

# Binding modes of peptidomimetics designed to inhibit STAT3

Ankur Dhanik<sup>1</sup>, John S McMurray<sup>2</sup>, Lydia E Kavrakı<sup>1\*</sup>

**1 Department of Computer Science, Rice University, Houston, Texas, USA**

**2 Department of Experimental Therapeutics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA**

\* E-mail: kavrakı@rice.edu

## Abstract

STAT3 is a transcription factor that has been found to be constitutively activated in a number of human cancers. Dimerization of STAT3 via its SH2 domain and the subsequent translocation of the dimer to the nucleus leads to transcription of anti-apoptotic genes. Prevention of the dimerization is thus an attractive strategy for inhibiting the activity of STAT3. Phosphotyrosine-based peptidomimetic inhibitors, that mimic pTyr-Xaa-Yaa-Gln motif and have strong to weak binding affinities, have been previously investigated. It is well-known that structures of protein-inhibitor complexes are important for understanding the binding interactions and designing stronger inhibitors. Experimental structures of inhibitors bound to the SH2 domain of STAT3 are, however, unavailable. In this paper we describe a computational study that combined molecular docking and molecular dynamics to model structures of 12 peptidomimetic inhibitors bound to the SH2 domain of STAT3. A detailed analysis of the modeled structures was performed to evaluate the characteristics of the binding interactions. We also estimated the binding affinities of the inhibitors by combining MMPB/GBSA-based energies and entropic cost of binding. The estimated affinities correlate strongly with the experimentally obtained affinities. Modeling results show binding modes that are consistent with limited previous modeling studies on binding interactions involving the SH2 domain and phosphotyrosine(pTyr)-based inhibitors. We also discovered a stable novel binding mode that involves deformation of two loops of the SH2 domain that subsequently bury the C-terminal end of one of the stronger inhibitors. The novel binding mode could prove useful for developing more potent inhibitors aimed at preventing dimerization of cancer target protein STAT3.

## Introduction

Development of effective therapeutics is the ultimate goal of cancer research [1–6], but it is a time-consuming and expensive process [7–10]. Structure-based computational techniques [11,12] such as virtual screening [13–15], docking [16,17], and molecular dynamics [18,19] have proven useful in the development of drugs. Even if there have not been many successful drug discovery stories based on computation alone, the use of structure-based computational techniques has helped gain better understanding of how a putative drug compound binds to its target receptor, and has reduced the drug development time and costs [20–22]. In this paper, we discuss computational modeling of binding interactions between a specific set of peptidomimetic inhibitors [23–26] and the Src-homology 2 (SH2) domain of STAT3 or Signal Transducer and Activator of Transcription 3 [27] (Figure 1). STAT3 is constitutively activated in a number of human cancer types such as lung cancer, breast cancer, multiple myeloma, and others [28–30]. The Jak-STAT pathway [31,32] describes the mechanism of action that leads to the transcription of anti-apoptotic genes. Upon extracellular signaling, a series of phosphorylations of cell surface receptors and Janus kinases (JAKs) inside the cell results in the phosphorylation of STAT3. A phosphorylated STAT3 then forms a dimer via its SH2 domain and the dimer translocates to the nucleus where it is involved in the transcription process.

Our focus in this work is on 12 peptidomimetic [23–26](mimic pTyr-Xaa-Yaa-Gln motif) inhibitors that target the SH2 domain of STAT3 with the aim of preventing the dimerization of STAT3, and subsequent translocation and transcription. The experimental structures of the peptidomimetics bound

to the SH2 domain are unavailable. However, the experimental binding affinities, which measure the thermodynamic stability of binding interactions between the peptidomimetics and the SH2 domain, have been derived using fluorescence polarization [33]. Our goal is to computationally model the binding modes which define how a conformation of a peptidomimetic binds to the conformation of the SH2 domain, analyze the binding interactions, estimate the binding affinities, and calculate the correlation between the estimated and the experimental binding affinities.

Our computational modeling approach combines molecular docking and molecular dynamics and derives inspiration from previous work [18, 24, 34–39]. Given a protein and an unbound ligand, molecular docking computes the preferred conformation and location of the ligand in the binding pocket of the protein. Many molecular docking programs exist (see representative examples [40–48]) and several docking studies have been performed with varied amount of success (e.g., [49–57]). Three major limitations however remain.

- A docking program typically computes the best conformation and placement of the ligand such that it minimizes an energy function specific to the docking program. The energy function approximates the free energy of binding and, in general, accuracy of the binding energy is sacrificed so that the computation of energy can be performed in minimal time. The approximate energy functions, therefore, may result in conformations that are not accurate [58–60].
- Most docking programs treat the protein as a rigid molecule or, at the very best, a molecule with limited flexibility. Thus, most of these programs perform what is known as flexible ligand docking to a rigid receptor. However, it is well known that more accurate modeling of binding interactions between a ligand and a receptor requires accounting for the flexibility of the receptor [61, 62].
- Docking of small ligands with 5 or 6 rotatable bonds is fairly accurate and computationally fast. However, docking of large ligands with many rotatable bonds, such as the peptidomimetic inhibitors in our dataset, is inaccurate and computationally expensive. A large number of rotatable bonds increases the dimensionality of the conformation space of the ligand which makes searching for the docked conformation extremely challenging and time-consuming [57, 63, 64].

Our modeling approach addresses the above limitations in the following way. Docking of a peptidomimetic is first done with an AutoDock [44, 65]-based incremental docking protocol that we have developed recently [66]. A molecular dynamics simulation of the docked conformation of the peptidomimetic in complex with the SH2 domain is then performed. Using molecular dynamics, we are able to treat both the ligand and the receptor as flexible and, more importantly, we analyze deformations in the structure of the complex in a simulated solvent environment. The physics-based force field used in the molecular dynamics simulations is more detailed and accurate as compared to the energy functions used in molecular docking. Molecular dynamics simulation thus also lends itself to calculation of more accurate binding affinity estimates [67, 68].

Using our modeling approach, we show that we were able to obtain various binding modes for the peptidomimetics. Not only did we obtain previously proposed binding modes [24, 69], but we also obtained a novel binding mode. The estimated binding affinities and the experimental binding affinities are well correlated which validates our modeling approach. By using the estimated binding affinities and conformational analysis of the molecular dynamics trajectories, we are able to differentiate between strong and weak binders. In the following section, we present details of our peptidomimetic dataset, explain our computational modeling approach, binding affinity calculations, and data analysis techniques. This is followed by a description of the results from the computational modeling of the peptidomimetics in complex with the SH2 domain. Finally we conclude with an overall discussion of our work.

## Methods

### Dataset

The 12 inhibitors used in this study were obtained from a series of 142 peptidomimetic compounds [23–26]. These 142 peptidomimetics mimic pTyr-Xaa-Yaa-Gln motif and were developed to bind to the SH2 domain with the purpose of inhibiting the activity of STAT3. The binding affinities (measured as  $IC_{50}$  values) of the 142 peptidomimetics were evaluated using fluorescence polarization [33]. The  $IC_{50}$  value gives the concentration of the peptidomimetic that is required to competitively inhibit the binding of FAM-Ala-pTyr-Leu-Pro-Gln-Thr-Val-NH<sub>2</sub> (FAM=5-carboxyfluorescein) to Stat3 by 50% [23]. The binding affinities of the 142 peptidomimetics were found to range from weak ( $IC_{50} = 100,000nM$ ) to strong ( $IC_{50} = 39nM$ ).

The molecular dynamics simulation, which is part of our modeling approach, and the method used for the estimation of binding affinities (see sections below) are computationally expensive. Therefore, we limited our modeling study to 12 peptidomimetics. The 12 peptidomimetics used in this study were chosen such that they represent a range of values of the experimental binding affinities as shown in Figure 2 and also represent a range of sizes varying from 9 torsional degrees of freedom to 22 torsional degrees of freedom. Each peptidomimetic was named such that the compound number represents the order in which the peptidomimetic appears in the original publications [23–26] where the 142 peptidomimetics were first described.

The structure of STAT3 was obtained from the Protein Data Bank [70] (PDB ID 1BG1). The structure contains residues 136 to 716 of Stat3, half a DNA duplex, and 127 water molecules per asymmetric unit [27]. Since we are interested in the modeling of the peptidomimetics bound to the SH2 domain, the structure of the SH2 domain corresponding to residues 585 to 688 (Figure 1) was isolated. The water molecules and the DNA duplex were ignored. Using the molecular builder of the Maestro software [71] (version 9.1), the 2-D chemical representations of the 12 peptidomimetics (Figure 2) were converted to 3-D structures of the unbound peptidomimetics.

### Modeling Approach

Our two-step computational modeling approach combined molecular docking and molecular dynamics. Molecular docking of a large ligand such as a peptidomimetic with many rotatable bonds is challenging. A large ligand spans a high-dimensional conformation space which makes exploration of docked conformation of the ligand challenging. Our recently developed Autodock-based incremental docking protocol has been shown to improve docking of large ligands [66]. Therefore, we first docked the 12 peptidomimetic inhibitors in our dataset to the SH2 domain of STAT3 using our incremental docking protocol [66], and subsequently performed molecular dynamics simulations of the docked conformations of the peptidomimetics in complex with the SH2 domain.

Starting from a fragment of the ligand, at each incremental step, our docking protocol explores a few rotatable bonds, then selects a small number of best partially docked fragments, grows the fragments by adding few more rotatable bonds and atoms, and docks again. The dock-select-grow-dock process is repeated until all the rotatable bonds in the ligand are explored. AutoDock [44, 65] is used in each step to explore only a few rotatable bonds and this makes the docking operation fast and accurate.

Each peptidomimetic in our dataset was docked to the SH2 domain of STAT3 using our incremental docking protocol. Since the phosphate group of the pTyr residue in each peptidomimetic is known to bind to the sub-pocket formed by residues Lys591, Arg609, Ser611, Glu612, and Ser613 [24, 69], at each incremental docking step we selected conformations with the lowest values of scoring function  $S$ , where

$$S = 0.25(P_d) + S_{AD}, \quad (1)$$

$P_d$  is the squared distance of the phosphorus atom (in pTyr) from coordinates  $(-8.42, 4.50, -6.09)$  that

represent approximate center of the sub-pocket, and  $S_{AD}$  is the binding affinity estimated by AutoDock’s energy function. The scoring function  $S$ , thus, penalizes large distance between the phosphate group and the sub-pocket. The details of the incremental docking of peptidomimetics and scoring function  $S$  are available in the Supporting Information (Section S1).

After all the rotatable bonds in a peptidomimetic were explored and all its atoms were docked, the docked conformation of the peptidomimetic with the lowest value of  $S$  (see, equation (1)) was selected. For each peptidomimetic in our dataset, molecular dynamics simulation of the selected docked conformation, in complex with the SH2 domain of STAT3, was performed. The *sander* module in the AMBER11 software package [72] was used for the simulation. The peptidomimetic inhibitor was described with generalized amber force field [73] (GAFF), and point charges were calculated for the atoms using *antechamber* module (from AmberTools software package [74] version 1.5) and AM1-BCC charge model. The protein was described with AMBER’s ff99SB force field. The complex was solvated in a 15Å octahedral box of TIP3P water and the whole system was neutralized by adding  $\text{Na}^+$  counterions according to the net charge of the peptidomimetic. Table S1 lists the number of atoms in each of the 12 molecular dynamics systems.

The complex was first minimized using 100 cycles of steepest descent minimization followed by 1900 cycles of conjugate gradient minimization. This was followed by 50ps of temperature equilibration where the temperature was raised from 100K to 300K using Berendsen [75] control with coupling parameter set to 2ps. Pressure equilibration was then performed for 200ps using Berendsen control with pressure relaxation time set to 2ps. Finally, a production simulation of 10ns was performed at constant temperature and pressure, and the trajectory was output at every 10ps. During the molecular dynamics simulation, SHAKE algorithm was used to constrain bonds involving hydrogen atoms and therefore forces for the bonds involving hydrogen atoms were not calculated. For computing electrostatic energies, Particle Mesh Ewald [76] (PME) method was used with the non-bonded cutoff set to the default value of 8Å. Plots showing variation of system properties (total energy, potential energy, temperature, and pressure) during the production simulation are available in the Supporting Information (Figures S1-S12) and reveal equilibrated and stable systems.

To evaluate the accuracy of our modeling approach, we performed a study where we compared the structures modeled using our approach with experimentally-derived structures. Since experimentally determined structures of the peptidomimetics in complex with the SH2 domain of STAT3 or any other protein from the STAT family are unavailable, the validation was done using a dataset of similar complexes derived from the PDBbind database [77]. The details and analysis of the validation study are available in the Supporting Information (Section S2). The analysis (Figures S13-S22, Table S2) shows that the modeled structures and experimental structures are spatially close and therefore we conclude that our modeling approach is well-suited for modeling of peptidomimetic-SH2 complexes that are described in this paper.

## Binding affinities

Trajectories obtained from molecular dynamics simulations were also used to estimate binding affinities. The binding affinity of each peptidomimetic in complex with the SH2 domain was obtained using Molecular Mechanics Poisson Boltzmann (or Generalized Born) Surface Area (MMPB/GBSA) calculations [78]. The binding affinity is given by

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{PB/GB} + \Delta G_{SA} - T\Delta S \quad (2)$$

where,  $T$  is the temperature,  $\Delta E_{MM}$  represents the molecular mechanics energy,  $\Delta G_{PB/GB}$  represents the polar part of solvation energy,  $\Delta G_{SA}$  represents the non-polar part of the solvation energy, and  $T\Delta S$  represents the energetic penalty due to loss of entropy upon binding.

To compute  $\Delta G_{bind}$ , values of  $\Delta E_{MM}$ ,  $\Delta G_{PB/GB}$ , and  $\Delta G_{SA}$  were computed averaged over the snapshots of the molecular dynamics production trajectory. The  $T\Delta S$  value was computed from a

normal mode analysis of the system using *nmode* module of the AmberTools [74] package. For entropy computation, every 10<sup>th</sup> snapshot of the molecular dynamics trajectory was used. Calculations of all of the above  $\Delta G_{bind}$  components were done using MMPBSA script provided by the AmberTools package. For computing the polar part of solvation energy using Poisson-Boltzmann calculations, the ionic strength was set to 0.1 mM. All other parameters needed by MMPBSA script were kept at their default values.

## Data analysis

Trajectory data obtained after the 10ns molecular dynamics simulation was analyzed in a variety of ways. Prior to analyzing the data, however, the following processing was done. Water and counterions were removed from the trajectory data. All atoms were moved such that the center of mass of the complex moved to the center of the simulation box and imaging was done to bring all atoms inside the primary unit cell. All conformations of the complex contained in the snapshots of the trajectory were then fitted to the conformation in the first frame of the production simulation. Mass-weighted root mean squared distance (RMSD) fitting was done. Note that since we output trajectory at every 10ps, we obtained 1000 snapshots or conformations of the complex from a 10ns molecular dynamics simulation.

Average (mass-weighted) root mean square fluctuations (RMSF) were computed for each peptidomimetic bound to the SH2 domain. The RMSF value represents the average value of the RMSD between the peptidomimetic conformation in the first frame of the molecular dynamics trajectory and the conformations in the subsequent frames. Thus, the RMSF value is indicative of the time-averaged fluctuation of the peptidomimetic conformation. Clustering of conformations of the peptidomimetic was done and conformations that are representative of the clusters were identified. Clustering was done using k-means [79] (k was set to 5) algorithm with RMSD as the similarity metric. Hydrogen bonds are critical for stabilizing the binding interactions [80–83] and were identified between each peptidomimetic and the SH2 domain. If a hydrogen bond was present in less than 50% of the conformations in the trajectory, it was ignored. For each peptidomimetic in complex with the SH2 domain, we computed the hydrogen bond occupancy of the residues of the SH2 domain. Hydrogen bond occupancy of a residue is defined as the fraction of conformations in the molecular dynamics trajectory that contain at least one hydrogen bond involving that particular residue. Computation of RMSF values and k-means clustering was done using *ptraj* module from the AmberTools [74] package. Hydrogen bonds were identified using *hbond* tool in the Chimera software package [84] version 1.6.

## Results

### Conformational analysis

Figure 3 shows the best docked conformation, of each of the 12 peptidomimetics, computed using the incremental docking protocol. These docked conformations were then solvated and subjected to 10ns of molecular dynamics simulations. Snapshots of the trajectories were output at every 10ps and therefore we obtained 1000 conformations for each of the 12 peptidomimetic-SH2 domain complexes. The RMSF value for each peptidomimetic is shown in Figure 4. The RMSF value quantifies the average spatial fluctuation of the peptidomimetic conformation in the 1000 snapshots. A low RMSF value is thus indicative of spatial stability of the conformation of the peptidomimetic bound to the SH2 domain. The RMSF values for weak binders such as comp13, comp15, and comp60 are higher as compared to the RMSF values of the strong binders such as comp70, comp121, comp134, comp135, and comp136. As an exception, comp140, another strong binder, shows surprisingly large RMSF value (1.82Å) that is comparable to the RMSF values of the weak affinity peptidomimetics.

Through clustering of the 1000 conformations, we obtained 5 representative conformations of each peptidomimetic-SH2 complex (Figure 5). All representative conformations have the phosphate group of

the pTyr residue or its surrogate in the location of the corresponding pTyr705 in the crystal structure of STAT3 [27]. The representative conformations of the strong binders such as comp70, comp121, comp134, comp135, and comp136 are spatially similar, while those of weak binders such as comp13, comp15, and comp60 show more spatial variation.

## Hydrogen bonds

Hydrogen bonds are critical to the binding interactions between the peptidomimetics and STAT3. Figure 6 shows all the residues that are involved in hydrogen bonds with at least one peptidomimetic. Residues Lys591, Arg609, Ser611, Glu612, and Ser613 are involved in the hydrogen bond interactions and form the phosphate-binding pocket (sub-pocket-1) where the phosphate group of pTyr residue (or its surrogate) in each peptidomimetic binds. Three sub-pockets in the binding site of the SH2 domain are also involved in hydrogen-bonding interactions. Residues Glu638, Tyr640, and Gln644 form sub-pocket-2, residues Gly656, Lys658, and Tyr657 form sub-pocket-3, and residues Trp623, Ile659, and Met660 flank sub-pocket-4.

The residues of the SH2 domain which participate in hydrogen bonds with a specific peptidomimetic and the hydrogen bond occupancy involving those residues and the peptidomimetic are shown in Figure 7. The occupancy plots in Figure 7 also show that the strong binders such as comp70, comp121, comp134, and comp136 form hydrogen bonds with more than 5 residues of the SH2 domain, and the weak binders such as comp13 and comp15 form hydrogen bonds with 3 and 4 residues respectively. Another weak-affinity peptidomimetic comp60 forms hydrogen bonds with 6 different residues but all of these residues surround the phosphate-binding pocket (sub-pocket-1). This means that, in the case of comp60, while the pTyr residue binds tightly to the sub-pocket-1, the rest of the peptidomimetic is not involved in stable hydrogen bond interactions. A couple of strong binders, comp135 and comp140, form hydrogen bonds with 4 residues each. Since the conformation of comp140 is unstable (as evident by the RMSF value) and we ignore hydrogen bonds if they are present in less than 50% of the conformations in the molecular dynamics trajectory, hydrogen bond interactions with fewer residues of the SH2 domain is expected. In the case of comp135, however, the RMSF value is low (1.22Å). We postulate that comp135 may have an alternate and more stable bound conformation similar to the conformation of comp134.

## Binding affinity

The binding affinity value reflects the thermodynamic stability of the binding interactions between a peptidomimetic and the SH2 domain of STAT3. In a computational modeling study such as this, a large positive correlation between the experimental binding affinities and estimated binding affinities is desired. A high correlation allows accurate prediction of strong and weak binders. We used binding energy function described by equation (2) to estimate the binding affinity values in four different schemes: A. entropic component ( $T\Delta S$ ) was ignored and the non-entropic component was computed using MMGBSA, B. entropic component ( $T\Delta S$ ) was ignored and the non-entropic component was computed using MMPBSA, C. entropic component ( $T\Delta S$ ) was included and the non-entropic component was computed using MMGBSA, and D. entropic component ( $T\Delta S$ ) was included and the non-entropic component was computed using MMPBSA. The experimental binding affinities were calculated from the  $IC_{50}$  values using the function

$$\Delta G_{exp} = R \times T \times \log K_i \quad (3)$$

where,  $K_i = \log \frac{IC_{50} \times 10^{-9}}{1.066}$  mol, R (Gas constant) = 0.001986 kcal K<sup>-1</sup>mol<sup>-1</sup>, T = 298K, and  $IC_{50}$  values are in nM. The Cheng-Prusoff estimation [85] of  $K_i$  from  $IC_{50}$  was done using a  $K_d$  of 150nM for FAM-Ala-pTyr-Leu-Pro-Gln-Thr-Val-NH<sub>2</sub> [86] and a concentration of 10nM in the fluorescence polarization assay used to evaluate the binding affinities of the 12 peptidomimetics [23].

Figure 8 (top) plots Pearson’s correlation coefficient ( $R$ ), that measures the correlation between the experimental and estimated binding affinities, versus the length (2ns, 4ns, 6ns, 8ns, and 10ns) of molecular dynamics simulation. Note that the binding affinities are computed averaged over the snapshots of the molecular dynamics simulation. From the figure, it is clear that, for all four affinity estimation schemes, the value of  $R$  increases with the increase in the length of molecular dynamics simulation. Out of the four schemes, the best correlation coefficient values were observed for the scheme D (cyan) which estimates affinity as a sum of the entropic component and the MMPBSA-based non-entropic component of the energy function. The maximum observed value of  $R$  is 0.63 which was computed using scheme D and 10ns molecular dynamics simulation trajectories.

The estimated binding affinities, for the 12 peptidomimetics, obtained using scheme D are shown in Figure 8 (bottom). For each peptidomimetic, multiple values of the binding affinities that correspond to different lengths of molecular dynamics simulation are shown. It is clear that the affinity values converge as the length of simulation increases. The affinity values, derived from the 10 ns molecular dynamics trajectories, correspond to the  $R$  value of 0.63 as described above. Since the  $R$  value is large, not surprisingly, weak binders such as comp13, comp15, and comp60 have higher estimated affinity values, the value for comp15 ( $IC_{50} > 100,000\text{nM}$ ) being the highest (-18.22 kcal/mol). Similarly, the binding affinity values for strong binders such as comp70, comp121, comp134, and comp136 are low, the value for comp70 ( $IC_{50} = 190\text{nM}$ ) being the lowest (-45.40 kcal/mol).

## Binding modes

The conformations in Figure 5 show the presence of two binding modes that have also been described in previous computational modeling studies [24, 69]: the bent mode and the extended mode. All or some of the representative conformations for comp70, comp135, comp140, and comp142 display the bent mode where the phosphate group sits in sub-pocket 1 and the peptidomimetic bends such that the Gln (or its derivative) residue of the peptidomimetic sits in sub-pocket 2. In the extended mode, as seen in all or some of the representative conformations for comp134, comp136, comp140, and comp142, the phosphate group sits in sub-pocket 1 and the backbone extends such that Gln (or its derivative) residue of the peptidomimetic sits in sub-pocket 3. Apart from the bent and the extended modes, a novel binding mode was observed. The five representative conformations of comp121 display what we term a wedged mode. In this mode, while the phosphate group binds to the sub-pocket-1, the other end of the peptidomimetic is wedged in a groove formed by two loops of the SH2 domain described by residues 623-629 and residues 656-668.

The binding modes are shown in detail in Figures 9, 10, and 11. Both cartoon and surface representations of the SH2 domain are shown. The labeled yellow residues of the SH2 domain are involved in hydrogen bond interactions and the hydrogen bonds are shown with dashed black lines. The surface of the SH2 domain is colored using the Coulombic surface coloring scheme in the Chimera software package. The surface is characterized by electrostatic potentials ranging from positive electrostatic potential (blue surface) to a negative potential (red surface). The bent mode is displayed by the peptidomimetic comp70 (Figure 9), the extended mode is displayed by comp134 (Figure 10), and the wedged mode is displayed by comp121 (Figure 11). The binding affinities of the three peptidomimetics, experimental as well as computed, are high (low  $\Delta G$  and  $IC_{50}$  values) and, as shown in Figure 4, the RMSF values for comp70 (0.98Å), comp134 (0.95Å), and comp121 (0.91Å) are the lowest out of the RMSF values for the 12 peptidomimetics. Thus, these three compounds present a strong evidence that there are three possible modes in which peptidomimetics can tightly bind to the SH2 domain.

As expected, all three binding modes include multiple hydrogen bonds connecting the phosphate group to sub-pocket-1. The amino acids forming sub-pocket-1 create a strong positive electrostatic potential which thus tightly binds the negatively charged phosphate group in all peptidomimetics. In the bent mode (Figure 9), the Gln residue of comp70 binds to the sub-pocket-2 and forms multiple hydrogen bonds with residues Tyr640 and Gln644 of the SH2 domain that flank sub-pocket-2. The binding interactions are

also stabilized by the hydrogen bonds formed between the carbonyl oxygen of the Haic group and residue Tyr657 of the SH2 domain. A similar interaction was observed between a carbonyl oxygen of pTyr-Asp-Lys-Pro-His and Tyr651 in the crystal structure of STAT1 [87]. In the extended mode (Figure 10), the carbonyl oxygen of the Leu at pTyr+1 position forms hydrogen bond with Tyr657 and the side chain amide group of the Gln-mimic residue at the C-terminus of the peptidomimetic forms hydrogen bonds with the main chain C=O of Gly656 and the backbone NH groups of Lys658 and Ile659. In the newly discovered wedged mode (Figure 11), the carbonyl oxygen of the Leu residue forms hydrogen bond with Trp623 which lies on the loop formed by residues 623-629 and the carbonyl oxygen of methanoproline is involved in a hydrogen bond with the side chain OH of Tyr657. The driving force for this binding mode appears to be hydrophobic contact between the C-terminal benzene ring and residues of loops 623-629 and 656-658 as well as a hydrogen bond between the benzylamide NH and the main chain C=O of Met660. Interestingly, the side chain amide group of Gln does not appear to interact directly with the protein.

## Discussion

Transcription factor STAT3 is an important target protein that is involved in a multitude of human cancers. In this work, we focused on a specific set of 12 peptidomimetic compounds that mimic the pTyr-Xaa-Yaa-Gln recognition motif and were designed to bind with the SH2 domain of STAT3 and prevent its dimerization which is a critical event leading up to the transcription of anti-apoptotic genes. Experimental binding affinities of the peptidomimetics were measured using fluorescence polarization and a range of affinity values were observed for the 12 peptidomimetics. Binding affinities for the peptidomimetics, expressed as IC<sub>50</sub> values, range from 39nM for a strong binder to over 100,000nM for a weak binder. Since experimental structures of the complexes formed between the peptidomimetics and the SH2 domain are unavailable, we used a computational strategy to model the complexes.

Our modeling strategy proceeded in two steps. In the first, we generated docked conformations of the peptidomimetics using a computational AutoDock-based incremental docking protocol that was developed by us for docking large compounds in a fast and accurate manner [66]. The peptidomimetics in our dataset are all large compounds with the number of rotatable bonds ranging from 9 to 22. In the second step of our modeling strategy, we selected the best docked conformation and then ran molecular dynamics simulations of the complex in a solvated box. Molecular dynamics simulations served multiple purposes. The flexibility of the SH2 domain was taken into account, fluctuations of the bound conformations over the length of molecular dynamics simulation were computed, and finally, rigorous estimates of binding affinities, as a sum of normal-mode analysis based entropic component and MMPB/GBSA based non-entropic component, were computed. Accurate estimates of binding affinities are very important for differentiating strong binders from weak binders, and therefore, a positive correlation between the experimental binding affinities and estimated binding affinities is desired. Our two-step modeling strategy resulted in a high positive correlation ( $R=0.63$ ) between the experimental and estimated affinities.

For each of the 12 peptidomimetics, we performed molecular dynamics simulations for a production length of 10ns. The trajectory data for each simulation was output at 10ps. Thus, we obtained 1000 conformations for each peptidomimetic in complex with the SH2 domain. The average fluctuation of the conformations of each peptidomimetic was measured as RMSF (root mean square fluctuation) values. The weak binders displayed larger fluctuation as compared to the strong binders. A clustering of the conformations showed the preferred binding modes of the peptidomimetics. Three strong binders, with IC<sub>50</sub> values equal to 190nM (comp70), 83nM (comp134), and 68nM (comp121), displayed three different but stable binding modes: the bent mode, the extended mode, and the wedged mode respectively. The peptidomimetics in these three binding modes showed very small ( $< 1.0\text{\AA}$ ) conformational fluctuations in the molecular dynamics simulations, a large number of stable hydrogen bond interactions with the residues of the SH2 domain, and the estimated binding affinities value were low in accordance with the

experimental binding affinities.

Previous modeling studies related to SH2 domain binding have proposed the bent and the extended binding modes [24, 69]. In this paper, we propose a new binding mode which we term the wedged mode. In the wedged mode, the peptidomimetic (comp121) binds to the SH2 domain such that the negatively charged phosphate group of the pTyr residue sits inside a pocket which has a positive electrostatic potential and the C-terminal benzyl group gets wedged in a cavity formed by two loops of the SH2 domain described by the residues 623-629 and 656-668 respectively. Apart from the stable hydrogen bond interactions with the residues in the phosphate-binding pocket, hydrogen bonds also exist between the peptidomimetic and residues on the two loops. The RMSF value for the 1000 conformations of the comp121 is 0.91Å and is the lowest among the RMSF values for the 12 peptidomimetics.

Despite the overall success of modeling strategy as described in this paper, there were exceptions to the observed trends. For example, in the case of comp140 which is a relatively strong binder ( $IC_{50} = 105\text{nM}$ ), we obtained a large RMSF value and estimated binding affinities that are comparable to those of weak binders. This anomaly could be attributed to an inaccurate starting docked conformation of the peptidomimetic. In the molecular dynamics simulation, an inaccurate starting docked conformation would result in trajectory that leads to inaccurate estimation of binding affinity. It should be noted that computational docking of large ligands such as peptidomimetics in our dataset is extremely challenging. Although our incremental docking protocol improves docking of large ligands, more work needs to be done in this area.

The computational modeling strategy described in this paper and the subsequent data analysis, nonetheless, reveals important aspects of the peptidomimetic binding to the SH2 domain of STAT3. Not only were we able to estimate binding affinities that were well correlated with experimental binding affinities, we were also able to identify binding modes, including a novel wedge mode, that result in stable binding interactions. A typical peptidomimetic drug design process that is based on a specific motif involves designing peptidomimetics with diverse chemical modifications. Accurate estimation of binding affinities using our method could help in predicting which modifications could lead to strong binding. The knowledge gained by this study could also be used to improve the design of the peptidomimetics by better targeting the sub-binding-pockets identified in this paper with structural modifications or conformational restraints. The proposed novel wedge binding mode could prove very useful in this regard.

## Acknowledgments

This work has been supported in part by a training fellowship from the Keck Center Computational Cancer Biology Training Program of the Gulf Coast Consortia (CPRIT RP101489), The Texas Higher Education Coordinating Board (NHARP 01907), The National Cancer Institute (RO1 CA096652), The John and Ann Doerr Fund for Computational Biomedicine at Rice University, and the Cyberinfrastructure for Computational Research funded by National Science Foundation under Grant CNS-0821727.

## References

1. Shay JW, Wright WE (2002) Telomerase: A target for cancer therapeutics. *Cancer Cell* 2: 257-265.
2. Isaccs JS, Xu W, Neckers L (2003) Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell* 3: 213-217.
3. Khalil IG, Hill C (2005) Systems biology for cancer. *Current Opinion in Oncology* 17: 44-48.
4. Reichert JM, Valge-Archer VE (2007) Development trends for monoclonal antibody cancer. *Nature Reviews Drug Discovery* 6: 349-356.

5. Moasser MM (2007) Targeting the function of the her2 oncogene in human cancer therapeutics. *Oncogene* 26: 6577-6592.
6. Bhasin D, Cisek K, Pandharkar T, Regan N, Li C, et al. (2008) Design, synthesis, and studies of small molecule stat3 inhibitors. *Bioorganic & Medicinal Chemistry Letters* 18: 391 - 395.
7. DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *Journal of Health Economics* 22: 151 - 185.
8. Projan SJ (2007) Why is big pharma getting out of antibacterial drug discovery? *Nature Reviews Drug Discovery* 6: 115-120.
9. Kamb A, Wee S, Lengauer C (2007) Why is cancer drug discovery so difficult? *Nature Reviews Drug Discovery* 6: 115-120.
10. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D (2011) The cost of drug development: A systematic review. *Health Policy* 100: 4-17.
11. Anderson AC (2003) The process of structure-based drug design. *Chemistry & Biology* 10: 787-797.
12. Kuntz IW (1992) Structure-based strategies for drug design and discovery. *Science* 257: 1078-1082.
13. Cavasotto CN, W Orry AJ (2007) Ligand docking and structure-based virtual screening in drug discovery. *Current Topics in Medicinal Chemistry* 7: 1006-1014.
14. Matter H, Sotriffer C (2011) Applications and success stories in virtual screening, Wiley-VCH Verlag GmbH & Co. KGaA. pp. 319-358. doi:10.1002/9783527633326.ch12.
15. Xing L, McDonald JJ, Kolodziej SA, Kurumbail RG, Williams JM, et al. (2011) Discovery of potent inhibitors of soluble epoxide hydrolase by combinatorial library design and structure-based virtual screening. *Journal of Medicinal Chemistry* 54: 1211-1222.
16. Shoichet BK, Kuntz ID (1991) Protein docking and complementarity. *Journal of Molecular Biology* 221: 327 - 346.
17. Kitchen DB, Decornez H, Furr JR, Bajorathe J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* 3: 935-949.
18. Mangoni M, Roccatano D, Di Nola A (1999) Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins: Structure, Function, and Bioinformatics* 35: 153-162.
19. Karplus M, McCammon A (2002) Molecular dynamics simulations of biomolecules. *Nature Structural Biology* 9: 646-652.
20. Terstappen GC, Reggiani A (2001) In silico research in drug discovery. *Trends in Pharmacological Sciences* 22: 23 - 26.
21. Kubinyi H (2006) Success Stories of Computer-Aided Design, John Wiley & Sons, Inc. pp. 377-424. doi:10.1002/0470037237.ch16.
22. Schneider G (2010) Virtual screening: an endless staircase? *Nature Reviews Drug Discovery* 9: 273-276.

23. Coleman DR, Ren Z, Mandal PK, Cameron AG, Dyer Ga, et al. (2005) Investigation of the binding determinants of phosphopeptides targeted to the SRC homology 2 domain of the signal transducer and activator of transcription 3. Development of a high-affinity peptide inhibitor. *Journal of medicinal chemistry* 48: 6661–6670.
24. Mandal PK, Limbrick D, Coleman DR, Dyer Ga, Ren Z, et al. (2009) Conformationally constrained peptidomimetic inhibitors of signal transducer and activator of transcription 3: Evaluation and molecular modeling. *Journal of medicinal chemistry* 52: 2429–2442.
25. Mandal PK, Ren Z, Chen X, Xiong C, McMurray JS (2009) Structure-affinity relationships of glutamine mimics incorporated into phosphopeptides targeted to the SH2 domain of signal transducer and activator of transcription 3. *Journal of medicinal chemistry* 52: 6126–6141.
26. Mandal PK, Gao F, Lu Z, Ren Z, Ramesh R, et al. (2011) Potent and selective phosphopeptide mimetic prodrugs targeted to the Src homology 2 (SH2) domain of signal transducer and activator of transcription 3. *Journal of medicinal chemistry* 54: 3549–63.
27. Becker S, Groner B, Müller CW (1998) Three-dimensional structure of the Stat3 $\beta$  homodimer bound to DNA. *Nature* 394: 145–51.
28. Bowman T, Garcia R, Turkson J, Jove R (2000) Stats in oncogenesis. *Oncogene* 19: 2474–2488.
29. Buettner R, Mora LB, Jove R (2002) Activated stat signaling in human tumors provides novel molecular targets for therapeutic intervention. *Clinical Cancer Research* 8: 945–954.
30. Hua Y, Jove R (2004) The stats of cancer - new molecular targets come of age. *Nature Reviews Cancer* 4: 97–105.
31. Darnell J, Kerr I, Stark G (1994) Jak-stat pathways and transcriptional activation in response to ifns and other extracellular signaling proteins. *Science* 264: 1415–1421.
32. Levy DE, Lee Ck (2002) What does Stat3 do? *Journal of Clinical Investigation* 109: 1143–1148.
33. Owicki JC (2000) Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *Journal of Biomolecular Engineering* 5: 297–306.
34. García-Nieto R, Pérez C, Gago F (2000) Automated docking and molecular dynamics simulations of nimesulide in the cyclooxygenase active site of human prostaglandin-endoperoxide synthase-2 (cox-2). *Journal of Computer-Aided Molecular Design* 14: 147–160.
35. Okimoto N, Futatsugi N, Fuji H, Suenaga A, Morimoto G, et al. (2009) High-performance drug discovery: Computational screening by combining docking and molecular dynamics simulations. *PLoS Computational Biology* 5: e1000528.
36. Xu X, Kasembeli MM, Jiang X, Tweardy BJ, Tweardy DJ (2009) Chemical probes that competitively and selectively inhibit stat3 activation. *PLoS ONE* 4: e4783.
37. Kaushik S, Sowdhamini R (2011) Structural analysis of prolyl oligopeptidases using molecular docking and dynamics: insights into conformational changes and ligand binding. *PLoS ONE* 6: e26251.
38. Park I, Li C (2011) Characterization of molecular recognition of STAT3 SH2 domain inhibitors through molecular simulation. *Journal of Molecular Recognition* 24: 254–265.

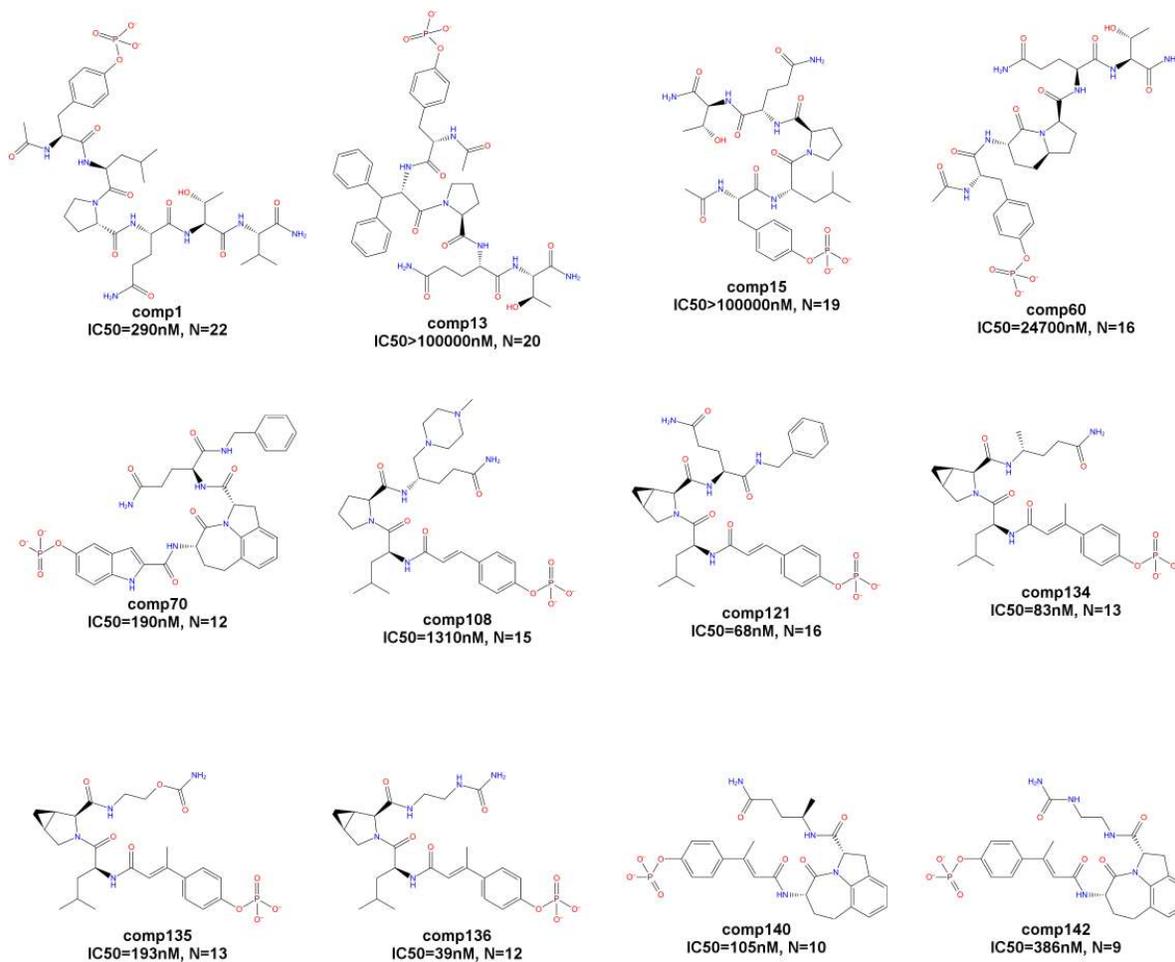
39. Raveh B, London N, Zimmerman L, Schueler-Furman O (2011) Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their Receptors. *PLoS ONE* 6: e18934.
40. Kuntz I (1982) A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* 161: 269–288.
41. Abagyan R, Totrov M, Kuznetsov D (1994) ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry* 15: 488–506.
42. Jones G, Willett P, Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology* 245: 43–53.
43. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* 261: 470–489.
44. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* 19: 1639–1662.
45. Jain AN (2000) Morphological similarity: a 3D molecular similarity method correlated with protein-ligand recognition. *Journal of Computer-aided Molecular Design* 14: 199–213.
46. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* 47: 1739–1749.
47. Plewczynski D (2010) VoteDock: consensus docking method for prediction of protein-ligand interactions. *Journal of Computational Chemistry* 32: 1–14.
48. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31: 455–461.
49. Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Structure, Function, and Bioinformatics* 57: 225–242.
50. Chang MW, Lindstrom W, Olson AJ, Belew RK (2007) Analysis of hiv wild-type and mutant structures via in silico docking against diverse ligand libraries. *Journal of Chemical Information and Modeling* 47: 1258–1262.
51. Onodera K, Satou K, Hirota H (2007) Evaluations of molecular docking programs for virtual screening. *Journal of Chemical Information and Modeling* 47: 1609–1618.
52. Rogers JP, Beuscher AE, Flajolet M, McAvoy T, Nairn AC, et al. (2006) Discovery of protein phosphatase 2c inhibitors by virtual screening. *Journal of Medicinal Chemistry* 49: 1658–1667.
53. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, et al. (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *Journal of Chemical Information and Modeling* 49: 1455–1474.
54. Chang MW, Ayeni C, Breuer S, Torbett BE (2010) Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PloS ONE* 5: e10926.

55. Li X, Li Y, Cheng T, Liu Z, Wang R (2010) Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *Journal of Computational Chemistry* 31: 2109–2125.
56. Li N, Thompson S, Schultz DC, Zhu W, Jiang H, et al. (2010) Discovery of selective inhibitors against EBNA1 via high throughput in silico virtual screening. *PLoS ONE* 5: e10126.
57. Plewczynski D, Lziniewski M, Augustyniak R, Ginalska K (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *Journal of Computational Chemistry* 32: 742–755.
58. Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. *Journal of Medicinal Chemistry* 44: 1035–1042.
59. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, et al. (2006) A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* 49: 5912–5931.
60. Dhanik A, Kavraki LE (2012) Protein-ligand interactions: computational docking, John Wiley Sons, Ltd. doi:10.1002/9780470015902.a0004105.pub2.
61. Cavasotto CN, Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *Journal of Molecular Biology* 337: 209 - 225.
62. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *Journal of Medicinal Chemistry* 47: 45–55.
63. Hetényi C, van der Spoel D (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Science* 11: 1729–1737.
64. Dhanik A, McMurray JS, Kavraki L (2011) On modeling peptidomimetics in complex with the SH2 domain of Stat3. In: *Proceedings of 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 3229–3332.
65. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, et al. (2009) AutoDock4 and AutoDock-Tools4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30: 2785–2791.
66. Dhanik A, McMurray JS, Kavraki L (2012) AutoDock-based incremental docking protocol to improve docking of large ligands. In: *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*. pp. 48–55.
67. Massova I, Kollman P (2000) Combined molecular mechanical and continuum solvent approach (mm-pbsa/gbsa) to predict ligand binding. *Perspectives in Drug Discovery and Design* 18: 113–135.
68. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) Validation and use of the mm-pbsa approach for drug discovery. *Journal of Medicinal Chemistry* 48: 4040–4048.
69. McMurray JS (2007) Structural basis for the binding of high affinity phosphopeptides to stat3. *Biopolymers* 90: 69–79.
70. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Research* 28: 235–242.

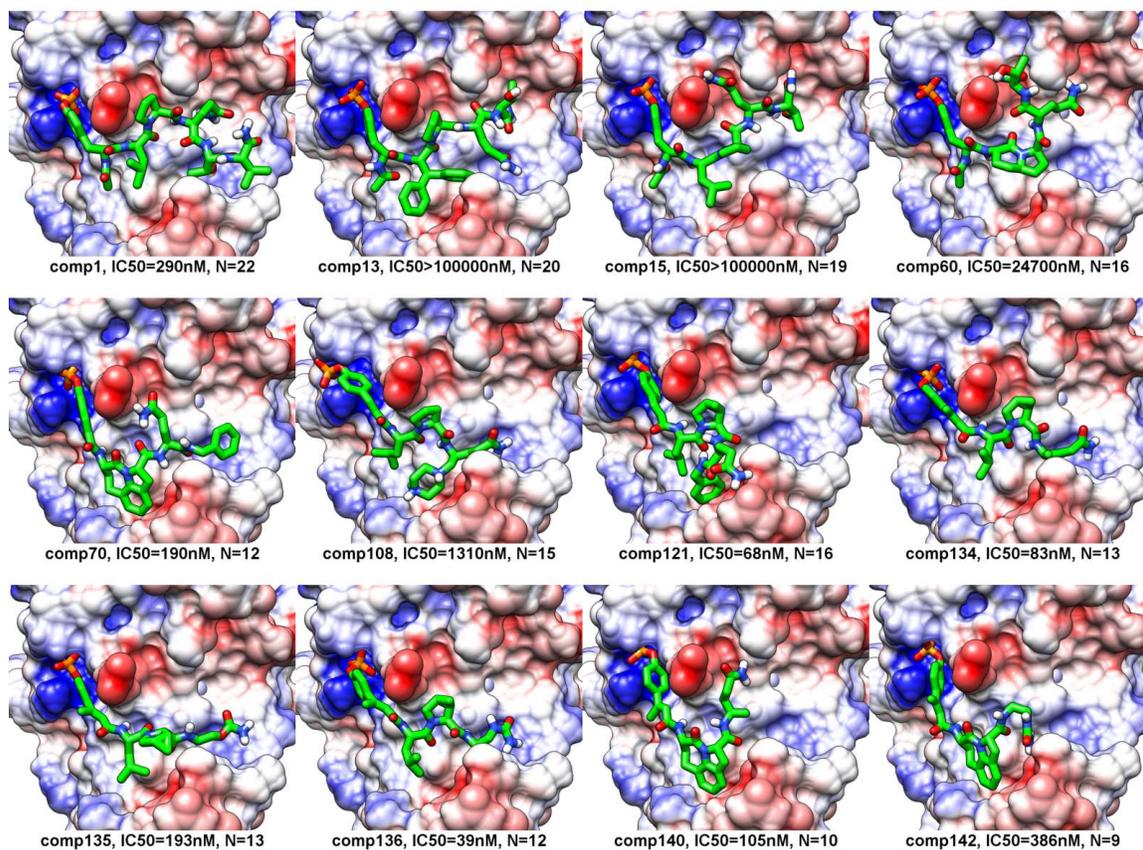
71. URL <http://www.schrodinger.com/products/14/12/>.
72. Case DA, Darden TA, Cheatham III TE, Simmerling C, Wang J, et al. (2010) Amber 11. University of California, San Francisco .
73. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry* 25: 1157–1174.
74. Ambergtools. <http://ambermd.org/>.
75. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81: 3684–3690.
76. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 98: 10089–10092.
77. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind database: methodologies and updates. *Journal of Medicinal Chemistry* 48: 4111–4119.
78. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, et al. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research* 33: 889–897.
79. Hartigan JA, Wong MA (1979) Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 28: pp. 100–108.
80. Weber P, Ohlendorf D, Wendoloski J, Salemme F (1989) Structural origins of high-affinity biotin binding to streptavidin. *Science* 243: 85–88.
81. Wade RC, Goodford PJ (1993) Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. ligand probe groups with the ability to form more than two hydrogen bonds. *Journal of Medicinal Chemistry* 36: 148–156.
82. Pierce AC, Sandretto KL, Bemis GW (2002) Kinase inhibitors and the case for ch...o hydrogen bonds in protein-ligand binding. *Proteins: Structure, Function, and Bioinformatics* 49: 567–576.
83. Panigrahi SK, Desiraju GR (2007) Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins: Structure, Function, and Bioinformatics* 67: 128–141.
84. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) Ucsf chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25: 1605–1612.
85. Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $IC_{50}$ ) of an enzymatic reaction. *Biochemical Pharmacology* 22: 3099–3108.
86. Schust J, Berg T (2004) A high-throughput fluorescence polarization assay for signal transducer and activator of transcription 3. *Analytical Biochemistry* 330: 114 - 118.
87. Mao X, Ren Z, Parker GN, Sondermann H, Pastorello MA, et al. (2005) Structural bases of unphosphorylated stat1 association and receptor binding. *Molecular Cell* 17: 761 - 771.



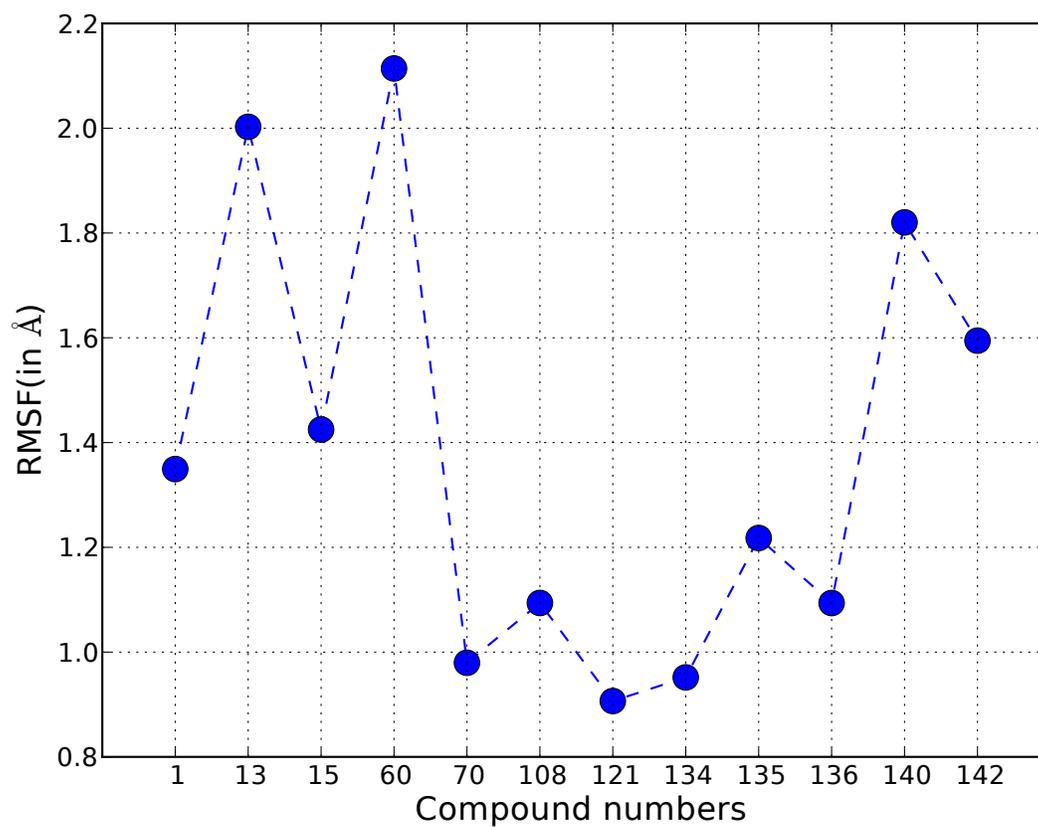
**Figure 1. STAT3 structure.** Protein Data Bank (ID 1BG1) structure of STAT3 is shown. The structure has four domains: a N-terminal four-helix bundle (in blue, residues 138-320), an eight-stranded  $\beta$ -barrel (in purple, residues 321-465), an  $\alpha$ -helical connector domain (in green, residues 466-585), and a SH2 domain (in yellow, residues 586 to 688).



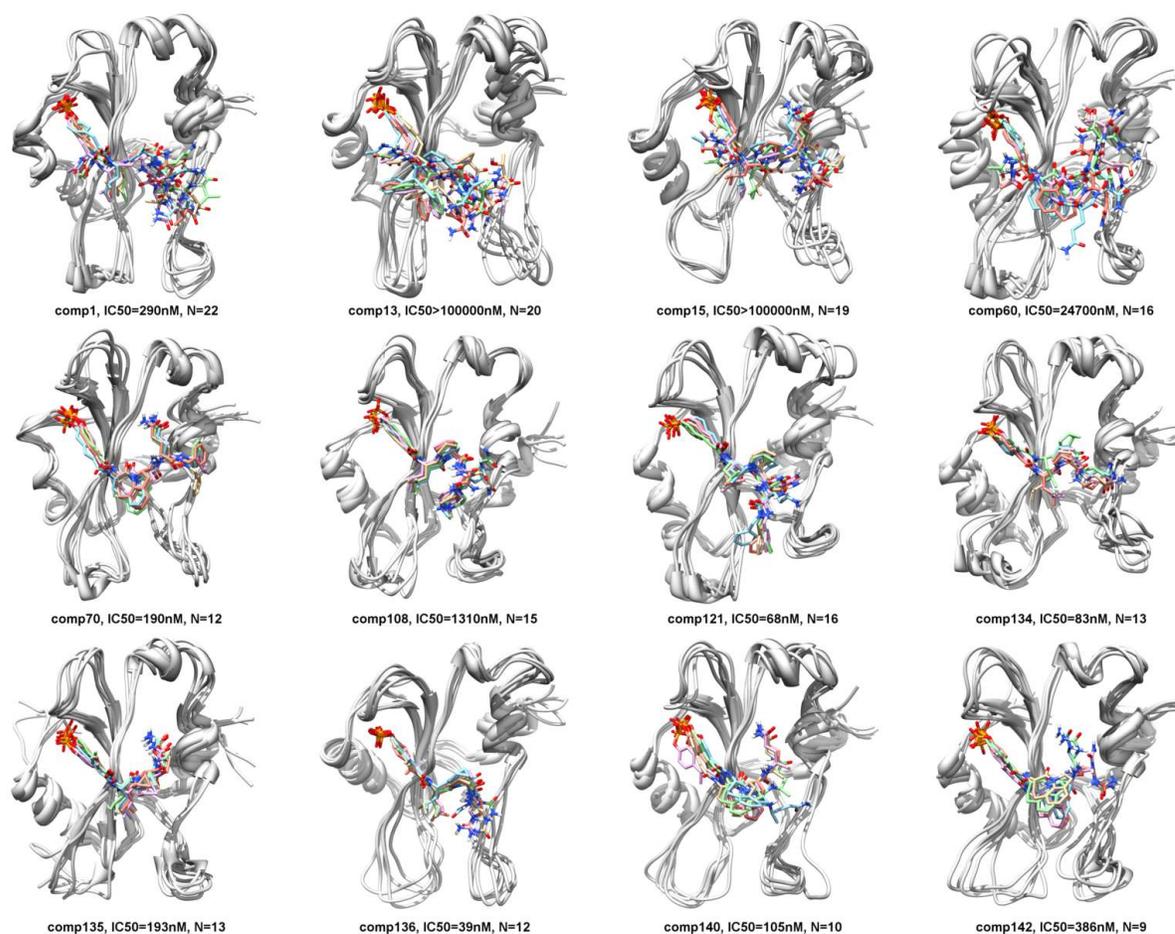
**Figure 2. 12 peptidomimetics.** 2-D chemical representations of the 12 peptidomimetics that form our dataset are shown.  $IC_{50}$  value represents the experimental binding affinity of each peptidomimetic derived using fluorescence polarization and N represents the number of rotatable bonds in each peptidomimetic.



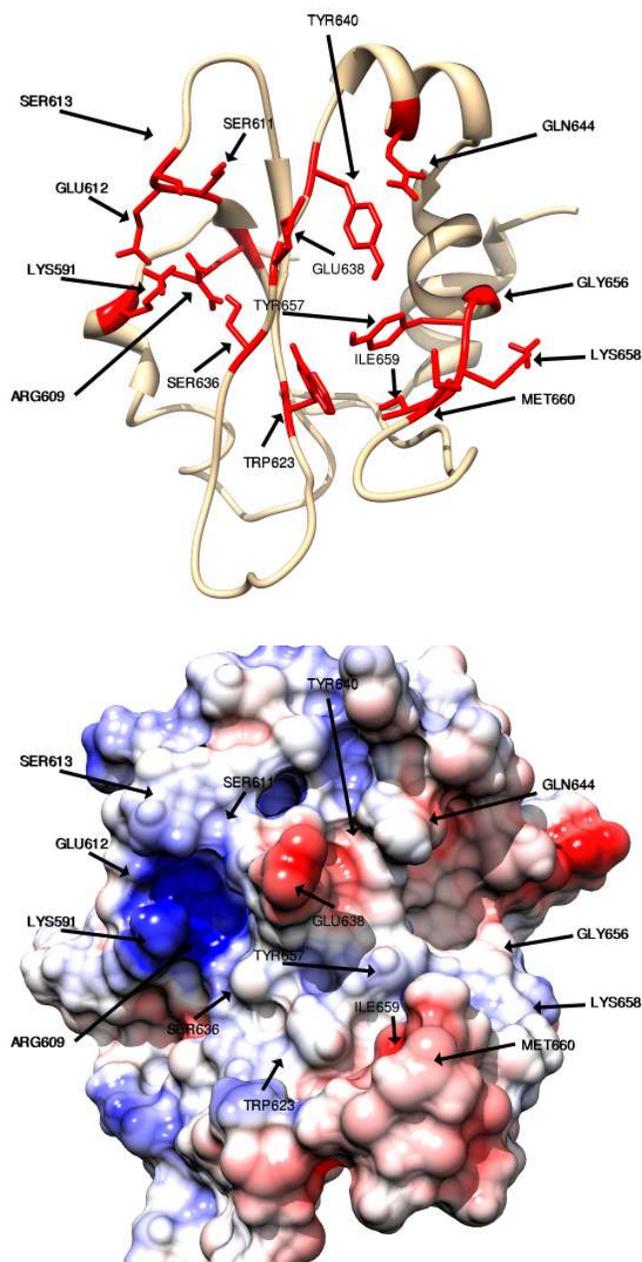
**Figure 3. Docked conformations.** Docked conformation of each peptidomimetic that was obtained using our incremental docking protocol is shown. The peptidomimetic conformation (in green) is shown in complex with the SH2 domain of STAT3 (in surface representation). The surface coloring shows the Coulombic electrostatic potential in different regions of the surface of the SH2 domain. The potential ranges from positive (in blue) to negative (in red). IC<sub>50</sub> value represents the experimental binding affinity of each peptidomimetic derived using fluorescence polarization and N represents the number of rotatable bonds in each peptidomimetic.



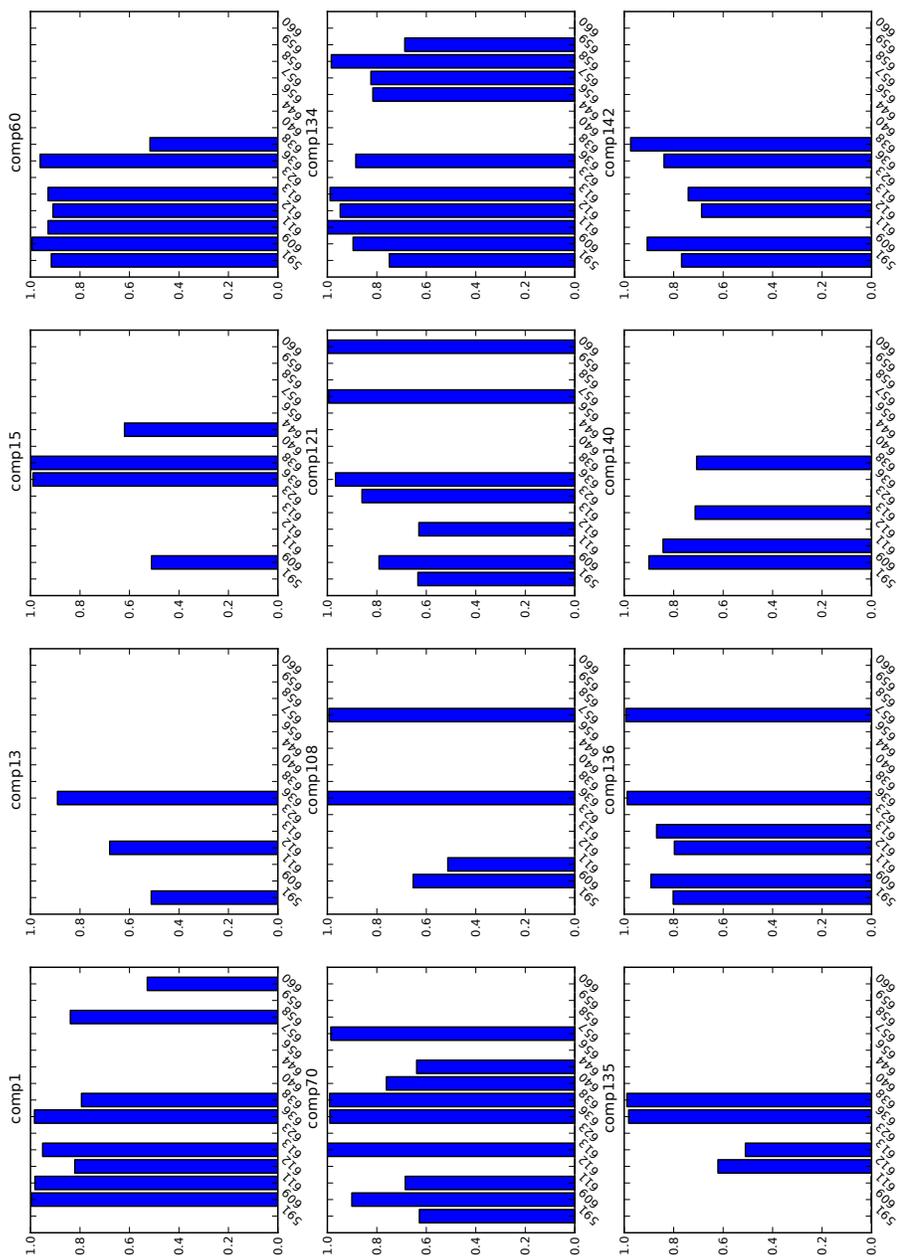
**Figure 4. Root mean square fluctuations.** Root mean square fluctuation (RMSF) of the 12 peptidomimetics in complex with the SH2 domain of STAT3 is shown. Each RMSF value was computed using 1000 conformations of the peptidomimetic derived from the 10ns molecular dynamics trajectory.



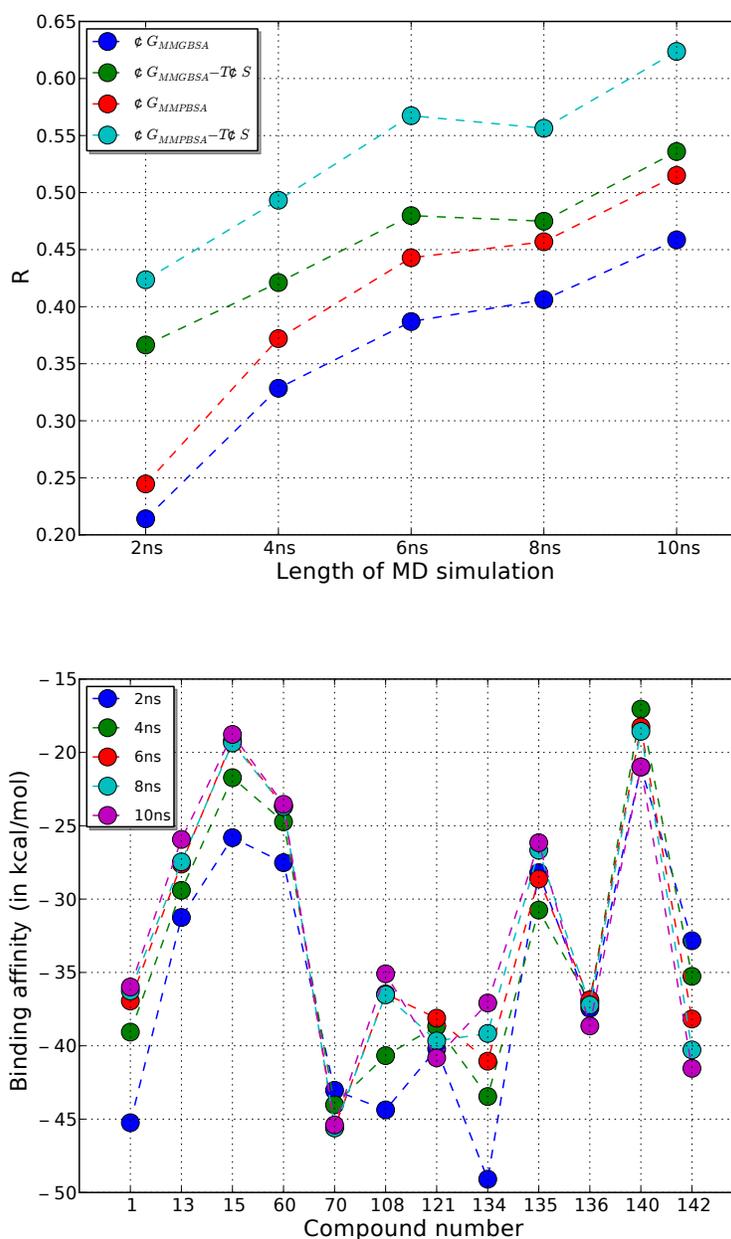
**Figure 5. Representative conformations after clustering.** For each peptidomimetic, 5 conformations in complex with the SH2 domain (in gray) of STAT3 are shown. The 5 conformations are the representatives of the 5 clusters obtained after k-means clustering of the 1000 conformations that were derived from 10ns molecular dynamics trajectory of each peptidomimetic-SH2 domain complex. IC<sub>50</sub> value represents the experimental binding affinity of each peptidomimetic derived using fluorescence polarization and N represents the number of rotatable bonds in each peptidomimetic.



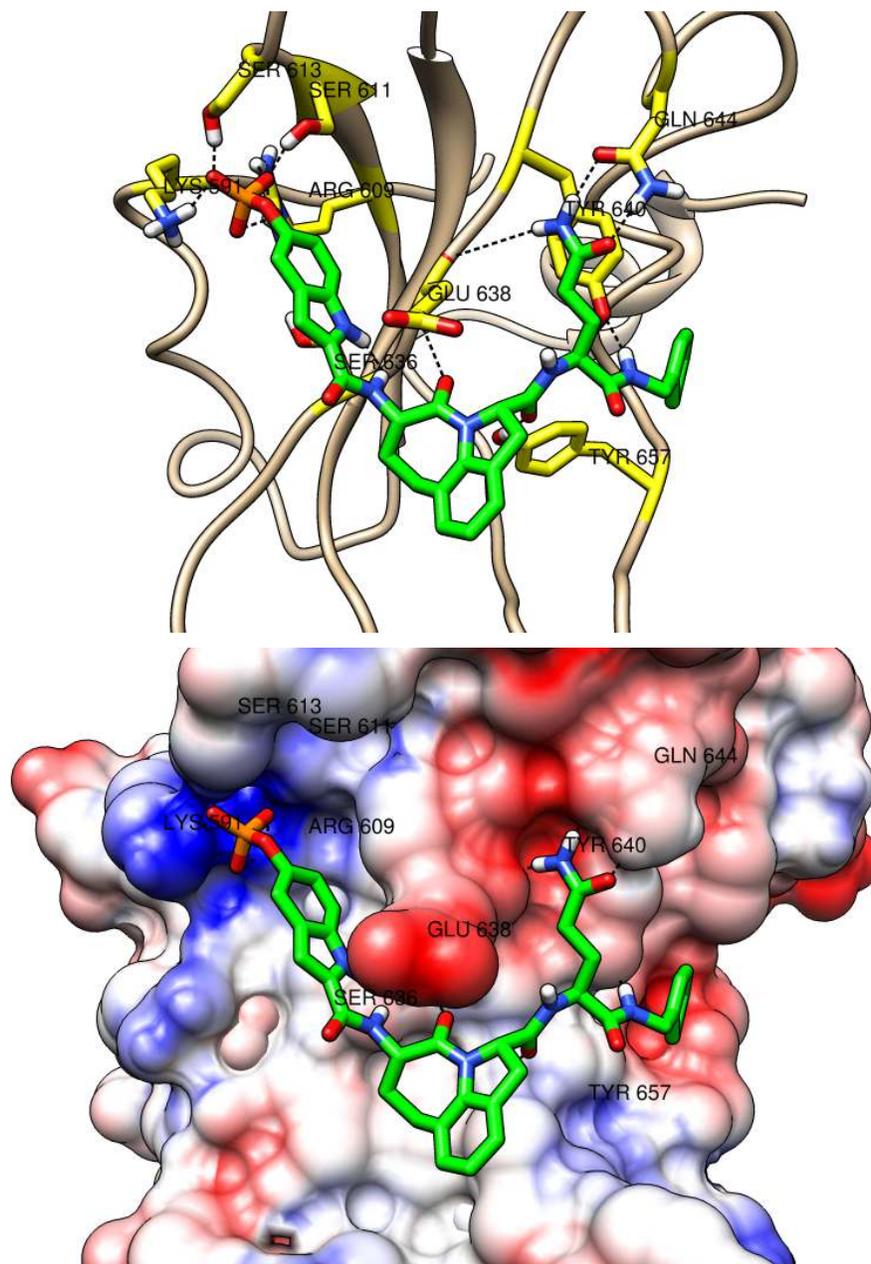
**Figure 6. Residues involved in hydrogen bonds.** The residues (labeled) of the SH2 domain that form hydrogen bonds with at least one of the 12 peptidomimetics are shown. The top figure shows a cartoon representation of the SH2 domain and the bottom figure shows a surface representation. The surface coloring shows the Coulombic electrostatic potential in different regions of the surface of the SH2 domain. The potential ranges from positive (in blue) to negative (in red). Note that a hydrogen bond is ignored if it is present in less than 50% of the conformations in the 10ns molecular dynamics trajectory.



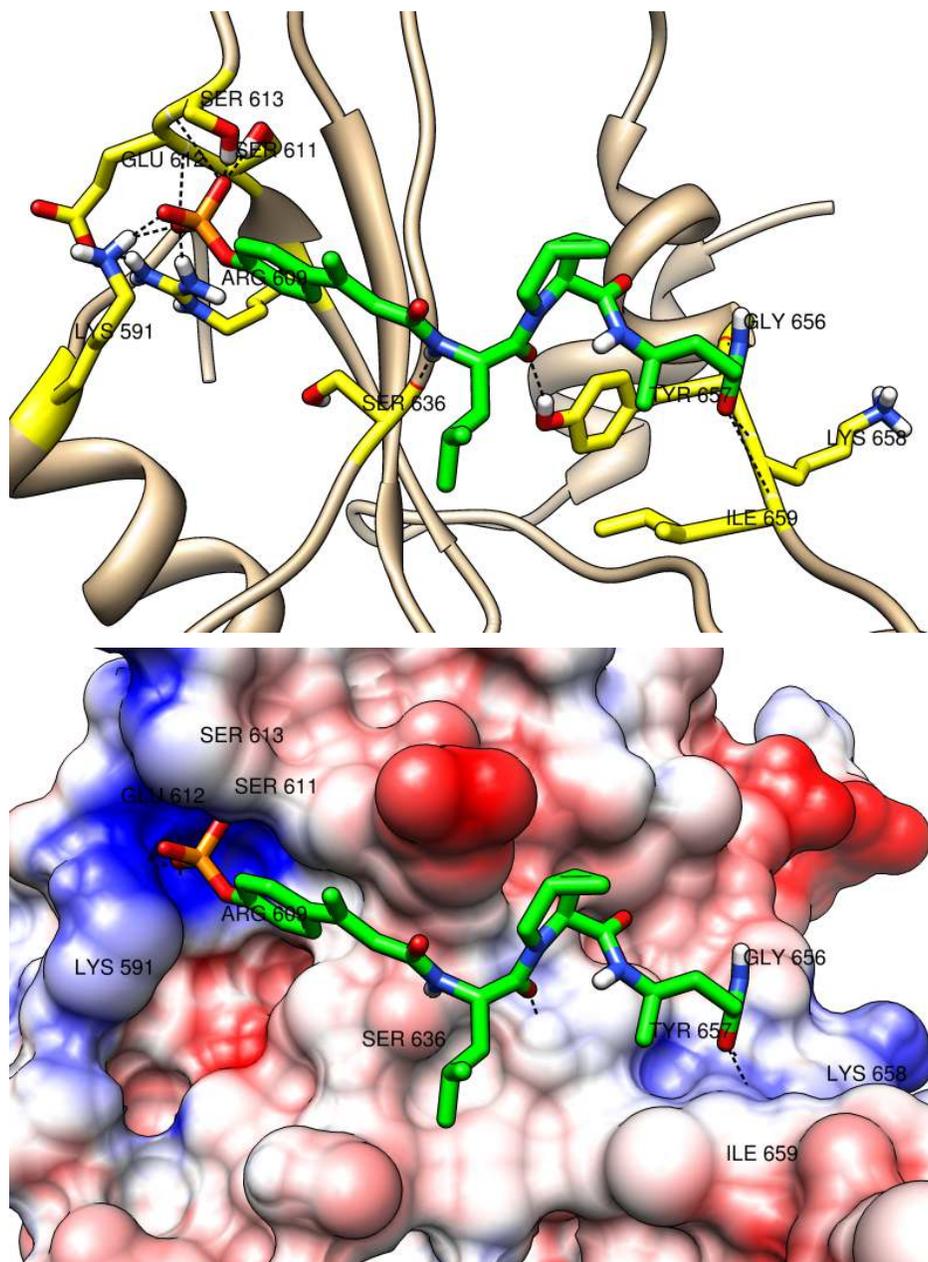
**Figure 7. Hydrogen bond occupancy.** Hydrogen bond occupancy plots for each peptidomimetic are shown. In each sub-plot, the x-axis represents the serial numbers of the residues of the SH2 domain of STAT3 and the y-axis represents the hydrogen bond occupancy value for a given residue. Hydrogen bond occupancy is computed as the fraction of conformations out of 1000 conformations of a peptidomimetic in which the given residue participates in a hydrogen bond. The 1000 conformations of each peptidomimetic were derived from the corresponding 10ns molecular dynamics trajectory. Note that a hydrogen bond is ignored if it is present in less than 50% of the conformations in the 10ns molecular dynamics trajectory.



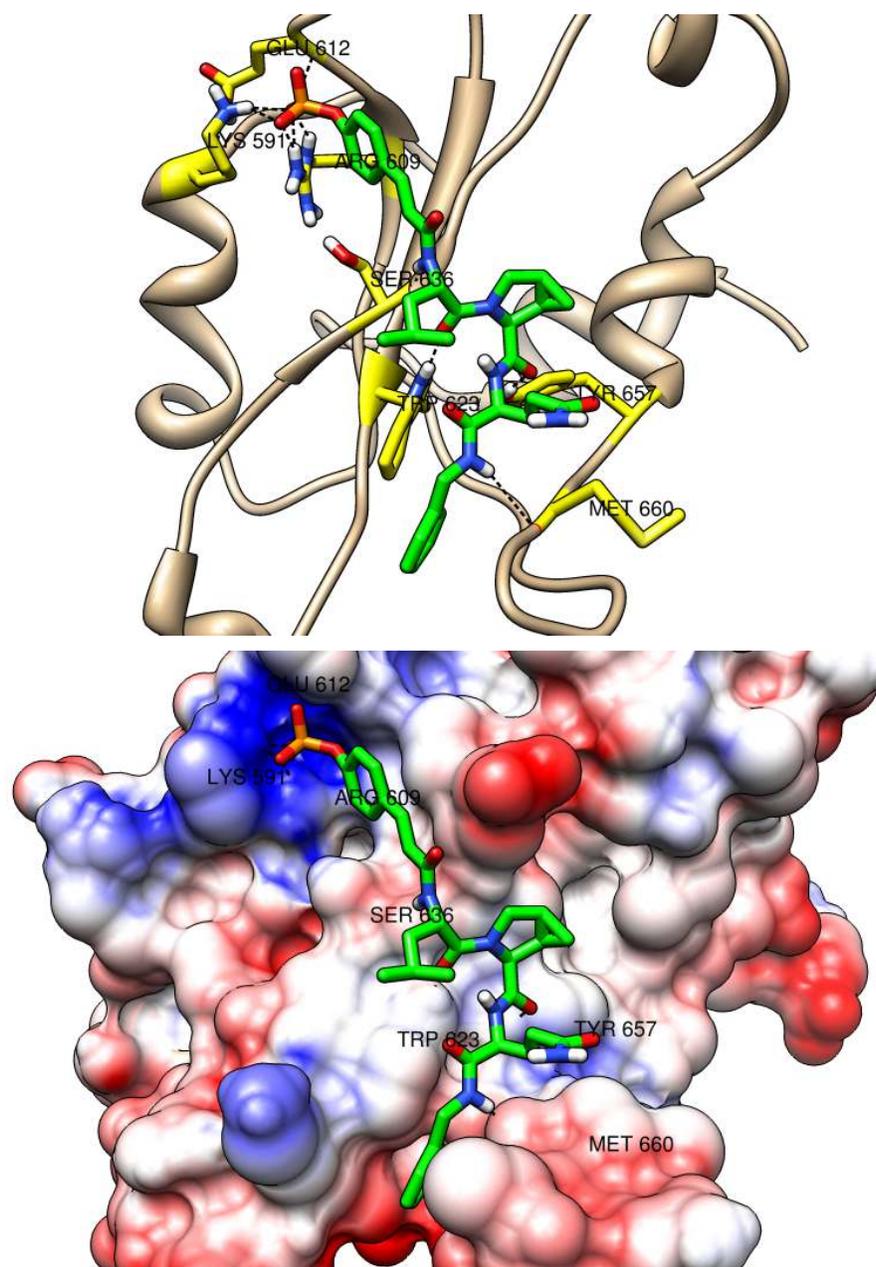
**Figure 8. Correlation between experimental and estimated affinities.** The top figure shows the variation of the Pearson correlation coefficient (R), computed between the experimental binding affinities and the estimated binding affinities of the 12 peptidomimetics, with the length of molecular dynamics simulation. The binding affinities were estimated using 4 different schemes.  $\Delta G_{MMGBSA}$  and  $\Delta G_{MMPBSA}$  represent non-entropic contribution to the binding affinity computed using the MMGBSA and MMPBSA methods in AmberTools software package.  $T\Delta S$  represents the entropic contribution computed using the *nmode* method in AmberTools. The bottom figure shows, for each peptidomimetic, the estimated binding affinity value computed using  $\Delta G_{MMPBSA} - T\Delta S$  scheme. Because the values computed using MMGBSA, MMPBSA, and *nmode* methods are averaged over the snapshots of the molecular dynamics trajectory, we also plot the variation of estimated binding affinity values with increasing length of the molecular dynamics simulation.



**Figure 9. Bent binding mode.** The bent binding mode of peptidomimetic comp70 (in green) is shown. The peptidomimetic is in complex with the SH2 domain of STAT3 which is shown in cartoon (top) and surface (bottom) representations. The residues of the SH2 domain which participate in hydrogen bonds are labeled. The top figure also shows the hydrogen bonds (dashed lines) that the residues (in yellow) participate in. The surface coloring shows the Coulombic electrostatic potential in the different regions of the surface of the SH2 domain. The potential ranges from positive (in blue) to negative (in red). The  $IC_{50}$  value for comp70 is 190nM.



**Figure 10. Extended binding mode.** The extended binding mode of peptidomimetic comp134 (in green) is shown. The peptidomimetic is in complex with the SH2 domain of STAT3 which is shown in cartoon (top) and surface (bottom) representations. The residues of the SH2 domain which participate in hydrogen bonds are labeled. The top figure also shows the hydrogen bonds (dashed lines) that the residues (in yellow) participate in. The surface coloring shows the Coulombic electrostatic potential in the different regions of the surface of the SH2 domain. The potential ranges from positive (in blue) to negative (in red). The  $IC_{50}$  value for comp134 is 83nM.



**Figure 11. Wedged binding mode.** The proposed novel wedged binding mode of peptidomimetic comp121 (in green) is shown. The peptidomimetic is in complex with the SH2 domain of STAT3 which is shown in cartoon (top) and surface (bottom) representations. The residues of the SH2 domain which participate in hydrogen bonds are labeled. The top figure also shows the hydrogen bonds (dashed lines) that the residues (in yellow) participate in. The surface coloring shows the Coulombic electrostatic potential in the different regions of the surface of the SH2 domain. The potential ranges from positive (in blue) to negative (in red). The  $IC_{50}$  value for comp121 is 68nM.