Assessing Uremic Toxin Binding Dynamics in Human Serum Albumin, Sudlow Site I

Reza Taheri

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Committee:

Jim Pfaendtner

Xiaosong Li

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Assessing Uremic Toxin Binding Dynamics in Human Serum Albumin, Sudlow Site I

Reza Taheri

Chair of the Supervisory Committee:
Jim Pfaendtner
Department of Chemical Engineering

In an effort to better understand the dynamics corresponding to the stabilization of uremic toxins such as indoxyl and p-cresyl sulfate, simulative modeling must be employed to observe molecular perambulations. In this investigation, the GROMACS molecular dynamics engine was used to perform a 250 ns simulation of human serum albumin complexed with these uremic toxins in its Sudlow Site I binding pocket. The trajectories were then taken and subjected to analyses designed to assess interaction energy and intermolecular distance between the amino acid residues that comprised the binding site. Intermolecular distance information was subjected to dimensionality reduction via primary component analysis, and then a mean-shift algorithm was employed to determine the most likely binding conformation, which was then cross referenced with the initial configuration from the experimentally-derived PDB file.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

For Mom, Dad, and Mags.

1.INTRODUCTION

As the single most prevalent cause of death amongst American adults, kidney failure provides the impetus for this thesis' investigation of protein-ligand binding dynamics. Closely affiliated with chronic kidney disease (CKD), are a collection of small, anionic molecules known as protein bound uremic toxins (PBUTs), that form tightly knit complexes with human serum albumin (HSA), a protein responsible for the transport of hydrophobic compounds throughout the vasculature which comprises approximately half of the blood plasma proteome [1-3]. Ordinarily these toxins are removed via proximal tubule secretion following complexion with the organic anion transporter 1 (OAT 1), but due to endocrine disruption and fibrosis of the renal epithelium, expression of OAT 1 protein is suppressed, and the toxins complex instead to HSA. Complexion to HSA ultimately inhibits removal from the vasculature via renal filtration, which ultimately leads to the further compounding of PBUT toxicity, thus furthering kidney degradation [4]. Removal of uremic toxins is possible via dialysis, however, this can only be achieved when the toxins are removed from the complexes formed with HSA [5]. The objective of the investigation then, is to identify what residual interactions are responsible for the stabilization of these PBUTs, so that we might tactfully disrupt them, and achieve the unbinding of these uremic toxins from HSA—thereby improving the efficacy of dialysis.

To examine the uremic toxins more closely, we are fortunate in regard to the fact that these compounds are relatively simplistic. The two toxins at the center of our investigation, indoxyl sulfate (ISO) and p-cresyl sulfate (PCS) are both small, endogenously-synthesized bacterial waste products, with shared aromatic and sulfate moieties [4,6]. Where IOS and PCS differ from one another, interestingly enough, is the fact that the former's aromatic moiety is

constituted by an indole group. This is noteworthy for two particular reasons; the first is that IOS is distinguished from PCS on the basis of its ability to both donate and accept hydrogen bonds due to the presence of the secondary amine–we will hope to see this manifest itself in the comparisons of hydrophilic contacts between the two toxins and binding pocket residues. The second reason, is that this same amine serves to withdraw electron density from the adjacent sulfate group; it is a facet we've seen manifest itself in the preliminary investigations of Smith & Pfaendtner, and it is what we expect to see in our dissections of Coulombic and Lennard-Jones based contributions to the interaction energy [7]. This is all to say that the selection of IOS and PCS was very deliberate on account of their disparate chemistries, and the expectation is to witness what affect the presence of a hydrogen-bond donating moiety might have.

The novelty of our investigation here hinges on the new regions in HSA which we have chosen for examination. Previous experimentation has favored Sudlow Site II as the primary binding locale for both IOS and PCS, however, we chose to include the first Sudlow binding site, Sudlow Site I, for its ability to accommodate two toxins, as opposed to just one [8]. Taking into account the aforementioned aromaticity of both the IOS and PCS, it will be quite interesting to observe any possible toxin-toxin interactions; likely stemming from pi-pi stacking. We anticipate that that the two ligands' interactions will be dwarfed by those established with the residues, but even still the event of uremic toxins interacting with one another is a largely unexplored phenomenon.

In order to examine these aforementioned protein-ligand interactions, we will make use of the highly modular molecular dynamics simulation. The integral components of a simulation include the atomic coordinate files, forcefields and topology [9]. In terms of coordinate files,

these documents serve the purpose of delineating where, in relation to one another, the atoms that comprise a given system are to be designated in a three dimensional space. Perhaps the most ubiquitous coordinate document within the sphere of biomolecular computation is that of the protein database file, or PDB [10]. These coordinate files are derived from X-ray crystallography, which is a benefit in our attempt to model real-world macromolecular phenomena. Aside from the requisite X, Y, and Z coordinate designations of each atom comprising a molecule, the columns in a PDB file are also used to delineate parameters such as occupancy and temperature factor which are integral in determining how kinetic energy manifests itself throughout the course of the run [11].

The next critical component of molecular dynamics are those of the topology and forcefields, which can be thought of as the amalgam of auxiliary position-based parameters, and the concordant energies between atoms in these positional conformations. Forcefields in classical molecular dynamics can be described as a series of equations, pictured below, which serve to account for both bonded and non bonded atomic interactions in the system [12].

$$U_{bonded} = \frac{k}{2} {r_{12}}^2 \tag{1.1}$$

$$U_C = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r_{12}}\right) \tag{1.2}$$

$$U_{LJ} = \frac{-A}{{r_{12}}^6} \tag{1.3}$$

$$U_{Angle} = k\theta^2 \tag{1.4}$$

$$U_{Dihedral} = \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] \tag{1.5}$$

There are three primary contributors to interatomic interactions: Coulombic, Lennard-Jones, and electrostatic repulsion. Coulombic interactions essentially serve as an attractive electrostatic force between two opposing charges that is inversely proportional to interatomic distance [13,14]. Next is that of the Lennard-Jones contribution, which account for attractive forces such as dipole-dipole, pi-pi stacking, or other non-electrostatic forces. The final is that of the electrostatic repulsion energies, which as the name might suggest, accounts for the ejection two like-charged species might encounter in close proximity to one another. Accounting for all three of these interactions in total generates a graph for interactive potential akin to a morse potential, which is visualized below in Figure 1.

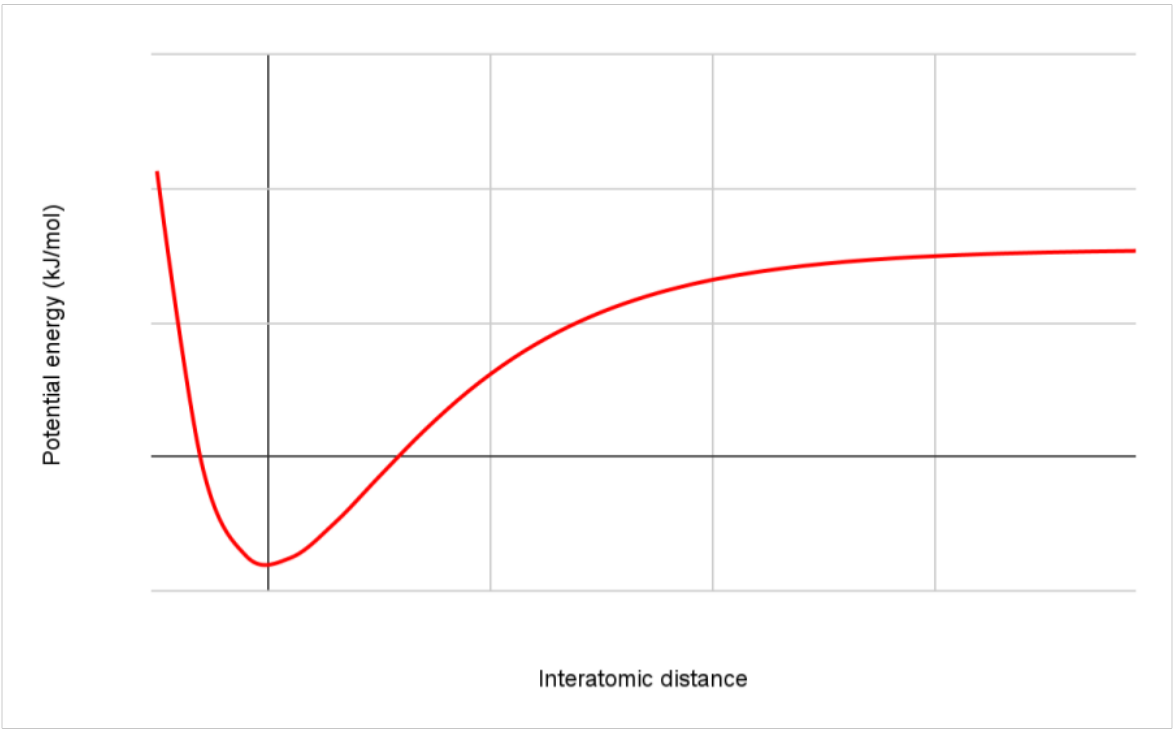

Figure 1. Interaction potential in kJ/mol as a function of interatomic distance between two atoms.

Topologies on the other hand, serve as a sort of catalog of all possible parameterizations of bonded interactions within a system, as dictated by the aforementioned forcefields [15]. Covalent bonds, the dihedrals they form, and one-four interaction potentials are all accounted for in the topology file, and this is of critical importance when one takes into consideration the nature of how atoms interact in bonded and non-bonded modalities. This is all to say that forcefield and topology files serve a vital role in dictating how atomic positions are manifested on the basis of interactive energies and bonding parameters.

Now that the essential components of a molecular dynamic simulation have been described, there are a few preliminary stages relating to workflow that must be outlined. Any successful molecular simulation is contingent upon the adequate relaxation of the involved structures in a system. This is due the fact that coordinates are often non preferential in regards to arrangement, and need some time allocated in order to relax, reducing maladaptive energetic contributions from phenomena such as steric impedance, the aforementioned electrostatic repulsion, and the like [17]. The aim is to reach a global energetic minimum by which to set an adequate precedent for the proceeding simulation, therefore obtaining energetic descriptions that either agree with experimentally validated data, or do not exist entirely outside of the realm of plausibility. So relaxation through the induction of thermal energy over the course of a timespan is what is occurring at the most basic level of equilibration (the formal term designated to the relaxation process), but there are differentiable options for approach. Isobaric, isochoric, and microcanonical ensembles are all unique equilibration methodologies that hold pressure, volume, and total systemic energy constant, respectively, and due to the modular nature of molsim

workflows, we are able to seek multiple viable routes to structural relaxation [18]. For the purposes of this investigation, we will primarily be utilizing the isobaric and isothermal ensembles to subtly bring our system up to temperature through annealing, and then subsequently to replicate a physiological environment wherein pressure is held constant.

Following equilibration, the achievement of our optimally relaxed structure, comes the extended-timescale run, a final, unfettered simulation that allows our system to carry out its physical properties at our chosen temperature and pressure. Following the final run, comes our analysis. The PLUMED analysis package will be of great use in this step, as we can utilize it to observe any specified measurement at an atomistic resolution over the course of the entire trajectory [19]. For the purposes of our investigation, we will be principally interested in residual contacts between the binding pocket of the protein and our ligand, as well as the distances between our centers of mass between those most prevalently contacting residues, as well as our ligands. Regarding the latter, we will take the information pertaining to the center of mass distances, and subject them to a form of dimensionality reduction known as principal component analysis (PCA), which simply takes our **N**-dimensional space (with **N** denoting our number of residues) and reducing it to a bi-dimensional distribution of indices, each representing a center of mass distance for each of the **N** residues at a given timestep from the course of the simulation trajectory [20]. Following this step, a mean shift algorithm will be employed to cluster the center-of-mass distributions into a discrete number of most probable binding modes [21].

Now that we have discussed all vital components at play, it is critical that we provide a summary of how they are to coalesce. Through a protein database file we will perform some brief alteration of the ligand molecules contained, and then proceed forward with our

equilibrations and production run. Following this, final trajectories will be subjected to analyses aimed at elucidating the nature of interaction between the ligand and the binding pocket, as well as between the toxins themselves. Energetic contributions from both Coulombic and Lennard-Jones potentials will provide further insight relating to the specific intermolecular interactions, which will couple well with the aforementioned ligand-residue contacts. With all said and done, this narrative will serve as a single step forward in the expansion of our understanding of how these toxins leaden essential physiological function, and hopefully prime the next generation of nephrologists on how to ameliorate the issue of Chronic Kidney Disease.

2.METHODS

PDB file code 2BXH was procured through the Chimera database, and reduced to a single chain from its original dimer. IOS and PCS ligands were constructed in GaussView, and the RMSD align module available through VMD molecular visualization package was utilized to superimpose these structures on the indoxyl sulfates sans hydrogens present in the original PDB file [22,23]. Alignment of PCS to IOS was performed on the basis of generally aligning the termini of each ligands' hydrophobic moieties, and then a second, atom-for-atom lineup of each's sulfates. Ligand coordinates were saved separately, and protein atoms were subjected to the H++ server for the sake of assigning residual charge states as well as accounting for missing hydrogens occurring in the original database file [7, 24-27].

The tLEaP module available through AMBER molecular dynamics package was then used to couple our system, in the absence of the ligands, to the amber14sb.ff forcefield parameters, as well as the TIP3P water specifications [28]. This generated AMBER coordinate

and topology files which were processed through the ParmEd module in order to generate a consolidated, GROMACS-compatible topology file [29]. It is also of vital importance that disulfide linkages are specified in the tLEaP input, which are delineated in the original PDB text. Ligand topology was obtained separately through .itp files available through the Pfaendtner Research Group's pbut_analysis directory on GitHub, and included into the GROMACS topology via the INCLUDE command. Similarly, position restraint files for both the protein and ligands were generated through GROMACS, and included using the same INCLUDE command [7].

Pre-equilibration continued with parameter specification and population of ancillary atomic species. Coordinate space was defined as cubic with edges 10.8 nm in length [30]. Solvation was accomplished with the eponymous GROMACS module, although it should be noted that the spc216.gro model was utilized in the command entry on account of TIP3P not being an available option. This is a largely irrelevant concern as the TIP3P specification from the tLEaP input file takes precedence in the topology, but nevertheless an important point to note as water model is a critical specification in the solvate command. Sodium counterions were introduced to neutralize the system's net negative charge, thus concluding the population stage of our simulation prep at approximately 124,000+ atoms. System relaxation began with a 5000-step energy minimization simulation with convergence specified to 10 kJ/mol. Velocities were seeded in an annealing simulation that was performed under NVT conditions; critical .mdp specifications included the v-rescale thermostat, and an established gradient of 5 to 298 K for the first 240 ps of the overall 250 ps timespan. Further relaxation proceeded under NPT conditions for another 500 ps, with the designated barostat set to Berendsen, targeted at 1 bar [7].

A third NPT equilibration was then conducted at a substantially longer timescale of 5 ns, and was coupled with a precise biasing scheme intended on enconsing our system in a reliably minimized energetic state, while keeping our ligands intact within Sudlow Site I. This biasing followed an upper wall scheme, which administered an inertial force–parameter *k*–of 150 kJ/(mol nm^2) at any instance in which the root mean square deviation (parameter *x*) of the ligand's carbon atoms as well as those of any residue within 6 Å of the ligand exceeded 0.1 nm, and was accomplished through the Plumed 2.7 plugin. When utilizing Plumed for any performance contingent upon RMSD, it is of critical importance that the required reference PDB files meet two critical parameters: **1)** the indices of the PDB and corresponding .gro files match exactly, and **2)** that occupancy and β-factor columns register nonzero values. Thermostat specification was maintained at v-rescale, however barostat settings were changed to Parrinello-Rahman in attempt to more accurately incorporate the influence of externally-applied osmotic pressure; temperature converged to 298 K at a 10 ps^{-1} frequency, and pressure to 1 bar at 1 ps^{-1}. Three production simulations of uniquely-seeded velocities were then initiated from the outputs of the aforementioned biased NPT runs, under the same ensemble conditions. Timespan was set to 250 ns, with identical thermostat and barostat specifications as the previous equilibration step [7].

As was done in Smith & Pfaendtner, the first 50 ns of the production trajectories were expurgated to allow a final bout of unbiased equilibration. VMD was used to generate a rudimentary list of all residues within 6 Å of each ligand, and these residues' atomic indices, as well as those of the ligands, were detailed in a Plumed input template–obtainable on the pbut_analysis GitHub repository–in order to collect atomic and hydrophilic coordination

numbers between residues and ligands. Atomic and hydrophilic coordination differ on the basis of distance, and as such atomic coordination registered at any contact indiscriminately within 4 Å (r = 0.4 d = 0, as specified in Smith & Pfaendtner), whereas hydrophilic registered those between 2.4 and 2.7 Å (r = 0.03 d = 0.27). Once these collective variables were garnered, a histogram template (available through the same repository) constructed through Python was used to visualize and identify the nine most prevalent residues along the course of the 200 ns production trajectories. These nine residues were then subject to analysis through the GROMACS energy module (requiring the craftful delineation of energy groups between said residues and their corresponding ligand), and it is here were we note the shortcomings of GROMACS 2020 in doing so; all energy raw energy .xvg files were generated through GMX 2018, and then visualized through the requisite energy analysis JupyterLab notebook, again available through pbut_analysis. It should be noted that for each analysis mentioned above, individual plumed input files and JupyterLab notebooks were dedicated to each toxin within Sudlow Site I [7].

Data collection for binding mode determination was performed through Plumed, with collective variables defined as the amalgamated center-of-masses for each toxin, as well as those of the nine residues delineated in the previous coordination number analysis. Dimensionality reduction was performed via principal component analysis (PCA) offered through scikit-learn, and a mean shift clustering approach was utilized to determine the modes which represent binding conformations as they appeared in the reduced, bidimensional representation. In an effort to better appraise the influence of the aforementioned clustering's bandwidth parameter on performance, a silhouette score was employed on all resulting data. PoseView was employed in

order to compare representative configurations from each binding mode, with the initial experimental configuration from which trajectories were seeded. In Figure 2. below, we see these initial configurations visualized, and will refer back to them once we've garnered the conformations of the most populated cluster representatives [7].

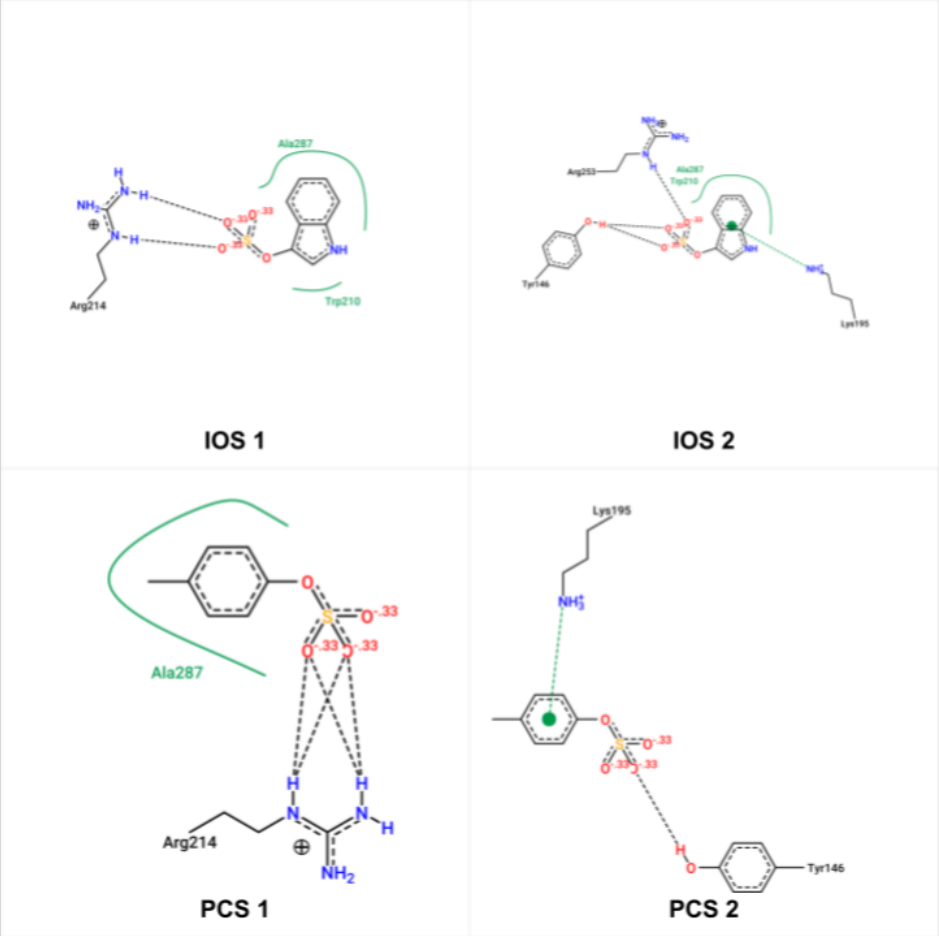

Figure 2. PoseView output of IOS experimentally-resolved structures, and those of PCS supplanted in their place using VMD's RMSD Align tool. PCS was aligned to IOS by overlapping sulfate and hydrophobic regions.

3.RESULTS AND DISCUSSION

Just as was done in Smith & Pfaendtner, energetic analyses were bifurcated between Coulombic and Lennard-Jones contributions. This design was followed in an effort to determine which moiety present on IOS and PCS played the most important role in stabilization within the binding pocket. As stated earlier, both IOS and PCS are compounds with a wide range of structural and chemical novelties, and as such there exist a wide range of capable intermolecular forces to be formed. An E_C/E_{LJ} calculation can be utilized to differentiate interactions as disparate as pi-pi stacking and salt bridging, and on the basis of this resolution we make our selection. Energetics for protein-ligand and ligand-ligand interactions were averaged over all three simulations for each toxin, and compiled below in Table 1.

Table 1. Average interaction energies for total protein-ligand and ligand-ligand complexes. Averages were taken over the course of all three production simulations.

	E_C (±SEM) [kJ/mol]	E_{LJ} (±SEM) [kJ/mol]	E_{tot} (±SEM) [kJ/mol]
ISO (1/2)	-137 (10.8)	-84 (5.65)	-222 (16.3)
ISO (2/2)	-138 (7.05)	-76 (3.90)	-214 (10.9)
PCS (1/2)	-107 (26.1)	-76 (10.9)	-183 (15.2)
PCS (2/2)	-128 (2.11)	-79 (8.26)	-207 (10.4)
ISO-ISO	-1.43 (1.94)	-7.69 (5.30)	-9.12 (6.81)
PCS-PCS	+0.624 (0.954)	-4.74 (5.64)	-4.12 (4.70)

These findings agree with the Sudlow Site II investigations of Smith & Pfaendtner in that Lennard-Jones contributions are dwarfed by those of the Coulombic. However, there was an expectation that IOS's Lennard-Jones contribution would noticeably outmatch that of PCS's, which was not replicated to as noticeable a degree in the SSII investigations. This expectation initially arose from the fact that indoxyl sulfate's hydrophobic moiety has a higher number of atomic constituents than p-cresyl sulfate's, which causes one to reasonably expect a greater LJ contribution. Then again, IOS possesses an indole group, and on account of this moiety being mirrored in tryptophan residues, we would expect to see the potential for pi-pi stacking intermolecular forces accounted for in the table above.

Consulting the coordination number of SSI toxins between the constituent residues, we do find evidence supporting the occurrence of intermolecular interactions between IOS and tryptophan. Toxin, one of two, arbitrarily assigned to residue 559 in our coordinate files exhibits the strongest interaction with tryptophan, as W214 sits atop the histogram detailing atomic contact numbers in Figure 3, below. What is curious however, is the fact that W214 ranks third in hydrophilic contact number betwixt toxin one, which obfuscates the attribution to these interactions solely to pi-pi stacking; this would suggest at the very least an indiscriminate mixture of pi-pi stacking and hydrogen bonding. Intermolecular distance for pi-pi stacking averages at some 3.37 Å, which eliminates any possible oversight when setting the parameters of our contact numbers [31]. Furthermore, we note that both PCS present in SSI exhibit notable contact between W214, both in atomic and hydrophilic conformations. On account of PCS's own aromaticity, pi-pi stacking with tryptophan cannot be ruled out entirely, and in all likelihood the dynamic exists in the same amalgamation between stacking and hydrogen bonding. Further

indication of p-cresyl sulfate's draw towards these interactions with aromatic, hydrogen-bond-capable residues seems to be bolstered by Y150 signaling strong atomic and hydrophilic contacts with both toxins in SSI.

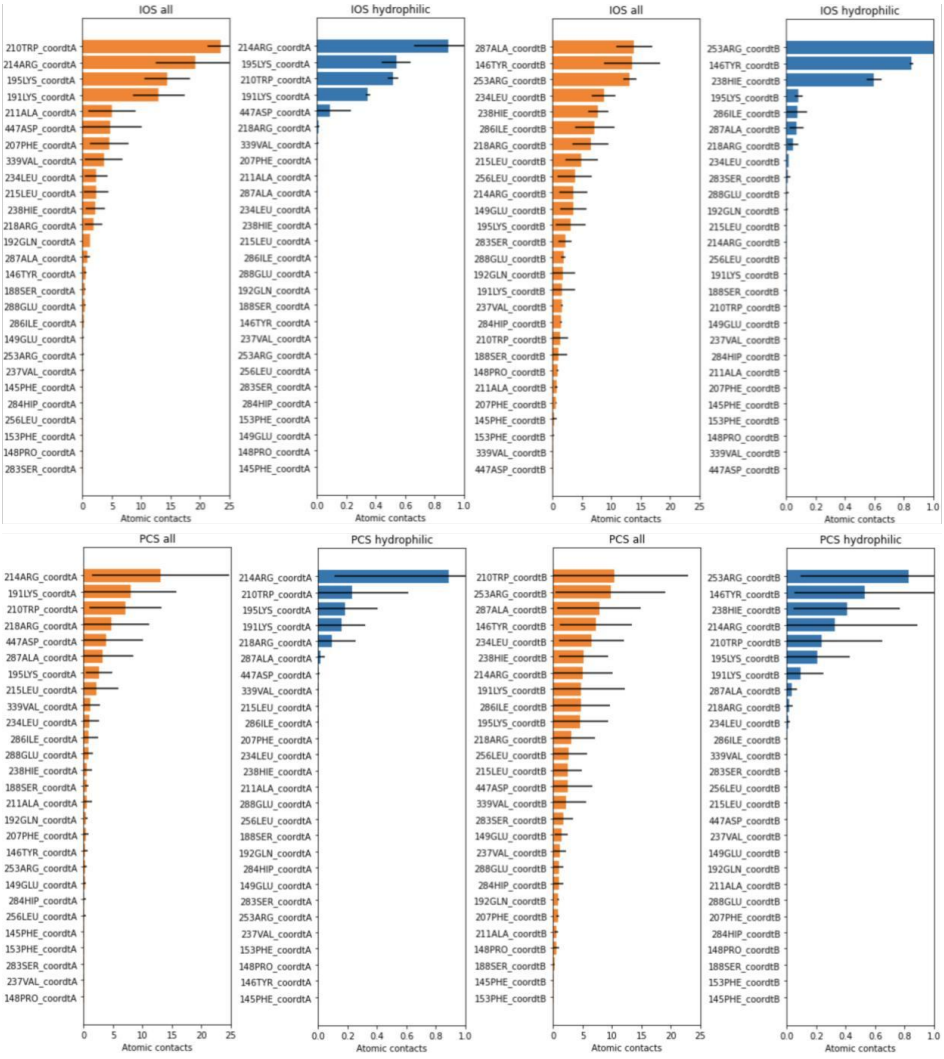

Figure 3. Atomic and hydrophilic contacts for toxins and binding pocket residues; atomic contacts in orange, and hydrophilic in blue. Moving left from right we have IOS toxin #1 (residue 559), then IOS #2 (560), and PCS #1 and #2 (again, residues 559 and 560, respectively).

Perhaps the most glaring lack of expected interaction is that between tryptophan and toxin two (again, residue 560 from the coordinates) and any of the SSI residues. Granted there is strong evidence for the expected interaction between Y150, but the fact that both PCS interact more conspicuously than a compound with an identical indole group is strange to say the least. Perhaps this could be linked to intermolecular interactions between the IOS in SSI divesting any notable interaction between toxin two and tryptophan, but further investigation is necessary. One final discrepancy between Smith & Pfaendtner and our own SSI investigations on the basis of contact number was that of inferred hydrogen-bonds between IOS and PCS. We note that averaged hydrophilic contacts between the IOS one, PCS one and two remain similar, however, IOS two formed a substantially higher amount of these contacts, and on this basis we can assume a higher tendency to establish hydrogen bonds with SSI residues. This is a reasonable finding on account of IOS's previously mentioned dual hydrogen donor/acceptor capabilities, but does strangely fly in the face of Smith & Pfaendtner's anomalous report of PCS forming more hydrogen bonds in SSI than IOS, despite only being able to act as an acceptor.

Returning to Table 1., briefly, we do note some findings worth discussing within the lens of ligand-ligand interactions in SSI. In comparison to residual interactions, obviously the magnitude of interaction is paltry, but perhaps not to such a degree that they can be entirely discounted; after all GROMACS's energy output resolution is as exiguous as 10^{-6} kJ/mol. Both IOS-IOS and PCS-PCS interactions share attractive Lennard-Jones contributions that dwarf, in magnitude, those of the Coulombic. This finding is congruent aromatic moieties facilitating pi-pi stacking. We note that these contributions are relatively larger in IOS than PCS, and this is likely due to the higher constituency of the former's aromatic region. What is most notable however,

are the dichotomous Coulombic contributions between the two pairs of toxins. PCS-PCS interactions display a lightly repulsive Coulombic contribution, contrasting starkly with IOS-IOS's mildly attractive contribution. What is so exciting about these findings, is that it comports with the two toxins' contrasting hydrogen bonding capabilities. As discussed earlier, p-cresyl sulfate lacks the ability to act as a hydrogen bond donor, unlike indoxyl sulfate, and as such we see this opposing electrostatic behavior manifest. IOS's electrostatic attraction most likely stems from hydrogen bonding between one toxin's secondary amine and the other's sulfate, whereas PCS's repulsion seems to be due to these bulky, negatively-charged sulfates repelling one another, even if it is slight in magnitude. Considering the ligands' conformation in the original coordinate file this makes sense; sulfates were about as far away from one another as possible. It is still interesting that hydrogen-bonding would take such a precedence over sulfate-sulfate repulsion in the case of indoxyl sulfate; perhaps it could be said that the repulsion has some diminishing effect on the former.

As was articulated in Smith & Pfaendtner, the inference that electrostatic contributions play the principal role in stabilizing both PCS and IOS ligands is sound based on the findings previously mentioned [7]. This is inaberrant as experimental evidence indicates ligand dissociation being directly proportional to ionic concentration (a point levied in the SSII investigations), but still further investigation is warranted.

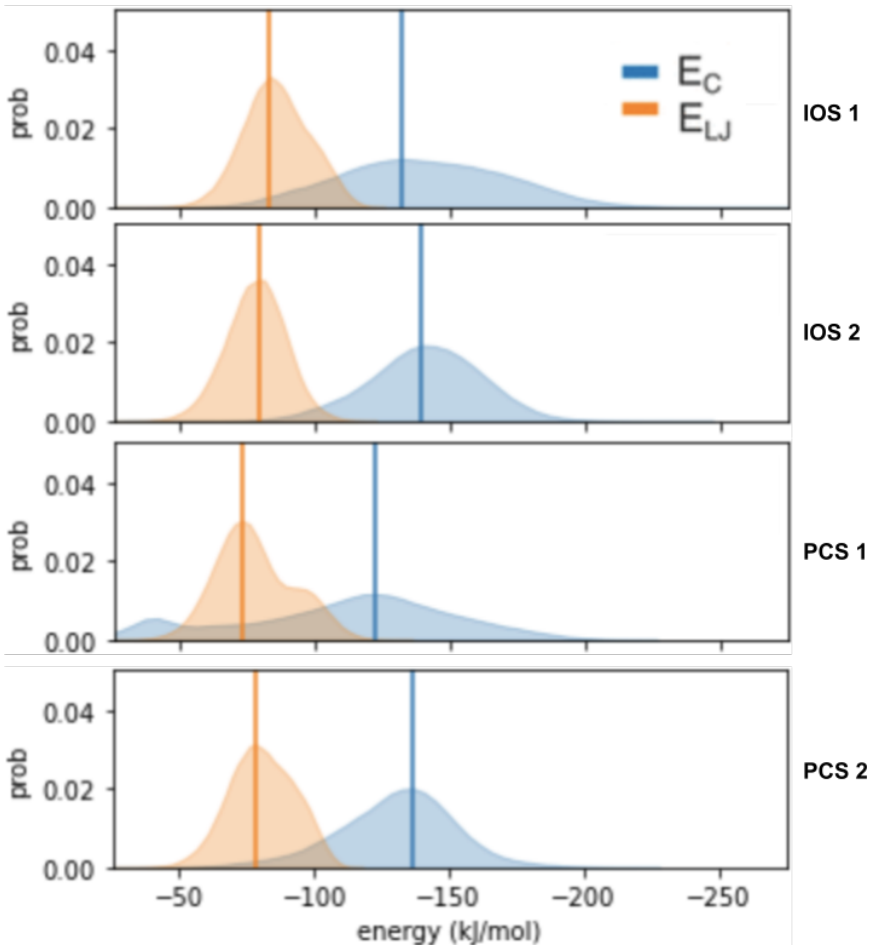

Figure 4. Protein-ligand Coulombic and Lennard-Jones contributions expressed as probability distributions, amalgamated over all three production runs. Average values are denoted with the solid lines.

Probability distributions were utilized above in Figure 4. for the sake of revealing some more of the subtleties involved in binding pocket dynamics than simple averages. As with the SSII investigations, the relatively diffuse shape of the Coulombic probability distribution indicates that there are more likely than not a multitude of relative energetic minima attributable to stable ligand conformations. Indeed, the Lennard-Jones contributions seem to model the opposite of

this inference in all toxins save PCS one, where there seems to be a minute secondary vertex shifted to the right of the global maximum–somewhere near -100 kJ/mol. It is curious as to what specific residual interactions might precipitate this peak, but due to its distinct character (especially in comparison to the shape of the Coulombic distributions) we might infer the presence of two discrete metastable conformations of PCS one as they pertain to hydrophobic interactions.

Moving along, we reexamine averaged protein-ligand energies at a resolution of individual residues. Based on the aforementioned contact number diagram, the top nine residues of interest were selected for each toxin, and analyzed via GROMACS' energy module, though this number was an incidental result of the actual selection being based on the top five residues of the atomic and hydrophilic contacts, accounting for repeats. Averaged Coulombic and Lennard-Jones contributions for each residue-ligand interaction are visualized below in Figure 5., with error bars representing a single standard of deviation. There is a particular virtue in delineating between electrostatic and hydrophobic interactions at the residual level, as this is the scale where we make the definitive assertion of what specific interactions constitute a given binding conformation, rather than the bulky statistical assessments that have preceded. This analysis should prove useful, too, in providing a cross-validation basis for the upcoming PoseView assessment of binding conformation.

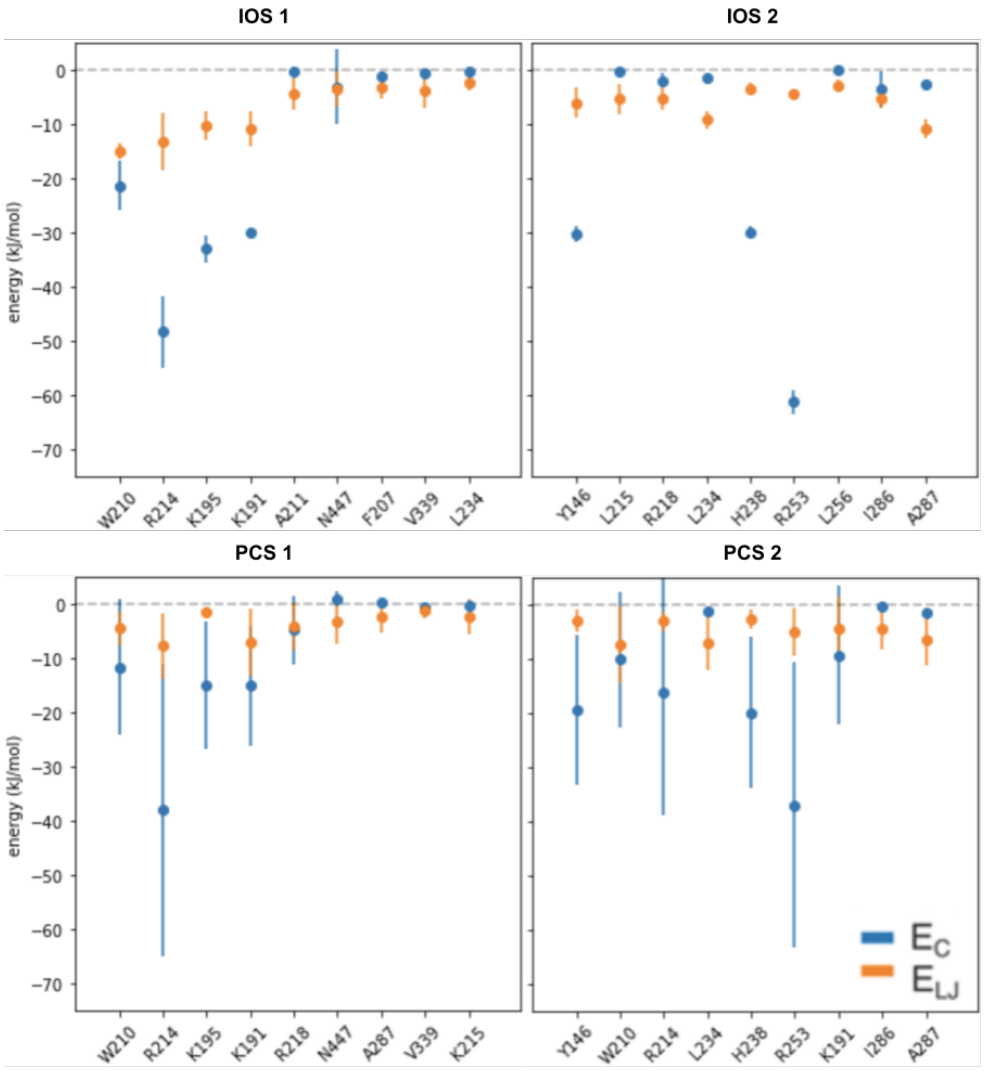

Figure 5. Coulombic and Lennard-Jones contributions of residual interactions with ligands, averaged over all production simulations. Error bars signify a standard of deviation in either direction from the average.

Excluding a small handful of exceptions, we witness a solid agreement with the findings presented in Table 1.; most residual Coulombic indices stand at a more attractive energetic value than those of the Lennard-Jones. Bar-none we witness the strongest interactions–Coulombic at that– with local arginine residues; this is in solid agreement with the SSII investigations, which

identified a salt bridge with residue R410 as the primary mediator of ligand stability in its respective binding pocket. IOS one in SSI (residue 559) is stabilized most strongly by residue R214, but is flanked by electrostatic contributions from K195 and K191, none of which, again, are Lennard-Jones contributing dominant. W210, although ranking fourth in terms of strongest interaction, does present a bit of an interesting finding on account of it being the highest average energy where Lennard-Jones and electrostatic interactions are relatively close in value. This is worth noting on account of the previously mentioned shared indole moiety, and we can infer, based on these findings, that there may be the established pi-pi interaction between W210 and this first indoxyl sulfate that is most nearly on-par with the more likely than the hydrogen bonding more strongly occurring. Obviously there are factors such as intermolecular distance which need to be considered before making a definitive assertion as to whether or not one interaction might take precedence over another, but still this is an interesting finding that falls in congruence with the expected outcome.

Regarding the second indoxyl sulfate, we note a dominant electrostatic contribution from the regional arginine, R253, with Y146 and H238's electrostatics coming in second and a close third, respectively. A salt bridge similar to the SSII and IOS one stabilization can be expected for R253's mode of interaction [7]. As far as the tyrosine is concerned we can surmise that the interaction energy is likely due to a hydrogen bond forming between the hydroxyl hydrogen of the phenol group, and the sulfate of the ligand; however there are the combinatorial possibilities of ligand acceptance of hydrogen bonding on the secondary amine of the indole, donation by the same amine coupled with the acceptance of said donation by the tyrosine's oxygen. Factoring phenomena such as steric hindrance into the picture however seems to indicate the first proposed

mode to be the most acceptable. Addressing the measurements of H238, we would almost expect a similar behavior to W210 from the first indoxyl sulfate, on account of the sidechain also possessing an aromatic ring. While we do note a nonzero attractive Lennard-Jones contribution, it is nowhere close to the value of the electrostatic contribution, which may be due to structural incongruities between its aromatic ring, and the ligand's.

Regarding the PCS interactions, we do note a handful of standout findings, along with some similarities to the IOS interactions from the preceding discussion. Again, as with IOS, the strongest interactions are formed with local arginines, with R214 and R253 ranking highest amongst toxins one and two, respectively. What is perhaps most interesting, and unique to PCS, is the number of shared residues in each of these interaction energy summaries. There are four total residues shared between PCS one and two, most of which rank somewhat high on the spectrum of interaction energy. IOS by comparison shares only one residue between its toxins, L234, and neither of the interaction energies formed with either ligand is particularly strong in comparison to the others. What is truly noteworthy about this finding though is how it reframes some of the information gathered in results previously mentioned, namely the net interactions determined for ligand-ligand interactions. IOS one and two exhibited a net attractive electrostatic interaction energy, whereas PCS was repulsive; one might wager that that if two ligands shared attractive interactions with a multitude of mutual residues like in the case of the latter that a net attractive force is to be expected, and a repulsive one if they were seemingly pulled in opposite directions like IOS. What the information seems to suggest instead however is that the residual attractive forces–which dwarf the ligand-ligand interactions in magnitude–bring the two p-cresyl sulfates in the same direction and therefore close proximity to one another, wherein a

countervailing electrostatic repulsive force arises to oppose and ultimately maintain a conformation that is energetically favorable.

Having expounded interaction energies, we now turn our attention to the corresponding residual distances. Distances between ligand center-of-masses and those of the nine aforementioned residues of interest were collected over the course of the 200ns trajectory, and subjected to principal component analysis in order to reduce the large dataset to a two dimensional conformational space. The amalgam of indices within this space was then used as an input for our unsupervised mean shift algorithm, which assigned modes based on prescribed clusters (as dictated by the silhouette score).

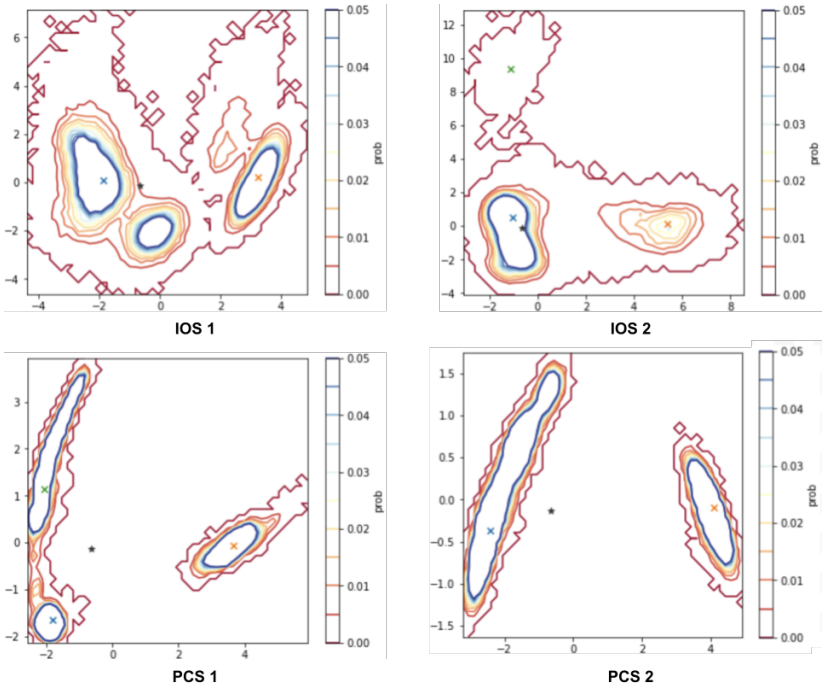

Figure 76. Probability density function of each SSI toxin based on binding mode clustering, plotted along the two orthogonal principal components. Multicolor ‘X’ symbols represent each mode index, whereas the star represents the experimental orientation from PDB 2BXH from which simulations were seeded.

Visualized above in Figure 6. are the results of the mean shift clustering, represented as a probability density ‘heat map’ within the principal component space. As we can see, there are a variety of metastable binding conformations able to be formed by each toxin, with two modes identified for IOS one and two, as well as PCS two, and three clusters for PCS one. Tabulated below are the eigenvalues for each toxin’s principal component, as well as the fractional distribution captured by each of the clusters. As far as noteworthy observations, we find a chief discrepancy separating the distributions of IOS and PCS being the overlap with the experimental structures from which their trajectories were seeded. Both IOS toxins share a solid overlap in their probability density functions with the frame belonging to the original IOS coordinates found in PDB file 2BXH; this is not only in good agreement with the SSII findings, but also augurs well for the reliability of the overall investigation in terms of not being wildly deviant from experimental findings. PCS, on the other hand, exhibits next to no overlap in this regard, which seems fair accounting for the fact that there is no X-Ray crystallographically-sourced structure for this compound, and that we had to improvise our own based on that of the IOS configurations. This is all to say that the case for PCS being a viable ligand for SSI is an ever-weakening one, which is supported by HNMR evidence suggesting that the toxin exhibits a proclivity not to the Sudlow Site, but to an unknown binding site, located elsewhere on HSA [32].

Table 2. Parameter summaries for PCA and mean-shift clustering analysis; eigenvalues corresponding to the primary components of each toxin, and cluster weights.

Toxin	Eigenvalue (PC 1)	Eigenvalue (PC 2)	Wt. (Clust. 1)	Wt. (Clust. 2)	Wt. (Clust. 3)
IOS 1	0.953	0.031	0.667	0.333	-
IOS 2	0.556	0.238	0.838	0.155	0.07
PCS 1	0.757	0.232	0.332	0.333	0.334
PCS 2	0.953	0.031	0.667	0.333	-

Moving along, we now must compare how each of the representative structures determined through our clustering analyses compare to the initial, experimentally determined structures. Visualized below in Figure 7., are the mode-sought structures, and even in light of the disparity hinted at above in Figure 6. in terms of the PCS toxins sharing little overlap with the experimental structure, there still exists a semblance of commonality between initial and optimized structures. The primary cluster structure of IOS one establishes a higher number of hydrophobic contacts in comparison to the experimental orientation; this could perhaps be indicative of a progression deeper towards the hydrophobic core of the binding pocket as was determined for SSII in Smith & Pfaendtner, however contact with residue K195 seems to obfuscate this assertion [7]. Intermolecular distance visualizations seem to indicate closer contact to polar sidechained residues in the first binding mode, namely R214, K195, and K19, but two hydrophobic residues also exhibit similarly close average center-of-mass distances, W210 and F207. Perhaps a stronger case could be made for IOS two's attraction to the hydrophobic core;

R253's electrostatic attraction still dominates, however interaction with the polar K195 is abandoned in favor of a hydrogen bond Y146's hydroxyl hydrogen. In terms of intermolecular distances, these values remain much more uniform in the case of IOS two, and considering that five of the nine residues of interest are hydrophobic, a stronger case might be made for this ligand's progression into a predominantly hydrophobic region of the binding pocket.

PCS exhibits a similar split in behavior amongst its two toxins. In the first PCS, more contacts with hydrophobic residues are indeed established in comparison to the experimentally resolved structure, however they are countervailed by the formation of contacts with different polar residues. Combined with the added unreliability of center-of-mass distances, and it becomes harder to make a cogent argument as to this particular ligand traversing further into the hydrophobic region of SSI. Perhaps Figure 7. suggests a distancing from the hydrophilic mouth of the binding pocket in that the interaction with the polar K191 is portrayed as one hydrophobic in nature, but again the presence of the electrostatic interaction with R214 makes this assessment murky. Regarding the second PCS, there is a similar argument as the second IOS in that a relatively stronger assertion can be made to this toxin's attraction towards the hydrophobic zone, but again the intermolecular distance data is too abstruse to say for sure. Additionally, another factor that must be considered is the aforementioned forces acting on the two PCS toxins in tandem; residual interactions seem to be pulling the two PCS together to such a degree that the two begin to repel each other in light of their close proximity. Depending on the direction of this attractive residual force, both PCS could be pulled toward or away from the hydrophobic region.

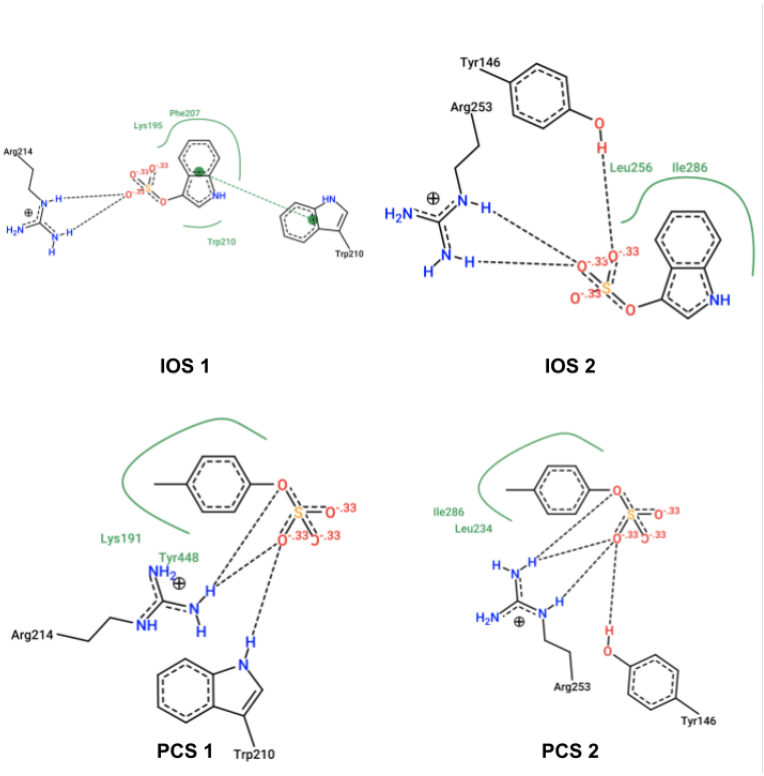

Figure 7. PoseView representations of the mode-sought conformations of each toxin within the SSI binding pocket. Hydrogen bonds are represented by dashed lines, whereas hydrophobic interactions are color-coded green.

4.CONCLUSION

In review, this investigation took an experimentally-resolved structure of HSA complexed with indoxyl sulfate, seeded two of our own virtually-designed toxins, and submitted the new structures to a series of equilibrative simulations, culminating in a 250 ns run. This trajectory was then taken and subjected to a preliminary analysis designed to identify which residues exhibited a propensity for intermolecular attraction between these four ligands, with which we

based all subsequent analyses on. Total interaction energy between protein and ligand was gathered over the abridged 200 ns trajectory, as well as the energies for the residues from the preceding step. Intermolecular distances between residues and ligands were gathered, and used for a machine learning-based analysis of most probably binding modes. All in all, this investigation was intended to serve as a demonstrable approach analyzing protein ligand dynamics in a virtual medium, and hopefully the garnered results would be worth cross referencing with experimental data.

In the ongoing campaign against chronic kidney disease and efforts to better understand the mechanism in which protein bound uremic toxins dissociate from their high affinity unions with human serum albumin, investigations such as these aim to make some form of contribution. Through rigorous analysis of our simulated output, we have unearthed a considerable amount of information pertaining to the particular dynamics associated with ligand stabilization in the Sudlow Site I binding pocket. From these findings, as with the SSII investigations from the past, a mechanism might be proposed to recreate these stabilizing conditions artificially, thus providing a means by which to liberate these toxins from their high-affinity complexes with human serum albumin.

BIBLIOGRAPHY

1. Holle, J.; Kirchner, M.; Okun, J.; Bayazit, A. K.; Obrycki, L.; Canpolat, N.; Bulut, I. K.; Azukaitis, K.; Duzova, A.; Ranchin, B.; Shroff, R.; Candan, C.; Oh, J.; Klaus, G.; Lugani, F.; Gimpel, C.; Büscher, R.; Yilmaz, A.; Baskin, E.; Erdogan, H.; Zaloszyc, A.; Özcelik, G.; Drozdz, D.; Jankauskiene, A.; Nobili, F.; Melk, A.; Querfeld, U.; Schaefer, F. Serum Indoxyl Sulfate Concentrations Associate with Progression of Chronic Kidney Disease in Children. *PLoS One* **2020**, *15* (10), e0240446. https://doi.org/10.1371/journal.pone.0240446.
2. He, Y.; Ning, T.; Xie, T.; Qiu, Q.; Zhang, L.; Sun, Y.; Jiang, D.; Fu, K.; Yin, F.; Zhang, W.; Shen, L.; Wang, H.; Li, J.; Lin, Q.; Sun, Y.; Li, H.; Zhu, Y.; Yang, D. Large-Scale Production of Functional Human Serum Albumin from Transgenic Rice Seeds. *Proceedings of the National Academy of Sciences* **2011**, *108* (47), 19078–19083. https://doi.org/10.1073/pnas.1109736108.
3. Pogostin, B. H.; Malmendal, A.; Londergan, C. H.; Åkerfeldt, K. S. PKa Determination of a Histidine Residue in a Short Peptide Using Raman Spectroscopy. *Molecules* **2019**, *24* (3), 405. https://doi.org/10.3390/molecules24030405.
4. Lim, Y. J.; Sidor, N. A.; Tonial, N. C.; Che, A.; Urquhart, B. L. Uremic Toxins in the Progression of Chronic Kidney Disease and Cardiovascular Disease: Mechanisms and Therapeutic Targets. *Toxins* **2021**, *13* (2), 142. https://doi.org/10.3390/toxins13020142.
5. Itoh, Y.; Ezawa, A.; Kikuchi, K.; Tsuruta, Y.; Niwa, T. Protein-Bound Uremic Toxins in Hemodialysis Patients Measured by Liquid Chromatography/Tandem Mass Spectrometry and Their Effects on Endothelial ROS Production. *Anal Bioanal Chem* **2012**, *403* (7), 1841–1850. https://doi.org/10.1007/s00216-012-5929-3.
6. Watanabe, H.; Noguchi, T.; Miyamoto, Y.; Kadowaki, D.; Kotani, S.; Nakajima, M.; Miyamura, S.; Ishima, Y.; Otagiri, M.; Maruyama, T. Interaction between Two Sulfate-Conjugated Uremic Toxins, p-Cresyl Sulfate and Indoxyl Sulfate, during Binding with Human Serum Albumin. *Drug Metab Dispos* **2012**, *40* (7), 1423–1428. https://doi.org/10.1124/dmd.112.045617.
7. Smith, J.; Pfaendtner, J. Elucidating the Molecular Interactions between Uremic Toxins and the Sudlow II Binding Site of Human Serum Albumin. *J. Phys. Chem. B* **2020**, *124* (19), 3922–3930. https://doi.org/10.1021/acs.jpcb.0c02015.
8. Li, S.; Tonelli, M.; Unsworth, L. D. Indoxyl and P-Cresol Sulfate Binding with Human Serum Albumin. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2022**, *635*, 128042. https://doi.org/10.1016/j.colsurfa.2021.128042.
9. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

10. Zardecki, C.; Dutta, S.; Goodsell, D. S.; Voigt, M.; Burley, S. K. RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. *J. Chem. Educ.* **2016**, *93* (3), 569–575. https://doi.org/10.1021/acs.jchemed.5b00404.
11. *Introduction to Protein Data Bank Format*. https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html (accessed 2022-08-15).
12. Chang, C.-E.; Huang, Y.-M.; Mueller, L.; You, W. Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools. *Catalysts* **2016**, *6*, 82. https://doi.org/10.3390/catal6060082.
13. Kumar, H.; Maiti, P. Introduction to Molecular Dynamics Simulation; 2011; pp 161–197. https://doi.org/10.1007/978-93-86279-50-7_6.
14. *Basics of Molecular Dynamics Simulations*; 2018.
15. Vermaas, J. V.; Hardy, D. J.; Stone, J. E.; Tajkhorshid, E.; Kohlmeyer, A. Topogromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *J Chem Inf Model* **2016**, *56* (6), 1112–1116. https://doi.org/10.1021/acs.jcim.6b00103.
16. Frenkel, D.; Smit, B. Understanding molecular simulation: From algorithms to applications; Academic Press: San Diego, 2002.
17. Gautam, B. *Energy Minimization*; IntechOpen, 2020. https://doi.org/10.5772/intechopen.94809.
18. McQuarrie, D. A.; Simon, J. D. *Physical Chemistry a molecular approach*; Viva Books: New Delhi u.a., 2015.
19. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Computer Physics Communications* **2009**, *180* (10), 1961–1972. https://doi.org/10.1016/j.cpc.2009.05.011.
20. Jolliffe, I. T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016**, *374* (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202.
21. Comaniciu, D.; Meer, P. Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Trans. Pattern Anal. Machine Intell.* **2002**, *24* (5), 603–619. https://doi.org/10.1109/34.1000236.
22. Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam,

M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

23. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38.
24. *H++ (web-based computational prediction of protonation states and pK of ionizable groups in macromolecules)*. http://newbiophysics.cs.vt.edu/H++/index.php (accessed 2022-08-15).
25. *Ramu Anandakrishnan, Boris Aguilar and Alexey V. Onufriev, "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation", Nucleic Acids Res., 40(W1):W537-541. (2012).*
26. *Myers J, Grothaus G, Narayanan S, Onufriev A, "A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules", Proteins, 63, 928-938 (2006).*
27. *Gordon JC, Myers JB, Folta T, Shoja V, Heath LS and Onufriev A., "H++: a server for estimating pKas and adding missing hydrogens to macromolecules", Nucleic Acids Res. Jul 1;33:W368-71. (2005).*
28. D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman (2022), Amber 2022, University of California, San Francisco.
29. Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons Learned from Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *J Comput Aided Mol Des* **2017**, *31* (1), 147–161. https://doi.org/10.1007/s10822-016-9977-1.
30. Bauer, P.; Hess, B.; Lindahl, E. GROMACS 2022.2 Manual. **2022**. https://doi.org/10.5281/zenodo.6637572.
31. Riwar, L.-J.; Trapp, N.; Kuhn, B.; Diederich, F. Substituent Effects in Parallel-Displaced π–π Stacking Interactions: Distance Matters. *Angewandte Chemie International Edition* **2017**, *56* (37), 11252–11257. https://doi.org/10.1002/anie.201703744.
32. Li, S.; Tonelli, M.; Unsworth, L. D. Indoxyl and P-Cresol Sulfate Binding with Human Serum Albumin. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2022**, *635*, 128042. https://doi.org/10.1016/j.colsurfa.2021.128042.