Computational Titration Analysis of a Multiprotic HIV-1 Protease-Ligand Complex.

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

Department of Biochemistry and Molecular Biology, University of Parma, 43100, Italy, Department of General and Inorganic Chemistry, University of Parma, 43100, Italy, Department of Medicinal Chemistry and Institute for Structural Biology & Drug Discovery, Virginia Commonwealth University, Richmond, VA, 23298-0540, U.S.A .

RECEIVED DATE (automatically inserted by publisher); E-mail: andrea.mozzarelli@unipr.it, glen.kellogg@vcu.edu

It has generally proven difficult to correlate solution binding data and crystallographic structural data with an in-depth understanding of the atomic-level *chemical* details of binding. This is especially true in cases where the binding site and/or the ligand are multiprotic, i.e., possessing multiple ionizable groups. While ionization states are governed by the pH of the media and the pK_as of the ionizable moieties, localized effects can significantly impact ionization, and each ionizable group is not independent of all others in the complex. The desired understanding is further obscured by experimental constraints: 1) there are relatively few cases where binding as a function of pH has been reported for structurally well-characterized protein-ligand complexes; 2) the pH at which solution binding is measured may not be the optimum pH for binding; and 3) the pH of crystallization may not be the same as that for solution measurements. Thus, the linkage between solution data and crystal structure may be fairly tenuous, and predictions, e.g., in virtual screening, of questionable value.

We recently explored this issue for ligands bound to influenza virus neuraminidase and proposed a new calculation paradigm that we termed 'computational titration'.[1] We were able to model a multiplicity of protonation states for the complexes and propose families of 'isoprotic' structures that correlate with the experimental free energy of binding data. The resulting mean HINT[2] energy scores produced relatively good free energy predictions. Because this paradigm requires building and optimization of tens to thousands of models with specific protonation for each protein-ligand complex, some amount of manual filtering was necessary and models that would obviously be of high energy were excluded. A more accurate method of energy evaluation would also include a statistical mechanics evaluation of *all* models.

With these considerations we have implemented a new computational tool that *in silico* builds and energy scores molecular models of all protonation states for a ligand-bound active site, including the energetic effects of bridging water molecules.[3] To evaluate this tool we have modeled the pH-dependent inhibition of HIV-1 protease by Glu-Asp-Leu, as recorded experimentally by Louis et al. between ca. pH 3.0 and 5.0.[4] A 2.0 Å resolution structure of the same complex was obtained on crystals grown at pH 4.2.[4] Figure 1 illustrates the environment of the ligand at its binding site nominally at "pH 7". There are four Asp (25α, 29α, 30α, 25β) residues interacting with the peptide, and the peptide itself has three carboxylate groups and an amine from the side chains and termini. Thus, there are eight potentially ionizable moieties at the binding site although one, the glutamate carboxylate of the peptide, is positioned largely out of the cavity and is directed into the solvent. Two potentially significant[3] water molecules (301 and 313) are also present at the active site. The close interactions between peptide and protein (Figure 1) suggest that at least some of the protein and/or ligand carboxylates must be protonated for viable ligand binding. In particular,[3,4,5] the Asp25α/Asp25β region is normally protonated in some manner.

The molecular models were prepared using the Sybyl modeling program as described previously.[1,2c,3] First, the coordinates for the complex were retrieved from the PDB (accession code 1a30), protons were added to all atoms (protein, peptide and water) and the resulting model was subjected to an energy minimization that allowed only the (non-experimentally determined) hydrogens to relax. Finally, the protein, peptide and water were separated for computational titration.

Analysis of this molecular model indicated 4374 unique protonation models. These models (Table 1) range from the most basic (all 8 sites *de*protonated, charge -7) to the most acidic (all 8 sites protonated, charge +1). The computational titration procedure built molecular models for each case, optimized the positions of polar hydrogens (–OH, $–NH_2$ and –SH), and reoriented the two water molecules as necessary with exhaustive torsional rotation algorithms[6] to improve protein/ligand/water hydrogen bonding. The reported HINT[2] score (H_{TOTAL}) for each protonation model is:

$$H_{TOTAL} = H_{protein\text{-}ligand} + H_{ligand\text{-}water} \quad (1)$$

where the two terms are from protein-ligand and ligand-water interaction scores.[3]

It is important to note that all of the 4374 models differ *only* in the presence or placement of protons. Thus, any of them would still fit within the original electron density envelopes of the crystallographic data. We might term these models to be 'isocrystallographic'. This is not to say, however, that the converse is necessarily true: some of these proton models likely correspond to crystallographically distinct structures, but analysis is beyond the scope of the present work. However, we would expect to see an indication of potential structural rearrangement manifested with highly unfavorable free energy binding scores.

Table 1. Computational Titration Results for Glu-Asp-Leu in HIV-1

model charge	no. models[a]	min. score[b] (kcal mol^{-1})	max. score[b] (kcal mol^{-1})	ave. score[b] (kcal mol^{-1})	Bolztmann[b,c] (kcal mol^{-1})
-7	1	6.04	6.04	6.04	6.04
-6	15	1.11	6.53	3.91	2.35
-5	98	-2.52	6.38	2.00	-0.44
-4	364	-5.80	5.51	0.30	-2.84
-3	840	-8.16	4.77	-1.19	-4.61
-2	1232	-8.47	3.51	-2.50	-5.35
-1	1120	-8.57	0.80	-3.62	-5.79
0	576	-8.37	-1.23	-4.58	-6.07
1	128	-7.94	-3.24	-5.39	-6.23

[a]Each carboxylate has three possible states: deprotonated (1) and protonated at either carboxy oxygen (2). [b]Scores converted to free energy with equation 2.[3] [c]Boltzmann-weighted free energy of ligand binding.

[†]Biochemistry and Molecular Biology, University of Parma.
[‡]General and Inorganic Chemistry, University of Parma.
[§]Medicinal Chemistry, Virginia Commonwealth University.

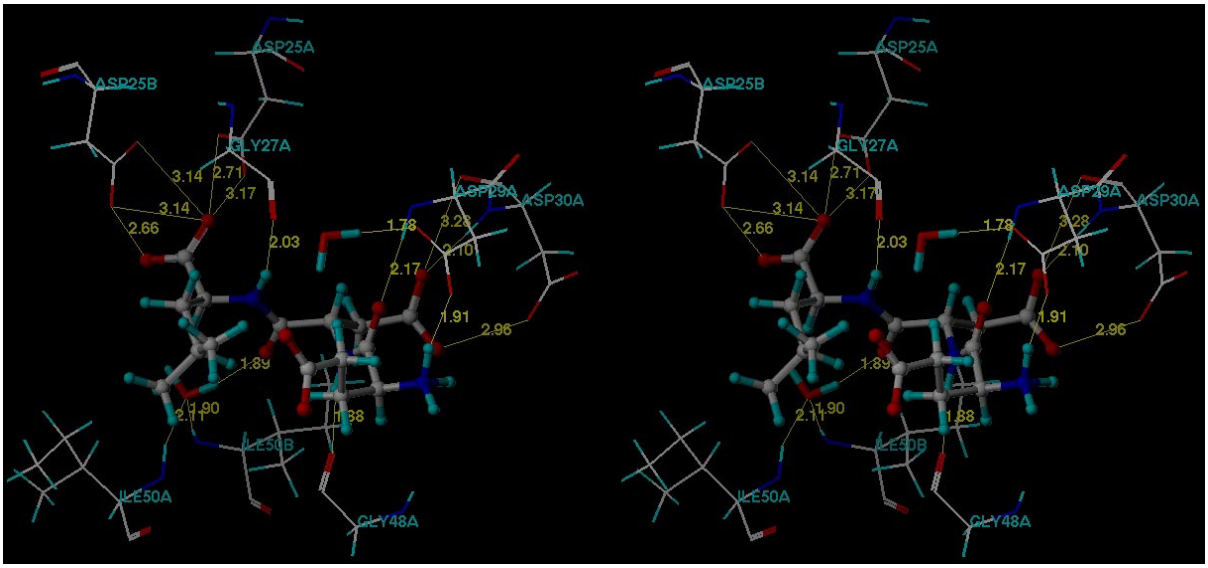


Figure 1. Stereoview of Glu-Asp-Leu (rendered ball & stick) bound in HIV-1 protease from 1a30.[4] Two (301 and 313)[3] bridging water molecules (rendered stick) are shown. Model is displayed at nominal pH 7, where all carboxylates are ionized and the primary amine is protonated. Distances (Å) between atoms for principal hydrogen bonds and unphysical interactions involving carboxylates, carbonyls and/or water molecules are indicated.

Figure 2 shows the results of the titration analysis (also Table 1). The scores for each site/model charge vary over a wide range, indicating both favorable and (highly) unfavorable protonation models within each isoprotic set. For each, an arithmetic and Boltzmann average of the scores converted to free energy with[3]

$$\Delta G_{binding} = -0.0016\, H_{TOTAL} - 4.873 \qquad (2)$$

is shown. Equation 2, derived from a set of 23 HIV-1 protease-ligand complexes (not including 1a30), has a standard error of ±0.84 kcal mol^{-1}.[3] The Boltzmann curve shows a particularly interesting trend: an initially rapid decrease in free energy for the first few protons added to the system as the highly negative active site is neutralized, followed by a leveling off as the site appears to compensate by adopting low energy torsional conformations for the resulting acids. This behavior parallels nicely the experimental pH-dependent binding.[4] The calculated and experimental titration curves were aligned (Figure 2) by scaling and shifting the pH axis of the experimental binding data to this presumed parallel relationship. Other alignment criteria and/or assumptions are possible. Here the difference between the measured and predicted $\Delta G_{binding}$ is about 0.6 kcal mol^{-1} throughout the pH range.

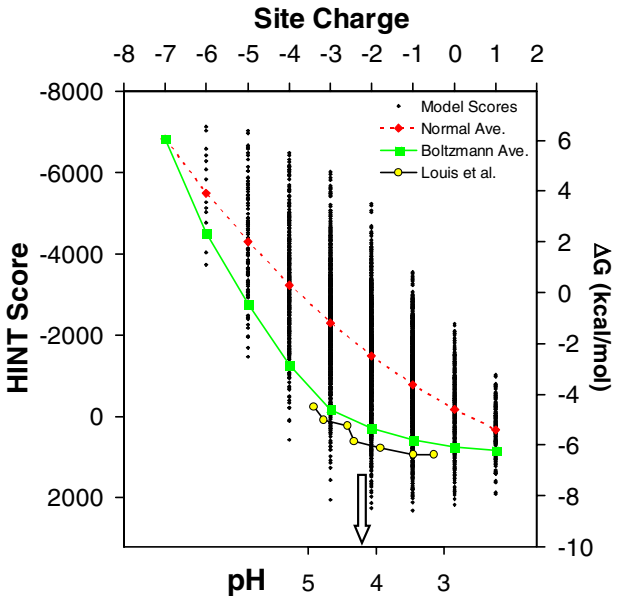


Figure 2. Composite of calculated and experimental titration results for Glu-Asp-Leu bound in HIV-1 protease. Louis et al.[4] data have been scaled and shifted on the pH/site charge axis to match Boltzmann curve behavior. Arrow indicates pH of crystallization for 1a30.[4]

The crystal model thus appears to correspond to a site charge of -3 to -2. Among the 2072 models of this site charge range, many of the more energetically reasonable ones have protonated two sites of Asp25α, Asp25β or the C-terminal carboxylate of the peptide, either Asp30α or the Asp carboxylate of the peptide, and the N-terminal amine of the peptide. Protonation of the peptide Glu carboxylate is largely irrelevant as it interacts mostly with solvent, and may be the fifth added proton for -2 charge models.

While this calculation should be applicable to nearly any protein-ligand complex with structural data, we are quite interested in further calibrations with pH-dependent binding data.

Acknowledgements. We thank the Italian Ministry of Instruction, University and Research and Virginia Commonwealth University for support.

Supporting Information Available: Detailed description and HINT scores of protonation-specific molecular models.

References

(1) Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E., *J. Med. Chem.* **2003**; *46*, 4487-4500.

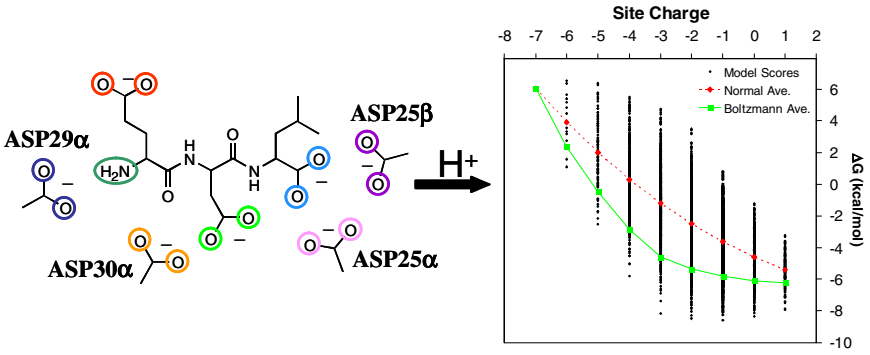
(2) (a) Wireko, F. C.; Kellogg, G. E.; Abraham, D. J., *J. Med. Chem.* **1991**, *34*, 758-767. (b) Kellogg, G. E.; Abraham, D. J., *Eur. J. Med. Chem.* **2000**, *35*, 651-661. (c) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A., *J. Med. Chem.* **2002**, *45*, 2469-2483.

(3) Fornabaio, M.; Spyrakis, F.; Mozzarelli, A.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E., *J. Med. Chem.* (in press).

(4) Louis, J. M.; Dyda, F.; Nashed, N. T.; Kimmel, A. R.; Davies, D. R., *Biochemistry* **1998**, *37*, 2105-2110.

(5) (a) Smith, R., Brerenton, I. M.; Chai, R. Y.; Kent, S. B. H., *Nature Struc. Bio*. **1996**, *3*, 946-950. (b) Wang, Y. X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A., *Biochemistry* **1996**, *35*, 9945-9950.

(6) Kellogg, G. E.; Chen, D. L., *Chem. Biodiversity* **2004**, *1*, 98-105.



ABSTRACT FOR WEB PUBLICATION

A new computational method for analyzing the protonation states of protein-ligand complexes with multiple ionizable groups is applied to the structurally characterized complex between the peptide Glu-Asp-Leu and HIV-1 protease. This complex has eight ionizable groups at the active site: four from the ligand and four Asp residues on the protein. Correlation, with an error of ca. 0.6 kcal mol^{-1}, is made between the calculated titration curve and the experimental titration curve. The analysis suggests that between four and five of the eight ionizable groups are protonated at the pH of crystallization.