Journal of Computer-Aided Molecular Design, 10 (1996) 233–246
ESCOM

J-CAMD 338

A structural and energetics analysis of the binding of a series of *N*-acetylneuraminic-acid-based inhibitors to influenza virus sialidase*

Neil R. Taylor** and Mark von Itzstein***

Department of Medicinal Chemistry, Victorian College of Pharmacy, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

Received 22 September 1995
Accepted 1 February 1996

Keywords: Calculated binding energies; Molecular mechanics; Continuum electrostatics; Sialidase; Influenza

Summary

A molecular dynamics/energy-minimisation protocol has been used to analyse the structural and energetic effects of functional group substitution on the binding of a series of C4-modified 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid inhibitors to influenza virus sialidase. Based on the crystal structure of sialidase, a conformational searching protocol, incorporating multiple randomisation steps in a molecular dynamics simulation was used to generate a range of minimum-energy structures. The calculations were useful for predicting the number, location, and orientation of structural water molecules within protein–ligand complexes. Relative binding energies were calculated for the series of complexes using several empirical molecular modelling approaches. Energies were computed using molecular-mechanics-derived interactions as the sum of pairwise atomic nonbonded energies, and in a more rigorous manner including solvation effects as the change in total electrostatic energy of complexation, using a continuum-electrostatics (CE) approach. The CE approach exhibited the superior correlation with observed affinities. Both methods showed definite trends in observed and calculated binding affinities; in both cases inhibitors with a positively charged C4 substituent formed the tightest binding to the enzyme, as observed experimentally.

Introduction

We have recently reported the design, synthesis, and biological evaluation of influenza virus sialidase inhibitors [1–3] employing a receptor-based design approach. Influenza virus sialidase is one of the two glycoproteins on the viral envelope considered important in the process of cell infection. The enzyme is a glycohydrolase which catalyses the cleavage of terminal *N*-acetylneuraminic acid (Neu5Ac) from glycoconjugates via a sialosyl cation intermediate [4–6]. It is comprised of a tetrameric head of four identical subunits of approximately 60 kDa and a stalk fixed at one end to the viral membrane. Each monomer of the enzyme is made up of six, four-stranded β-sheets arranged as if on the blades of a propeller. Substrate binding occurs in a deep pocket on the upper surface of each monomer. This pocket is lined by amino acid residues which appear to be highly conserved across all strains of the virus, despite low overall sequence homologies [7]. There are two calcium-ion binding sites in the enzyme, one approximately 12.0 Å from the binding pocket and one on the four-fold axis of the tetramer. Calcium, although not essential, enhances activity [8].

The three-dimensional crystal structure of the complex of influenza virus sialidase and the transition-state analogue 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid (Neu5Ac2en, **1**) reveals a very good fit of charge and shape complementarity [9]. It appears that most of the polar sites on the carbohydrate ligand form strong electrostatic interactions with polar groups in the sialidase active site, and the hydrophobic region of the ligand makes good contact with two hydrophobic amino acid side chains in the binding pocket. The fit is not perfect however, and with the aid of molecular modelling tech-

*This paper is based on a presentation given at the 14th Molecular Graphics and Modelling Society Conference, held in Cairns, Australia, August 27–September 1, 1995.

**Presently on a visiting postdoctoral fellowship in the Department of Biomolecular Structure, Glaxo Research & Development Ltd, Greenford, Middlesex UB6 OHE, U.K.

***To whom correspondence should be addressed.

niques, analogues of Neu5Ac2en were designed with the aim of improving the binding affinity [1,6,10]. In particular, a pocket in the enzyme active site surrounded by negatively charged amino acid side chains was identified. It was postulated that introduction of a basic group at the C4 position of Neu5Ac2en should result in favourable electrostatic interactions with the negatively charged pocket [10]. A number of compounds were synthesised and several were found to have very high affinity [1]. Subsequent modelling experiments showed that for some of these compounds the major structural features of the binding could be successfully predicted and that the electrostatic enhancement to the binding energy could be qualitatively reproduced employing routine molecular mechanics calculations [6]. This success prompted further investigation and led to the development of a conformational searching routine for the prediction of the structures of complexes of C4-modified Neu5Ac2en compounds and influenza virus sialidase, and two methods for estimating relative binding energies. In this paper we give an account of these models, the important features of the calculated structures, and the trends observed between the calculated binding energies and the experimentally determined affinities.

Influenza virus sialidase provides an ideal target protein for the development of a computational model to study ligand binding. It has a high-affinity binding site which has been well-characterised by X-ray crystallography [9,11], and a simple model of independent ligand association applies. The active site is lined with an unusually large number of charged amino acid residues that form stable electrostatic and hydrogen-bonded interactions with the substrate. Binding is further enhanced because the active site is located in a deep cleft in the protein surface. The number of water molecules that can solvate the active site residues is limited due to the arrangement of these residues in the highly concave binding site. The free energy cost associated with desolvating active-site functional groups is thereby minimised. We believe that the free energy change associated with the binding event is dominated by electrostatic factors which can be reasonably estimated using currently available molecular modelling software.

In order to assess the relative binding strengths of different ligands to the active site of sialidase it was first necessary to determine possible binding geometries using conformational searching. Conformational searching methods generally fall into one of the following classes: systematic searching about torsion angles, molecular dynamics, random or Monte Carlo searching, distance-geometry, and ruled-based methods. These techniques have been extensively reviewed elsewhere [12,13]. The availability of the high-resolution crystal structure of the complex of sialidase and Neu5Ac2en, and the assumption that structurally related compounds bind in a grossly similar manner, directed us towards the implementation of a molecular-dynamics-based method. Our structure searching was performed using a technique in which the random starting velocities for the dynamics were continually reset, and the final structures obtained using constrained energy minimisation. We have been exploring whether this variation to the traditional approach of conformational analysis using molecular dynamics searches conformational space efficiently (by avoiding both the necessity of running long trajectories and using large numbers of starting approximations). The conformational flexibility of a large number of protein residues and some solvent molecules was explicitly included in the dynamics calculations.

Energy scoring is usually performed using one of three methods. Computationally, the most rigorous approach to the calculation of binding energy is the free energy perturbation (FEP) method. This method employs classical statistical mechanics to relate the free energies of two states to the ensemble average of the potential energy difference of the states. The approach is described in detail elsewhere [14–16]. In one paper, the FEP approach was combined with the Langevin dipoles technique for modelling electrostatic interactions in the protein–ligand complex [17]. Unfortunately, computational requirements generally restrict the use of the FEP method. Other approaches reported in the literature define empirical free energy functions to account for the various contributions to molecular association [18–24]. Free energy changes can be calculated from a variety of sources, such as changes in solvation, nonbonded electrostatic and van der Waals forces, hydrophobic forces, conformational strain energy, conformational entropy of protein residues and ligand, and translational and rotational entropy. Although these techniques show much promise they are highly sensitive to parameter sets, which are necessarily large. Finally, several rule-based, empirical scoring functions, designed for the searching of large 3D databases, have been described [25–28]. These techniques concentrate more upon the rapid scoring of a large number of candidate drug molecules rather than the accurate evaluation of structurally related species.

In our work a fast and accurate model for predicting relative binding affinities was sought, one that does not require lengthy parameterisation, is simple to implement, does not require vast quantities of CPU time, and is easily transferable from one enzyme–inhibitor complex to the next. We chose an empirical approach using classical forcefield calculations. According to classical molecular mechanics, the total potential energy of a molecule is the sum of interactions for bond stretching and bending, torsional forces, cross terms, and nonbonded van der Waals and Coulombic forces. Energies of intermolecular interaction were obtained by extracting the pairwise atomic nonbonded terms from the energy-optimised en-

zyme–ligand complexes calculated from the conformation searching method. This formed the basis of our molecular-mechanics-derived interaction (MMI) model.

The reliability of our modelling procedure is particularly sensitive to our ability to accurately simulate the electrostatic contributions to the binding enthalpy. The screened Coulombic potential used in energy minimisation is suitable for simulating molecular geometries, but is less reliable for the accurate calculation of energies. In order to determine the electrostatic contributions to the binding energies more accurately, we employed the continuum-electrostatics (CE) method. The software package DelPhi, which is part of the InsightII suite of programs (Biosym Technologies, Inc., San Diego, CA) was used. This program calculates electrostatic energies by solving the Poisson equation using the method of finite differences [29–31]. The program maps a macromolecule onto a three-dimensional grid by distributing atomic partial charges over the grid points and dielectric values at the midpoints of the lines between the grid points. A finite difference formula is then employed to calculate the potential at each point according to the Poisson equation. Calculations are repeated until the convergence criteria are satisfied. This approach can be interpreted as generating electrostatic free energies for molecules by simulating the effects of solvation using macroscopic dielectrics and then including microscopic Coulombic energies. The DelPhi package has been recently shown to be useful for computing solvation energies for small molecules [32,33]. By calculating the electrostatic energies of isolated and bound molecules, the energy change occurring with molecular association can be estimated [31].

Materials and Methods

Energy minimisation of the protein–substrate complex

The starting point of the calculations described in this report was an all-atom energy-minimised version of high-resolution crystal coordinates of the complex *N*-acetyl-neuraminic-acid (Neu5Ac)–sialidase (from A/Tokyo/3/67 influenza virus). A complete description of the steps taken to obtain the energy-optimised structure is given elsewhere [6] and only the main parts are described here. Energy minimisation was carried out using CVFF [34] within the Discover package (Biosym) on a SGI workstation (Silicon Graphics, Mountain View, CA). The forcefield parameters were all assigned using the automatic procedures within the Insight II version (Biosym). The carboxylic acid of Neu5Ac was ionised, as were the amino acid side chains of arginine, lysine, aspartate, and glutamate. A simple harmonic function for bond stretching was used and all nondiagonal terms were excluded. A cut-off distance of 11.5 Å was used, with a fifth-order polynomial switching function applying after 10.0 Å, and a buffer region extending to 15.0 Å. The effects of hydration were modelled by partially solvating the enzyme–Neu5Ac complex and by using a distance-dependent dielectric constant. The active site of sialidase was solvated by a 15.0-Å shell of water molecules and all of the charged surface residues were solvated by a 3.0-Å shell of water molecules, using the SOAK option of Insight II. A distance-dependent dielectric constant ($\varepsilon = cR$) was incorporated into the Coulombic potential [35]. The value of $c = 1$ was considered the most appropriate, because it produces the best results for short-range electrostatic energies (results not shown). The crystal coordinates of the complex were subjected to a multi-stage energy-optimisation procedure involving atom constraints [6]. The constraints were removed in steps, in order to minimise the adverse effects of strained interactions arising from the refinement [34]. Calculations were done using the algorithms steepest descents (for 500 iterations) and conjugate gradients (down to a maximum atomic root mean square (rms) derivative of 0.01 kcal/mol A^2).

Inhibitors

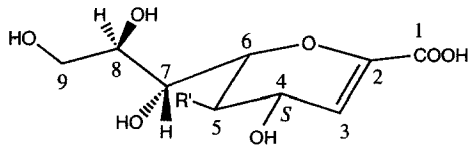
Table 1 lists Neu5Ac2en (**1**) and its analogues (**2–24**) of which the in vitro experimental binding affinities have been previously determined [2]. Each of the functional groups was constructed using Build of Insight II, and the atomic potential types and partial charges were assigned using the automatic methods within that module. The CVFF forcefield was not parameterised for all of the atom types that occur in the 24 inhibitors. Those forcefield parameters not included in the CVFF data libraries were generated by the program using equivalence tables and simple rule-based routines. It was clear that most of the C4 substituents containing an sp^3 nitrogen atom should be protonated (assuming physiological conditions) and the groups assigned an overall positive charge. These included compounds **2–5**, **7**, **10**, **12**, **13**, **15**, **19–21**, and **23**. The hydroxylamines, **6** and **11**, were left unprotonated. The K_i value listed for the 4-guanidino compound **3**, a slow-binding inhibitor, is an initial rate value at a given time and is therefore somewhat arbitrary. The observed time-dependent binding properties are complex, with a log K_i varying from –6.4 after 10 s to –10.5 after 150 min [36]. Accordingly, correlations of the measured and calculated binding strengths are discussed with this compound both included and excluded. For the remaining compounds the K_i values have an error of roughly half an order of magnitude.

For this study we assumed that the loss of translational and rotational entropy which accompanies complex formation, along with the entropic contributions associated with cavity creation, do not affect the rank ordering of binding affinities for ligands of comparable size and shape. By examining a range of closely related compounds which we believe to bind to sialidase in a like manner, we further assumed that the conformational

TABLE 1
EXPERIMENTAL BINDING AFFINITIES FOR Neu5Ac2en (**1**) AND A SERIES OF C4-SUBSTITUTED ANALOGUES

Inhibitor	C4 substituent[a]	K_i (μM)[b]
1	OH	4
2	NH_2	0.04
3	$N{=}C(NH_2)_2$	0.001
4	$NHCH_3$	1
5	$NHCH_2CH{=}CH_2$	6
6	$N(OH)CH_2CH{=}CH_2$	8
7	NH_2 (*R*)	0.1
8	$NHC(=O)CH_3$	200
9	$NHC(=O)NH_2$	55
10	$NHC(=O)CH_2NH_2$	150
11	$N(OH)CH_2CH_2OH$	8
12	$N(CH_3)_2$	3
13	$NHCH_2CH_2OH$	4
14	$NHC(=O)CH{=}CH_2$	70
15	$N{=}C(NH_2)_2$ (*R*)	0.075
16	=O	55
17	$=NOCH_3$	180
18	$=NNH_2$	50
19	$N(CH_2CH_2OH)_2$	180
20	$NHNH_2$	0.1
21	$N(CH_2CH{=}CH_2)_2$	8
22	$N(CH_3)_3^+$	10
23	$N(CH_2CH{=}CH_2)CH_2CH_2OH$	50
24	$=NCH_2CH{=}CH_2$	600

[a] Stereochemistry at C4 is (*S*), unless indicated otherwise.
[b] See Ref. 2.



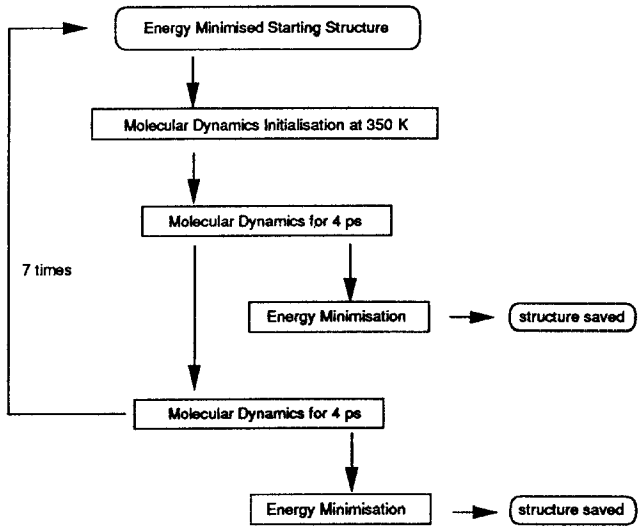
10 11
1 R' = $NHC(O)CH_3$

strain energy of enzyme and ligand in the bound state minus the conformational strain energy of enzyme and ligand in their separated states remains approximately constant for the set. Finally, our approaches require that the free energy gains from enzyme–ligand interactions are much greater than the free energy losses associated with freezing rotatable bonds. Neglecting these factors is not overly significant, because our objective was to reproduce trends in the measured affinities rather than to make quantitative predictions of absolute free energies of binding (for this reason an in-depth statistical analysis of the correlations that we obtained between the experimental binding affinities and calculated interaction energies has not been included in this paper).

Conformational searching

Modes of intermolecular interaction for each enzyme–inhibitor complex about the energy-minimised protein-crystal-structure coordinates were explored using the molecular dynamics/energy-minimisation procedure summarised in Scheme 1. For the energy optimisation of sialidase–inhibitor complexes, our starting coordinates for the protein atoms were those of the energy-minimised crystal structure described above. Beginning with an optimised structure for the protein substantially reduces computing time. All protein residues and crystallographic water molecules containing an atom within 16.0 Å of the bound ligand were free to move. Computing time was further reduced by a reduction in the number of degrees of freedom in the system. Our model thus consisted of the drug candidate, 115 amino acids, 25 crystallographic water molecules, a calcium ion, and solvent molecules. The explicit solvent molecules included in these experiments were reduced to a single 9.0-Å layer of water molecules on the ligand in the active site, and these were restrained by a weak harmonic restraint (force constant of 0.05 kcal/mol A^2). A solvent density slightly lower than that of bulk solvent was used in order to provide the ligand and active-site amino acids greater access to conformational space.

Each ligand was docked into the active site of sialidase with a position and conformation as similar to that of bound Neu5Ac2en (**1**) as possible. All substituents listed in Table 1 were assumed, not unreasonably, to occupy the C4 pocket in the active site of sialidase, as shown in Fig. 1. For the C4 groups of **6**, **11**, **12**, **19**, **20**, and **23** several starting structures were tested. In **20** the two possible modes of protonation of the C4 substituent were examined. The other five species contain tertiary nitrogen atoms. In these examples the larger of the two groups on the nitrogen atom was positioned in the C4 pocket and the second group was directed either towards Asp^{151} or between the side chains of Gly^{119} and Arg^{118}.



Scheme 1. Flow chart for the molecular dynamics/energy-minimisation protocol.

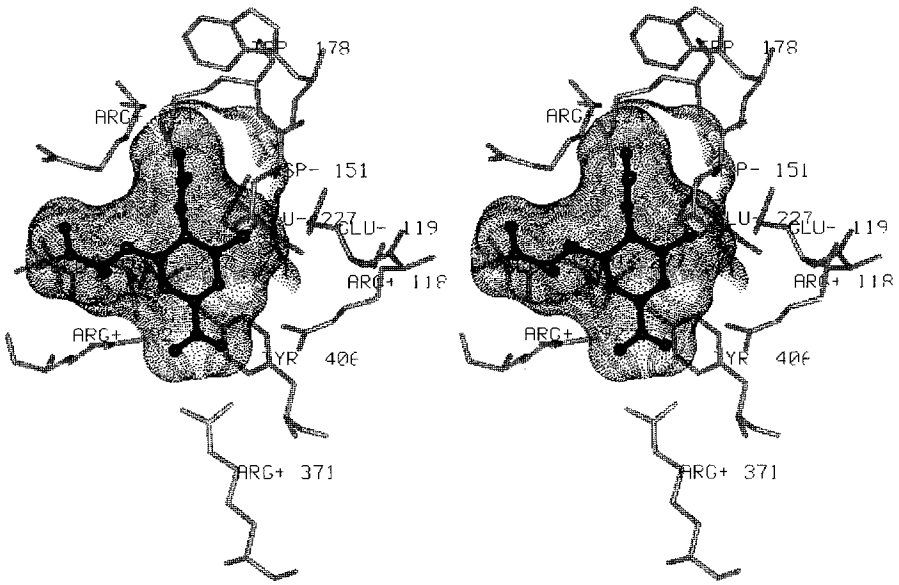


Fig. 1. Stereoview of part of the active site of influenza virus sialidase with bound Neu5Ac2en (**1**). A cavity can be observed between the molecular surface of the inhibitor and the molecular surface around some amino acids adjacent to the C4 hydroxyl group in the ligand.

The initial step in the conformational searching routine is an energy minimisation to relieve steric strain. In all computations the cut-off distance of 8.5 Å was used with the switching function applying beyond 7.0 Å. The energy-minimisation stages were performed by firstly using steepest descents (for 200 iterations) and secondly, conjugate gradients (for 2000 iterations). The molecular dynamics program was run at 350 K using the leapfrog algorithm within DISCOVER, with each short dynamics simulation starting with the original energy-minimised complex. Conformational space was not effectively searched using lower temperatures (less than 300 K), and higher temperatures (greater than 400 K) resulted in the sugar template drifting out of its bound position in the active site (results not included). Temperature initialisation was performed over 2 ps, using a time step of 1 fs, and the dynamics stages ran for 4 ps each, using a time step of 2 fs. Two low-energy conformations are obtained out of every set of randomly selected velocities using this method (see Scheme 1). For each complex a total of fifteen low-energy structures were generated (which include the minimised structure prior to any dynamics), corresponding to 70 ps of simulation time. Calculating larger numbers of structures did not usually generate new conformations (results not shown). A total of approximately 12 h of Silicon Graphics Power Challenge Server (CPU R8000) time was required for each complex.

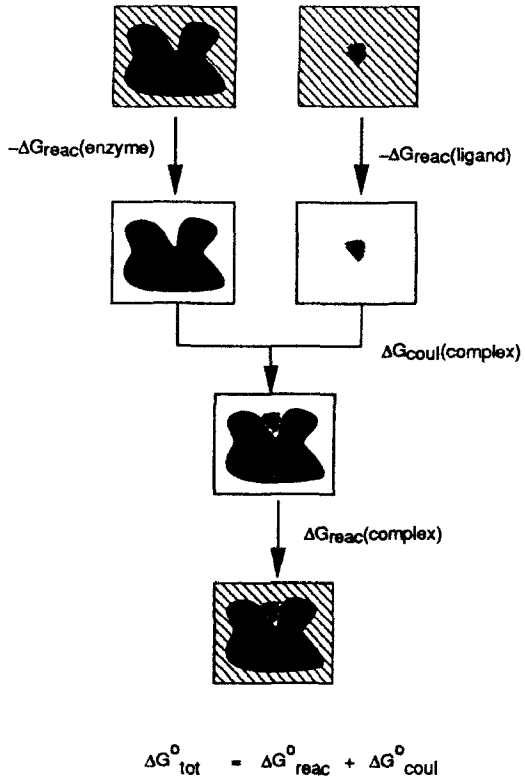
Binding energies from pairwise atomic intermolecular interactions

In the molecular-mechanics-derived interaction (MMI) approach, binding energy is defined as the sum of all pairwise atomic nonbonded intermolecular interactions between protein and ligand. In CVFF these interactions are described by the van der Waals and Coulombic terms, as shown in Eq. 1:

$$\text{Nonbond energy} = \sum (A_i/r^{12} - B_i/r^6) + \sum q_i q_j/\varepsilon r \quad (1)$$

where A_i, B_i = parameters for nonbond repulsive and dispersive interactions; r = atomic separations; q_i = atomic partial charge parameters; ε = dielectric constant. Intermolecular interactions were summed between the entire ligand and each active-site amino acid residue and each active-site structural water molecule. Performing the summation in this manner meant we could analyse the individual contribution to the binding of each active-site residue. In these nonbonded energy summations, nonbond cut-offs were not used but the distance-dependent dielectric still applied. Two values representing the binding affinity were obtained for each ligand from the structure-searching protocol, a maximum binding energy, $E(bind)_{max}$, and the arithmetic mean for the 15 conformations, $E(bind)_{av}$. A Boltzmann-weighted average was not investigated, because conformational space is rigorously sampled.

Tests were applied to ensure that the results were not overly sensitive to nonbond cut-off distances in the energy-optimisations stages, to the use of atom constraints, to the choice of starting conformations, or to residues excluded in the nonbond pairwise summation list. A series of energy optimisations, starting from the enzyme X-ray coordinates, and using the longer cut-off distance of 12.0 Å, produced lower interaction energies in just 30% of cases, with average and maximum differences of 5% and 9%, respectively, for those examples. Furthermore, the inclusion



Scheme 2. Flow chart of the thermodynamic process for calculating the change in total electrostatic free energy upon bimolecular association.

of all protein atoms in the nonbonding summation gives a minor systematic difference of approximately 3%. Finally, repeating the simulations for certain complexes and generating fifteen new conformations showed us that the calculated energy values were essentially reproducible.

Binding energies from continuum modelling

In the continuum electrostatics (CE) approach, values for the binding energy were determined by applying the thermodynamic process illustrated in Scheme 2. These steps are formulated in Eqs. 2–4:

$$\Delta G^{\circ}_{tot} = \Delta G^{\circ}_{coul} + \Delta G^{\circ}_{reac} \tag{2}$$

$$\Delta G^{\circ}_{coul} = G^{\circ}_{coul}(EI) - G^{\circ}_{coul}(E) - G^{\circ}_{coul}(I) \tag{3}$$

$$\begin{aligned} \Delta G^{\circ}_{reac} &= \Delta G^{\circ}_{reac}(EI) - \Delta G^{\circ}_{reac}(E) - \Delta G^{\circ}_{reac}(I) \\ &= [G^{\circ}_{solvent}(EI) - G^{\circ}_{solute}(EI)] \\ &\quad - [G^{\circ}_{solvent}(E) - G^{\circ}_{solute}(E)] \\ &\quad - [G^{\circ}_{solvent}(I) - G^{\circ}_{solute}(I)] \end{aligned} \tag{4}$$

The energy of bimolecular association, ΔG°_{tot}, is defined in this case as the sum of: (i) the change in Coulombic free energy, ΔG°_{coul}, which is a measure of all protein to ligand charge–charge, charge–dipole, and dipole–dipole interactions; and (ii) the change in reaction field free energy, ΔG°_{reac}, which corresponds to the electrostatic energy of transfer from one medium of continuum dielectric to another [31]. The subscripts 'coul' and 'reac' refer to the Coulombic and reaction field energies, respectively; 'solvent' denotes an external, aqueous dielectric constant of 78.6; and 'solute' denotes an internal, protein dielectric constant of 2.0; EI = enzyme–inhibitor complex; E = enzyme; I = inhibitor. The calculated low-energy structures from the conformational searching routine were used for the geometries of the complex, enzyme, and ligand in the thermodynamic process (water molecules comprising the solvent cap were excluded from the energy calculations). The various terms in Eqs. 3 and 4 were then computed using the DelPhi package from Biosym. The parameters for the atomic radii and partial charges for the protein were taken from the standard CVFF set [34]. For the ligands, the charges were determined using the automatic routine within the Builder module of InsightII. The method of focusing, which involves two DelPhi calculations, was used to minimise the errors at the boundaries

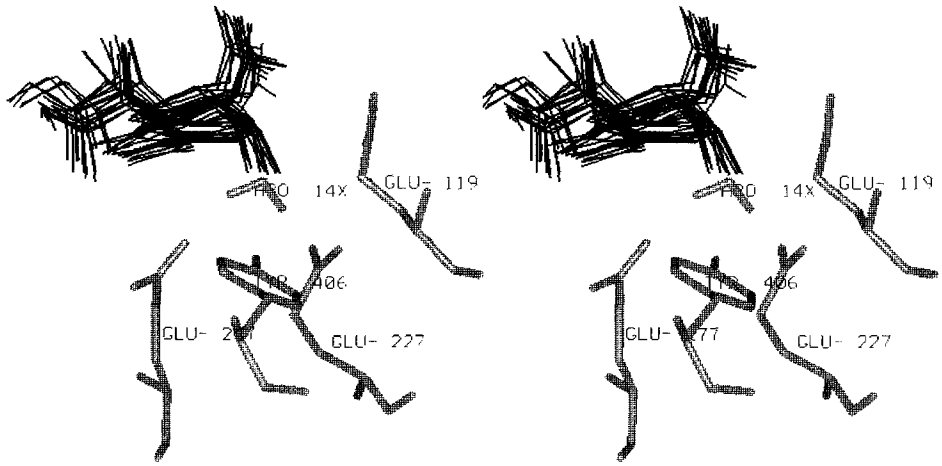


Fig. 2. The overlap in the calculated structures for 7 obtained from superimpositions of the protein atoms. The amino acid side-chain positions are those found in the lowest-energy structure.

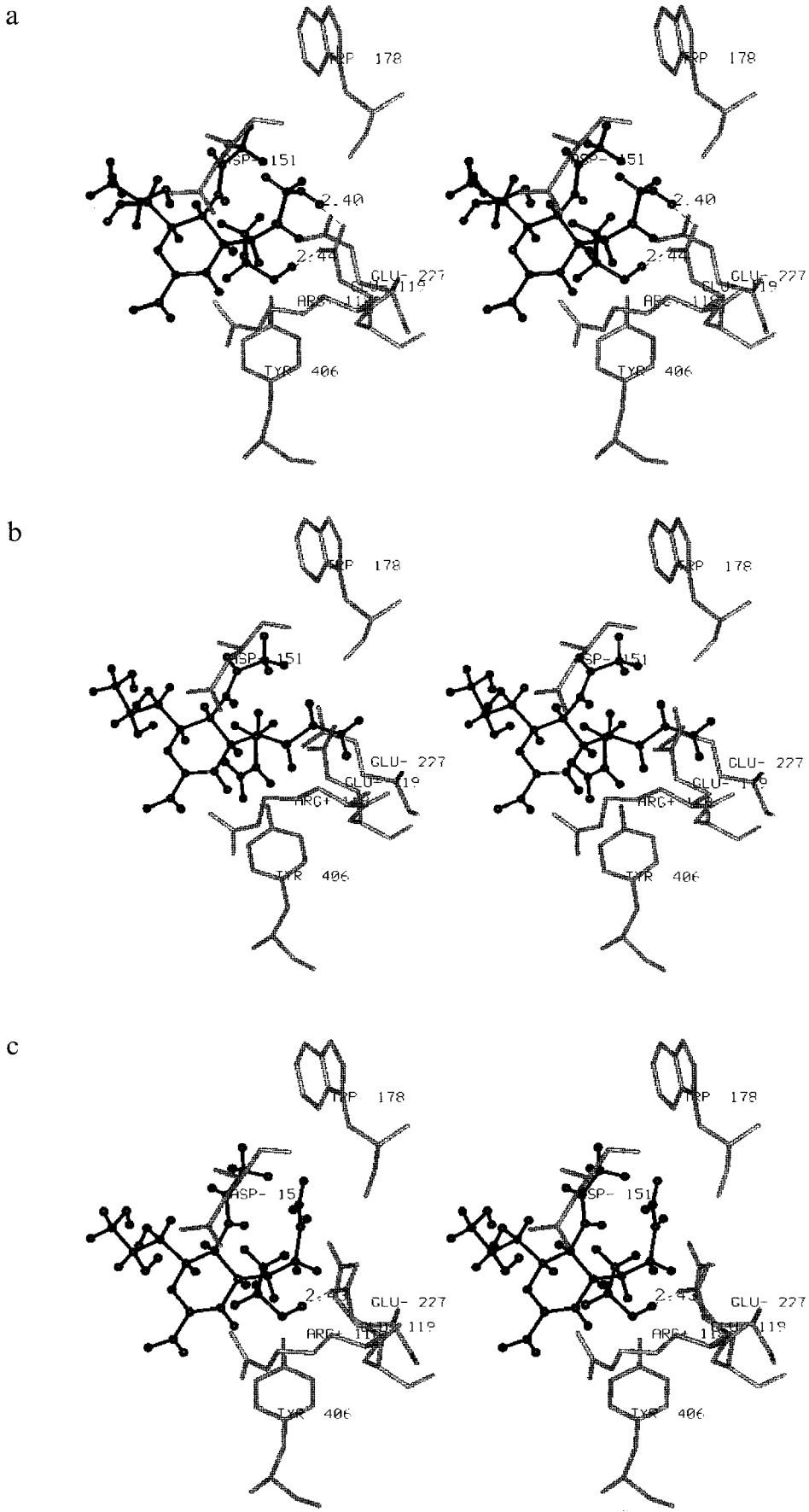


Fig. 3. The most favourable binding modes predicted for the three bulkiest C4 substituents: (a) **19**; (b) **21**; and (c) **23**.

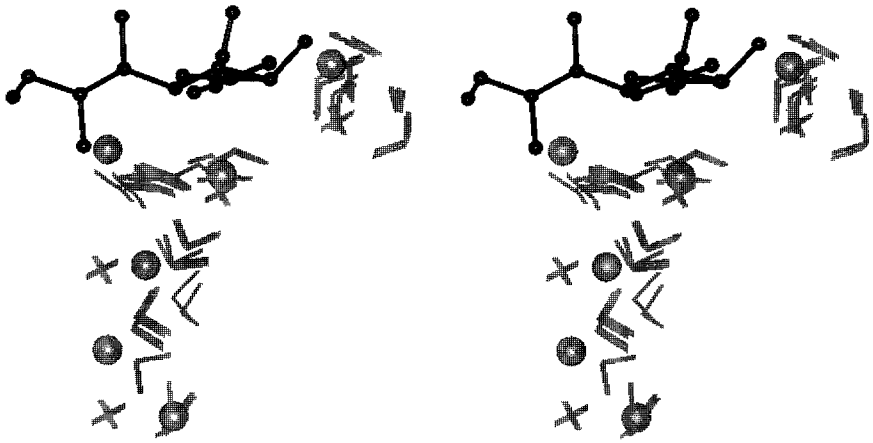


Fig. 4. A superimposition of: (i) water molecules observed in an early crystal structure of sialidase (crosses); (ii) the positions of the water molecules in all 15 calculated structures for compound **2**, obtained from superimpositions of protein atoms (stick models); and (iii) the observed water molecules in the active site of highly refined native sialidase (spheres). See text for details.

of the grid (DelPhi Manual, Biosym Technologies Inc., San Diego, CA, 1992). In this method, the values at the boundary points are interpolated from the results obtained from a calculation with a larger grid. In this work, the grid extended 15.0 Å beyond the molecule in the initial run and 10.0 Å beyond it in the final calculation. A large grid of 91 points along the edge was used in both cases, giving a resolution of 0.87 Å per grid point. A test case using the finer grid scale of 0.65 Å per grid point resulted in only minor energy differences.

TABLE 2
CALCULATED BINDING ENERGIES (KCAL/MOL) FROM THE MMI METHOD FOR THE SERIES OF 4-SUBSTITUTED DERIVATIVES OF Neu5Ac2en (**1**)

Inhibitor	$E(bind)_{max}$	$E(bind)_{av}$
1	−200	−187
2	−236	−211
3	−233	−223
4	−214	−202
5	−228	−210
6	−206	−186
7	−229	−215
8	−191	−181
9	−192	−179
10	−236	−227
11	−208	−194
12	−206	−182
13	−230	−216
14	−202	−182
15	−231	−213
16	−177	−166
17	−189	−173
18	−189	−171
19	−233	−210
20	−239	−216
21	−202	−189
22	−202	−188
23	−220	−202
24	−185	−177

As before, the binding affinities for each ligand were defined as both the average and maximum values of a set of low-energy conformations. Four combinations of calculated ΔG°_{reac} and ΔG°_{coul} values were investigated. Two different values for the internal dielectric in the calculation of ΔG°_{coul}, 1.0 and 2.0, were tested, and two different approaches in the calculation of ΔG°_{reac}, with the external dielectric constant changing from 78.6 to 1.0, as well as from 78.6 to 2.0, were examined. An internal dielectric constant of 1.0 was tested because the CVFF atom-centred charge parameters were developed using this value. In addition to the plots of ΔG°_{tot} versus $\log K_i$, the individual terms ΔG°_{reac} and ΔG°_{coul} were tested for correlations with the experimental results. Finally, a set of comparisons were made in which the change in reaction field associated only with each ligand in isolation was considered. That is, the term $\Delta G^\circ_{reac}(I)$ was combined with ΔG°_{coul} for the calculation of $\Delta G^\circ_{tot}(I)$. This is shown in Eq. 5:

$$\Delta G^\circ_{tot}(I) = \Delta G^\circ_{coul} + \Delta G^\circ_{reac}(I) \quad (5)$$

For a set of 15 low-energy structures, approximately 24 h of Silicon Graphics Power Challenge Server (CPU R8000) time was required using the selected parameters.

Results and Discussion

Structural features of inhibitor binding

The pocket in the active site of sialidase targeted by Neu5Ac2en (**1**) and its C4-modified analogues is illustrated in Fig. 1. It is clear from the diagram that substituents with basic moieties at C4 should form favourable electrostatic interactions with the negatively charged side chains of Glu^{119}, Asp^{151}, and Glu^{227}. The size limitations of the cavity are not as obvious and C4 functional groups with up to four heavy atoms in a chain have been synthesised and tested for activity.

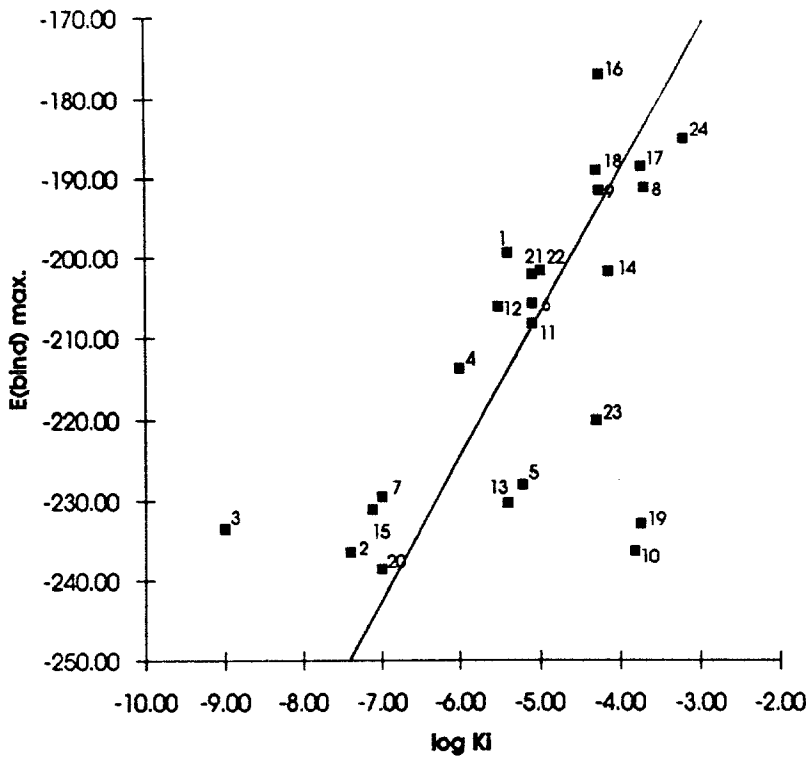


Fig. 5. Experimental binding affinities (μM) versus the maximum binding energy (kcal/mol) calculated using the MMI method.

In an earlier paper, the complex set of interactions for compounds **1**, **2**, and **3** with active-site residues of sialidase has been described [6]. In summary, the C4-hydroxyl in **1** was found to form a strong hydrogen bond with either a water molecule, Glu^{119}, or Asp^{151}; the C4 primary amine in **2** formed strong hydrogen bonds with each of the charged groups and a water molecule was usually involved; and this was also found for the C4-guanidino in **3**, which similarly formed favourable contacts with Glu^{227} and the backbone oxygen atom of Trp^{178}.

Compound **7** was calculated to fit into the active site with its ring in the same position as its epimer **2**. Figure 2 shows that the fifteen calculated ligand conformations overlap very closely, indicative of a tight-binding inhibitor. The amine group in **7** fitted between the hydroxyl of Tyr^{406} and the acidic group of Glu^{119} and Glu^{227}. Close hydrogen-bonding contacts were made with these residues, and also with two water molecules 14X and 6X. In approximately half of the calculated minima, the amine formed a short hydrogen bond with Glu^{277}. The hydrogen-bonding interactions in the calculated structures are comparable with those calculated for compound **2**, which is consistent with their equal affinities. By contrast, compound **15** did not fit into the active site, nor did its epimer **3**. The network of hydrogen bonds between the guanidino group and the protein was similar to the network described for compound **7**, however the ring remained 1–2 Å further out from its position in the crystal structure of the Neu5Ac2en–sialidase complex.

The compounds with the three bulkiest substituents at C4 (**19**, **21**, and **23**) were calculated to bind in different modes to one another. In the low-energy structures for all three inhibitors the six-membered ring was significantly displaced away from the C4 pocket, towards the opposite side of the active site. By far the most stable structures calculated for **19** had both hydroxyl groups forming short hydrogen bonds to Glu^{119} with the proton on the nitrogen pointing towards the floor of the active site. For compound **21**, the most favourable binding mode calculated had one of the olefin groups in the C4 pocket and the other directed between residues Arg^{118} and Glu^{119}. In this arrangement, the proton on the nitrogen atom was directed towards the back of the active site. The result of testing three starting approximations for compound **23** was a set of lowest-energy structures with the olefin occupying the C4 pocket, the hydroxyl group forming a hydrogen bond with Glu^{119}, and the proton pointing into the floor of the active site. Figure 3 illustrates these three different binding modes.

The calculated structures for the hydrazonium analogue (**20**) displayed close similarities to the structures predicted for **2** [6]. In the lowest-energy structures, based solely upon enzyme–ligand interactions, the terminal nitrogen atom was protonated. The amino-acetamide substituent at C4 in compound **10** was of particular interest. The primary amine of **10** was assumed protonated in the calculations and was found to interact very favourably with Glu^{227} and the backbone oxygen of Trp^{178}. For the

three methyl-substituted amine derivatives (**4**, **12**, and **22**) and the two large groups in **5** and **13**, the patterns for binding were as expected. In the two hydroxylamine species (**6** and **11**) the hydroxyl groups on the nitrogen atom pointed in different directions in the lowest-energy structures. In **6**, the OH group formed a hydrogen bond to Glu^{119} and in **11**, the group was hydrogen-bonded to Asp^{151}. Finally, none of the remaining seven compounds (**8**, **9**, **14**, **16**, **17**, **18**, and **24**) formed favourable hydrogen bonding or electrostatic interactions with the protein residues that line the C4 pocket.

The crystallographic water molecules included in the calculations exhibited very interesting behaviour. Figure 4 shows the starting positions for some of the bound water molecules in the vicinity of the inhibitor taken from an early crystal structure of N2 sialidase (as determined by Dr. Peter Colman, Division of Biomolecular Structure, CSIRO, Parkville, Australia). The calculated positions of these water molecules, obtained from the simulation of compound **2**, are also presented (in the simulations for the other inhibitors, very similar binding patterns were generated). Finally, the waters of crystallisation in native N9 sialidase, taken from a highly refined structure, are illustrated (also from Dr. Peter Colman). Four of these water molecules, those closest to the bound ligand, can also be seen in the crystal structure of sialidase from influenza B [37]. The calculations appear to have successfully predicted the positions of all of the bound water molecules in that region of the active site. Finally, the spread in the calculated positions of any single water molecule gives us information on the tightness of binding. For example, the water inside the C4 pocket can be seen to bind in a variety of modes, suggesting it may be displaced, whereas the water directly beneath the acetamido group of Neu5Ac is of the opposite type, and therefore more tightly bound.

Binding energies from the molecular-mechanics-derived interactions (MMI)

Table 2 lists the binding energies determined from the summation of intermolecular nonbonded energies. Figures 5 and 6 illustrate the graphs of calculated energies versus experimental K_i values. There appears to be an apparent trend between calculated and experimental results in both plots. However, the correlation coefficients for $E(bind)_{max}$ and $E(bind)_{av}$ versus experimental K_i values are low: R^2= 0.37 and 0.35, respectively. This result is due to three outliers; when they are excluded the R^2 values become 0.69 and 0.63, respectively. The three outlying results involve compounds **3**, **10**, and **19**. It was noted above that the K_i value used for **3**, with a guanidino group at C4, has a large degree of ambiguity. If its binding affinity after 10 s was used, the data point would be consistent with the other results. The poor result for compound **10**, which contains a charged amino-acetamide group, is more difficult to explain. Favourable charge–charge interactions between the primary amine on the C4 substituent and Glu^{227} produced the low binding energy. We reasoned

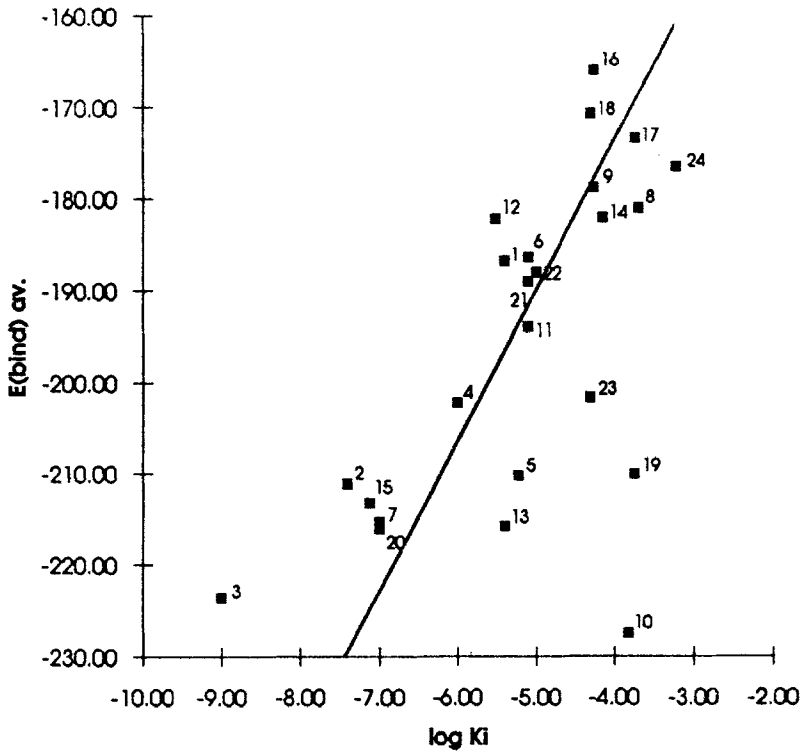


Fig. 6. Experimental binding affinities (μM) versus the average binding energy (kcal/mol) calculated using the MMI method.

TABLE 3
RESULTS OBTAINED FROM THE CE METHOD FOR CALCULATING RELATIVE BINDING ENERGIES (EQS. 2–4)

Inhibitor	Maximum values			Average values		
	ΔG°_{reac}	ΔG°_{coul}	ΔG°_{tot}	ΔG°_{reac}	ΔG°_{coul}	ΔG°_{tot}
1	–	−117	−13	106	−105	1
2	–	−170	−20	135	−145	−10
3	–	−170	−21	147	−160	−13
4	–	−149	−12	131	−138	−7
5	–	−158	−12	138	−141	−3
6	–	−108	−3	108	−100	8
7	–	−166	−23	137	−150	−13
8	–	−104	0	106	−96	10
9	–	−107	0	106	−98	9
10	–	−174	−16	152	−162	−10
11	–	−121	−5	115	−110	5
12	–	−139	−6	128	−125	3
13	–	−158	−16	135	−144	−9
14	–	−105	4	105	−91	14
15	–	−170	−21	136	−149	−14
16	–	−95	2	97	−87	10
17	–	−106	−3	98	−91	7
18	–	−102	4	99	−91	9
19	–	−158	−13	139	−141	−2
20	–	−175	−25	143	−155	−11
21	–	−135	−1	131	−123	8
22	–	−139	−7	131	−128	2
23	–	−148	−13	134	−132	2
24	–	−97	6	100	−90	11

Energies are in kcal/mol.

that energy barriers prevent this low-energy binding mode occurring in practice. As mentioned above, compound **19**, with a charged 2,2'-dihydroxy-diethylamine at C4, formed an unusual bidentate binding interaction with Glu[119]. The binding of this compound produced a number of changes in the geometry of the active site which is not accounted for in the MMI approach. This is because the method simply measures the relative strengths of direct protein–ligand interactions. Furthermore, the binding of **14** results in a loss of many additional degrees of bond rotational freedom relative to other ligands, and this may be significant.

The compounds containing the positively charged substituents (**2**, **3**, **5**, **7**, **10**, **13**, **15**, **19**, and **20**) form a group having the lowest (strongest) calculated intermolecular binding energy. All but the four compounds with bulky groups at C4 (**5**, **10**, **13**, and **19**) exhibit very tight binding in vitro. The remaining five groups with a positive charge at C4 (**4**, **12**, **21**, **22**, and **23**, and compounds **6** and **11**) comprise the majority of the other low-energy complexes calculated. The compounds with a relatively small, positively charged group (**4** and **12**) have lower K_i values than the compounds with a large group (**6**, **11**, **14**, **21**, **22**, and **23**). Thus, there are two clear trends in the calculated results: (i) the approach is roughly able to reproduce the favourable binding of inhibitors with positively charged functional groups at the C4 position; and (ii) the binding of bulky substituents at C4 can be overestimated. The higher than expected calculated binding energies for large groups is believed to be the result of neglecting entropic effects associated with freezing rotatable bonds and energetic costs associated with deforming the geometries of active-site residues. Both of these effects were initially assumed not to play important roles in the binding of the ligands to sialidase; clearly this is not the case for the bulkiest C4 groups. With regard to those compounds containing a hydroxyl or alkene functionality, no obvious trends could be discriminated. The binding of inhibitors containing these groups was dominated by charge and size factors.

From inspection of the different structures obtained for each compound and the corresponding energies of interaction, it was apparent that similar binding energies could be achieved by quite different electrostatic and hydrogen-bonding arrangements. The moieties involved were commonly the hydroxyls on the ligands, negatively charged Glu and Asp residues, and water molecules bridging between enzyme and ligand. It appears that a degree of degeneracy for intermolecular interaction exists and this might lower the entropy penalty associated with bimolecular association. A related phenomenon is the favourable entropy of residual motions in complexes described by Searle and Williams [22].

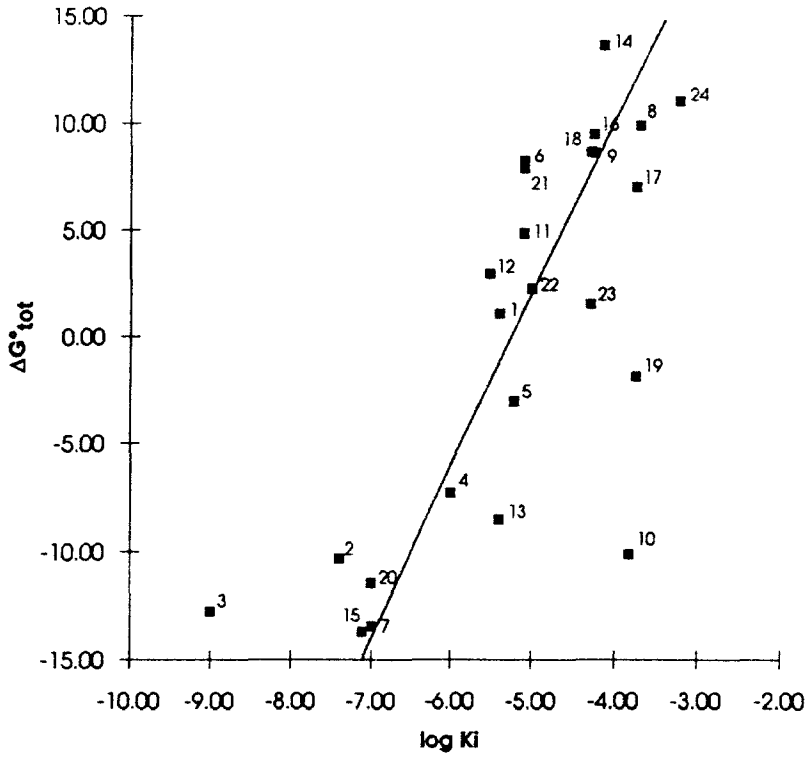


Fig. 7. Experimental binding affinities (μM) versus the average total binding energy (kcal/mol) calculated using the CE method.

Binding energies from the continuum electrostatics (CE) method

Table 3 lists the maximum and average binding-energy values for the enzyme–inhibitor complexes calculated using Eqs. 2–4. Figure 7 illustrates the plot of the average ΔG°_{tot} energies versus experimental affinities. A clear trend can be observed between the experimental and calculated results. Improved correlation coefficients of $R^2 = 0.56$, for all 24 compounds, and $R^2 = 0.80$, minus the three outliers, were obtained. The correlations for the maximum values were slightly lower than those obtained for the averages: $R^2 = 0.54$ for 24 compounds, and $R^2 = 0.77$ for 21 compounds. In the influenza-virus-sialidase–inhibitor system, electrostatic interactions certainly dominate the binding energy. However, contributions from other factors, such as changes in attractive and dispersive forces and entropic effects, would still be expected to be significant. With this in mind, very high correlations would not be expected.

The major trend in the results is the same as that observed after application of the MMI method: ligands with a positively charged group at C4 bind tightest. At the low end of the calculated binding-energy scale there is a group of eight inhibitors each of which contain a positively charged C4 substituent. Compounds **7** and **15**, the two epimers at C4, were predicted to form the most stable complexes. Those ligands with bulky groups at C4 were distributed more evenly above and below the line of best fit than observed previously in Figs. 5 and 6. Furthermore, one of the three outliers, **19**, has a closer fit with experiment. These observations indicate that application of the Poisson equation and inclusion of solvation effects within the CE method has resulted in improved correlations between experiment and theory over the MMI approach.

None of the alternative methods of matching calculated values with the experimental data gave improved correlations. The trend between log K_i and calculated energy was very poor in the experiment in which ΔG°_{reac} was determined using an internal dielectric constant of $\varepsilon = 1.0$ (results not shown). For the results in which ΔG°_{coul} was calculated using an internal dielectric of $\varepsilon = 1.0$, the fit between experiment and theory was closely related to the fit obtained between ΔG°_{coul} and log K_i. In these cases the favourable attraction between a positively charged ligand substituent and the acids which line the C4 pocket was overcompensated.

The correlation of ΔG°_{coul} versus ΔG°_{reac} was unexpectedly high, $R^2 = 0.96$. This result becomes clearer bearing in mind that the geometries for the enzyme and ligand remain constant throughout the determination of ΔG°_{coul} and ΔG°_{reac}. Coulombic energy is proportional to the products $q_i q_j$, where q_i is the partial charge on atom i, and i is not equal to j. Reaction field energy is similarly dependent upon these products, although with different distance-dependence factors. Clearly, the distance dependence in the equations is not dominating, and similarly, the self terms, $q_i q_i$, included in the calculation of the reaction fields, do not have much overall effect.

The results obtained using Eq. 5 produced a similar fit with experiment as that described for Eq. 2. It is therefore conceivable that one can use Eq. 5 in favour of Eq. 2 to model ligand-binding energetics. However, computing time is not significantly decreased using this approach, because Coulombic energies must still be calculated using the finite difference solutions to the linear Poisson equation.

It has often been suggested that the assignment of macroscopic dielectric constants to protein interiors may be a poor representation of electrostatic shielding effects in forcefield calculations. It is well-understood that the internal dielectric constant of a protein is a function of position; the dielectric constant near the protein surface has a greater value than at the centre of the protein [38]. Furthermore, dielectric properties near protein surfaces have a greater sensitivity to neighbouring groups than at the hydrophobic cores [38]. Despite these obvious deficiencies, an internal dielectric constant of 2.0 takes into account conformational polarisation, facilitates improved estimations of electrostatic interactions, and the CE approach has been able to reproduce the trends observed in the measured binding affinities of the sialidase inhibitors with a reasonable degree of accuracy.

Conclusions

In this paper we described a molecular dynamics/energy-minimisation protocol developed for the study of protein–ligand interactions. Beginning with a crystal structure of a protein–ligand complex, we sought to extend our understanding of the intermolecular interactions and to predict the binding of structurally related species. We obtained a variety of low-energy structures from a molecular-dynamics-based conformational searching method in which random velocities were continually reset. Two major strengths of the approach were (i) the explicit inclusion of ligand and amino acid side-chain flexibility; and (ii) fast and efficient conformational searching. Calculated results were visualised as individual structures using interactive computer graphics and superimposed sets of conformers. A range of enzyme–ligand-binding modes, not influenced by the bias of the researcher, was found to provide insights into new possibilities for intermolecular interactions and the design of more potent inhibitors. The approach was also useful for predicting the numbers, positions, and orientations of bound water molecules in the sialidase–inhibitor system.

Two methods of estimating relative binding energies were tested: a molecular mechanics derived interaction (MMI) method and a continuum electrostatics (CE) approach. The MMI method achieved some level of success considering the simplicity of the approach. Trends between experimental and calculated results could be distinguished. Application of the CE method resulted in a marked improvement in the correlation between theory and experiment. For both models, assumptions that strain energy and bond-rotational entropy do not significantly affect the relative binding affinities appeared to hold for all ligands, except for those with very bulky substituents. In general, there was little difference between the use of average and maximum binding energies calculated from each method. It is interesting to compare the maximum energies obtained from the MMI approach with those obtained from the CE method. In several cases the lowest energies obtained came from different enzyme–inhibitor conformations. It appears that the hydrogen-bonding requirements of the carbohydrate-based inhibitors studied can be satisfied by the multitude of polar groups in the active site of sialidase in various modes with very similar energies.

Acknowledgements

This work was financed by Glaxo Australia and we thank them for their support. The authors are indebted to many colleagues at the Victorian College of Pharmacy, Monash University, for their contributions to the synthesis, isolation, characterisation, and testing of the compounds covered in this research. Thanks are extended to Dr M. Hann for many useful discussions and assistance with the manuscript and to Dr. P. Colman for supplying the X-ray crystal coordinates of the various influenza virus sialidases.

References

1 von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., 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. and Penn, C.R., Nature, 363 (1993) 418.
2 Holzer, C.T., von Itzstein, M., Jin, B., Pegg, M.S., Stewart, W.P. and Wu, W.-Y., Glycoconjugate J., 10 (1993) 40 .
3 von Itzstein, M., Wu, W.-Y. and Jin, B., Carbohydr. Res., 259 (1994a) 301.
4 Chong, A.K., Pegg, M.S. and von Itzstein, M., Biochem. Int., 24 (1991a) 165.
5 Chong, A.K., Pegg, M.S., Taylor, N.R. and von Itzstein, M., Eur. J. Biochem., 207 (1992) 335.
6 Taylor, N.R. and von Itzstein, M., J. Med. Chem., 37 (1994) 616.
7 Colman, P.M., Pept. Protein Rev., 4 (1984) 215.
8 Chong, A.K., Pegg, M.S. and von Itzstein, M., Biochim. Biophys. Acta., 1077 (1991b) 65.
9 Varghese, J.N., McKimm-Breschkin, J., Caldwell, J.B., Kortt, A.A. and Colman, P.M., Proteins, 14 (1992) 327.
10 von Itzstein, M., Dyason, J.C., Oliver, S.W., White, H.F., Wu, W.-Y., Kok, G.B. and Pegg, M.S., J. Med. Chem., 39 (1996) 388.
11 Varghese, J.N. and Colman, P.M., J. Mol. Biol., 221 (1991) 473.
12 Howard, A.E. and Kollman, P.A., J. Med. Chem., 31 (1988) 1660.
13 Leach, A.R., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry II, VCH, New York, NY, 1991, p. 1.

14 Wong, C.F. and McCammon, J.A., J. Am. Chem. Soc., 108 (1986) 3830.
15 Van Gunsteren, W.F. and Berendsen, H.J.C., J. Comput.-Aided Mol. Design, 1 (1987) 171.
16 McCammon, J.A., Science, 238 (1987) 486.
17 Lee, F.S., Chu, Z.-T., Bolger, M.B. and Warshel, A., Protein Eng., 5 (1992) 215.
18 Andrews, P.R., Craik, D.J. and Martin, J.L., J. Med. Chem., 27 (1984) 1648.
19 Novotny, J., Bruccoleri, R.E. and Saul, F.A., Biochem., 28 (1989) 4735.
20 Williams, D.H., Aldrichim. Acta, 24 (1991) 71.
21 Horton, N. and Lewis, M., Protein Sci., 1 (1992) 169.
22 Searle, M.S. and Williams, D.H., J. Am. Chem. Soc., 114 (1992) 10690.
23 Searle, M.S., Williams, D.H. and Gerhard, U., J. Am. Chem. Soc., 114 (1992) 10697.
24 Murphy, K.P., Xie, D., Garcia, C., Amzel, L.M. and Freire, E., Proteins, 15 (1993) 113.
25 Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comp. Chem., 13 (1994) 380.
26 Lawrence, M.C. and Davis, P.C., Proteins, 12 (1992) 31.
27 Böhm, H.-J., J. Comput.-Aided Mol. Design, 8 (1994) 243.
28 Miller, M.D., Kearsley, S.K., Underwood, D.J. and Sheridan, R.P., J. Comput.-Aided Mol. Design, 8 (1994) 153.
29 Gilson, M.K. and Honig, B., Nature, 330 (1987) 84.
30 Gilson, M.K., Sharp, K. and Honig, B., J. Comp. Chem., 9 (1988) 327.
31 Gilson, M.K. and Honig, B., Proteins, 4 (1988) 7.
32 Alkorta, I., Villar, H.O. and Perez, J.J., J. Comp. Chem., 14 (1993) 620.
33 Sitkoff, D., Sharp, K.A. and Honig, B., J. Phys. Chem., 98 (1994) 1978.
34 Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M. and Hagler, A.T., Proteins, 4 (1988) 31.
35 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
36 Pegg, M.S. and von Itzstein, M., Biochem. Mol. Biol. Int., 32 (1994) 851.
37 Burmeister, W.P., Ruigrok, R.W.H. and Cusack, S., EMBO. J., 11 (1991) 49.
38 Smith, P.E., Brunne, R.M., Mark, A.E. and Van Gunsteren, W.F., J. Phys. Chem., 97 (1993) 711.