J Mol Model (2009) 15:323–328
DOI 10.1007/s00894-008-0398-5

ORIGINAL PAPER

# Homology modeling and molecular dynamics study on N-acetylneuraminate lyase

**Hui-Ying Chu · Qing-Chuan Zheng · Yong-Shan Zhao · Hong-Xing Zhang**

Received: 22 July 2008 / Accepted: 1 September 2008 / Published online: 5 December 2008


**Abstract** With homology modeling techniques, molecular mechanics and molecular dynamics methods, a 3D structure model of N-acetylneuraminate lyase from human (hNAL, EC 4.1.3.3) was created and refined. This model was further assessed by Profile-3D and PROCHECK, which confirms that the refined model is reliable. Furthermore, the docking results of the substrates (sialic acid and KDO) into the active site of hNAL indicate that hNAL can cleave the sialic acid and KDO. Thr51 and Tyr143 may be the key amino acids residues as they have strong hydrogen bonding interactions with the substrates, which is in good agreement with the experimental results by Izard et al. (Structure 2:361–369. doi:10.1016/S0969-2126(00)00038-1 (1994)). From the docking studies, we also suggest that Asp176 and Ser218 only form hydrogen bonds with sialic acid, therefore, they may help sialic acid interact with hNAL steadly.

**Keywords** Docking · Homology modeling · N-acetylneuraminate lyase

## Introduction

N-acetyl-neuraminic acid (sialic acid) is an interesting high-carbon sugar with important pharmaceutical implications. It plays a prominent role in numerous biological functions, including virus infections [1, 2]. N-acetylneuraminate lyase (NAL, EC 4.1.3.3) is induced by sialic acid and is controlled as effective for sialic acid of biosynthetic origin as it is for exogenous sialic acid of which may accumulate inside the bacterial cell. NAL has been found in pathogenic as well as non-pathogenic bacteria, and mammalian tissues. In addition, sialic acid is found as an essential sugar residue in many complex carbohydrates associated with inflammatory diseases and cancer metastasis [3].

The NAL from human (hNAL) which consists of 320 amino acids, has a molecular mass of 35.2 kDa. Recent studies suggest that NAL catalyses the reversible aldol reaction of sialic acid, and has been extensively used in the synthesis of sialic acids and its analogues, for example the 3-deoxy-manno-oct-2-ulosonic acid (KDO) [4–7]. Sialic acid, the most preferred substrate, cleaves to N-acetyl-mannosamine and pyruvate, and KDO cleaves to arabinose and pyruvate [8]. Although the investigations of substrate binding mechanisms have been performed by biochemical methods recently, up to now, there are no theoretical efforts being made and the three-dimensional (3D) structure of the hNAL remains to be elucidated. Knowledge of the 3D structure is, however, essential for understanding the catalysis mechanism.

To our best knowledge, the homology modeling is an efficient method for the 3D structure construction of protein [9–11]. In the present study, the 3D feature of hNAL were obtained by a homology modeling procedure based on the crystal structure of NAL from *Escherichia coli*, deposited in Protein Data Bank (PDB code 1NAL) [12]. With homology modeling techniques and molecular dynamics (MD) simulations, a 3D model of hNAL was created and refined. This model can be used to carry out binding study by flexible docking with sialic acid and KDO. The docked complex would be used to identify the key residues for revealing further the ligand reaction mechanism.

H.-Y. Chu · Q.-C. Zheng · Y.-S. Zhao · H.-X. Zhang (✉)
State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130023, People's Republic of China
e-mail: zhanghx@mail.jlu.edu.cn

## Theory and methods

All simulations were performed on SGI O3800 workstations using InsightII software package developed by Accelrys Inc. [13]. Primary sequence of hNAL was obtained from the UniprotKB-Swiss-Prot (http://ca.expasy.org/sprot/, Accession No. Q9BXD5). The consistent-valence forcefield (CVFF) was used for energy minimization (EM) and MD simulations.

### 3D model building

The homology module [14] was used to build the initial model of hNAL.

The first step was searching a number of related sequences to find a related protein as a template. BLAST search algorithm [15] was used for the search on line (http://www.ncbi.nih.gov/). The sequences were imported into the ClustalW program (version 1.83). Program Modeler was performed to build the 3D structure of hNAL. Modeler is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints [16–18]. For the remaining side chains, library values of rotamers were adopted. Through the procedure mentioned above, an initial model was thus completed.

The initial model was improved by EM. After 300 steps of conjugate gradient (CG) minimization performed, the MD simulation was carried out to examine the quality of the model structures by checking their stability via performing 300 ps simulations at a constant temperature 298K. The explicit solvent model TIP3P water was used, and the homology solvent model was constructed with in layers with 10 Å. Finally, CG energy minimization of full protein was performed until the root mean-square (rms) gradient energy was lower than $1\times10^{-5}$ kcal $mol^{-1}$ $Å^{-1}$. All calculations mentioned above were accomplished by using Discover3 software package [19]. In this step, the quality of initial model was improved.

After the optimization procedure, the structure was checked using Profile-3D [20] and PROCHECK [21]. The Profile-3D method measured the compatibility of an amino acid sequence with a known 3D protein structure. This is especially useful in the final phase of the protein structure modeling.

### Binding-site analysis

The binding-site module is a suite of programs in InsightII for identifying and characterizing protein active sites, binding sites, and functional residues from protein structures and multiple sequence alignments. In this study, ActiveSite-Search was used to identify protein active sites and binding sites by locating a cavity in the hNAL structure. When the search was completed, the largest site was automatically displayed on the structure. And then, by using Asite-Display, other sites were also obtained. The results can be used to guide the protein-ligand docking experiment.

### Docking ligands to hNAL

Affinity, which uses a combination of Monte Carlo type and simulated annealing procedure to dock, is a suite of programs for automatically docking a ligand (guest) to a receptor (host) [22]. By means of the 3D structure of sialic acid and KDO which were obtained from Gaussian03 program [23], the automated molecular docking was performed by using docking program Affinity. A key feature is that the "bulk" of the receptor, defined as atoms which are not in the binding site specified, is held rigid during the docking process, while the binding site atoms and ligand atoms are movable. The potential function of the complex was assigned by using the CVFF and the cell multiple approaches was used for nonbonding interactions. To account for the solvent effect, the centered enzyme-ligand complex was solvated in layers of TIP3P water molecules with 5 Å. Finally, the docked complex of hNAL with sialic acid, and KDO was selected by the criteria of interacting energy combined with the geometrical matching quality. The complex was used as the staring conformation for further energetic minimization and geometrical optimization before the final model was generated. The global structure with the lowest energy was chosen for computing intermolecular binding energies. The Ludi program was used to characterize the affinity and the binding preference of a ligand to the protein.

## Results and discussion

### Homology modeling of hNAL

The sequences analysis and alignment with known proteins from GenBank show a high similarity of the N-acetylneuraminate lyase subfamily from various bacteria. In the result of the BLAST searching, the high sequence identity between the hNAL and 1NAL is 30% which allows for rather straightforward sequence alignment (see Fig. 1). It was reported that the sequence of hNAL is that of the precursor protein which appears to have a signal peptide sequence. By using SignalP software, we can predicate that a possible cleavage site for a signal peptidase is present between the amino acid residue Asn22 and Gly23. In this paper, the model is made up of residues from 23 to 320. Automated homology model building was performed using protein structure modeling program Modeler. All the side

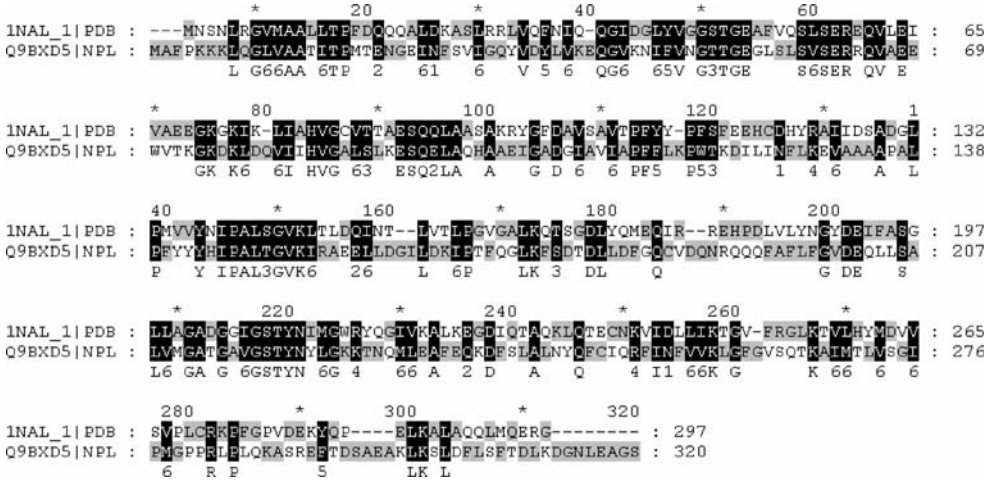


**Fig. 1** Sequence alignment of hNAL and 1NAL

chains of model protein were set by rotamers. With this procedure, the initial model was complete. This model was refined by molecular mechanism (MM) optimization and MD simulations, and then the final stable structure of hNAL was obtained as displayed in Fig. 2. From Fig. 2, we can see that this enzyme has 14 helices and 6 sheets. The conformation with the lowest energy was chosen and the 3D structure was superimposed with 1NAL. Their root mean square deviation (RMSD) value is 0.48 Å, indicating a good overall structure alignment with 1NAL. The final structure with the lowest energy was checked by Profile-3D (Fig. 3) and the self-compatibility score for this protein is 108.32, which is higher than the low score 61.02 and closed to the top score 135.59. Note that compatibility scores above zero correspond to 'acceptable' side chain environment. From Fig. 3, we can see that all residues are reasonable. And then, the structure of hNAL was evaluated using PROCHECK. Among the 298 residues, no residue was found in the disallowed regions of Ramachandran plot (Fig. 4). The statistical score of the Ramachandran plot shows that 89.3% are in the most favored regions, 9.9% in the additional allowed regions, and 0.8% in generously allowed regions. The above results indicate that the homology model is reliable.

## Identification of binding site in hNAL

In order to investigate the interaction between hNAL and substrates, the binding pocket was defined as a subset that contains residues in which any atoms within 5 Å from substrate. The active sites were obtained using binding site module. It was reported that N-acetylneuraminate lyase subfamily has the NAL motif and this motif is involved in substrate binding and catalysis [12]. In hNAL, the conserved NAL motif forms a deep pocket, and the hNAL share a common α/β-barrel motif. There are ten residues (Thr52, Gly53, Tyr143, Ile145, Ala147, Leu148, Lys173, Gly199, Asp201, Glu202, Ser218, Thr219), which are highly conserved in the sequences of NAL proteins from animals and bacteria. It is obvious that the chosen site is in good agreement with the conclusion drawn by Izard et al. [12]. This site is composed of 21 residues (Phe47, Thr51-Thr-Gly53, Ile111, Tyr143-His-Ile145, Ala147-Leu148, Lys173, Ser175-Asp-Thr177, Phe198-Gly-Val-Asp-Glu202, Ser218-Thr219). Based on the experiment and our theoretical predicted results, in this study this site is chosen as the more favorable binding site to dock the ligand.

## Docking study

As reported that hNAL catalyses the reversible aldol reaction. Since the active site of NAL is composed of the $(\alpha/\beta)_8$-barrel, it represents an ideal structurally motif for

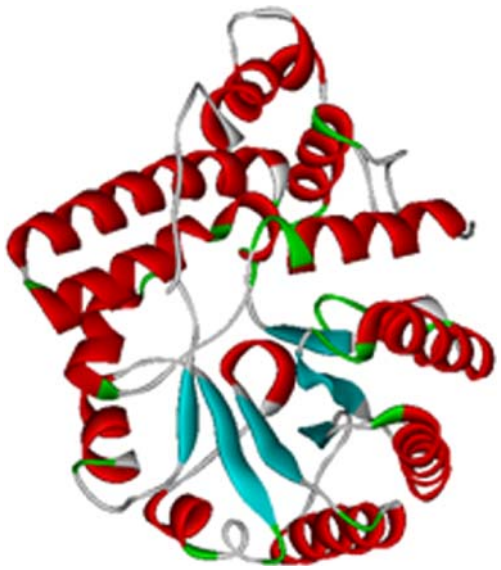

**Fig. 2** The final 3D-structure of hNAL. The α-helix is represented by red color, the β-sheet is represented by cyan color and the β-turn is represented by green color

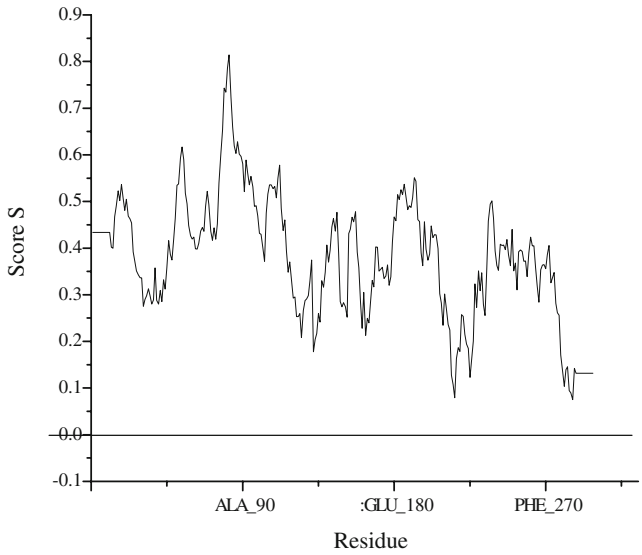


**Fig. 3** The 3D profiles verified results of hNAL model, residues with positive compatibility score are reasonably folded

directed evolution. As mentioned before, NAL from *E.coli* has different substrate preferences over sialic acid, and KDO, with the relative activates of 100 and 3.7%, respectively [8]. We tried to explain the substrate specificity of hNAL based on its 3D structure and docking experiment. The 3D structures of the two substrates are shown in Fig. 5.

To understand the interaction between hNAL and sialic acid, the complex of hNAL and sialic acid was generated by Affinity module and the binding 3D conformation of the complex is described in Fig. 6. As is well known, hydrogen

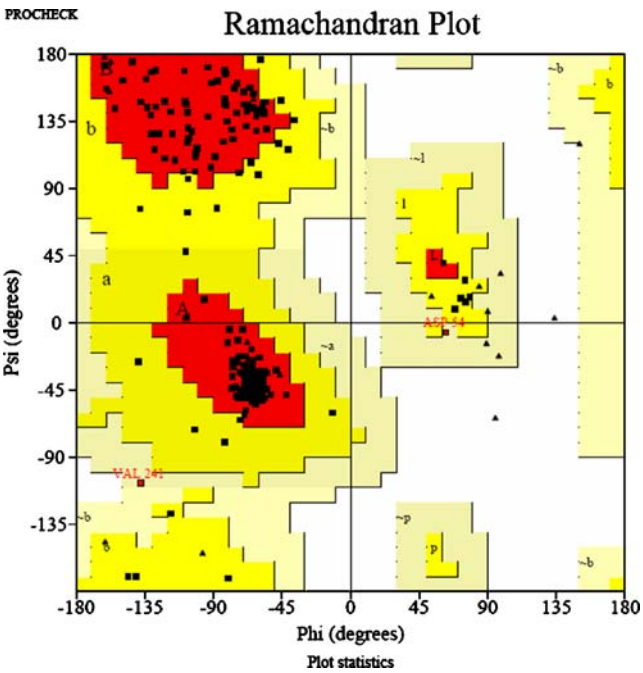


**Fig. 4** Ramachandran plot of hNAL obtained by PROCHECK

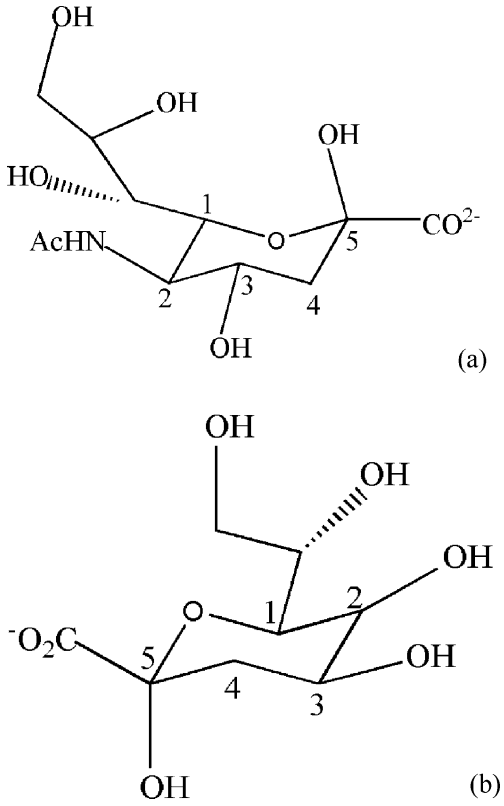


**Fig. 5** The structure of the substrates: (**a**) sialic acid, (**b**) KDO

bonds play an important role for structure and function of biological molecules, especially for the enzyme catalysis. To determine the key residues that comprise the active site of the model, the interaction energies of the substrate with each of the residues in the active site were calculated. Significant binding-site residues in the models were identified by the total interaction energy between the substrate and each amino acid residues in the enzyme. This identification, compared with a definition based on the distance from the substrates, can clearly show the relative significance for every residue. Table 1 gives the interaction

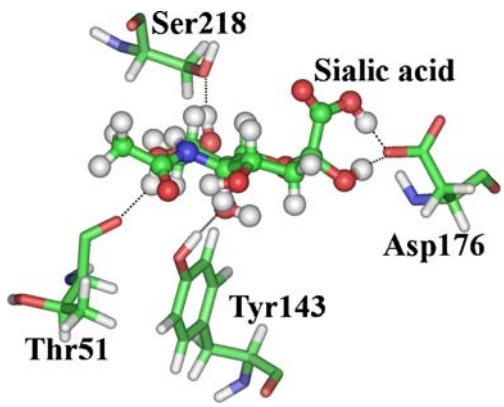


**Fig. 6** The hydrogen bonding interaction in complex of hNAL with sialic acid

**Table 1** The total energy ($E_{total}$), Van Der Waals ($E_{vdw}$) and electrostatic energy ($E_{ele}$) between individual residues of hNAL and sialic acid or KDO (kcal $mol^{-1}$)

| Residue | Sialic acid | | | KDO | | |
|---|---|---|---|---|---|---|
| | $E_{vdw}$ | $E_{ele}$ | $E_{total}$ | $E_{vdw}$ | $E_{ele}$ | $E_{total}$ |
| Total | −24.97 | −72.45 | −97.42 | −20.02 | −45.66 | −65.68 |
| Thr51 | −1.57 | −3.65 | −5.22 | −0.02 | −5.21 | −5.23 |
| Thr52 | −3.18 | −0.29 | −3.47 | −2.36 | −1.01 | −3.37 |
| Tyr143 | −1.21 | −5.90 | −7.11 | −1.36 | −5.59 | −6.95 |
| Ile145 | −2.81 | 0.27 | −1.94 | −2.17 | 0.01 | −2.16 |
| Leu148 | −3.63 | 0.10 | −3.53 | −2.40 | 0.08 | −2.32 |
| Asp176 | 9.99 | −47.47 | −37.28 | 4.64 | −16.35 | −11.71 |
| Val200 | −3.77 | 1.09 | −2.68 | −2.79 | 0.88 | −1.91 |
| Asp201 | −0.58 | −5.41 | −5.99 | −0.81 | −7.70 | −8.51 |
| Glu202 | −0.19 | −1.14 | −1.33 | −0.10 | −0.51 | −0.61 |
| Ser218 | −0.34 | −5.00 | −5.34 | 0.56 | −4.32 | −3.76 |

energies including the total, van der Waals and electrostatic energies with total energies lower than −0.6 kcal $mol^{-1}$. There are five hydrogen bonds between hNAL and sialic acid: the carboxyl O of Asp176 interacts with the hydroxyl H and carboxyl H of 5' C in sialic acid, which has two strong hydrogen bonds (1.37 Å, 1.38 Å, −37.28 kcal $mol^{-1}$). The hydroxyl O of 1'C in the sialic acid is tightly bound to the hydroxyl H of Thr51 by a hydrogen bond (1.08 Å, −5.22 kcal $mol^{-1}$). Another important finding from the docking simulations suggests that the hydroxyl H of Tyr143 with the hydroxyl O of 1'C in sialic acid (1.61 Å, −7.11 kcal $mol^{-1}$) and the hydroxyl O of Ser218 with hydroxyl H of 1'C in sialic acid (1.63 Å, −5.34 kcal $mol^{-1}$) also form hydrogen bonds. These interactions determine a stable binding mode for enzyme-substrate complex ($E_{vdw}$= −24.97 kcal $mol^{-1}$, $E_{ele}$=−72.45 kcal $mol^{-1}$, $E_{total}$= −97.42 kcal $mol^{-1}$). Through the interaction analysis, we can know that Thr52 (−3.47 kcal $mol^{-1}$), Ile145 (−1.94 kcal $mol^{-1}$), Leu148 (−3.53 kcal $mol^{-1}$), Val200 (−2.68 kcal $mol^{-1}$), Asp201 (−5.99 kcal $mol^{-1}$), Glu202 (−1.33 kcal $mol^{-1}$) are important anchoring residues for sialic acid, and have main contribution to the substrate interaction. All the residues mentioned above are all conserved in the N-acetylneuraminate lyase subfamily except Val200. Thus, the residues of Thr51 and Tyr143 may be important residues, which are in good agreement with the suggestion by Izard et al. [12], and the Asp176 and Ser218 may help sialic acid interact with hNAL steady. On the basis of these observations, sialic acid binds to hNAL with higher ligand/receptor interaction energy of −97.42 kcal $mol^{-1}$.

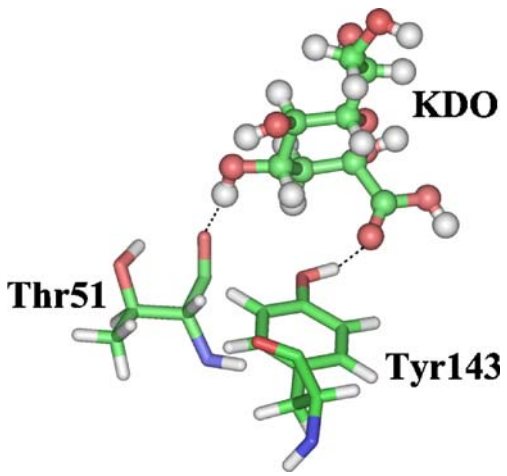


**Fig. 7** The hydrogen bonding interaction in the complex of hNAL with KDO

In the case of affinity docking simulations, orientation within the active site pocket of KDO is opposite to those of the hNAL and sialic acid, but the position is similar (see Fig. 7). There are only two hydrogen bonds between hNAL and KDO. The amino acids that recognize and interact with KDO are Thr51 and Tyr143. The weak intermolecular hydrogen bonds, which are the carbonyl O of Thr51 with the hydroxyl H of 2'C in KDO (1.99 Å, −5.23 kcal $mol^{-1}$) and the hydroxyl H of Tyr143 with hydroxyl O of 3' C in KDO (1.92 Å, −6.95 kcal $mol^{-1}$) are presented in the complex. The absence of critical hydrogen bond interactions with Asp176 and Ser218, and weak hydrogen bond interaction of KDO with Thr51 and Tyr143, presumably lead to the interaction energy of KDO with the active-site residues of hNAL (−65.68 kcal $mol^{-1}$) that is lower than that in the complex of hNAL with sialic acid. Through the interaction analysis (In Table 1), we know that Thr52 (−3.37 kcal $mol^{-1}$), Ile145 (−2.16 kcal $mol^{-1}$), Leu148 (−2.32 kcal $mol^{-1}$), Asp176 (−11.71 kcal $mol^{-1}$), Val200 (−1.91 kcal $mol^{-1}$), Asp201 (−8.51 kcal $mol^{-1}$), Ser218 (−3.76 kcal $mol^{-1}$) are important anchoring residues for KDO, and have a main contribution to the substrate interaction. All the residues mentioned above are all conserved in the N-acetylneuraminate lyase subfamily except Val200.

For the docked structures, the interaction energies order are sialic acid > KDO. For the Ludi scores, the interaction energies order are sialic acid > KDO (In Table 2). These results qualitatively validate the modeling and docking studies described for this subfamily.

**Table 2** The interaction energies (kcal $mol^{-1}$) between the substrates and hNAL

| Substrate | Evdw (kcal $mol^{-1}$) | Eele ($kcalmol^{-1}$) | Etotal ($kcalmol^{-1}$) | Ludi Score | No. of H-bonds | Relative Activeity (%)* |
|---|---|---|---|---|---|---|
| Sialicacid | −24.97 | −72.45 | −97.42 | 278 | 5 | 100 |
| KDO | −20.02 | −45.66 | −65.68 | 147 | 2 | 3.7 |

* Relative activity is proposed by Wada et al. [8]

Compared with sialic acid and KDO, they have the same main moiety, but the only difference is the side chain of 2'C, which is acetic amino (AcNH) group in sialic acid and hydroxyl group in KDO. They can both be cleaved by hNAL. The AcNH group in the sialic acid is a strong attractive group and can induce the conformational changes in the active site residues of hNAL. We can also find that as a result of AcNH group existing, leading to the orientation of substrate is opposition in all. In the case of KDO, it essentially has poor interactions with hNAL, mainly because it lacks AcNH group and loosely binds to the enzyme, not well fixed in the substrate-binding pocket of the active site. Thus, this explains why the complex of KDO and hNAL has lower interaction. The results are in good agreement with the experiment reported by Wada et al. [8].

## Conclusion

The 3D structure of N-acetylneuraminate lyase from human (hNAL) has not been known. In this investigation, the 3D structure of hNAL was built by homology modeling, which was based on the known crystal structure of N-acetylneuraminate lyase from *E.coli* (PDB code 1NAL). And then energy minimization and molecular dynamics were used to refine the structure. With this model, a flexible docking study was performed and the docking results indicate that the AcHN group in sialic acid can stabilize the position and orientation in the active site of hNAL. Thr51 and Tyr143 may be the key amino acids residues interacting with the substrates, and Asp176 and Ser218 may help sialic acid interact with hNAL steady. These results will offer further experimental studies of structure-function relationships.

**Acknowledgement** This work is supported by the Natural Science Foundation of China, Key Projects in the National Science & Technology Pillar Program, and Specialized Research Fund for the Doctoral Program of Higher Education(Grant Nos. 20573042, 2006BAE03B01, and 20070183046)

## References

1. Kageshita T, Hirai S, Kimura T, Hanai N, Ohta S, Ono T (1995) Cancer Res 55:1748–1751
2. Vimr E, Lichtensteiger C (2002) Trends Microbiol 10:254–257. doi:10.1016/S0966-842X(02)02361-2
3. Sears PP, Wong CH (1998) Cell Mol Life Sci 54:223–252. doi:10.1007/s000180050146
4. Lin CC, Lin CH, Wong CH (1997) Tetrahedron Lett 38:2649–2652. doi:10.1016/S0040-4039(97)00442-5
5. Kragl U, Goedde A, Wandrey C, Lubin N, Auge C (1994) J Chem Soc, Perkin Trans 1:119–124. doi:10.1039/p19940000119
6. Machajewski TD, Wong CH (2000) Angew Chem Int Ed 39:1352–1374. doi:10.1002/(SICI)1521-3773(20000417)39:8<1352::AID-ANIE1352>3.0.CO;2-J
7. Kim MJ, Hennen WJ, Sweers HM, Wong CH (1988) J Am Chem Soc 110:6481–6486. doi:10.1021/ja00227a031
8. Wada M, Hsu CC, Franke D, Mitchell M, Heine A, Wilson I et al. (2003) Bioorg Med Chem 11:2091–2098. doi:10.1016/S0968-0896(03)00052-X
9. Zheng QC, Li ZS, Sun M, Zhang Y, Sun CC (2005) Biochem Biophys Res Commun 333(3):881–887. doi:10.1016/j.bbrc.2005.05.169
10. Xu W, Cai P, Yan M, Xu L, Ouyang PK (2007) Chem. J. Chin Univ 28(5):971–973
11. He YP, Hu HR, Xu LS et al (2005) Chem. J. Chin Univ 26 (2):254–258
12. Izardl T, Lawrence MC, Malby RL, Glenn GL, Colman PM (1994) Structure 2:361–369. doi:10.1016/S0969-2126(00)00038-1
13. Insight II User Guid, San Diego: Biosym/MSI (2000)
14. Insight II Homology User Guide, San Diego: Accelrys Inc (2000)
15. Altschul SF, Madden TL, Schfer AA, Zhang JZ, Miller W, Lipman DJ (1997) Nucleic Acids Res 25:3389–3402. doi:10.1093/nar/25.17.3389
16. Sali A, Overington JP (1994) Protein Sci 3(9):1582–1596
17. Sali A, Potterton L, Yuan F, Vlijmen H, Karplus M (1995) Proteins 23(3):318–326. doi:10.1002/prot.340230306
18. Sali A (1995) Curr Opin Biotechnol 6(4):437–451. doi:10.1016/0958-1669(95)80074-3
19. InsightII Discover3 User Guide, San Diego: Biosym/MSI (2000)
20. Insight II Profile-3D User Guide, San Diego: Biosym/MSI (2000)
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–219. doi:10.1107/S0021889892009944
22. InsightII Affinity User Guide San Diego: Biosym/MSI (2000)
23. Frisch MJ, Trucks GW Schlegel HB Gaussian 03 (Revision A.1) Gaussian Pittsburgh (2003)