Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 2. Computational Titration and pH Effects in Molecular Models of Neuraminidase−Inhibitor Complexes

Micaela Fornabaio,[†,§] Pietro Cozzini,[‡] Andrea Mozzarelli,*[,†,§] Donald J. Abraham,[||] and Glen E. Kellogg*[,||]

Department of Biochemistry and Molecular Biology, Department of General and Inorganic Chemistry, National Institute for the Physics of Matter, University of Parma, 43100 Parma, Italy and Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Richmond, Virginia 23298-0540

Received May 28, 2003

One factor that can strongly influence predicted free energy of binding is the ionization state of functional groups on the ligands and at the binding site at which calculations are performed. This analysis is seldom performed except in very detailed computational simulations. In this work, we address the issues of (i) modeling the complexity resulting from the different ionization states of ligand and protein residues involved in binding, (ii) if, and how, computational methods can evaluate the pH dependence of ligand inhibition constants, and (iii) how to score the protonation-dependent models. We developed a new and fairly rapid protocol called "computational titration" that enables parallel modeling of multiple ionization ensembles for each distinct protonation level. Models for possible protonation combinations for site/ligand ionizable groups are built, and the free energy of interaction for each of them is quantified by the HINT (Hydropathic INTeractions) software. We applied this procedure to the evaluation of the binding affinity of nine inhibitors (six derived from 2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid, DANA) of influenza virus neuraminidase (NA), a surface glycoprotein essential for virus replication and thus a pharmaceutically relevant target for the design of anti-influenza drugs. The three-dimensional structures of the NA enzyme−inhibitor complexes indicate considerable complexity as the ligand−protein recognition site contains several ionizable moieties. Each computational titration experiment reveals a peak HINT score as a function of added protons. This maximum HINT score indicates the *optimum* pH (or the optimum protonation state of each inhibitor−protein binding site) for binding. The pH at which inhibition is measured and/or crystals were grown and analyzed can vary from this optimum. A protonation model is proposed for each ligand that reconciles the experimental complex structure with measured inhibition and the free energy of binding. Computational titration methods allow us to analyze the effect of pH in silico and may be helpful in improving ligand binding free energy prediction when protonation or deprotonation of the residues or ligand functional groups at the binding site might be significant.

Introduction

Computational methods provide very useful and powerful tools for modeling of proteins, for evaluation of the strength of protein−ligand interactions, and for understanding structure−activity relationships. These tools can assist in the discovery and optimization of lead compounds with enormous advantages for pharmaceutical development by reducing synthetic efforts and time. To this goal, the computational prediction of the binding affinity between a protein and its putative ligands is the most crucial issue. A wide variety of methods have been developed to address the problem, ranging from very elementary "back of the envelope" calculations, e.g., the Peter Andrews model,[1] to sophisticated and more computationally demanding free energy calculation methods such as those pioneered by Peter Kollman[2,3] and others such as Åqvist[4,5] and Jorgensen.[6,7] For more details on binding free energy calculation methods, see reviews by Ajay and Murcko,[8] Wang et al.,[9] Gohlke and Klebe,[10] Lazaridis,[11] and Cozzini et al.[12] However, many of these approaches provide, at most, a very approximate evaluation of the entropic component of binding free energy, while most of the others are very difficult and expensive to undertake. To more properly take into account the entropic components of free energy, a "natural" force field based on the experimentally determined log $P_{o/w}$ (partition coefficient for 1-octanol/water) was developed. log $P_{o/w}$ is an equilibrium measurement of the ratio of these concentrations and is thus related to free energy.[13,14] The software model based on this force field, HINT (Hydropathic INTeractions),[15−18] was recently used to evaluate the free energy of binding of 53 protein−ligand complexes of known three-dimensional structure. A rather good correlation between the

* To whom correspondence should be addressed. A.M.: phone: +39-0521-905138, fax: +39-0521-905151, e-mail: biochim@unipr.it; G.E.K.: phone: 804- 828-6452, fax: 804-827-3664, e-mail: glen.kellogg@vcu.edu.
† Department of Biochemistry and Molecular Biology, University of Parma.
‡ Department of General and Inorganic Chemistry, University of Parma.
§ National Institute for the Physics of Matter, University of Parma.
|| Virginia Commonwealth University.

10.1021/jm0302593 CCC: $25.00
Published on Web 09/06/2003

HINT scores and the experimental dissociation constants was observed.[19] Earlier investigations of protein−protein associations in native and mutant hemoglobins[20−22] also showed very good correlations between HINT scores and the free energy of dimer−dimer association.

An important aspect of the modeling of ligand−protein complexes, the dependence of ligand affinity on the ionization state of residues and/or ligand functional groups involved in binding, was only partially addressed in our previous work.[19] For one of the complexes examined, a phosphonate-containing ligand bound to penicillopepsin, the experimental binding free energy was available at three distinct pH values[23] and a striking correlation was observed with the HINT score as a function of pH.[19] This interesting observation suggested that a more in-depth investigation of protein−ligand complexes where ionization state and the associated *local* pH are expected to play significant roles in binding free energy was potentially valuable. This is a key issue in many drug design projects because virtual screening experiments of databases against protein targets are almost exclusively carried out with no or very limited attempts to optimize the ionization state of residues surrounding the binding site and/or functional groups on the ligands themselves. Predictions of local pH are confounded because the protein and ligand affect each other as the binding occurs.

The description of the protonation equilibria and tautomerism of proteins and ligands and the prediction of the actual protonation state and pK_a of ionizable groups have been the subject of several investigations (e.g., refs 24−34, 35 (and references therein), 36−38). It should also be pointed out that there is not a *single* global model for each protein−ligand system, i.e., a well-determined ionized state for each defined residue, because protons are not static and the ionization state of residues is a group function.[39] In other words, there may be multiple energetically accessible states for each complex. Proton transfer between molecules and, in particular, proton migration across hydrogen bonds has been identified as one of the fundamental mechanisms for biological processes,[40] and the energy barriers for these transfers are expected to be small.

In this paper we address two fundamental questions: (1) how can the complexity arising from different possible ionization states of residues involved in protein−ligand binding be properly modeled and (2) can computational methods evaluate the dependence of the dissociation constants of ligands on pH? We developed a protocol, called computational titration, that includes a careful modeling of the ionization states and resonance forms for the ligands and protein residues at the binding site and the evaluation of the free energy of binding for modeled complexes by HINT. This procedure is applied to a collection of influenza virus neuraminidase−inhibitor systems. Influenza virus neuraminidase (NA) is a tetramer of identical subunits (MW 240 kDa) and is one of the two surface glycoproteins of influenza virus (the other being hemagglutinin).[41,42] It cleaves terminal sialic acid residues from glycolipids or glycoproteins and is essential for spreading progeny virus particles during infection. Inhibition of NA stops virus infection.[43,44] NA is a design target for anti-influenza agents,[44−58] and specific drugs are already on the market.[59−62] Importantly, the apo-NA and NA complexed with several different ligands have been structurally characterized by X-ray crystallography.[44,51,53,57,63−66] The NA active site is a cavity, invariant for all known strains of influenza virus, formed by 11 conserved residues and defined by a relatively high number of ionizable residues, namely, Asp151, Glu119, Glu227, Glu276, Arg118, Arg292, Arg371. In addition, many of the reported inhibitors have pH-dependent functional groups such as amine/ammonium and carboxylic acid/carboxylate. Thus, this system is quite complex and not immediately amenable to simple modeling approaches.

Results

We applied the HINT model[17] in evaluating the interactions between neuraminidase and nine inhibitors (Scheme 1). The inhibitors are DANA (2,3-didehydro-2-deoxy-*N*-acetylneuraminic acid) (**1**),[57] five DANA derivatives, 4-amino-DANA (**2**),[57] 9-amino-DANA (**3**),[57] 4,9-diamino-DANA (**4**),[57] 4-guanidino-DANA (also known as Zanamivir) (**5**),[53] dihydropyran-phenetyl-propyl-carboxamide (**6**),[53] and three smaller ligands that are benzoic acid derivatives, BANA106 (4-(acetylamino)-3-hydroxy-5-aminobenzoic acid) (**7**),[44] BANA108 (4-(acetylamino)-3-aminobenzoic acid) (**8**),[44] and BANA105 (4-(acetylamino)-3-hydroxy-5-nitrobenzoic acid) (**9**).[44] Crystallographic and in vitro inhibition data for the nine neuraminidase−ligand complexes in this study are highlighted in Table 1.

DANA (**1**) is a transition-state analogue inhibitor of neuramindase (NA). The five ligands related to DANA derive from the substitution of hydroxyl groups at the positions 4 and/or 9 with aminic groups (**2**−**4**), a guanidinum group (**5**), an aminic group, and a phenetyl-propyl-carboxamide (**6**). Two of the benzoic acid-derived ligands have an aminic group at position 6 (**7** and **8**), while the third has a nitro group. Thus, these ligands present a number of functional groups that can assume different ionization states. The protein active site also presents several polar ionizable residue types, i.e., aspartates, glutamates, arginines, etc. As a representative case, the complex between NA and DANA is shown in Figure 1. As protons are normally undetectable by X-ray crystallography, there is experimental uncertainty in the "correct" assignment of the protonation state of the ionizable protein residues and ligand groups. Furthermore, the protonation state under which the crystals were grown does not necessarily correspond with the protonation state under which the measurements in solution were carried out, and both measurements may not correspond to the "optimum" pH for binding.

Model Building and Computational Titration. Model building is more complex than just assigning the correct ionization states of residues/functional groups. The assignment of the spatial positions of these hydrogens is equally important, as the strength of hydrogen bonds is affected by the geometry of the interacting atoms. This is singularly difficult because protons are *not* static and may be shared between two, three, or more functional groups. Our working hypothesis and guiding principle in this work is that there is not **one** *definitive* model. Nevertheless, our challenge is to apply static molecular modeling tools to this environment.

Scheme 1. Chemical Structure of the Neuraminidase Ligands Analyzed.

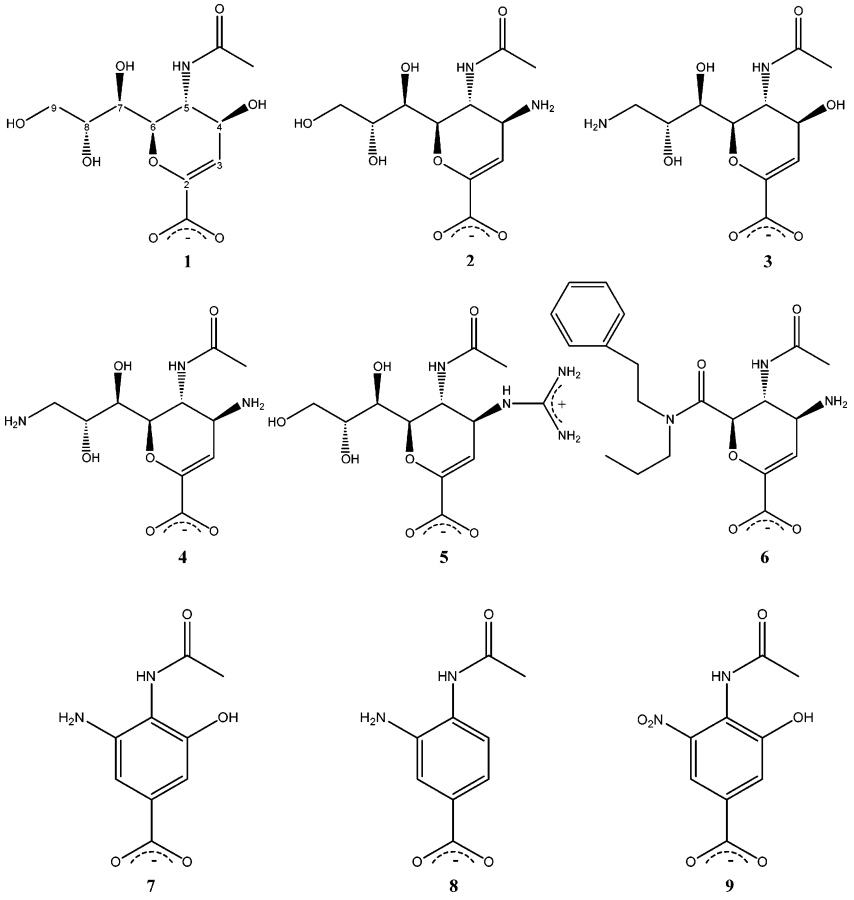


Table 1. Crystallographic and Inhibition Data for the Neuraminidase−Ligand Complexes Analyzed

PDB (ligand)	NA	crystallographic resolution (Å)	pH_{cryst}	K_i (μM)	pH_{Ki}	ΔG (kcal mol^{-1})
1f8b (**1**)	virus A-N9	1.8[57]	5.9	4[74]	5.5	−7.4
1f8c (**2**)	virus A-N9	1.7[57]	5.9	0.04[74]	5.5	−10.1
1f8d (**3**)	virus A-N9	1.4[57]	5.9	400[57]	5.5	−4.6
1f8e (**4**)	virus A-N9	1.4[57]	5.9	15[57]	5.5	−6.6
1a4g (**5**)	virus B	2.2[53]	7.5−8.0	0.004[a 53,45]	5.5	−11.8
1a4q (**6**)	virus B	1.9[53]	7.5−8.0	3.6[a 53]	na	−7.8
1ivc (**7**)	virus A-N2	2.4[44]	6.3	>10[4a 44,75]	6.0	−2.7[70]
1ive (**8**)	virus A-N2	2.4[44]	6.3	>10[4a 44,75]	6.0	−2.7[70]
1ivd (**9**)	virus A-N2	1.8[44]	6.3	750[a 44]	6.0	−4.6

[a] IC_{50}; na = not available.

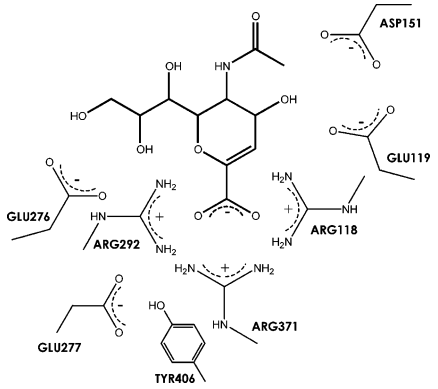
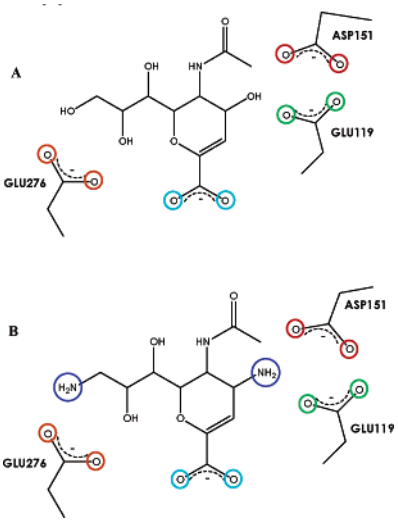


Figure 1. Schematic representation of the active site of neuraminidase in the presence of DANA (**1**).

With these issues in mind, we developed a new method, called "computational titration", to assist in modeling ionization states in the biological environment and, thus, more correctly evaluating ligand−protein interactions. Our working procedure includes two phases: model building and hydropathic analysis. The starting point is the PDB file for the complex. As hydrogens are not present in PDB files, they were added and energy minimized, i.e., keeping the heavy atom framework intact.[19] Then, the "essential" hydrogen atoms present in the active sites, that is the hydrogen atoms bound to polar heavy atoms (N, O, P, S), were carefully examined as these can be trapped in local minima. Manual adjustment of the R−XH torsions were performed if necessary to maximize the hydrogen-bond interactions. The other important point of the model building phase is examination of the ionization states of the residues and ligands as a function of added protons on ionizable residue of ligand and protein active sites participating in binding, i.e., the "computational titration". The computational titration procedure introduces one proton at a time into the molecular model for a complex. The

Scheme 2. NA-DANA (**1**) (A) and NA-4,9-diamino DANA (**4**) (B) Complexes[a]



[a] The protonation sites are highlighted.

analogy is with an experimental titration where a solution is acidified. As there are many possible sites of protonation, as well as several protons to be potentially added, *all* possibilities that we deemed to be viable were modeled in our procedure. We are thus considering the relationship between distinct protonation patterns, "pH" values and energy of binding. As a representative case, consider the interactions involved in the binding between the protein active site residues and the ligand DANA as shown in Figure 1. The majority of these interactions are conserved in all complexes and include three arginines (Arg118, Arg292, Arg371) that bind to the carboxylate of the ligand, the acetylamino methyl group that interacts with the hydrophobic pocket, and the acetylamino oxygen and nitrogen atoms that interact with Arg152 (not highlighted), the hydroxyl groups of the glycerol side chain of the ligand (O8 and O9) that are hydrogen bonded to Glu276 (these interaction are not present in complexes with **6**−**9** that do not have the glycerol side chain), and the hydroxyl group at the 4 position of the ligand that is at the entrance of a pocket formed by Asp151, Glu119, and Glu227. Of the last three residues, Glu227 is not as close to the ligand as Asp151 and Glu119. Another residue conserved in the active site of all neuraminidase−inhibitor complexes is Tyr406. This tyrosine is very close to both the oxygen atoms of the ligand carboxylate and to those of Glu277. Careful examination of the protein−ligand structures suggests that Tyr406 has a preferential interaction with Glu277, as the Glu277 OE2 is the closest hydrogen-bond acceptor atom to the OH of the tyrosine. The relevance of this observation will be part of the Discussion section.

For the NA DANA-based ligand complexes, the ionizable neuraminidase residues Asp151, Glu119, and Glu276 were initially targeted for titration because they would appear to have the most significant contributions to the ligand binding by forming hydrogen bonds. For the NA−DANA, five titration levels were explored (see Scheme 2A), corresponding to the addition of zero, one, two, three, or four protons to the molecular model. The "0H" case represents the situation where Asp151, Glu119, and Glu276 are unprotonated. The "1H" case represents the situation in which a single proton is added to one of these three residues (Asp151, Glu119, or Glu276). However, the carboxylate moiety of Glu276 makes very good hydrogen bonds with the two hydroxyl groups of the ligands, so that protonation of this residue would appear to be unfavorable for binding, and we did not protonate Glu276 in the initial stages of the titration. There are then four different possibilities for the "1H" case by protonation of Asp151 and Glu119, depending on which of the four carboxylate oxygen atoms the added proton is attached and all were modeled. The "2H" case is the situation in which two protons are added into the model, i.e., both Asp151 and Glu119 are simultaneously protonated. Again, there are four combinations, and all were modeled. In the "3H" case, both the Asp151 and Glu119 are protonated and a proton must now be added to either OE1 or OE2 of Glu276, yielding eight possible combinations. Finally, in the "4H" case, the last hydrogen added to the model protonates the carboxylate group of the ligand. For steric reasons, the additional proton could only be added to one of the two oxygens of this group, again yielding eight model combinations. As the negatively charged carboxylate group of the ligand strongly interacts with the trio of positively charged arginines (Arg118, Arg292, and Arg371), its protonation is unfavorable for binding and is thus the last stage of the "computational titration". It should also be noted that further complexity would be added by examining the protonation state of the arginines, i.e., a large number of additional models would be accessible by "titrating" the three protons of the arginines (and sharing them with the ligand carboxylate). However, arginine is rarely neutral because of its high pK_a, and accordingly, it was modeled in the protonated form in this study. The models thus derived for each ligand are summarized in Table 2 and detailed explicitly in Table 3 (Supporting Information).

For ligands that themselves possess ionizable groups interacting with polar protein residues, e.g., 4,9-diamino DANA (**4**) that has two aminic groups at positions 4 and 9, the computational titration becomes more complex, requiring the preparation of an even larger number of models. Seven titration levels were explored for this complex, corresponding to the addition of 0−6 hydrogens into the model (Scheme 2B highlights the protonation sites for the "computational titration"). While the "0H" case is the situation in which both the ionizable groups of the ligand and the protein are deprotonated, requiring only one model, the "1H" case, where only one proton is present either on the ligand or on the protein, leads to six molecular models. In the "2H" case the two additional protons can be both added to the protein or to the ligand or one proton to one of the two ionizable groups of the ligand and the other one to the protein (Asp 151 or Glu119), yielding 13 molecular models. In the "3H" case one hydrogen is always added to the amine group at C9 position of the ligand, and either the amine group at C4 position can be protonated and one of the protein residues protonated, or both the two protein residues can be neutral, yielding eight possible models. In the "4H" case both the amines are ionized and Asp151 and Glu119 are both neutral and four models were built. In the "5H" case there is an additional hydrogen on one of the two oxygens of Glu276, yielding again four possible models. Finally, in the "6H"

Table 2. HINT Scores and Free Energies of Binding for Neuramindase-Binding Ligands in Computational Titration

complex (Lig)	$\Delta G_{binding}$ (kcal mol^{-1})	added protons	model count	model notes	mean HINT score
1f8b (**1**)	−7.4	0	1	a	4461
		1	4		4699 ± 170
		2	4	b−d	4877 ± 51
		3	8		4285 ± 46
		4	8		2610 ± 36
1f8c (**2**)	−10.1	0	1		5302
		1	5	a	5762 ± 509
		2	8		6126 ± 358
		3	4	b−d	6383 ± 82
		4	8		5687 ± 68
		5	8		3799 ± 73
1f8d (**3**)	−4.6	0	1		4302
		1	5	a	4879 ± 286
		2	8		5415 ± 282
		3	4	b, c	6184 ± 39
		4	4	d	5083 ± 52
		5	4		3477 ± 26
1f8e (**4**)	−6.6	0	1		5029
		1	6		5696 ± 458
		2	13	a	6392 ± 342
		3	8		7654 ± 350
		4	4	b, c	7864 ± 85
		5	4	d	6412 ± 51
		6	4		4603 ± 67
1a4g (**5**)	−11.8	0	2	c	5439 ± 15
		1	9	a	5804 ± 126
		2	12		6110 ± 89
		3	4	b, d	6200 ± 54
		4	8		5415 ± 40
		5	8		3455 ± 29
1a4q(**6**)	−7.8	0	1	c	5304
		1	5	a	5771 ± 518
		2	8		6212 ± 359
		3	4	b, d	6437 ± 95
		4	4		4374 ± 116
1ivc (**7**)	−2.7	0	1		3053
		1	5	a	3432 ± 155
		2	8	b−d	3809 ± 140
		3	4		4240 ± 161
		4	4		4044 ± 161
		5	8		2844 ± 71
1ive (**8**)	−2.7	0	1		2380
		1	5	a	2737 ± 143
		2	8	b−d	3048 ± 77
		3	4		3317 ± 65
		4	4		3164 ± 63
		5	8		2256 ± 33
1ivd (**9**)	−4.6	0	1	a	3728
		1	4	b−d	3679 ± 146
		2	4		3675 ± 44

[a] This protonation level is assumed to be neutral, i.e., at pH 7.0. [b] This protonation level is assumed to be the condition where the K_i measurement was made. [c] This protonation level is assumed to be the condition where the X-ray crystals were grown. [d] This protonation level is proposed as best representing the local pH environment of the complex at the binding site.

case the ligand carboxylate is also protonated and, again, four molecular models were created (see Tables 2 and 3).

In the complex between NA* [*The amino acid numbering used throughout this report corresponds to that of neuraminidase from virus A. It needs to be pointed out, however, that in the PDB files for complexes 1a4q (**6**) and 1a4g (**5**) the residue numeration corresponds to that of neuraminidase from virus B.] and dihydropyran-phenetyl-propyl-carboxamide (**6**), the protonation sites for the titration were the amino and carboxylate groups of the ligand and Asp151 and Glu119 at the protein active site. The tertiary amide of this ligand was not protonated. Glu276 appears to interact weakly with the hydrophobic side chain of the ligand but strongly interacts with the Arg224 protein residue. Protonation of this glutamate does not impact the HINT score and was not considered in our models.

The three benzoic acid derivative ligands (**7**−**9**) are considerably less bulky than **1**−**6**, and thus, the binding pocket has a number of water molecules that may be influencing the pH effect. We will explore the implications of this observation in the Discussion. The possible titratable groups of these complexes are the amino and carboxylate groups of the ligands and Asp151, Glu119, and Glu227. While the anilino NH_2 of **7** and **8** is a relatively weak base compared to the amines of **2**−**4**, its pK_a is not dissimilar from that of aspartate and glutamate. BANA105 (**9**) is a weakly binding benzoic acid derivative with an NO_2 group. This particular ligand has only the carboxylate as an ionizable/titratable group. In general, molecular mechanics poorly represents the actual electronic structure of the NO_2 group. HINT parameters, as they are indexed by the Tripos force field atom typing, are similarly inadequate for representing this group.[67] Thus, the interactions between the NO_2 and the surrounding residues are probably underestimated by our modeling and analysis and should at best be considered a lower limit. We report the associated scores for the protonation models of **9** but do not otherwise consider **9** in other analyses.

Hydropathic Analysis of Models. For each specific protonation model of individual NA−inhibitor complexes (Table 2), the HINT score was calculated as described previously.[17,19,21,22] At each "pH", i.e., the number of protons dropped into the molecular models, a mean HINT score value was obtained by averaging the HINT scores obtained for each of the individual models. These mean values for the eight complexes are listed in Table 2 and plotted with standard errors, where applicable, in Figures 2a−h for the NA-**1** to NA-**8** complexes, respectively, as a function of added protons. The error bars indicate the variability of HINT score for each protonation model, corresponding in large measure to the uncertainty of placing a single proton or set of protons into the molecular models. All these graphs show a very similar behavior. They all show a smooth and fairly consistent increase in HINT score as a function of "pH", reaching what may be regarded as a peak HINT score, i.e., a maximal binding free energy. This peak is at the level of two added protons for **1**, at the level of three added protons for **2**, **3**, **6**, **7,** and **8,** while the maximum HINT score corresponds to four protons added into the model for ligand **4**. The peak is not so apparent for **5**, for which the maximum HINT score value is almost the same for the addition of two or three hydrogens. Finally **6**, **7,** and **8** all show the maximum HINT score at the level of three protons. There is generally a consistent rise up to this peak point and a somewhat sharp drop-off for the next added protons. Taking into account the error bars on some of these figures suggests that alternative interpretations of optimum protonation level are possible. This behavior is reminiscent of an experimental pH titration curve and suggests that the optimum binding would occur at the pH coinciding with the number of added protons at the

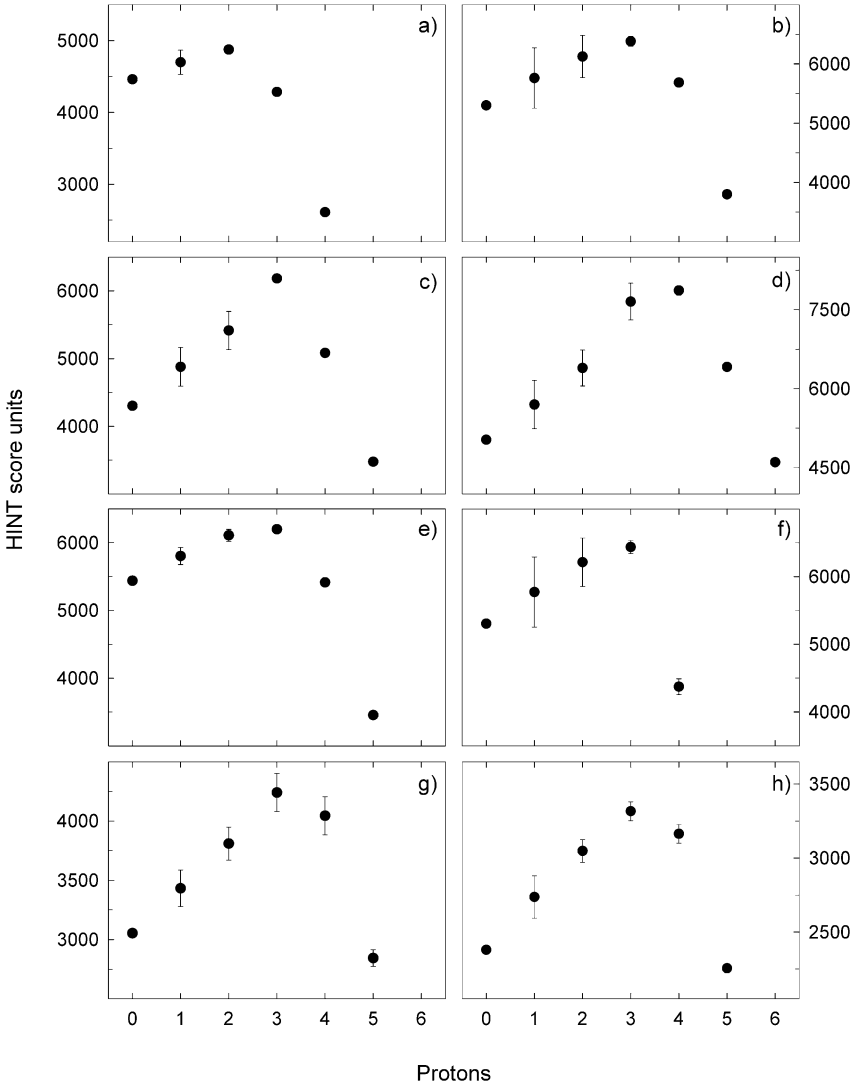


Figure 2. Computational titration of NA−ligand complexes. Plots a−h correspond to ligands **1**−**8**, respectively. Data points are mean HINT score values for different models at a defined protonation level. The error bar represents the standard deviation of the mean.

maximum value of HINT score. The relation between actual solution pH, local "site" pH, and protonation level will be addressed in the Discussion.

In previous publications[19,22,68] we demonstrated a linear relation between HINT score and free energy of binding. In the present work, the mean HINT scores at constant "pH" values for the complexes were correlated with experimentally determined free energies of binding. These data are shown for the "0H" case (Figure 3). The linear fit of the data is relatively good, exhibiting $r = 0.88$ and standard error = 1.7 kcal mol^{-1}, following the equation

$$\Delta G = -0.0025\mathrm{H}_{\mathrm{TOTAL}} + 4.524 \qquad (1)$$

where $\mathrm{H}_{\mathrm{TOTAL}}$ is the total HINT score. However, a more physically and chemically appropriate correlation would include the HINT score corresponding to the pH at which the solution binding measurements were made (see Discussion, Figures 4 and 7).

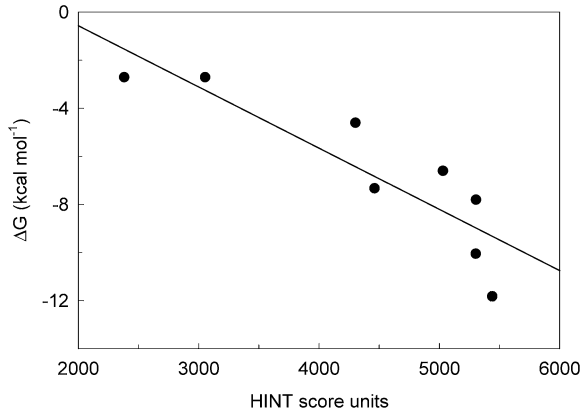


Figure 3. Dependence of the experimental ΔG on HINT score units for NA−inhibitor complexes for the "0H" case, i.e., when accessible ligand and protein residues are deprotonated. The line through data points represents the best squares fit.

Discussion

The influenza virus is the cause of a major respiratory tract disease affecting millions of people each year. Because of the rapid rate of mutation of influenza viral antigens, vaccines are frequently ineffective and must be reformulated each year. Neuraminidase (NA), one of the two major surface antigens of influenza virus, offers an attractive site for therapeutic intervention in

influenza infections. Accordingly, a large number of studies have been reported on this enzyme,[44−58] describing application of drug design approaches to the development of anti-influenza agents.

HINT, the computational tool used to calculate the strength of interactions between NA and inhibitors in this work, is a model based on experimental log *P* values that calculates a "score" for each atom−atom interaction in a biomolecular complex. This score has both a character defining the particular type of interaction, i.e., hydrogen-bond, acid−base, hydrophobic, base−base, acid−acid, hydrophobic−polar, and a magnitude representing the strength of the interactions. In effect, each atom−atom interaction is a partial δg, a fraction of the total $\Delta G_{\text{interaction}}$, so that the total HINT score between the two molecules can be directly correlated to the free energy of binding.[17,18]

In this work we investigate the problem of modeling ionization states of protein residues and/or functional groups on the ligand in the evaluation of protein−ligand interactions using the HINT model. Protonation or deprotonation of titratable groups can cause changes in binding affinities, enzymatic activities, and structural properties of proteins and often represents a key event in enzymatic reactions. The problem of the determination of ionization state under physiological conditions is of great value to computational chemists, and several approaches have been proposed for pK_a prediction, determination of protonation and redox equilibria in proteins, prediction of ionization states of titratable residues in proteins, and ionization/tautomerization of small molecules at binding sites.[24−34,35−38]

Despite the relatively large number of previous computational experiments on the neuraminidase system,[45,47,49−53,56−58,69−71] in most of the studies virtually no attention has been paid to the *actual* ionization state of active site residues and/or ligand functional groups. This is especially surprising because very few of the assays and crystallizations from which the modeling was derived were performed at pH 7. Thus, as we built and optimized our models for NA−ligand complexes, we attempted to account for the various protonation states of the residues and ligands. As we analyzed these models, we developed a protocol for these procedures that we termed "computational titration". Our primary goal in the present study was to facilitate and improve our predictions of free energy for ligand binding in cases where protonation/deprotonation of ligand functional groups and/or protein residues may be significant and, by extension, improve general scoring methods for virtual screening. However, the results from the computational titration itself are interesting, and our ability to examine the effects of pH in silico merit further discussion in the following subsections. We demonstrate, with the arguments to follow, that we have built models for the NA−ligand complexes in our study that are simple to understand but robust in their performance.

Which Protonation Level Is "Correct" for the Models? By "dropping" one proton at a time into the models, we gradually acidified, in silico, the environment of the protein−ligand complexes at the binding site. The "titration" graphs (Figure 2) all show a peak in HINT score at a particular "pH level" (i.e., number of protons added into the model). While the observed peaks are characteristic of the particular bound ligand, questions regarding these peaks are relevant: (i) Is the peak "pH level" the *optimum* "pH level"? (ii) Does optimum pH actually correspond to the pH of K_i measurements? (iii) Does optimum pH correspond to the crystallization conditions and/or the associated X-ray crystallographic molecular model?

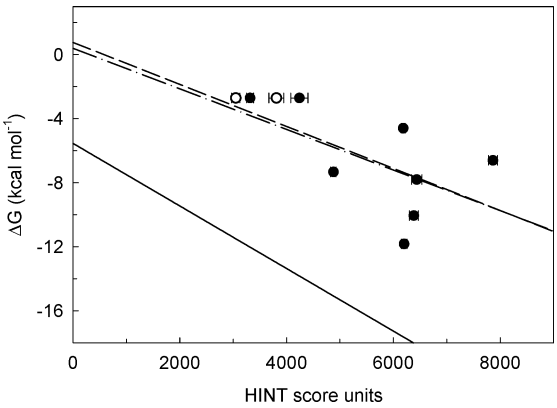


Figure 4. Dependence of the experimental ΔG on HINT score units for NA−inhibitor complexes, determined either at the maximum of the computational titration curves (Table 2) (closed circles for **7** and **8**, dash regression line) or at the protonation level assumed to be the pH of experimental K_i measurement (Table 2) (open circles for **7** and **8**, dash−dot regression line). The solid line represents the regression for ΔG vs HINT score for 53 protein−ligand complexes.[19]

Since the peaks represent the maximum HINT score value for each protein-ligand complex, they should represent, with a fair level of confidence, the protonation level for optimum binding. By analogy, the actual pH level corresponding to the proton count peak should be the *optimum* pH. The second question is more difficult to definitively answer because experimental binding data in response to pH for these complexes are not available. However, some clues can be obtained when the previously reported fit line for the HINT score vs ΔG for 53 protein−ligand complexes[19] is superimposed with the peak HINT scores and free energies of binding for these neuraminidase−inhibitor complexes (Figure 4). The regression equation for these NA−complex data at peak HINT score is

$$\Delta G = -0.0013\text{H}_{\text{TOTAL}} + 0.760 \qquad (2)$$

where $r = 0.58$ and the standard error of estimate is ± 2.9 kcal mol^{-1}.

The HINT scores for these NA complexes would seem to be higher than expected from their measured K_is. This suggests that the optimum pH level may not correspond to the pH at which the K_is were measured, and a more complex relation between protonation model and free energy will be required, as shown below. Finally, structural data of these complexes at different pH values are also not available. X-ray crystallographic experiments can only very rarely locate any hydrogen atoms in solved structures, and even knowing the pH at which the crystals were grown is not sufficient to determine the *actual* (local) protonation state of the residues in the crystal structure. However, in a few instances, structural information can imply protonation state. Careful analyses of measured interatomic dis-

tances from high-resolution crystallographic measurements coupled with molecular mechanics models of the likely structures can also imply protonation states in some instances.[57]

Correlation of Protonation Level with Actual pH. Our initial interest in the effect of residue ionization state on free energy predictions was piqued by a series of data reported by Bartlett and colleagues[23] that (a) showed K_i as a function of pH for a complex between penicillopepsin and a phosphonate-containing inhibitor and (b) proposed active site models, including protonation of two aspartates, to explain the pH vs K_i observation. Previously,[19] we analyzed this complex by building molecular models corresponding to the three different protonations and found a correlation between the actual pH and the HINT score of the models. Interestingly, and perhaps fortuitously with these models, 1 pH unit corresponded to 1 added proton, i.e., $\Delta pH/\Delta proton = -1$. In a sense, each added proton may be thought of as adding one hydrogen bond between the ligand and protein. For the neuraminidase−inhibitor complexes in this study, binding data as a function of pH are not available. However, we can make a few assumptions that should assist us in correlating which model protonation level corresponds to the measured K_is for the complexes.

First, we will assume that (neutral) pH 7 corresponds to the cases where all carboxylic acids are deprotonated, while the primary amines and guanidiniums are protonated, as should be expected at pH 7. For **1** and **9** this would be zero added protons; for **2**, **3**, **5**, **6**, **7,** and **8** this would be one added proton; and for **4** this would be two added protons. Second, referring to the plots of Figure 2, note that the slopes of the initial (rising) portion of the plots, where each added proton is presumed to add a hydrogen bond, can be averaged to a value of ca. 400 ± 200 HINT score units (0.8 ± 0.4 kcal mol^{-1})[19,20,22] per added proton, which is consistent with the usual strength of a hydrogen bond. Third, despite the lack of correlative data between the systems, we will use the relationship $\Delta pH/\Delta proton = -1$ observed in the penicillopepsin system because the pH range of interest (5.5−6.0, where the inhibition assays were performed) is close to our benchmark of pH 7.0.

With these assumptions in place, we propose that one additional proton relative to the pH 7 cases (above) corresponds to the complexes of measured inhibition for **7**, **8,** and **9** (where inhibition was measured at pH 6.0), and two additional protons relative to the pH 7 cases (above) corresponds to the complexes of measured inhibition for **1**−**6** (where inhibition was measured at pH 5.5). Similarly, we can propose that the crystallographic measurements of the complexes with **1**−**4** and **7**, **8,** and **9** were made at this same protonation level, while crystallographic measurements of complexes with **5** and **6** (pH 7.5−8.0) were likely made on the neutral complexes or conceivably singly deprotonated ("0H" case) complexes. This is particularly relevant for **6** (below). Figure 4 plots the scores of these calibrated models with the free energy of binding. The least-squares regression of these data shows

$$\Delta G = -0.0013 H_{TOTAL} + 0.396 \qquad (3)$$

where $r = 0.61$ and the standard error of estimate is ±2.8 kcal mol^{-1}. However, there is evidence of a further pH effect in the complexes of **3** and **4** that may be related to "local" protonation variations around the N-9 amine substitutions of these compounds. A later section of the Discussion addresses this issue.

The model described above is a pragmatic response to a paucity of data. Clearly, having experimental binding *and* structural data measured at different pH values would make possible a more thorough understanding of the relationship between the optimum pH level from computational titration and the actual pH values measured for the K_is and the crystal growth. At best, this would be a difficult and costly set of experiments to perform, and at worst, due to potential crystallization difficulties and solubility issues, this would be nearly impossible.

Buffering Effect of Water. The benzoic acid derivative ligands (**7**−**8**, Scheme 1) are not strong inhibitors of influenza virus neuraminidase. Even though these aromatic inhibitors occupy the same site as the DANA analogues (**1**−**6**) in the active site of NA, there are differences in the orientation of the ring and the side-chain groups. They are smaller than **1**−**6**, and the concomitant void space in the binding site is filled with water molecules that likely influence the effect of the in silico acidification in a number of ways. The most interesting observation is the "softer" roll-off in HINT score after the peak in Figure 2g and 2h for **7** and **8**, respectively. In effect, this additional (postoptimum) protonation seems to be a less significant perturbation on these complexes. This behavior can be rationalized with two factors. First, these inhibitors are smaller and thus interact less strongly with the protein. A second, more intriguing, factor is that the water molecules in the binding site appear to be playing a buffering role with respect to nearby protonations. Water molecules can reorient to present either hydrogen-bond donors or hydrogen-bond acceptors to their neighbors while maintaining a relatively consistent overall binding energy. Figure 5 illustrates three protonation models for complexes of **7**, where the water molecules in the binding site adapt somewhat compensating orientations. The present study does not incorporate the energetic effect of binding site water molecules, but it is an important contribution to free energy that will be examined in a future report.

Other Considerations for Model Building and Free Energy Predictions. We address three points here: (1) consideration of intramolecular as well as intermolecular interactions in model building and optimization, (2) the relationship between crystallographic models obtained under particular conditions and coordinate sets optimized for other conditions, and (3) the meaning of the equations relating HINT score and free energy.

Tyr406 is a conserved residue of the neuraminidase active site that is close both to the ligand carboxylate group oxygens and to the carboxylate oxygens of Glu277. Thus, Tyr406 could potentially interact either with the ligand or with the protein by forming a hydrogen bond with the tyrosine OH atom depending on the rotation of the CZ−OH bond torsion angle. To evaluate this type of situation, both possibilities must be modeled and scored. In this case, analysis of the complex structure

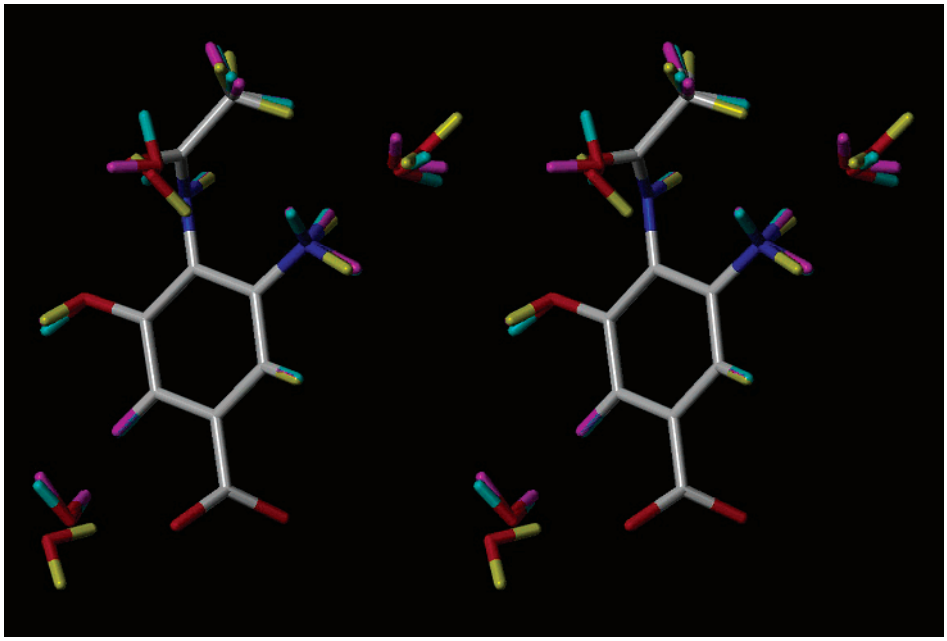


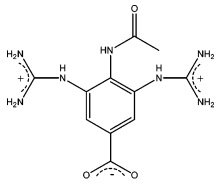
Figure 5. Stereoview of protonation models for the NA-**7** complex in the presence of "buffering" water molecules at the binding site. In the model with yellow hydrogens all ionizable groups are deprotonated; in the model with magenta hydrogens the 9-N amine on the ligand is protonated to NH_3^+ and acid protein residues Glu119, Asp151, and Glu227 are charged; in the model with cyan hydrogens the 9-N amine on the ligand is protonated to NH_3^+ and acid protein residues Glu119, Asp151, and Glu227 are protonated. Water molecules in the model were placed by using GRID (version 21, www.moldiscovery.com)[76] and optimized with HINT.

shows that among the four oxygens, Glu277 OE2 is the closest to Tyr406 OH and makes the strongest hydrogen bond. In all of the models in this present work, Tyr406 was modeled with a (protein) intramolecular interaction rather than being directly involved in ligand binding. In general, our methodology of free energy scoring dictates that residues of the protein and functional groups of the ligand be tested for intramolecular as well as intermolecular interactions.

While our model building procedure restricts the optimization of heavy (non-hydrogen) atoms[19] and their positions are generally taken as described by the crystallographic data, there are situations that may be encountered in computational titration where heavy atom movement *is* required. In other words, if adding a specific proton to a heavy atom induces a change in hybridization, etc., of that atom, how can molecular models for these situations be optimized while retaining the essential experimental features of the crystal structure? Also consider the correspondence, or lack thereof, between crystallographic models from crystals grown at a particular pH and the solution binding data obtained at a different pH. This was discussed above but is a potential source of (possibly large) errors in free energy predictions.

It is interesting that in all of our studies we calculated linear equations, e.g., eqs 1 and 2 in this work, relating HINT score to free energy in which the slopes (ΔHINT score/Δ*G*) have been remarkably consistent. However, the *y*-intercepts of these equations have more variability (e.g., Figure 4). We interpret this as follows. The HINT methods we have described in here, and previously,[19] are quite robust with respect to ΔΔ*G*, which is the change in free energy within a series of ligands bound at a single protein. With an invariant algorithmic approach, the method appears to be quite reliably accounting for both enthalpic and entropic contributions to ΔΔ*G* regardless of the biomolecular system. On the other hand, there must be some systematic factors, related to each biomolecular system, that manifest themselves in the *y*-intercept portion of the predictive equations. The nature and origin of these factors is under investigation. Thus, this methodology can, in general and in the absence of calibration data, predict binding free energy with an accuracy of around ±2.5 kcal mol^{-1} for any system, but when characterized complexes can be used as a learning set for a single system, the HINT technology can be used with greater accuracy, with prediction errors as low as ±1 kcal mol^{-1}. These errors are, for the most part, consistent with those observed in other systems with other scoring technologies. Even complex calculations such as those using free energy perturbation[2] or linear interaction energy[4] methods require considerable calibration of the parameters with respect to the studied systems before achieving accurate free energy estimates.

Why Are 3 and 4 such Poor Inhibitors? Compounds **3** and **4**,[57] which have the 9-OH of DANA (**1**) replaced by 9-NH_2, exhibit surprisingly low affinities for neuraminidase. These two compounds were designed,[57] at least in part, to explore the similarly (and unexpected) low affinity exhibited by 3,5-diguanidino-4-(*N*-acetylamino)benzoic acid (**10**).[51] Indeed, the 9-N



10

amine or ammonium of **3** and **4**, as the case may be, is

in the same position as the 9-OH and makes satisfactory hydrogen bonds with the site protein residues.[57] There is adequate evidence for this sort of isosteric substitution enhancing $\Delta G_{binding}$ in other systems by significant amounts. For example, one of the two rather small differences between methotrexate and folate is just such a change, and methotrexate has a $\Delta G_{binding}$ that is 4.4 kcal mol^{-1} *more* stable than folate with respect to dihydrofolate reductase binding.[72] The HINT analysis in this report also suggests that the N-9 substitution should, in general, improve binding. In addition, simple molecular mechanics (Tripos force field) binding energies ($\Delta G_{binding} = E_{complex} - E_{protein} - E_{ligand}$) follow the *expected* trend that ammonium ions in positions 4 and/or 9 of DANA have more favorable binding energies than either amines or hydroxyls in these positions.

While most other recent modeling studies on neuraminidase inhibitors have not included compounds **3** and **4**,[56,58,70,71] Smith et al.[57] analyzed the neuraminidase/ligand protonation problem for the inhibitors DANA (**1**), 4-amino- (**2**), 9-amino- (**3**), and 4,9-diamino-DANA (**4**) using a combination of molecular mechanics and Poisson−Boltzmann analyses to ascertain protonation at the active site. In addition, electrostatic arguments were proposed to explain the unexpectedly poor binding of compounds **3** and **4**.[57] Molecular mechanics calculations minimized the structures of DANA, 4-amino-, 9-amino-, and 4,9-diamino-DANA bound in the NA active site in each of the possible ionization states for the amines and with protonation/deprotonation of Glu119. Their analysis of these models, in comparison to the crystallographic data, suggested that Glu119 is neutral, i.e., in the acid form, and that in all cases the inhibitor amines are protonated. The effects of protonation at other active site residues, e.g., Asp151 and Glu276, were not explicitly reported. Furthermore, calculated binding and solvation free energies for each of the possible ionization states of the ligand amino groups were obtained by solving the Poisson−Boltzmann equation. It was shown that the solvent-screened interaction energy increases when the amine groups are charged, completely in contrast to the electrostatic interaction energy behavior. In summarizing their analysis, the energy cost of partially desolvating the protein and charged ligands is greater than the (considerable) gains in interaction energy for these ligand−protein complexes.[57] As these authors state, however, the omission of the contributions of nonelectrostatic energies may lead to nontrivial errors.

There is no a priori reason to discount the importance of desolvation energy as a contribution to free energy,[57] but certainly the charge of the 2,4-diaminopteridine ring of methotrexate is far more deeply buried in dihydrofolate reductase than is the 9-N charge of compounds **3** and **4** in neuraminidase. The binding site of neuraminidase for these ligands is quite open; in fact, perhaps a third of the ligand remains solvent exposed, which would argue for a lesser influence from desolvation energy rather than a larger one. We have puzzled over these results and would now like to offer two alternative explanations for why compounds **3** and **4** have such poor binding despite crystallographic evidence for very good interactions.

First, the hydrogen-bonding partner of the O8 hydroxyl may be uncertain at the temperature and pH of the binding assay: (A) There may be an intramolecular hydrogen bond between the hydroxyl at C8 and the oxygen of the dihydropyran ring of the ligand combined with ring fluxionality. Clearly, the X-ray results at 1.4 Å resolution do not indicate this, but the diffraction data was collected at low temperature, while the binding assay was performed at room temperature. Even without moving any heavy atoms, a simple rotation about the C8−O8 bond to maximize this interaction changes the overall binding score by +0.8 kcal mol^{-1}. (B) The hydroxyl O8 may be interacting with Glu277 instead of Glu276. This could change the overall binding score by +0.8−1.0 kcal mol^{-1}.

Second, because the binding site is so water exposed, there may be uncertainty in the actual protonation state of the site. In particular, it may be possible that Glu276 is also protonated to the neutral acid and that this state is stabilized by nearby water molecules. This may coincide with the release of an additional water molecule to bulk, thus impacting the entropy. This would be difficult to verify from the crystal structures because the resolutions for **1**−**4** vary (Table 1), and it has been reported that the number of water molecules located in a crystallographic analysis strongly depends on the quality of the structural determination.[73] As indicated above, the ability of water molecules to act as both hydrogen-bond donors and acceptors can stabilize many ionization state ensembles. Thus, we propose that the binding site protonation levels for compounds **3** and **4** are four and five protons, respectively (Table 2). Analyzing the resulting hydrogen-bonding network and understanding the water contribution to this effect is beyond the scope of the present work; however, Figure 6 indicates that a chemically reasonable model is possible at this protonation level (four) for **3**. This assumption, while simple to visualize and model, is actually another way of stating the importance of water in stabilizing structures or, in other words, solvation/desolvation. Figure 7 indicates this model for ΔG as a function of HINT score. The linear regression of these data combined with that of the other NA ligands follows the equation

$$\Delta G = -0.0020\mathrm{H}_{TOTAL} + 4.052 \qquad (4)$$

where $r = 0.81$ and the standard error of estimate is ±2.1 kcal mol^{-1}.

Multiplicity of Models in Computational Titration. To identify "viable" models in this work, we had to apply a certain degree of common sense and chemical intuition to, for example, design an order of protonation. For instance, we always protonated the carboxylate of the ligand in the last step, after other more energetically accessible sites had already been occupied. This is a pragmatic solution to the plethora of states and potential protonation models that exist. The actual number of states for the six proton 1f8e (compound **4**) system, when no preferences for models are preassigned, are 1, 10, 41, 88, 104, 54, and 16, respectively, for zero, one, two, three, four, five, and six added protons. The multiplicity beyond the expected 6!/ [n! (6 − n!)] arises from the fact that four of the six ionizable moieties are carboxylic acids where there are two protonation target

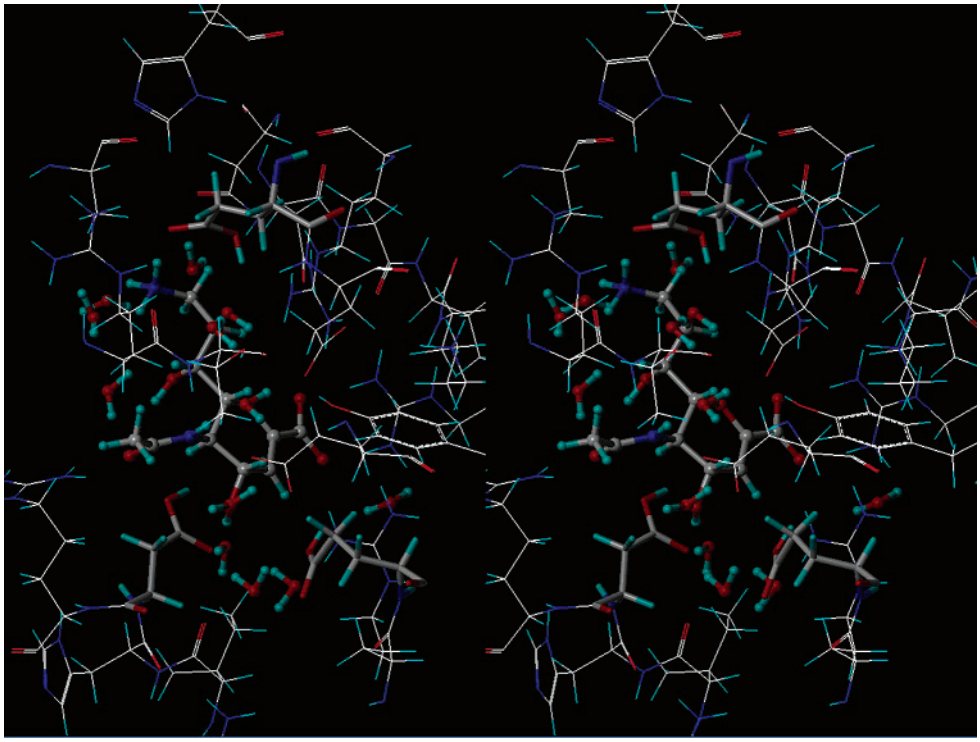


Figure 6. Stereoview of the ligand binding region for one of the four NA-**3** complex models where the protonation level is 4 (Table 2). Glu119, Asp151, and Glu276 as well as the 9-N nitrogen of the ligand are protonated. The water molecules (from crystallography[57]) have been optimized in this environment with HINT.

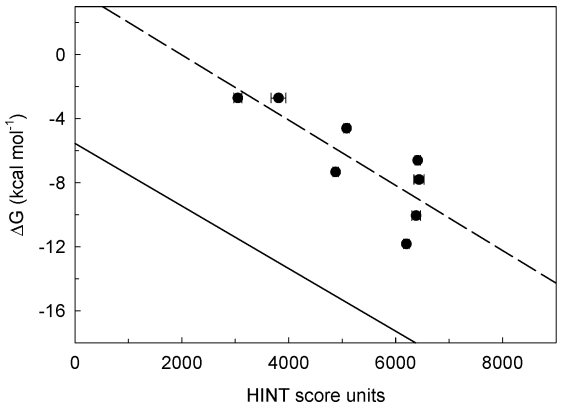


Figure 7. Dependence of experimental ΔG on HINT score units for NA−inhibitor complexes, determined at the protonation level proposed to best represent the local pH environment (Table 2), with the corresponding regression line (dash line). The solid line represents the regression for ΔG vs HINT score for 53 protein−ligand complexes.[19]

atoms (see Scheme 2B). This multiplicity obviously exceeds the limit for manual modeling of each case. A related point here is that while *all* of the protonation states could *potentially* be occupied, the ones with more favorable energetics are preferentially occupied. Thus, the intuition applied to designate the set of models built and analyzed for each protonation level also biases the energetics reported in our averages toward these more favorable states. A better approach, which we are currently exploring, would be to automate the build process so that all models (or a random sample thereof) are built and analyzed. The energy for each protonation level would then be calculated with application of a statistical mechanics partition function.

Summary

The accuracy of simple predictive methods for free energy of binding are compromised by a variety of chemical and physical effects due to solvation and desolvation. In this work we proposed that a simple protocol, computational titration, allows facile analysis of multiple ionization state models in protein−ligand complexes. We used the HINT model to score and evaluate the models, but other computational scoring algorithms may be equally appropriate. An important feature of the approach is that each ionization state may be represented by a number of molecular models, each corresponding to a particular placement of protons at the multiple "titratable" atoms of the protein−ligand system. This is a unique and important feature because this is a fairly realistic representation of the reality of "fluid" protons in solvated systems of this nature. Our operational procedure is to then average these states to acquire a single score for each protonation level. We are exploring alternative methods such as statistical mechanics (Boltzmann) weighting for these calculations in the future.

Our future efforts are also focusing on including the effect of water molecules in the active site in predictions of free energy. We previously showed with protein−protein systems that water bridging between interacting molecules may contribute significantly to the overall free energy. Overall, we are building a platform for rapid and comprehensive virtual screening incorporating computational titration and related methods.

Material and Methods

Model Building and Computational Titration. The three-dimensional coordinates of the nine neuraminidase−ligand complexes were obtained from the Protein Data Bank

(www.rcsb.org). These X-ray structures were imported into Sybyl (version 6.8, www.tripos.com), and the resulting atom and/or bond types were compared to the ligand chemical structures. Hydrogen atoms, not present in these PDB files, were added to the complexes using Sybyl tools. While the added hydrogens are internally correct for their heavy atom parents, the automated build algorithms do not consider inter- or intramolecular steric effects; thus, the hydrogens (only) in the complexes were energy-minimized with the Powell algorithm and a gradient of 0.1 kcal (mol Å)$^{-1}$ for 1000 cycles to remove potential bad contacts. Furthermore, because polar hydrogen atoms are often oriented away from their intermolecular hydrogen-bonding acceptor atoms, even after energy minimization due to local minima in the energy landscape, the *essential* hydrogens, i.e., the hydrogens bound to polar heavy atoms, were carefully examined and, when necessary, R−XH torsions were manually rotated in order to obtain an orientation that maximizes hydrogen-bond formation.

Additional protons, incorporated into the models with computational titration, were added—one at the time—at the possible protonation sites on the ligand and amino acid residues of the protein at the binding site, using Sybyl. The atom types, bond orders, etc., were adjusted as necessary, and again, manual optimization of hydrogen bonding was executed. To allow relaxation around the most favorable orientations, the hydrogens of the ligand and protein active site were again energy minimized using the Powell method with a gradient of 0.05 kcal (mol Å)$^{-1}$ for 40 cycles. For this step a "hot" radius of 4 Å and "interesting" radius of 8 Å were defined. In the "hot" region all protons were allowed to move; in the interesting region all atoms were held rigid but their properties could affect the "hot" atoms. It is important to note that these procedures do not significantly affect the heavy atom positions, i.e., the original crystallographic coordinates are essentially unaltered.

Hydropathic Analysis. The hydropathic analysis consists of partitioning the protein and the ligand molecular models, with the software HINT (version 2.35S, www.tripos.com). In effect, by calculating the log $P_{o/w}$ (partition coefficient for water/1-octanol) of the molecules, the empirical hydrophobic atom constants a_i (partial log P) and the partial solvent-accessible surface area S_i are assigned to every atom of both protein and ligand. The partition method for proteins is "dictionary", by which HINT calculates the log $P_{o/w}$ using a table of parameters based on residue type and solvent condition.[16] Nominally, the "neutral" solvent condition was chosen for the cases in which the ionization states of the residues were not changed (i.e., considering Lys, Arg, Glu, and Asp charged). For the cases in which the ionization state of a protein residue was modified, as during the computational titration, the "inferred" solvent option was chosen where the protonation state of each protein residue was automatically chosen on the basis of its atoms. The partition method used by HINT for the ligands is "calculate", an adaptation of the Hansch and Leo's CLOG-P method.[13,14] As noted before,[19−21] the "essential" method, in which only the polar hydrogen atoms are treated explicitly, provides the most consistent methodology and was also used in this work.

HINT calculates a score for each atom−atom interaction in a biomolecular association taking into account all possible noncovalent atom−atom interactions. The positive signed (favorable) contributions to the HINT score are acid−base, hydrogen-bond, and hydrophobic−hydrophobic interactions, and the negative signed (unfavorable) contributions are represented by acid−acid, base−base, and hydrophobic−polar interactions. In our protocol, the HINT score was calculated for each model created through the addition of protons to possible titration sites. HINT scores for all of the models built at each protonation level, i.e., corresponding to a particular number of protons into the models, were averaged, and a mean HINT score value was calculated. The related standard error is defined as the standard deviation of the mean, i.e., the standard deviation divided by the square root of the number of samples.

Acknowledgment. We gratefully acknowledge the support of the National Institutes of Health (D.J.A., Grant 5R01HL32793-15) and Virginia Commonwealth University for partial support of this research, the Italian Instruction, University and Research Ministry Grant PRIN01 (A.M.), and the National Institute for the Physics of Matter (A.M).

Supporting Information Available: Table of enumerated HINT scores and free energies of binding for neuraminidase ligands in computational titration (11 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Andrews, P. Functional groups, drug-receptor interactions and drug design. *TIPS* **1986**, *7* (4), 148−151.
(2) Kollman, P. A. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(3) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(4) Åqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−391.
(5) Åqvist, J.; Luzhkov, V. B.; Brandsal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358−365.
(6) Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(7) Rizzo, R. C.; Udier-Blagovic, M. M.; Wang, D. P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H., Jr.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of activity for nonnucleoside inhibitors with HIV-1 reverse transcriptase based on Monte Carlo simulations. *J. Med. Chem.* **2002**, *45*, 2970−2987.
(8) Ajay, Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.
(9) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein−ligand, protein−protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.
(10) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.
(11) Lazaridis, T. Binding affinity and specificity from computational studies. *Curr. Org. Chem.* **2002**, *6*, 1319−1332.
(12) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods. *Curr. Med. Chem.*, in press.
(13) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley and Sons Inc.: New York, 1979.
(14) Abraham, D. J.; Leo, A. J. Extension of the Fragment Method to Calculate Amino Acid Zwitterion and Side Chain Partition Coefficients. *Protein-Struct. Funct. Genet.* **1987**, *2*, 130−152.
(15) Wireko, F. C.; Kellogg, G. E.; Abraham, D. J. Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors. *J. Med. Chem.* **1991**, *34*, 758−767.
(16) Kellogg, G. E.; Joshi, G. S.; Abraham, D. J. New Tools for Modeling and Understanding Hydrophobicity and Hydrophobic Interactions. *Med. Chem. Res.* **1992**, *1*, 444−453.
(17) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: is Log$P_{o/w}$ more than the Sum of its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651−661.
(18) Kellogg, G. E.; Burnett, J. C.; Abraham, D. J. Very Empirical Treatment of Solvation and Entropy: a Force Field Derived from Log $P_{o/w}$. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 381−393.
(19) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, intuitive calculations of free energy of binding for protein−ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem.* **2002**, *45*, 2469−2483.

(20) Abraham, D. J.; Kellogg, G. E.; Holt, J. M.; Ackers, G. K. Hydropathic Analysis of the Non-Covalent Interactions between Molecular Subunits of Structurally Characterized Hemoglobins. *J. Mol. Biol.* **1997**, *272*, 613−632.
(21) Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational Methodology for Estimating Changes in Free Energies of Biomolecular Association upon Mutation. The Importance of Bound Water in Dimer-Tetramer Assembly for β37 Mutant Hemoglobins. *Biochemistry* **2000**, *39*, 1622−1633.
(22) Burnett, J. C.; Botti, P.; Abraham, D. J.; Kellogg, G. E. Accessible Method for Estimating Free Energy Changes Resulting from Site-Specific Mutations of Biomolecules: Systematic Model Building and Structural/Hydropathic Analysis of Deoxy and Oxy Hemoglobins. *Protein-Struct. Funct. Genet.* **2001**, *42*, 355−377.
(23) Bartlett, P. A.; Hanson, J. E.; Giannousis, P. P. Potent Inhibition of Pepsin and Penicillopepsin by Phosphorus-Containing Peptide Analogues. *J. Org. Chem.* **1990**, *55*, 6268−6274.
(24) Gilson, M. K.; Honig, B. H. Calculation of electrostatic potentials in an enzyme active site. *Nature* **1987**, *330*, 84−86.
(25) Bashford, D.; Karplus, M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **1990**, *29*, 10219−10225.
(26) Gilson, M. K. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* **1993**, *15*, 266−282.
(27) Yang, A.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. On the calculation of pK_a in proteins. *Proteins* **1993**, *15*, 252−265.
(28) Dixon, S. J.; Jurs, P. C. Estimation of pK_a for organic oxyacids using calculated atomic charges. *J. Comput. Chem.* **1993**, *14*, 1460−1467.
(29) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.
(30) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144−1149.
(31) Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. Computing ionization states of proteins with a detailed charged model. *J. Comput. Chem.* **1996**, *17*, 1633−1644.
(32) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. The determinants of pK_as in proteins. *Biochemistry* **1996**, *35*, 7819−7833.
(33) Sham, Y. Y.; Chu, Z. T.; Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(34) Duarte, H. A.; Carvalho, S.; Paniago, E. B.; Simas, A. M. Importance of tautomers in the chemical behavior of tetracyclines. *J. Pharm. Sci.* **1999**, *88*, 111−120.
(35) Briggs, J. M.; Antosiewicz, J. Simulation of pH-dependent properties of proteins using mesoscopic models. *Rev. Comput. Chem.* **1999**, *13*, 249−311 and references therein.
(36) Ullmann, G. M.; Knapp, E. W. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.* **1999**, *28*, 533−551.
(37) Crnogorac, M. M.; Ullmann, G. M.; Kostic N. M. Effects of pH on protein association: modification of the proton-linkage model and experimental verification of the modified model in the case of cytochrome *c* and plastocyanin. *J. Am. Chem. Soc.* **2001**, *123* (44), 10789−10798.
(38) Nielsen, J. E.; McCammon, J. A. On the evaluation and optimization of protein X-ray structures for pK_a calculations. *Protein Sci.* **2003**, *12*, 313−326.
(39) Dixon, H. B.; Clarke, S. D.; Smith, G. A.; Carne, T. K. The origin of multiply sigmoid curves of pH-dependence. The partitioning of groups among titration p*K* values. *Biochem J.* **1991**, *278*, 279−284.
(40) Bountis, T., Ed. *Proton Transfer in Hydrogen-Bonded Systems*; NATO ASI Series B: Physics 291; Plenum Press: New York, 1992; pp 1−355.
(41) Air, G. M.; Laver, W. G. The neuraminidase of influenza virus. *Proteins* **1989**, *6*, 341−356.
(42) Colman, P. M. Influenza virus neuraminidase: enzyme and antigen. In Krug, R. M., Ed. *The influenza viruses*; Plenum Press: New York, 1989; pp 175−218.
(43) Liu, C.; Eichelberger, M. C.; Compans, R. W.; Air, G. M. Influenza type A virus neuraminidase does not play a role in viral entry, replication, assembly, or budding. *J. Virol.* **1995**, *69*, 1099−1106.
(44) Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144−51.
(45) 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.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418−423.
(46) Colman, P. M. Influenza virus neuraminidase: structure, antibodies, and inhibitors. *Protein Sci.* **1994**, *3*, 1687−1696.
(47) Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Air, G. M.; Luo, M. A strategy for theoretical binding constant, K_i, calculations for neuraminidase aromatic inhibitors designed on the basis of the active site structure of influenza virus neuraminidase. *Proteins* **1995**, *23*, 264−277.
(48) Singh, S.; Jedrzejas, M. J.; Air, G. M.; Luo, M.; Laver, W. G.; Brouillette, W. J. Structure-based inhibitors of influenza virus sialidase. A benzoic acid lead with novel interaction. *J. Med. Chem.* **1995**, *38*, 3217−3225.
(49) von Itzstein, M.; Dyason, J. C.; Oliver, S. W.; White, H. F.; Wu, W. Y.; Kok, G. B.; Pegg, M. S. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *J. Med. Chem.* **1996**, *39*, 388−391.
(50) Wade, R. C. "Flu" and structure-based drug design. *Structure* **1997**, *5* (9), 1139−1145.
(51) Sudbeck, E. A.; Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Air, G. M.; Laver, W. G.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Luo, M. Guanidinobenzoic acid inhibitors of influenza virus neuraminidase. *J. Mol. Biol.* **1997**, *267*, 584−594.
(52) Taylor, N. R. Inhibition of Sialidase. In Gubernator, K. et al., Ed. *Structure-based ligand design*; WILEY−VCH: New York, 1998; Vol. 6, pp 105−119.
(53) Taylor, N. R.; Cleasby, A.; Singh, O.; Skarzynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 2. Crystallographic and molecular modeling study of complexes of 4-amino-4H-pyran-6-carboxamides and sialidase from influenza virus types A and B. *J. Med. Chem.* **1998**, *41*, 798−807.
(54) Varghese, J. N. Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* **1999**, *46*, 176−196.
(55) Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Ali, S. M.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E. A.; Finley, J.; Luo, M.; Air, G. M.; Laver, G. W. Potent inhibition of influenza sialidase by a benzoic acid containing a 2-pyrrolidinone substituent. *J. Med. Chem.* **1999**, *42*, 2332−2343.
(56) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.* **2001**, *44*, 961−971.
(57) Smith, B. J.; Colman, P. M.; Von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Sci.* **2001**, *10*, 689−696.
(58) Yi, X.; Guo, Z.; Chu, F. M. Study on molecular mechanism and 3D-QSAR of influenza neuraminidase inhibitors. *Bioorg. Med. Chem.* **2003**, *11*, 1465−1474.
(59) Dunn, C. J.; Goa, K. L. Zanamivir: a review of its use in influenza. *Drugs* **1999**, *58*, 761−784.
(60) Freund, B.; Gravenstein, S.; Elliott, M.; Miller, I. Zanamivir: a review of clinical safety. *Drug Saf.* **1999**, *21*, 267−281.
(61) Silagy, C. A.; Campion, K. Effectiveness and role of zanamivir in the treatment of influenza infection. *Ann. Med.* **1999**, *31*, 313−317.
(62) Smith, B. J.; McKimm-Breshkin, J. L.; McDonald, M.; Fernley, R. T.; Varghese, J. N.; Colman, P. M. Structural studies of the resistance of influenza virus neuramindase to inhibitors. *J. Med. Chem.* **2002**, *45*, 2207−2212.
(63) Colman, P. M.; Varghese, J. N.; Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **1983**, *303*, 41−44.
(64) Varghese, J. N.; McKimm-Breschkin, J. L.; Caldwell, J. B.; Kortt, A. A.; Colman, P. M. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **1992**, *14*, 327−32.
(65) Burmeister, W. P.; Henrissat, B.; Bosso, C.; Cusack, S.; Ruigrok, R. W. Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure* **1993**, *1*, 19−26.
(66) White, C. L., Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A sialic acid-derived phosphonate analogue inhibits different strains of influenza virus neuraminidase with different efficiencies *J. Mol. Biol.* **1995**, *245*, 623−634.
(67) Gussio, R.; Zaharevitz, D. W.; McGrath, C. F.; Pattabiraman, N.; Kellogg, G. E.; Schultz, C.; Link, A.; Kunick, C.; Leost, M.; Meijer, L.; Sausville, E. A. Structure-based design modifications of the paullone molecular scaffold for cyclin-dependent kinase inhibition. *Anticancer Drug. Des.* **2000**, *15*, 53−66.
(68) Cashman, D. J.; Rife, J. P.; Kellogg, G. E. Which Aminoglycoside Ring is most Important for Binding? A Hydropathic Analysis of Gentamicin, Paromomycin, and Analogues. *Bioorg., Med. Chem. Lett.* **2001**, *11*, 119−122.

(69) Taylor, N. R.; von Itzstein, M. A structural and energetics analysis of the binding of a series of *N*-acetylneuraminic-acid-based inhibitors to influenza virus sialidase. *J. Comput. Aided Mol. Des.* **1996**, *10*, 233−246.
(70) Muegge, I. The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4. *Med. Chem. Res.* **1999**, *9*, 490−500.
(71) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method. *J. Med. Chem.* **1999**, *42*, 5142−5152.
(72) Cocco, L.; Temple, C., Jr.; Montgomery J. A.; London, R. E.; Blakley, R. L. Protonation of methotrexate bpund to the catalytic site of dihydrofolate reductase from *Lactobacillus Casei*. *Biochem. Biophys. Res. Commun.* **1981**, *100*, 413−419.
(73) Karplus, P. A.; Faerman, C. Ordered water in macromolecular structure. *Curr. Opin. Struct. Biol.* **1994**, *4*, 770−776.
(74) Holzer, C. T.; von Itzstein, M.; Jin, B.; Pegg, M. S.; Stewart, W. P.; Wu, W. Y. Inhibition of sialidases from viral, bacterial and mammalian sources by analogues of 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid modified at the C-4 position. *Glycoconjugate J.* **1993**, *10*, 40−44.
(75) Lentz, M. R.; Webster, R. G.; Air, G. M. Site-directed mutation of the active site of influenza neuraminidase and implications for the catalytic mechanism. *Biochemistry* **1987**, *26*, 5351−5358.
(76) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules *J. Med. Chem.* **1985**, *28*, 849−857.

JM0302593