; Kobayashi, T.; Pshezhetsky, A. V.; Sakuraba, H. Novel missense mutations in the human lysosomal sialidase gene in sialidosis patients and prediction of structural alterations of mutant enzymes. *J. Hum. Genet.* **2002**, *47*, 29–37.

[28] Wang, D.; Zaitsev, S.; Taylor, G.; d’Azzo, A.; Bonten, E. Protective protein/cathepsin A rescues N-glycosylation defects in neuraminidase-1. *Biochim. Biophys. Acta* **2009**, *1790*, 275–282.

[29] Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

[30] Ohlson, T.; Wallner, B.; Elofsson, A. Profile–profile methods provide improved fold-recognition: a study of different profile–profile alignment methods. *Proteins* **2004**, *57*, 188–197.

[31] Xiang, Z. Advances in homology protein structure modeling. *Curr. Protein Pept. Sci.* **2006**, *7*, 217–227.

[32] Meindl, P.; Tuppy, H. 2-Deoxy-2,3-dehydroasialic acids. II. Competitive inhibition of *Vibrio cholerae* neuraminidase by 2-deoxy-2,3-dehydro-N-acetylneuraminic acids. *Hoppe-Seyler’s Z. Physiol. Chem.* **1969**, *350*, 1088–1092.

[33] von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418–423.

[34] Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **1997**, *119*, 681–690.

[35] Yoshida, T.; Munei, Y.; Hitaoka, S.; Chuman, H. Correlation analyses on binding affinity of substituted benzenesulfonamides with carbonic anhydrase using ab initio MO calculations on their complex structures. *J. Chem. Inf. Model.* **2010**, *50*, 850–860.

[36] Munei, Y.; Shimamoto, K.; Harada, M.; Yoshida, T.; Chuman, H. Correlation analyses on binding affinity of substituted benzenesulfonamides with carbonic anhydrase using ab initio
MO calculations on their complex structures (II). *Bioorg. Med. Chem. Lett.* **2011**, *21*, 141–144.

[37] Hitaoka, S.; Matoba, H.; Harada, M.; Yoshida, T.; Tsuji, D.; Hirokawa, T.; Itoh, K.; Chuman, H. Correlation analyses on binding affinity of sialic acid analogues and anti-influenza drugs with human neuraminidase using ab initio MO calculations on their complex structures – LERE-QSAR analysis (IV). *J. Chem. Inf. Model.* **2011**, *51*, 2706–2716.

[38] Yoshida, T.; Hitaoka, S.; Mashima, A.; Sugimoto, T.; Matoba, H.; Chuman, H. Combined QM/MM (ONIOM) and QSAR approach to the study of complex formation of matrix metalloproteinase-9 with a series of biphenylsulfonamides – LERE-QSAR analysis (V). *J. Phys. Chem. B* **2012**, *116*, 10283–10289.

[39] UniProt database. http://www.uniprot.org/ (accessed Aug 6, 2012).

[40] Vinogradova, M. V.; Michaud, L.; Mezentsev, A. V.; Lukong, K. E.; El-Alfy, M.; Morales, C. R.; Potier, M.; Pshezhetsky, A. V. Molecular mechanism of lysosomal sialidase deficiency in galactosialidosis involves its rapid degradation. *Biochem. J.* **1998**, *330*, 641–650.

[41] Gaskell, A.; Crennell, S.; Taylor, G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **1995**, *3*, 1197–1205.

[42] Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

[43] PSI-BLAST - Protein Similarity Search. http://www.ebi.ac.uk/Tools/sss/psiblast/ (accessed Aug 6, 2012).

[44] EMBOSS Stretcher - Pairwise Sequence Alignment. http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/ (accessed Aug 6, 2012).

[45] Larsson, P.; Wallner, B.; Lindahl, E.; Elofsson, A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.* **2008**, *17*, 990–1002.

[46] Gibrat, J.-F.; Madej, T.; Bryant, S. H. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **1996**, *6*, 377–385.

[47] MOE (The Molecular Operating Environment) Version 2011.10, software available from Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7. http://www.chemcomp.com.

[48] Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

[49] Lüthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356*, 83–85.

[50] Hata, K.; Koseki, K.; Yamaguchi, K.; Moriya, S.; Suzuki, Y.; Yingsakmongkon, S.; Hirai, G.; Sodeoka, M.; von Itzstein, M.; Miyagi, T. Limited inhibitory effects of oseltamivir and zanamivir on human sialidases. *Antimicrob. Agents Chemother.* **2008**, *52*, 3484–3491.

[51] Woods, J. M.; Bethell, R. C.; Coates, J. A. V.; Healy, N.; Hiscox, S. A.; Pearson, B. A.; Ryan, D. M.; Ticehurst, J.; Tilling, J.; Walcott, S. M.; Penn, C. R. 4-Guanidino-2,4-dideoxy-2,3-dehydro-N-acetylneuraminic acid is a highly effective inhibitor both of the sialidase (neuraminidase) and of growth of a wide range of influenza A and B viruses in vitro. *Antimicrob. Agents Chemother.* **1993**, *37*, 1473–1479.

[52] Magesh, S.; Moriya, S.; Suzuki, T.; Miyagi, T.; Ishida, H.; Kiso, M. Design, synthesis, and biological evaluation of human sialidase inhibitors. Part 1: selective inhibitors of lysosomal sialidase (NEU1). *Bioorg. Med. Chem. Lett.* **2008**, *18*, 532–537.

[53] Russell, R. J.; Haire, L. F.; Stevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hay, A. J.; Gamblin, S. J.; Skehel, J. J. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **2006**, *443*, 45–49.
Hitaoka, S.; Harada, M.; Yoshida, T.; Chuman, H. Correlation analyses on binding affinity of sialic acid analogues with influenza virus neuraminidase-1 using ab initio MO calculations on their complex structures. *J. Chem. Inf. Model.* 2010, 50, 1796–1805.

Weiner, S.; Kollman, P.; Case, D.; Singh, U.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 1984, 106, 765–784.

Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* 2003, 119, 3599–3605.

Maseras, F.; Morokuma, K. IMOMM: a new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* 1995, 16, 1170–1179.

Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* 2003, 119, 3599–3605.

Maseras, F.; Morokuma, K. IMOMM: a new integrated *ab initio* + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* 1995, 16, 1170–1179.

Klatt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* 1995, 99, 2224–2235.

Chandler, D.; Andersen, H. C. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. *J. Chem. Phys.* 1972, 57, 1930–1937.

Gō, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U.S.A.* 1983, 80, 3696–3700.

Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 1993, 215, 617–621.

Scott, A. D.; Phillips, C.; Alex, A.; Flocco, M.; Bent, A.; Randall, A.; O’Brien, R.; Damian, L.; Jones, L. H. Thermodynamic optimisation in drug discovery: a case study using carbonic anhydrase inhibitors. *Chem. Med. Chem.* 2009, 4, 1985–1989.

Case, D. A.; Cheatham, T. E., III.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005, 26, 1668–1688.

Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 2000, 21, 1049–1074.

Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* 2004, 25, 1157–1174.

Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993, 26, 283–291.

Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983, 22, 2577–2637.

Yoshida, T.; Kadota, Y.; Hitaoka, S.; Kori, E.; Horikawa, Y.; Taguchi, M.; Tsuji, D.; Hirokawa, T.; Chuman, H.; Itoh, K. Expression and molecular dynamics studies on effect of amino acid substitutions at Arg344 in human cathepsin A on the protein local conformation. *Biochim. Biophys. Acta* 2009, 1794, 1693–1699.