J Mol Model (2010) 16:975–991
DOI 10.1007/s00894-009-0606-y

ORIGINAL PAPER

Site-directed fragment-based generation of virtual sialic acid databases against influenza A hemagglutinin

Mohammed Noor Al-qattan · Mohd Nizam Mordi

Received: 5 August 2009 / Accepted: 30 September 2009 / Published online: 25 October 2009

Abstract In this study fragment-based drug design is combined with molecular docking simulation technique, to design databases of virtual sialic acid (SA) analogues with new substitutions at C2, C5 and C6 positions of SA scaffold. Using spaces occupied by C2, C5 and C6 natural moieties of SA when bound to hemagglutinin (HA) crystallographic structure, new fragments that are commercially available were docked independently in all the pockets. The oriented fragments were then connected to the SA scaffold with or without incorporation of linker molecules. The completed analogues were docked to the whole SA binding site to estimate their binding conformations and affinities, generating three databases of HA-bound SA analogues. Selected new analogues showed higher estimated affinities than the natural SA when tested against H3N2, H5N1 and H1N1 subtypes of influenza A. An improvement in the binding energies indicates that fragment-based drug design when combined with molecular docking simulation is capable to produce virtual analogues that can become lead compound candidates for *anti*-flu drug discovery program.

Keywords Fragment-based molecular design · Hemagglutinin · Influenza A · Molecular docking · Sialic acid analogues

M. N. Al-qattan · M. N. Mordi (✉)
Centre for Drug Research, Universiti Sains Malaysia,
11800 Minden, Penang, Malaysia
e-mail: mnizam@usm.my

Introduction

Influenza A virus is an enveloped negative strand RNA virus which belongs to the *Orthomyxoviridae* family. HA and NA are two surface glycoproteins that determine the viral antigenicity. SA is the natural ligand for both of the glycoproteins and located at the host cell membrane [1–3]. The influenza A infection is initiated by attaching viral HA to host cell SA moiety and followed by endocytosis [4]. The infectious cycle ends up with the hydrolysis of O-glycosidic bond that connects SA to the host cell membrane by NA [5], thus releasing the progeny viruses. The amino acids of both of HA1 and NA binding sites are mutagenically conserved for most influenza A strains [6, 7]. Pyranose ring is the scaffold of SA molecule, to which different functional groups are connected through different carbon atoms, *i.e.*, C2, C4, C5, and C6 (Fig. 1).

NA is attractive for drug design because of the relative deep active site in which low molecular weight inhibitors can make multiple favorable interactions. Rational modifications of SA based on the NA crystal structure has been carried out and successfully produced new class of *anti*-influenza therapy that inhibit NA and preventing progeny viral dissemination [8]. Currently Zanamivir and Oseltamivir are the clinically approved NA inhibitors.

On the other hand, HA is less attractive due to its relatively shallow sialic acid-binding sites. HA is arranged as homotrimer on the viral surface. Each monomer has two binding sites, the first occurs at the domain 1 (HA1) which is responsible for viral-host cell sticking while the second occurs at the interface between HA1 and domain 2 (HA2) [9–11]. Since HA1 is not an enzyme, the interaction between SA and HA1 is uncomplicated when compared to NA, for examples, protein's conformational changes are

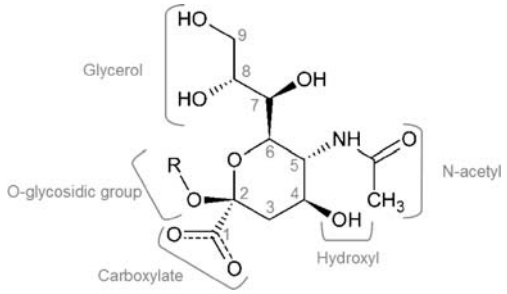

Fig. 1 Natural SA (R=H), SA analogue, *i.e.*, methyl-α-Neu5Ac ($R=CH_3$)

limited and bond breakage/formation does not take place at HA1 binding site [5].

Various SA analogues have been synthesized to inhibit HA1 and the inhibitory potency of the analogues have been studied using various techniques such as viral hemadsorption inhibition assays [12, 13], viral hemagglutination inhibition assays [14], and nuclear magnetic resonance (NMR) [10, 15, 16]. Thus far, no effective HA inhibitor has been designed, which could be attributed to the low affinity between many SA analogues due to shallow binding site of HA1. Due to the current pandemic of H1N1 and possibility of H5N1 resurgence, and the resistance that has been developed to current treatment [17], HA1 can be used as an alternative target to combat influenza A virus.

Earlier, we have reported that computer simulation can be used to reproduce the crystal conformation of SA and the experimental affinities of SA analogues against HA1 using molecular docking technique (Fig. 2) [18]. In this paper, we demonstrate the design of high affinity virtual SA analogues against influenza A HA1 using molecular docking and fragment-based molecular design techniques. The natural SA functional groups at C2, C5, and C6

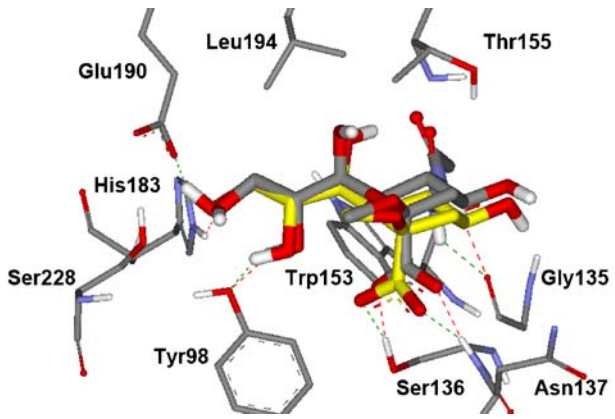

Fig. 2 Superimposed conformations of the crystal (gray) and docked (yellow) methyl-α-Neu5Ac at HA1 binding pocket of H3N2 (X-31) influenza A virus. Intermolecular H-bonds are shown as dashed lines

positions will be replaced by commercially available molecular fragments to generate single-substituted SA analogues databases of high affinity.

Methods

Overview

Databases of virtual SA analogues were generated based on the HA1 crystal structure to guide the natural SA functional groups substitutions (site-directed). Commercially available molecular fragments have been used as substituent for the analogue candidates in order to facilitate the synthetic accessibility (fragment-based).

SA analogues were assembled from two molecules. The first molecule comprises the crystal SA scaffold minus the functional group to be substituted by a molecular fragment. The second molecule comprises the docked (oriented) fragment to be used for substituting one of the natural SA functional groups. The fragments were docked against the crystal binding site of natural SA functional group in order to adopt the best binding conformation and affinity (orientation). Therefore, both of the crystal SA scaffold's molecule and the oriented fragment's molecule are in best binding conformations and affinities toward the HA1 binding pocket. Finally, both molecules were connected with the aim to preserve the initial binding conformations of both molecules. Empirical algorithm were developed to guarantee reliable automatic intermolecular connections. The work flow of the methods is shown in Fig. 3.

Fragments preparation

The fragments are commercially available from ChemBridge EXPRESS-Pick™ (http://chembridge.com/chembridge/data.html) and was downloaded as a single SDF formatted file after obtaining login username and password. The library has 4541 molecular fragments with physicochemical characteristics listed in Table 1. The library was cleaned from non-structural information, then processed using babel-1.1 software (http://smog.com/chem/babel/) to produce individual fragments files in Mdl Mol format. As some of the fragments have multiple bonded halogens or supplied as salts, the halogens were removed to get the pure fragments while the salt atomic coordinates were removed to exclusively extract the fragments atomic coordinates.

Generally the fragments molecules in the downloaded database were not at minimal energy conformation. Thus after salts and halogens were removed, an in-house Tool Command Language (TCL) programming script was used to add the missing hydrogen atoms and optimize the

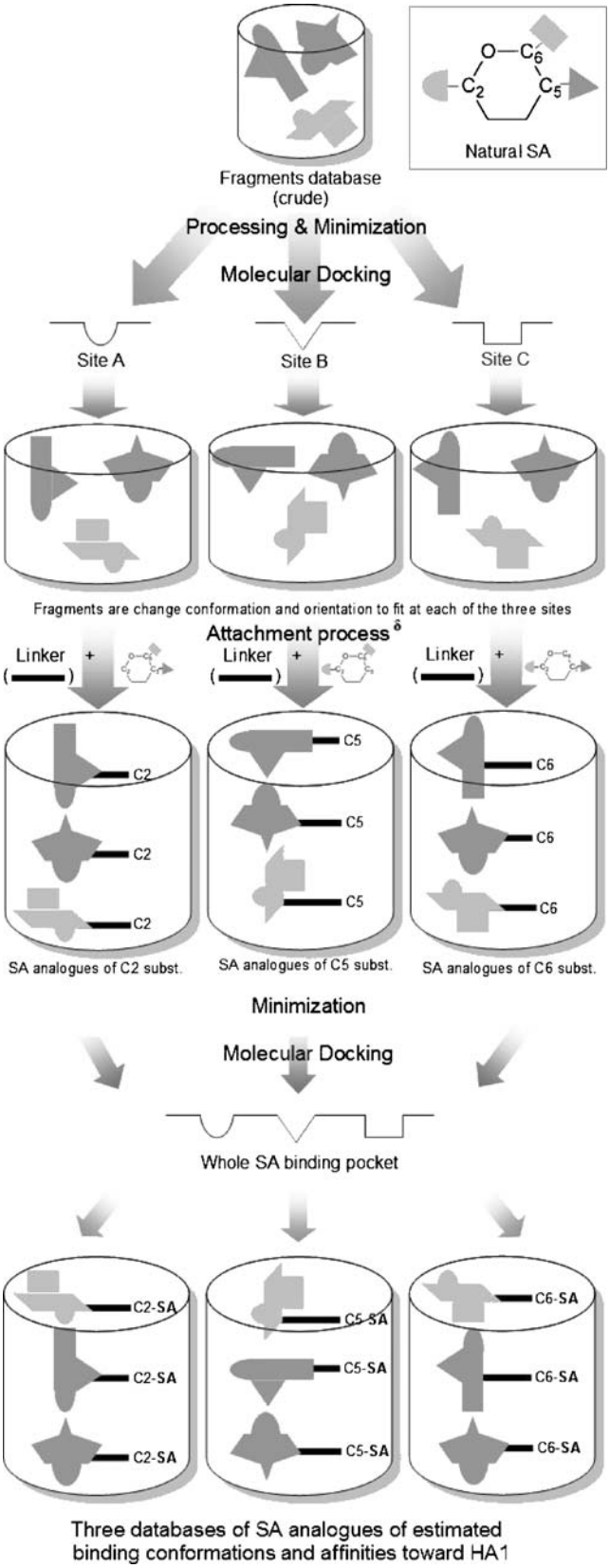

Fig. 3 Generation of SA analogue databases using site-directed fragment-substitution approaches. δ The attachment between the oriented fragments and the SA scaffold can be without linker molecules in some cases

conformation using HyperChem 6.01. Geometry optimization was performed with steepest descent followed by Polak Ribiere methods using AMBER force field [19]. Normalization of the atomic coordinates and addition of partial atomic Gasteiger charges were carried out using Vega ZZ 2.2.0 (http://www.vegazz.net/). PDBQ files were subsequently prepared using Autotors (An AutoDock utility which is used to define the rotatable bonds within the ligand, if any, and unite the charges of nonpolar hydrogens with carbon atoms bearing them).

Fragments docking (orientation)

The crystal structure for HA of influenza A (H3N2) X-31 serotype in complex with methyl-α-Neu5Ac has been used (PDB ID=1HGH). The three sites within and around HA1 binding pocket which represent the binding sites for C2-, C5-, and C6-functional groups of methyl-α-Neu5Ac and other crystallographically studied SA analogues has been chosen as fragments docking sites (Fig. 4). Site A comprises the amino acid residues that form the groove next to the natural SA binding pocket where large C2-substitutions of some SA analogues were found to extend. Site B comprises the binding sites of C5-N-acetyl group and C4-hydroxyl group while site C comprises the binding site of C6-glycerol. The extension of site B to involve the SA C4-functional group was to provide larger volume for the docked fragments to manipulate conformation while settling against the docking site.

The prepared fragments were docked using AutoDock3.05 against the predefined sites of HA1. Lamarckian Genetic Algorithm (LGA) using pseudo-Solis and Wets local search method [20] was used to explore the fragments conformational space to find the global minimal conformations and orientations. Due to the need to dock the whole fragments library three times (once to each of the three sites), docking parameters were chosen to allow fast docking calculations. The initial atomic coordinates, quaternion and dihedral angles were random and the step size for molecular transition movement was set at 1 Å, and for quaternion and dihedral torsions were 50 degrees. As LGA was used, the number of individuals within the population was defined at 150, while number of generations was 27,000, and number of maximum energy evaluations was 250,000. LGA was allowed to run 10 times before the best docked conformation was chosen. Other docking parameters were kept as default. As the fragments docking stage involve 13,623 individual docking experiment (4541 fragments X 3 docking sites), batch preparation of DPF files were conducted using in-house Unix Bash-shell programming script. Three databases of differently aligned fragments were obtained from the docking steps.

Table 1 Physicochemical properties of 4541 fragments of no bonded halogens and salts to be used for C2, C5, and C6 SA substitutions

Characteristic	Lower limit	Upper limit	average
Molecular weight (MW)	60	300	~215
Calculated logarithm of octanol/water partition coefficient (clogP)	−8.45	3	~1.26
Topological polar surface area (tPSA) (Å^2)	3.24	119.93	~47.11
Logarithm of intrinsic water solubility (logSW) (Mole/Litre)	−3.46	0.99	~1.65
Hydrogen-bond donor (HBD)	0	4	~1.09
Hydrogen-bond acceptor (HBA)	0	6	~2.35
Number of torsional degrees of freedom (TDOF)	0	3	~1.81

Generation of three databases of SA analogues from the oriented fragment databases

The three databases of oriented fragments obtained from the previous step were used to generate three databases of SA analogues. This was carried out by substituting the natural functional groups of crystal methyl-α-Neu5Ac at atoms C2, C5, and C6 with the oriented fragments. Thus, one hydrogen atom of the oriented fragment was removed and substituted by a covalent bond to the SA scaffold's atom (C2, C5 or C6) to generate a single SA analogue molecule within the binding pocket of HA1. Linker molecules were incorporated to connect distantly oriented fragments to the corresponding scaffold atoms.

The fragment-scaffold attachment process was carried out automatically using an in-house Unix Bash-shell programming script which requires pre-prepared scaffold molecules, linker molecules, and the three databases of oriented fragments. The linker molecules have been derived from the natural functional groups of the scaffold atoms of C2, C5, and C6 in order to preserve some important intermolecular interactions between SA analogues and HA1 (Table 2). In order to be at local minima, each linker molecule was connected individually to the corresponding crystal SA scaffold atom and the conformational search was performed strictly for linker atoms to achieve the most stable alignment within or around the HA1 binding pocket.

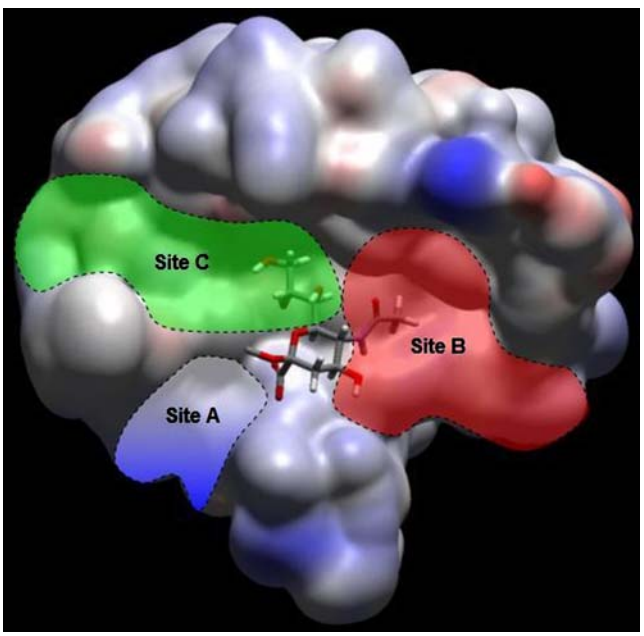

Fig. 4 Part of crystallographic structure of influenza A hemagglutinin (PDB ID=1HGH) shows methyl-α-Neu5Ac (sticks representation) within the binding pocket of HA1. C2-functional group extends over site A, C5-functional group extends to site B, and C6-functional group extends to site C

Fragments-scaffolds attachment algorithm

Empirical algorithm was developed and implemented into a computer program to properly choose the fragment's heavy atom to be connected to the corresponding SA scaffold's atom, and the linker molecule to be incorporated in the connection (if needed). For each pair of oriented fragment and SA scaffold's atom, the program screens all the available modes of connection by incorporating all the available linker molecules. Each connection mode is scored and the best scored mode is established. The fragment attachment process involved several steps and are discussed below.

Step 1: Determination of anchor atoms (AnAs). Anchor atoms (AnAs) are defined as all atoms that are available to make a covalent bond to the oriented fragment. They are comprised of scaffold's atom as well as the terminal atoms of its linker molecules. Therefore, 12 AnAs were available for connecting to the oriented fragment at site A (C2 scaffold atom + terminal atoms of the 11 linker molecules). Similarly, 4 AnAs were available for connecting to the oriented fragment at site B (C5 scaffold atom + terminal atoms of the 3 linker molecules), and 7 AnAs were available for connecting to the oriented fragments at site C (C6 scaffold atom + terminal atoms of 6 linker molecules). During the process of fragment's

Table 2 Molecular linkers used to attach the oriented fragments to the corresponding SA scaffold atoms. During attachment process, $^{\delta}$ the CH_2 within bracket will substitute the NH_2 group, while $^{\kappa}$ the H within bracket will substitute OH group whenever it is necessary (see step 6 of fragments-scaffolds attachment algorithm)

Scaffold atom	Linkers
C2	-O-
	-O-CH_2-
	-O-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-
	-O-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-CH_2-
$C5^{\delta}$	-NH(CH_2)-
	-NH-CH_2-
	-NH-CH_2-CH_2-
$C6^{\kappa}$	-CH_2-
	-CH_2-CHOH(H)-
	-CH_2-CHOH(H)-CH_2-
	-CH_2-CHOH(H)-CH_2-CH_2-
	-CH_2-CHOH(H)-CH_2-CH_2-CH_2-
	-CH_2-CHOH(H)-CH_2-CH_2-CH_2-CH_2-

connection to SA scaffold, each of the available AnAs has specific spatial position (*i.e.*, X, Y, and Z coordinates) which basically determines it's fitness for the connection.

Step 2: Specifying the best fragment hydrogen atoms that could be substituted by covalent bonds to AnAs and measuring the corresponding inter-atomic distances. This step was employed to determine which fragment's hydrogen atom to be substituted by covalent bond to the anchor atom (AnA) for all the available AnAs. A systematic search was carried out to check the distances between each of the available AnAs and all hydrogen atoms from the fragment molecule. The nearest hydrogen atom to each of the AnAs was specified and regarded as the best atom to be substituted by covalent bond to that AnA. The distances between the AnA and both of the fragment's nearest hydrogen and the fragment's heavy atom (HvA) that carries the nearest hydrogen were measured for all the available AnAs (Fig. 5).

Step 3: Converting the measured distances into deviation values in order to calculate the crude score for each AnA. The distances between AnAs and the fragment's nearest hydrogens as well as the distances between AnAs and the fragment's HvAs that carry the nearest hydrogens cannot be used directly to select the best AnA to be bonded to the fragment. Accordingly, the distances were converted to deviation values from typical distances. The typical distances have been retrieved from real inter-atomic bond lengths and represent the best locations for AnA in order to establish a covalent bond with the fragment's HvA. The typical distance between AnA and the fragment's HvA in order to establish a covalent bond between them was stated to be of 1.5 Å (which is about the average distance of C-C, C-O, C-N, and C-S bonds), while the typical distance between AnA and the fragment's hydrogen in order to substitute the latter by the former was stated to be of 0.5 Å (which is 1.5 Å minus the average distance of C-H, O-H, N-H, and S-H bonds). As a result, each AnA had two deviation values; one is for fragment's nearest hydrogen atom while the other is for the fragment's HvA that carries the nearest hydrogen atom. The smaller the deviation values are; the better is the spatial location of AnA to establish a covalent bond with that particular fragment's HvA (Fig. 6). Both deviation values were used in the selection of best AnA from all the available AnAs to be connected to the fragment (*i.e.*, the best connection mode). Using the second deviation value (*i.e.*, of fragment's best HvA) as the only selection criterion between the available AnAs, frequently produces binding modes incompatible with the hybridization state of the fragment's HvA. On the

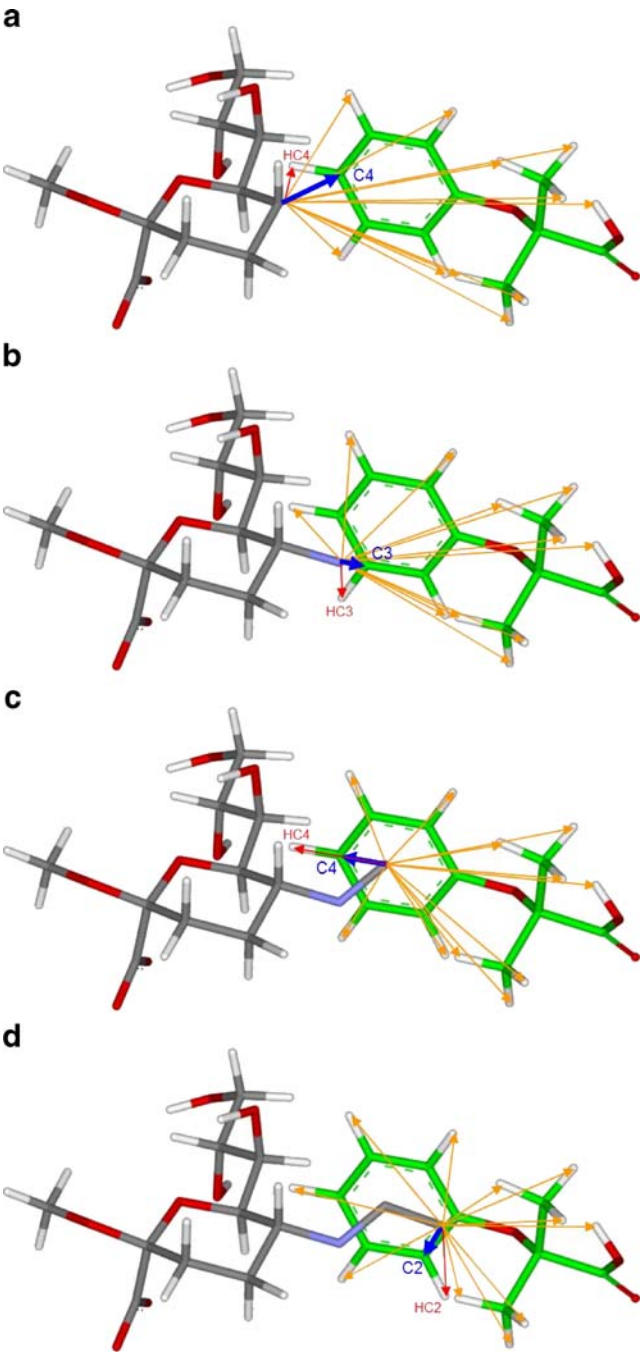

Fig. 5 An example of the systematic search which is performed to find the nearest hydrogen atom and it's corresponding heavy atom to each of (**a**) scaffold atom (AnA1), (**b**) linker 1 (AnA2), (**c**) linker 2 (AnA3), or (**d**) linker 3 (AnA4). The distance between AnA and the nearest fragment's hydrogen atom is represented as red arrow while the distance between AnA and the best fragment's HvA that carry the nearest hydrogen is represented as blue arrow

other hand, summation of both deviation values for each AnA produces a crude score (CS) which can effectively be used to discriminate the most suitable AnA to be involved in the attachment (Fig. 7).

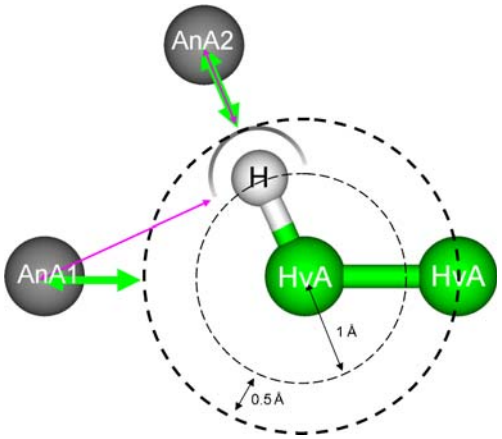

Fig. 6 The deviations from typical distances are shown between AnAs and both of fragment's Heavy Atom (HvA) (green arrows) and fragment's hydrogen atom (H) (pink arrows). The radiuses of circled orbits around fragment HvA represent the approximate fixed inter-atomic distances between HvA and HvA (outer bold circle of 1.5 Å) and between HvA and H (inner thin circle of 1 Å). The gray hemisphere with radius of 0.5 Å around fragment H represents the sites where AnA could be presented to substitute the hydrogen, with increasing opportunity toward the darker region. With respect to AnA1 and AnA2, the deviations from the typical AnA-HvA distance are equal for both atoms. However the deviation from the typical AnA-H distance is smaller for AnA2 compared to AnA1. Therefore, AnA2 has a more suitable location to substitute H than AnA1

Step 4: Addition of bad contact penalty term to the crude score. When an AnA comes too close to any of the fragment's HvAs (other than the HvA that carry the nearest hydrogen) a bad contact is said to occur. As a result, such AnA should have poorer score than other AnAs that make fewer or no bad contacts. As the typical distance to establish a covalent bond between any AnA and HvA is considered to be 1.5 Å, any AnA trespasses this distance without establishing a covalent bond with the fragment's HvA will have a positive penalty value proportional to the actual distance to be added to its crude score value. Therefore, an empirical mathematical function (equation below) was used to measure the degree of bad contact made by each AnA.

$$Bad-contact\ penalty(i)=\sum_{j=1}^{N}1.5-Dij \qquad (1)$$

Where i is an AnA, and j is a fragment HvA that does not carry the nearest hydrogen. D_{ij} is the distance between i and j which is less than 1.5 Å. N is the total number of fragments heavy atoms minus one. For each AnA, the value of bad-contact penalty was added to the value of crude score to get a value of "penalized score (PS)". As the molecular linkers are principally chains of sequenced AnAs, the penalty value scored by one AnA was

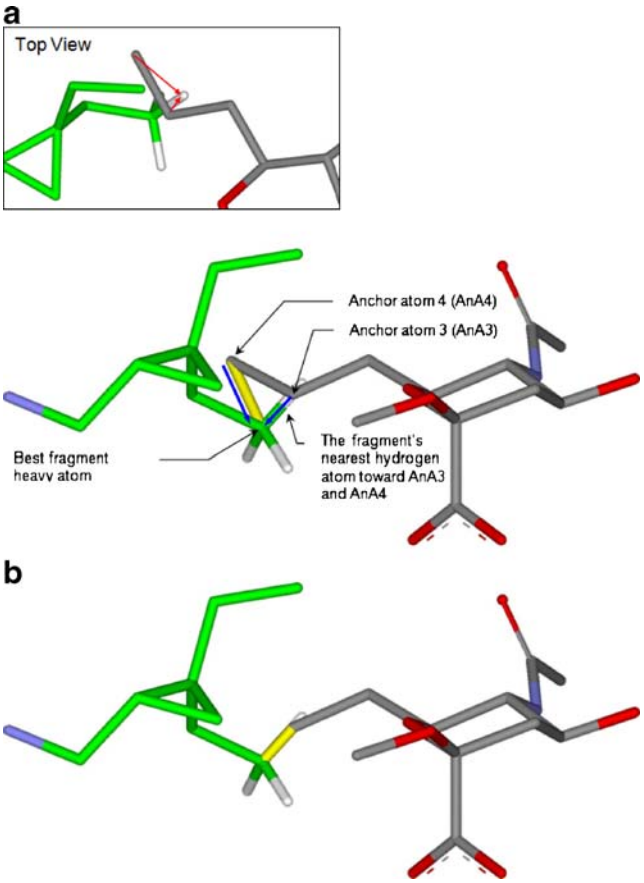

Fig. 7 Attachment of an oriented fragment to the SA scaffold C6 using (**a**) the deviation value of fragment's heavy atom (HvA) alone, or (**b**) combined with the deviation value of fragment's nearest hydrogen atom. Fragment's carbon atoms are colored green (only the hydrogen atoms of the best HvA are shown). The distance between AnA and the nearest fragment's hydrogen atom is represented as red arrow while the distance between AnA and the best fragment's HvA that carry the nearest hydrogen is represented as blue arrow. The bond to be constructed is colored yellow

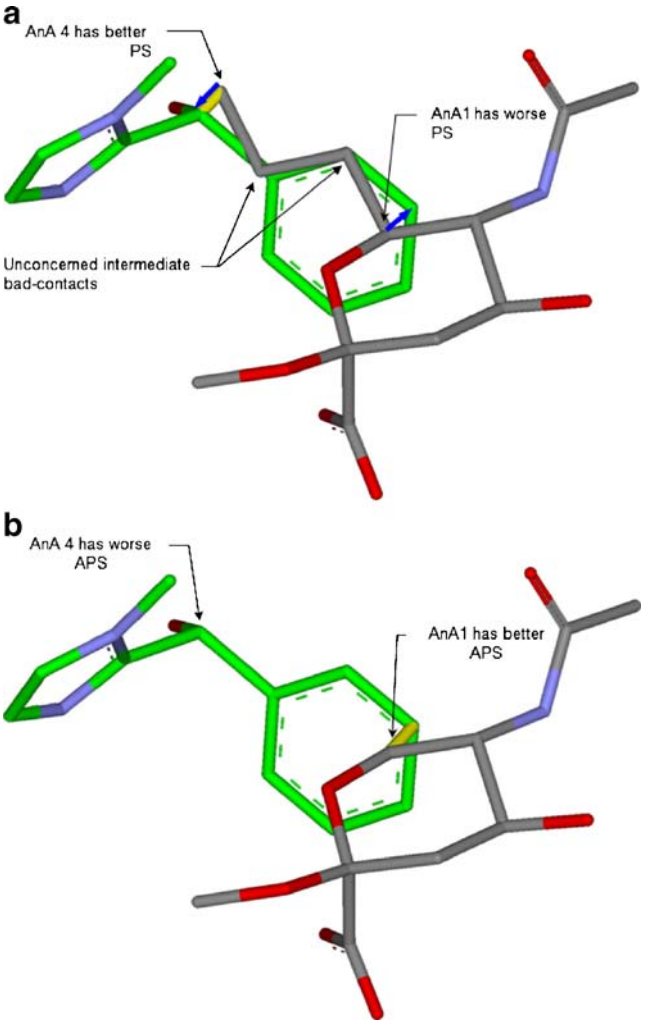

Fig. 8 Attachment of an oriented fragment to the SA scaffold C6 using (**a**) the penalized score (PS), and (**b**) the accumulated penalized score (APS)

bequeathed to the next AnAs (not to the previous AnAs) which results in penalty accumulation. For each AnA, the value of accumulated bad-contact penalty was then added to the crude score to get the "accumulated pPenalized score (APS)". The effect of penalty accumulation on the fragment attachment process is illustrated in Fig. 8.

Step 5: Hammering the unreasonable growth of molecular linkers. Since the use of large molecular linkers for fragments attachment will contribute unfavorably to the final molecular weights of the designed SA analogues, the growth of AnAs chain was resisted by favoring the use of AnAs of smaller sequence. According to this feature, for a group of AnAs in a fragment attachment process, any AnA of smaller sequence will substitute the AnA of larger sequence if the difference in APS values between them is not more than 0.5. The effect of incorporating this feature in the fragment attachment program is illustrated in Fig. 9.

Step 6: Structural refinement. This step involves sub-steps to control the structural property of the generated SA analogues. The first sub-stage includes prevention of construction of unusual binding modes. Carbon, oxygen, and nitrogen atoms represent all types of AnAs that can be used for fragments attachment, while carbon, oxygen, nitrogen and sulfur represent all types of fragments HvAs that can bear hydrogen(s) and are available for being bonded to AnAs. Therefore, several bonding modes can be established during the analogues design. Binding modes like oxygen-oxygen, oxygen-nitrogen, and nitrogen-nitrogen are considered unreasonable [21] and have been excluded during fragments attachment process. The second sub-stage includes the evaluation of scaffold's C4-hydroxyl

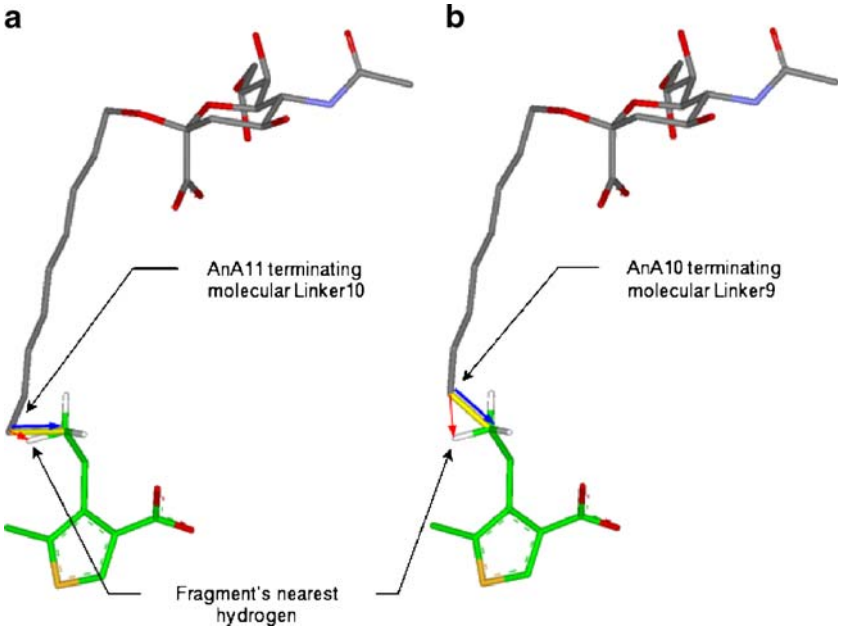

Fig. 9 Attachment of an oriented fragment to the SA scaffold C2 using the accumulated penalized score (APS) with (**a**) no AnAs chain hammering, and (**b**) with AnAs chain hammering

inclusion. Hence the fragments orientation has been performed against site B which environs the binding sites of SA C5-N-acetyl and C4-hydroxyl groups, the substitution of C5-functional group by the oriented fragment may generate atomic clashes between the fragment and C4-functional group. Therefore, a rapid search was employed to target any trespass of the equilibrium van der Waals distances between the SA C4-oxygen atom and all of fragment's heavy atoms (Fig. 10). The third sub-stage includes the evaluation of SA C8-hydroxyl inclusion. Both of crystallographic [9, 10] and docking results showed that SA C8-hydroxyl group makes favorable interaction with Tyr98 through hydrogen bond and possibly with Leu226 through van der Waals interaction. Previous inhibitory assays showed that changing the epimerization state of C8-hydroxyl, conversion to C8-oxo, or to C8-dimethyl reduces the affinity of SA analogue toward HA by five to seven folds [13], while our

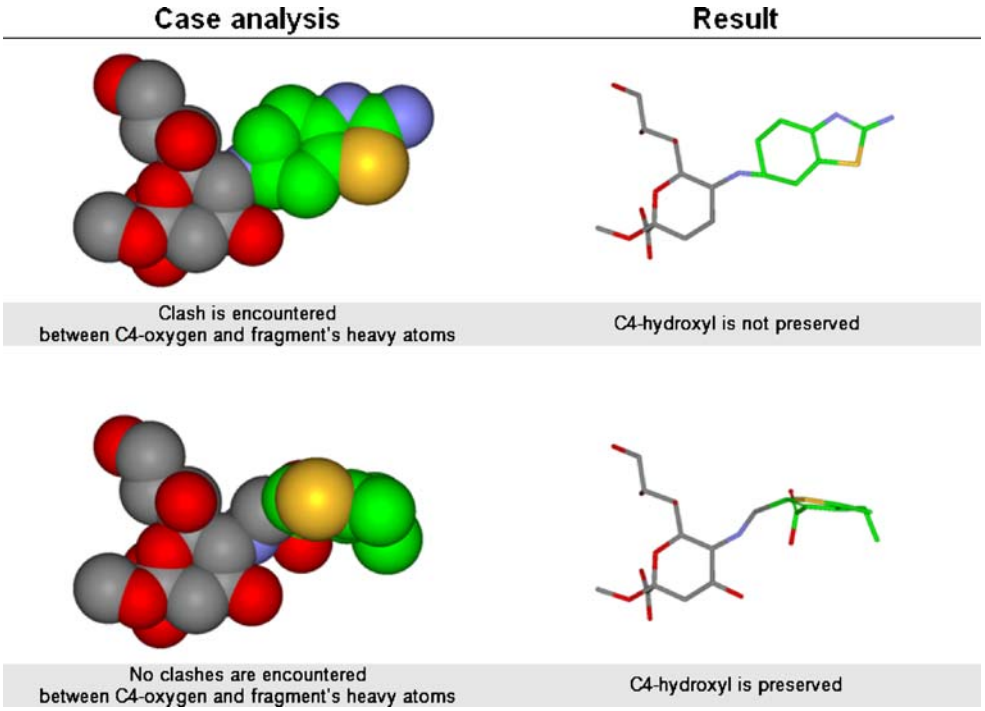

Fig. 10 Two different cases of fragments attachment to C5 of SA scaffold. In the first case a clash occurred between C4-oxygen and one of the fragments heavy atoms which results in dropping of the C4-hydroxyl in the generated SA analogue. While in the second case no clashes was encountered and the C4-hydroxyl is preserved. Fragment's carbon atom are colored green and Corey-Pauling-Koltun (CPK) molecular models are used to represent the atomic van der Waals radiuses

molecular docking of methyl-α-8-deoxy-Neu5Ac showed slight unfavorable change in the estimated free energy of binding (data not shown). Since the linker molecules used for connecting the oriented fragments to SA scaffold's C6 were derived from the natural glycerol side-chain, checking for the re-inclusion of hydroxyl group to AnA3 was recommended and has been introduced to the attachment program. The re-inclusion process used a protocol equivalent to the protocol used in evaluating the C4-hydroxyl preservation. The fourth and final sub-stage includes changing the atomic type of AnA when necessary. Since the linker molecules used for connecting the oriented fragments to SA scaffold's C5 were derived from the natural C5-N-acetyl group, a single nitrogen atom represents the smallest linker molecule that could be incorporated (AnA2). In some cases of fragments attachment to SA scaffold's C5 when AnA2 has the best APS, the best fragment HvA which bears the nearest hydrogen to AnA2 could be a non-carbon atom (*i.e.*, oxygen or nitrogen atoms), therefore, establishing a single bond between them gives unusual binding mode [21], while using another AnA of poor APS gives improper attachment. Therefore, the type of AnA2 was allowed to be changed from nitrogen to carbon whenever it is necessary to avoid the formation of unusual binding modes. Similar protocol has been used with AnA2 of C2-derived analogues where the oxygen atom was allowed to be flipped to carbon atom.

The fragment attachment process illustrated previously was used to generate three databases of single substituted SA analogues (*i.e.*, databases of C2-derived, C5-derived, and C6-derived SA analogues). Each database contains 4541 SA analogues, thus the total number of single-substituted SA analogues generated by site-directed fragment-based approaches was 13,623 analogues.

Docking of the three databases of SA analogues

The SA analogues databases generated in the previous stage were subsequently subjected to automatic molecular preparations using in-house developed programming scripts (see fragments preparation) then followed by docking to the whole HA1 binding pocket. Fast docking parameters (similar to the parameters used in fragments orientation) were used for SA analogues docking, however the initial molecular position, quaternion and dihedral were not random. Based on the docking results, the first 500 analogues of lowest estimated free energy of binding (EFEB) were redocked by mediumly slow docking using parameters that enables exhaustive conformational search followed by slow docking for the first 10 analogues in term of low EFEB and intramolecular interaction energy (IntraMIE). The parameters used in slow docking experiments were identical to the parameters that have been used in validation tests [18].

Results and discussion

Fragments orientation

The fragments orientation is a critical prerequisite for successful SA analogues generation because it determines the best fragments binding conformations and affinities at the binding sites of natural SA functional groups, which fragment's heavy atom suitable to be connected to SA scaffold and the best linker molecule to be incorporated. The oriented fragment should preserve its initial conformation after being connected to SA scaffold molecule in order for binding energy of the oriented fragment to be inherited to the generated SA analogue.

Several software for structure-based ligand design use the assembly of previously oriented fragments at the binding site as an approach to design complement ligand molecules of improved binding energies. The initial fragments orientation can either be performed by molecular dynamic simulation as in CONCERTS software [22], by using the well known DOCK program as in BUILDER software [23], or by pharmacophoric complementarity against the binding site as in LIGBUILDER [21], LUDI [24], GROW [25], LEAPFROG [26], and PRO_LIGAND [27]. However there is no software for structure-based ligand design that uses AutoDock to do fragments orientation.

In this study, we show that fragment orientation can be carried out using AutoDock3.05 with EFEB ranged from −1.8 to −7.5 kcal mol−1. It also provides better estimation of fragments binding conformations and affinities compared to DOCK software, due to an improved searching algorithm for ligand sampling (*i.e.*, LGA) and a well developed semi-empirical scoring function that composed of hydrogen bonding, desolvation, entropic terms as well as the electrostatic and van der Waals interaction terms. The DOCK's scoring functions only taking accounts of electrostatic and van der Waals interactions. In addition, AutoDock3.05 is attractive for fast calculation since it consumes less computational power compared to the molecular dynamic simulation and less prone to entrapment at local minima.

Molecular features of the designed SA analogues

In the C2-, C5- and C6-designed SA databases, the analogues share the same SA scaffold, *i.e.*, the pyranose

ring and functional groups with the exception of the substituted one. Thus, analogues within each database differ from each other in the type of molecular fragment used for substitution and the size of molecular linker incorporated in the connection. Within a database, these differences are responsible for the changes in molecular physicochemical properties and binding affinities toward HA1 binding pocket. The physicochemical features for all SA analogues within each of the three databases are summarized in Table 3.

Binding affinities of the three databases of SA analogues against HA1 binding pocket

The docking results showed that the database of C2-derived SA analogues has a mean EFEB value of −2.91± 1.38 kcal mol^{-1}, which was higher than the mean values for the databases of C5-derived (−5.06±0.97 kcal mol^{-1}) and C6-derived (−6.32±1.04 kcal mol^{-1}) SA analogues. The difference between the mean values of EFEB for the three databases was attributed mainly to the difference in the average number of torsional degrees of freedom (TDOF) between the three types of designed SA analogues. According to the semi-empirical scoring function of AutoDock3.05, the increase in number of TDOF for a given molecule is associated with an increase in the penalty value and consequently an increase in the value of EFEB (become more positive). The average number of TDOF differ between the three types of SA analogues is due to the diversity of molecular linkers used for fragment's connection which is affected by the location of the oriented fragments in relation to the SA scaffold. The average number of TDOF for C2-derived SA analogues was about 19 rotatable bonds compared to 13 rotatable bonds for C5-derived and 10 rotatable bonds for C6-derived analogues. Therefore, the mean penalty value (as well as the mean EFEB) decreases in the order of C2-derived > C5-derived > C6-derived SA analogues.

Direct correlations were obtained between numbers of TDOF and values of EFEB for C2-derived, C5-derived, and C6-derived SA analogues, which indicates that the reduction in estimated affinities toward the binding site can be attributed mainly to the increase in number of rotatable bonds within the designed analogues, as well as the improper binding between the analogues and the protein.

Table 3 Statistical summary of molecular weight (MW), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), calculated logarithm of octanol-water partition coefficient (cLogP), and number of torsional degrees of freedom (TDOF) of all the designed SA analogues

	Minimum	Maximum	Mean	Std. Deviation
C2-derived SA analogues				
MW	434.00	732.00	595.81	40.67
HBD	5.00	10.00	6.27	0.98
HBA	9.00	17.00	13.46	1.12
cLogP	−5.54	3.62	−0.861	1.41
TDOF	10.00	26.00	19.30	2.38
C5-derived SA analogues				
MW	344.00	602.00	481.47	41.65
HBD	3.00	11.00	5.12	1.24
HBA	7.00	16.00	11.37	1.33
cLogP	−7.84	0.82	−1.95	1.05
TDOF	9.00	21.00	13.17	1.76
C6-derived SA analogues				
MW	343.00	614.00	469.86	40.70
HBD	2.00	7.00	3.32	1.00
HBA	7.00	14.00	10.57	1.11
cLogP	−5.59	3.71	−0.427	1.01
TDOF	5.00	19.00	9.70	2.03

Stability of the docked analogues against the natural SA binding site of HA1

Since the pyranose ring is shared by all the designed analogues and is located within HA1 at approximately the same position in all the crystallized SA analogues [9, 10, 28], it can be used to evaluate the stability of the docked analogues against the natural SA binding site. During docking experiments, the stability of the designed SA analogue against the natural SA binding site of HA1 was monitored by calculating the RMSD value of pyranose ring between the analogue and the crystal methyl-α-Neu5Ac. A large value of RMSD indicates that the docked analogue either binds away from the natural SA binding site or with different orientation than the crystal methyl-α-Neu5Ac.

According to the mean values of RMSD for pyranose rings, the stability of designed SA analogues against the crystal SA binding site decreased in the order of C5-derived SA analogues (1.38±1.24 Å) < C6-derived SA analogues (1.94±1.58 Å) < C2-derived SA analogues (3.32±3.33 Å). The differences in SA scaffold's stability between the three types of SA analogues can be attributed to the variation in scaffolds structures, effects of attached fragments, reliability of attachment modes, size of linker molecules, and docking limitations.

The SA scaffold used in the generation of C5-derived analogues has the C5-N-acetyl group (which is known to form one hydrogen bond with protein) removed while the C6-glycerol side chain (which forms more than 3 hydrogen bonds with the protein) was preserved. On the other hand, for C6-derived SA analogues, the C5-N-acetyl group was kept while the C6-glycerol has been removed. Theoretically,

these differences should make the SA scaffold of C5-derived analogues more stable at the binding site than the C6-derived analogues, as the C6-glycerol moiety that involves more hydrogen interaction was preserved. However, the EFEB calculated using AutoDock3.05 for the SA scaffold of C5-derived analogues was −2 kcal mol^{-1} compared to −2.75 kcal mol^{-1} for the SA scaffold of C6-derived analogues. This was attributed to the difference in entropic penalty value that has been calculated for the two scaffolds which is proportional to the number of scaffold's rotatable bonds as C5-derived analogues has nine rotatable bonds while five rotatable bonds are present in the scaffold molecules of C6-derived analogues. Thus, lower penalty values for C6-derived analogues. With respect to C2-derived SA analogues, the SA scaffold has structural features of natural SA which provided better interaction with the protein compared to C5-derived and C6-derived SA analogues. However, the long sized linkers used in analogues design and the fast docking parameters used in conformational sampling adversely contributed on the stability of C2-derived SA analogues against the natural SA binding site during docking simulation.

The site-directed molecular design approach employed in this study assumed that the SA scaffold molecule should preserve its crystal position after being attached to the oriented fragment in order to obtain highly active SA analogue. Thus, any analogue with large RMSD of pyranose ring should, theoretically, have low binding affinity against the natural SA binding site as the favorable binding energy of the SA scaffold is reduced or lost due to the changing of the crystal position. The correlation between RMSD values of pyranose ring and the values of EFEB for the designed analogues (Fig. 11) shows consistency with the previous assumption. The correlation coefficient (r) for C5-derived and C6-derived SA analogues was 0.37 and 0.39, respectively. Relatively, lower correlation was obtained for C2-derived SA analogues (r=0.26), suggesting many of these analogues located away from the binding site or with different pyranose orientation compared to the crystal pyranose.

Importance of preserving the docked fragments conformations before and after attachment to SA scaffolds

The preservation of fragment's initial docked conformation after being attached to SA scaffold indicates accomplishment of low energy conformation, orientation, and position within the docking site. It also signifies a reliable connection mode established between the docked fragment and the crystal SA scaffold. However, minor changes in fragment's docked conformation after the attachment to SA scaffold is regarded as normal responses to establish the connection and to reduce internal constrain. On the other hand, major conformational changes usually imply problems at fragment's orientation stage, fragment's attachment mode, or docking simulation, and usually worsen the activity of the generated analogues.

Fragments of C2-derived SA analogues showed the highest deviation with mean RMSD of 7.85 Å, suggesting ineffective docking due to high flexibility of linker molecules used in fragments connection. Lower deviations were observed for fragments of C6-derived analogues with mean RMSD of 2.65 Å and fragments of C5-derived SA analogues with mean RMSD of 3.03 Å, suggesting more stable oriented fragments with minimal conformational changes were produced. Positive correlations obtained between RMSD of the oriented fragment and the EFEB of the generated analogue with r equals to 0.40, 0.43, and 0.47 for C2-derived, C5-derived, and C6-derived SA analogues, respectively (Fig. 12) support the necessity of stable oriented fragments to produce analogues with low EFEB.

Correlation between activities of the oriented fragments and activities of the generated analogues

Direct correlations were observed between values of EFEB of the oriented fragments and their corresponding SA analogues and are shown in Fig. 13. The correlation coefficients are decreased in the order of C6-derived SA analogues (r=0.51) > C5-derived SA analogues (r=0.43) > C2-derived SA analogues (r=0.35). The results show that binding energy of the generated SA analogues was a combination between binding energies of the oriented (docked) fragments and the crystal SA scaffolds.

A lower degree of correlation obtained for C2-derived SA analogues suggests many of the oriented fragments were able to interact at site A with good affinity during the orientation stage, but failed to preserve the same interactions after being attached to SA scaffold. The fragments changed their binding conformations, positions, or orientations as observed by high mean value of RMSD between the oriented conformation and the attached conformation with a mean RMSD of 7.85 Å. In addition, binding orientation of the SA scaffold for many of C2-derived analogues was also changed after the fragments attachment with a mean RMSD of pyranose ring of 3.32 Å. As a result, for many C2-derived SA analogues, the favorable binding energies of the oriented fragment and the crystal SA scaffold were changed resulting in dramatic change of total analogue's binding energy.

In contrast to the oriented fragments at site A, the fragments oriented at sites B and C relatively preserve their conformations, positions and orientations after attachment to SA scaffolds (mean RMSD are 3.03 Å and 2.65 Å,

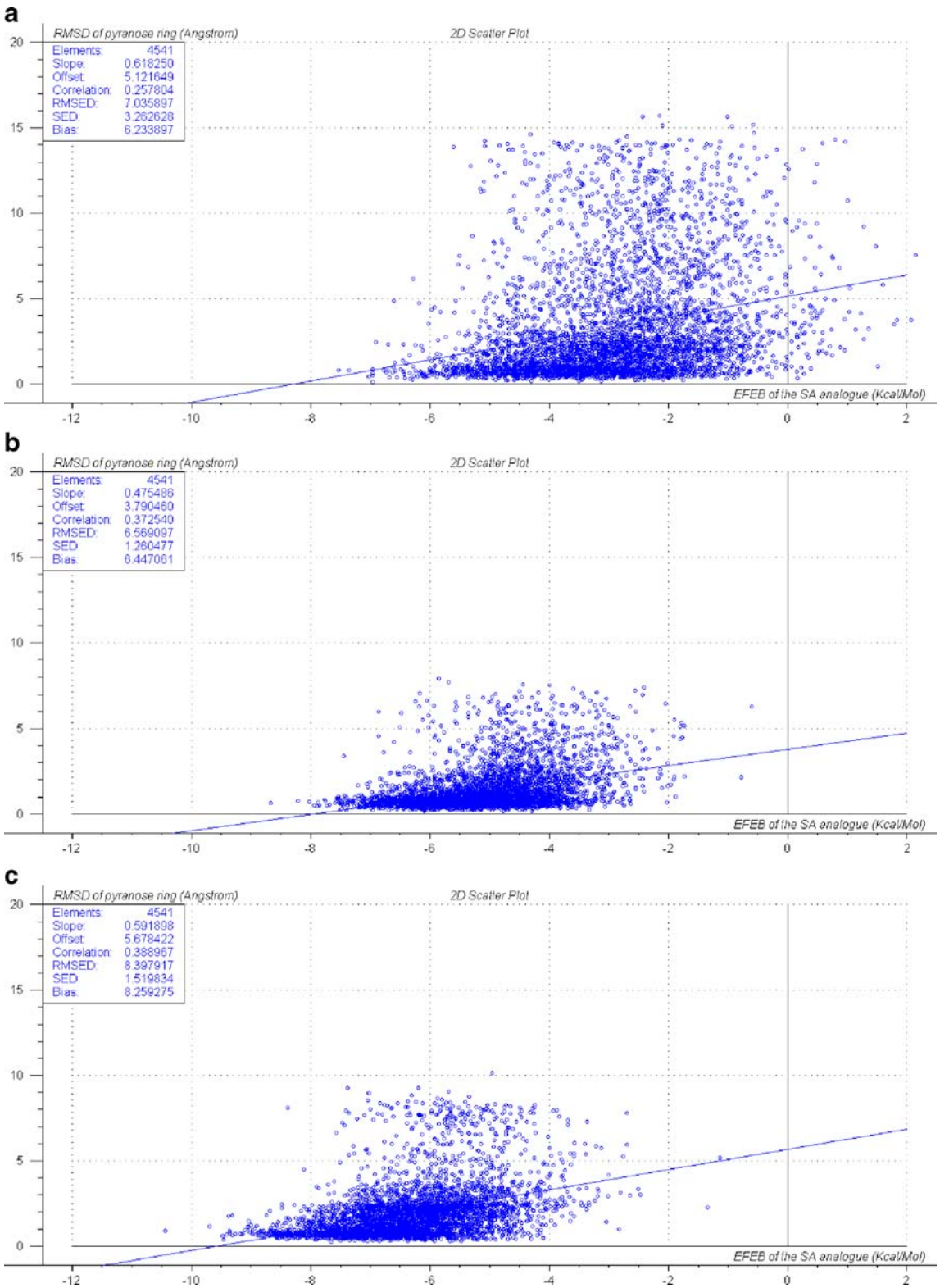

Fig. 11 2D-scatter plots show the correlation between RMSD value of pyranose ring (Å) and the value of EFEB (kcal mol^{-1}) for (**a**) C2-derived, (**b**) C5-derived, and (**c**) C6-derived SA analogues

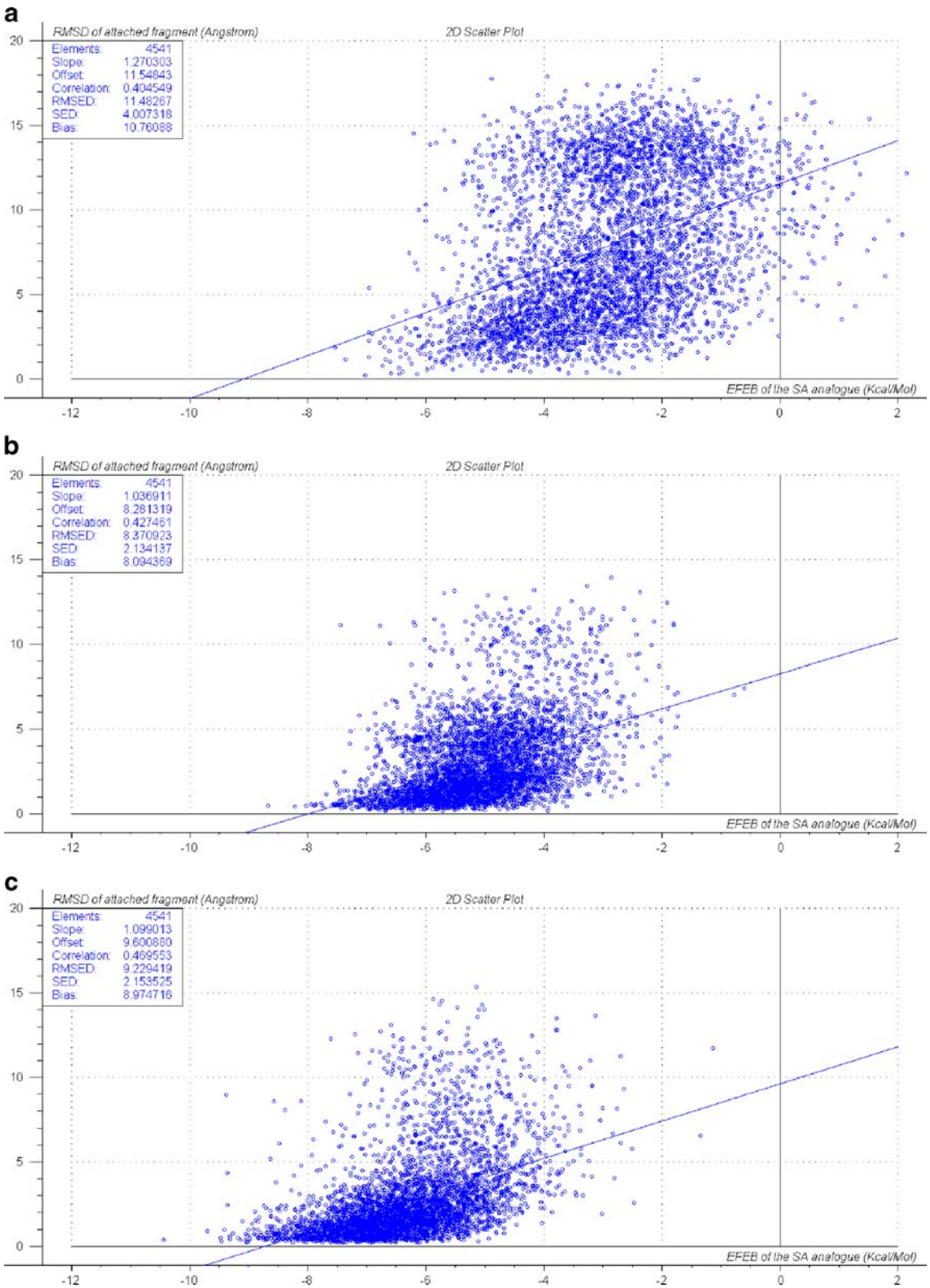

Fig. 12 2D-scatter plots show the correlation between values of RMSD of attached fragments (Å) and values of EFEB of the generated analogues (kcal mol^{-1}) for (**a**) C2-derived, (**b**) C5-derived, and (**c**) C6-derived SA analogues

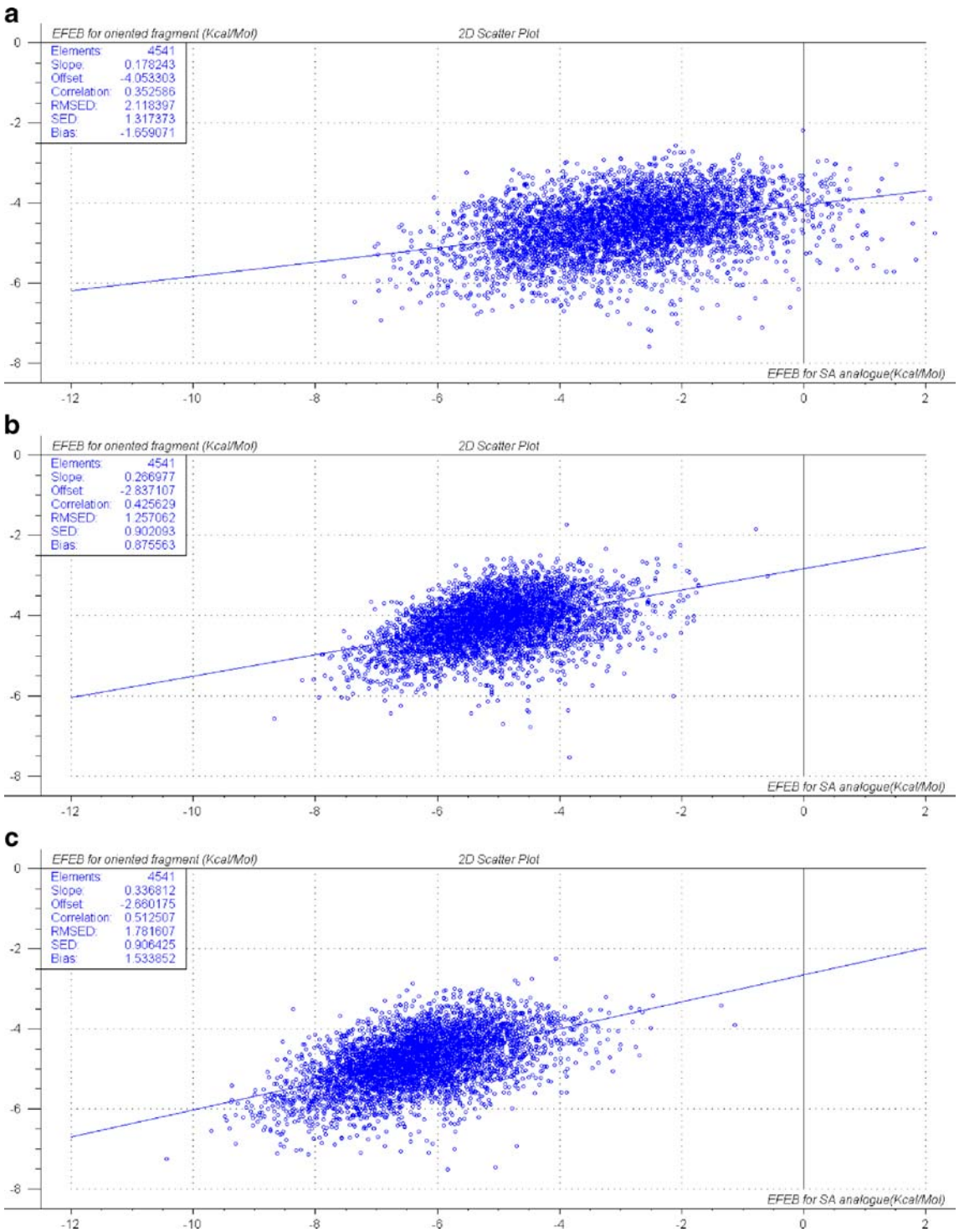

Fig. 13 2D-scatter plots show the correlation between values of EFEB (kcal mol^{-1}) for the oriented fragment and the correspondingly generated analogue for all (**a**) C2-derived, (**b**) C5-derived, and (**c**) C6-derived SA analogues

respectively). Similarly, the SA scaffold for C5-derived and C6-derived analogues were observed to be stable during docking simulation with a mean RMSD values of pyranose ring of 1.38 Å and 1.94 Å, respectively. Thus, it can be concluded that the preservation of fragment's oriented conformation and SA scaffold's crystal conformation is responsible for the additive merge of their binding energies for most C5-derived and C6-derived SA analogues.

Examples of active SA analogues designed by site-directed fragment-based approaches

An example of C2-derived SA analogue

According to results of slow docking experiments, the C2-derived SA analogue of SA-C2-2055 has an estimated affinity of about 1120 fold higher than the estimated affinity of methyl-α-Neu5Ac. SA-C2-2734 was generated from fragment no. 2055 within the ChemBridge fragments library database. This fragment has an EFEB of −5.9 kcal mol^{-1} and was ranked within the best oriented fragments at site A (rank no. 166 out of 4541). The fragment has been connected to the SA scaffold C2 using linker3 molecule. The docking results showed that the orientation of SA scaffold for SA-C2-2055 is similar to methyl-α-Neu5Ac crystal orientation with RMSD of pyranose ring of 0.42 Å. On the other hand, the attached fragment failed to preserve its initial oriented conformation with a RMSD value of 2.12 Å (Fig. 14a). The changing of fragment's oriented conformation to obey the empirically established connection with SA scaffold usually reduces the binding energy of the fragment. Thus, the EFEB for SA-C2-2055 is −8.50 kcal mol^{-1} (K_i=0.58 μMol), is slightly higher when compared to EFEB of its oriented fragment of −5.9 kcal mol^{-1} and the crystal SA scaffold of −4.30 kcal mol^{-1}.

An example of C5-derived SA analogue

The estimated affinity of the active C5-derived SA analogue of SA-C5-875 was about 1800 fold higher than the estimated affinity of methyl-α-Neu5Ac. The fragment that has been used to design this analogue was no. 875 within the ChemBridge database and is ranked 22 during the orientation against site B. Molecular linker2 has been used for connecting the fragment to SA scaffold. According to molecular docking results, both of the SA scaffold and the molecular fragment incorporated in this analogue were able to preserve their initial conformations with the RMSD values for the pyranose ring and the attached fragment of 0.40 Å and 0.42 Å, respectively (Fig. 14b). The results suggest a proper initial fragment orientation has been achieved at the orientation site, and a reliable attachment mode which did not disrupt the initial conformations of the two connected molecules. Thus, the binding energies of both of the oriented fragment (−6.77 kcal mol^{-1}) and the crystal SA scaffold (−2 kcal mol^{-1}) were preserved and additively combined to form an active C5-derived SA analogue of EFEB equals to −8.79 kcal mol^{-1} (K_i=0.3 μMol).

An example of C6-derived SA analogue

The estimated affinity of the active C6-derived SA analogue of SA-C6-1432 was about 11,500 fold higher than the estimated affinity of methyl-α-Neu5Ac. The molecular fragment used to design SA-C6-1432 was ranked 447 during fragments orientation against site C. The fragment was connected to C6 of SA scaffold using linker2. The deviation from the crystal SA binding site was small for this analogue with RMSD of pyranose ring equals to 1.27 Å, while the attached fragment was able to preserve its initial oriented conformation with RMSD equals to 0.88 Å and is shown in Fig. 14c. Although the SA scaffold showed deviation from the crystal position, the binding energies of the oriented fragment (−4.95 kcal mol^{-1}) and the crystal SA scaffold (−2.75 kcal mol^{-1}) were merged to form a complete C6-derived analogue with EFEB of −9.85 kcal mol^{-1} (K_i=60 nMol).

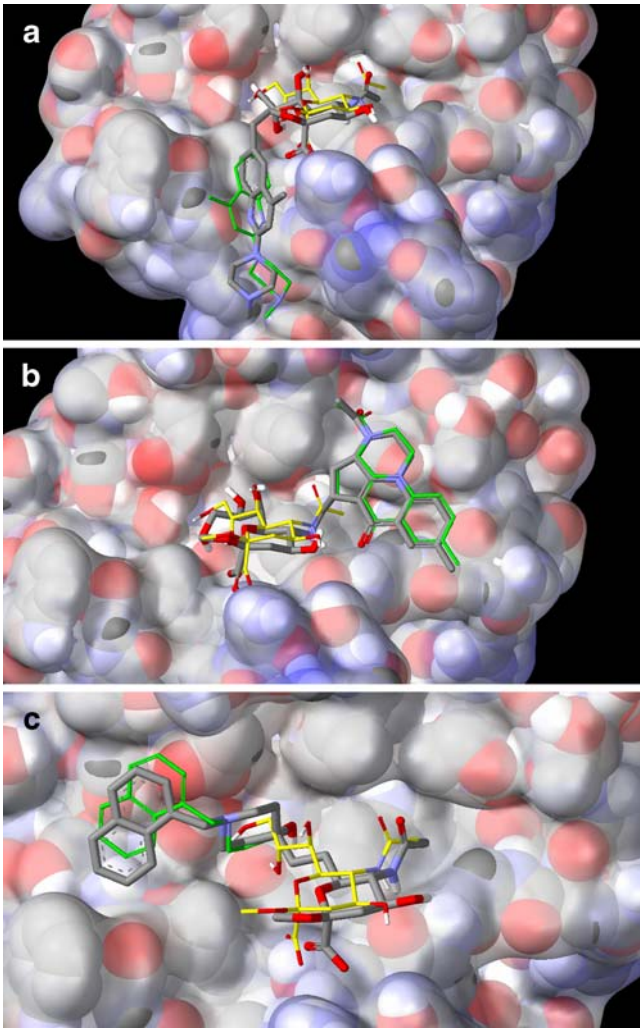

Fig. 14 The superposition of the crystal conformation of methyl-α-Neu5Ac (yellow), the docked conformation of active SA analogue (gray), and the docked conformation of the fragment which has been used to generate the analogue (green). (**a**) C2-derived SA analogue generated from fragment no. 2055, (**b**) C5-derived SA analogue generated from fragment no. 875, and (**c**) C6-derived SA analogue generated from fragment no. 1432

Activity against various strains of influenza A HA1

As the simulation in this study used HA1 from influenza A H3N2 strain, attempts were made to dock the generated analogues with influenza A strains of H5N1 that caused bird flu and H1N1 that caused swine flu. One of the C6-derived SA analogue was tested and could bind HA of H3N2 Aichi 1968 (X-31) virus with EFEB of around −11 kcal mol^{-1} (K_i=20 nMol). About the same affinity was also observed for the same analogue with the HA of H1N1 1918 virus (which has 79% sequence similarity with H1N1 2009), while EFEB against the HA of H5N1 was about −8.30 kcal mol^{-1} (K_i=0.83 μMol). These results suggest the designed analogues could bind to the three strains of influenza A studied, suggesting high similarity of HA1 binding pocket between the strains.

Conclusions

Our results showed that new analogues could be designed using site-directed fragment-based and molecular docking simulation. The Chembridge fragments docking against the binding pockets of natural SA functional groups helped in orienting the fragments molecules at these sites to establish the best interactions with the protein. Three databases of oriented fragments were generated representing interactions to binding sites of C2-, C5- and C6-functional groups of SA and were used to substitute the natural SA functional groups without disrupting the SA scaffold crystal conformation, thus, generating three databases of C2-, C5- and C6-derived SA analogues. Empirical algorithm was developed to connect the oriented fragments to the corresponding SA scaffold atoms using the most suitable linker molecules. The docking results of the three designed databases of SA analogues showed that many of the generated analogues could bind the HA1 binding pocket with higher affinity (up to 30,000 fold) than the natural SA. The mean value of estimated affinity for the SA analogues increased in the order of C2-derived < C5-derived < C6-derived SA analogues. C2-derived SA analogues have lower affinity as the attached fragments frequently change conformation and affinity when attached to SA scaffold. Good correlation between the EFEB of the oriented fragments and the EFEB of their corresponding SA analogues indicates favorable binding energies of the docked molecule (oriented fragment) and the crystallized molecule (SA scaffold) were additively merged within the generated SA analogue. Since the designed analogues preserve the conformation of the SA scaffold, they offer opportunity to be active against different serotypes of influenza A HA. Finally, as the synthetic accessibility of the SA analogues was partially considered during the molecular design process, some of the analogues are synthesizable and can be tested against influenza A HA which may aid in discovering new *anti*-flu drugs act by inhibiting viral attachment to host cells.

Acknowledgments We would like to acknowledge Mr. Andreas Meyer from ChemBridge Corporation. Mr Al-Qattan appreciates Universiti Sains Malaysia for supporting this project through a research fellowship scheme.

References

1. Varki NM, Varki A (2007) Diversity in cell surface sialic acid presentations: implications for biology and disease. Lab Invest 87:851–857
2. Schauer R, Kamerling JP (1997) Chemistry, biochemistry and biology of sialic acids. In: Montreuil J, Vliegenthart JFG, Schachter H (eds) New comprehensive biochemistry, vol 29b. Elsevier Science BV, Amsterdam, The Netherlands, pp 243–402
3. Schauer R (1982) Chemistry, metabolism and biological functions of sialic acids. Adv Carbohydr Chem Biochem 40:131–234
4. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Ann Rev Biochem 69:531–569
5. Air GM, Laver WG (1989) The neuraminidase of influenza virus. Proteins: Struc Funct Genet 6:341–356
6. Ward CW, Dopheide TA (1981) Amino acid sequence and oligosaccharide distribution of the haemagglutinin from an early Hong Kong influenza virus variant A/Aichi/2/68 (X-3 1). Biochem J 193:953–962
7. Nobusawa E, Aoyama T, Kato H, Suzuki Y, Tateno Y, Nakajima K (1991) Comparison of complete amino acid sequences and receptor-binding properties among 13 serotypes of hemagglutinins of influenza A viruses. Virology 182:475–485
8. Von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Phan TV, Smythe ML, White HF, Oliver SW, Colman PM, Varghese JN, Ryan DM, Woods JM, Bethell RC, Hotham VJ, Cameron JM, Penn CR (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature 363:418–423
9. Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC (1988) Structure of the influenza virus hemagglutinin complexed with its receptor, sialic acid. Nature 333:426–431
10. Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, Park SJ, Skehel JJ, Wiley DC (1992) Binding of influenza virus hemagglutinin to analogs of its cells urface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. Biochemistry 31:9609–9621
11. Sauter NK, Glick GD, Crowther RL, Park SJ, Eisen MB, Skehel JJ, Knowles JR, Wiley DC (1992) Crystallographic detection of a second ligand binding site in influenza virus hemagglutinin. Proc Nat Acad Sci USA 89:324–228
12. Pritchett TJ (1987) Receptor analogue inhibitors of influenza virus adsorption and infection. (thesis), Univ of California, Los Angeles
13. Kelm S, Paulson JC, Rose U, Brossmer R, Schmid W, Bandgar BP, Schreiner E, Hartmann M, Zbiral E (1992) Use of sialic acid analogues to define functional groups involved in binding to the influenza virus hemagglutinin. Eur J Biochem 205:147–153
14. Toogood PL, Galliker PK, Glick GD, Knowles JR (1991) Monovalent sialosides that bind tightly to influenza A virus. J Med Chem 34:3138–3140
15. Sauter NK, Bednarski MD, Wurzburg BA, Hanson JE, Whitesides GM, Skehel JJ, Wiley DC (1989) Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with

millimolar dissociation constants: a 500 MHz proton nuclear magnetic resonance study. Biochemistry 28:8388–8396
16. Machytka D, Kharitonenkov I, Isecke R, Hetterich P, Brossmer R, Klein RA, Klenk HD, Egge H (1993) Methyl alpha-glycoside of N-thioacetyl-D-neuraminic acid: a potential inhibitor of influenza A virus. A 1H NMR study. FEBS Lett 334:117–120
17. Hurt AC, Ernest J, Deng Y-M, Iannello P, Besselaar TG, Birch C, Buchy P, Chittaganpitch M, Chiu S-C, Dwyer D, Guigon A, Harrower B, Kei IP, Kok T, Lin C, McPhie K, Mohd A, Olveda R, Panayotou T, Rawlinson W et al (2009) Emergence and spread of oseltamivir-resistant A(H1N1) influenza viruses in Oceania, South East Asia and South Africa. Antivir Res 83:90–93
18. Mohammed Noor Al-qattan, Mohd Nizam Mordi, Docking of sialic acid analogues against influenza A hemagglutinin : A correlational study between experimentally measured and computationally estimated affinities. (Accepted for Publication, J Mol Mod)
19. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner PJ (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784
20. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 19:1639–1662
21. Wang R, Gao Y, Lai L (2000) LigBuilder: a multi-purpose program for structure-based drug design. J Mol Model 6:498–516
22. Pearlman DA, Murcko MA (1996) CONCERTS: dynamic connection of fragments as an approach to de novo ligand design. J Med Chem 39:1651–1663
23. Lewis RA, Roe DC, Huang C, Ferrin TE, Langridge R, Kuntz ID (1992) Automated site-directed drug design using molecular lattices. J Mol Graphics 10:66–78
24. Böhm H-J (1992) The computer program LUDI A new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6:61–78
25. Moon JB, Howe WJ (1991) Computer design of bioactive molecules: A method for receptor-based de novo ligand design. Proteins: Struct, Funct, Genet 11:314–328
26. Höltje H-D, Sippl W, Rognan D, Folkers G (1997) Molecular modeling: basic principles and applications, 3rd edn. VCH Verlagsgesellschaft GmbH, Weinheim, Germany, pp 192
27. Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, Waszkowycz B, Westhead DR (1995) PRO_LIGAND: an approach to de novo molecular design. 1. Application to the design of organic molecules. J Comput Aided Mol Des 9:13–32
28. Ha Y, Stevens DJ, Skehel JJ, Wiley DC (2003) X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. Virology 309:209–218