Pose Prediction Accuracy in Docking Studies and Enrichment of Actives in the Active Site of GSK-3β

Pravin Kumar Gadakar, Samiron Phukan, Prasanna Dattatreya, and V. N. Balaji*

Jubilant Biosys, Limited, #96, Industrial Suburbs, 2nd Stage, Yeshwanthpur, Bangalore - 560 022, India

Received November 9, 2006

We present molecular docking studies on the inhibitors of GSK-3β kinase in the enzyme binding sites of the X-ray complexes (1H8F, 1PYX, 1O9U, 1Q4L, 1Q5K, and 1UV5) using the Schrödinger docking tool Glide. Cognate and cross-docking studies using standard precision (SP) and extraprecision (XP) algorithms have been carried out. Cognate docking studies demonstrate that docked poses similar to X-ray poses (root-mean-square deviations of less than 2 Å) are found within the top four ranks of the GlideScore and E-model scores. However, cross-docking studies typically produce poses that are significantly deviated from X-ray poses in all but a couple of cases, implying potential for induced fit effects in ligand binding. In this light, we have also carried out induced fit docking studies in the active sites of 1O9U, 1Q4L, and 1Q5K. Specifically, conformational changes have been effected in the active sites of these three protein structures to dock noncognate ligands. Thus, for example, the active site of 1O9U has been induced to fit the ligands of 1Q4L, 1Q5K, and 1UV5. These studies produce ligand docked poses which have significantly lower root-mean-square deviations relative to their X-ray crystallographic poses, when compared to the corresponding values from the cross-docking studies. Furthermore, we have used an ensemble of the induced fit models and X-ray structures to enhance the retrieval of active GSK-3β inhibitors seeded in a decoy database, normally used in Glide validation studies. Thus, our studies provide valuable insights into computational strategies useful for the identification of potential GSK-3β inhibitors.

INTRODUCTION

Glycogen synthase kinase 3β (GSK-3β, also known as human TPKI, tau protein kinase I) is a unique multifunctional serine/threonine kinase that is inactivated by phosphorylation. In response to insulin binding, PKB/AKT phosphorylates GSK-3β on serine 9, which prevents the enzyme from phosphorylating glycogen synthase.[1−3] Our analysis on Jubilants' PathArt database revealed that GSK-3β has been linked to a diverse array of diseases like muscle hypertrophy, cancer, bipolar mood disorder, schizophrenia, and diabetes. Unphosphorylated glycogen synthase is active and able to synthesize glycogen. Thus it plays a key role in the transduction of regulatory and proliferative signals arising out at the cell membrane in insulin signaling pathway leading to potential modulation of blood glucose levels.[4] GSK-3β inhibitors have been discovered and developed as potential therapeutics against type II diabetes.[5] In addition to insulin signaling, GSK-3β participates in the **WNT (Wingless Integration Gene)** signaling pathway, where it forms a complex with axin, β-catenin, and adenomatous polyposis coli (APC) protein.[4] In the presence of WNTs, GSK-3β is unable to phosphorylate β-catenin, which leads to stabilization of β-catenin. The WNT pathway inactivates GSK-3β via the proteins, Dishevelled and FRAT, which disrupt the interaction of GSK-3β with axin, β-catenin, and APC.[6] In addition, there is considerable interest in the clinical role of GSK-3β inhibitors because they may mimic the effect of insulin or reduce the hyperphosphorylation of tau protein observed in Alzheimer's disease.[7]

Considerable interest has been devoted to the elucidation of the GSK-3β structure and the interactions with its inhibitors and substrate analogs. To date, 13 highly and moderately resolved (resolution of 1.95−2.8 Å) X-ray complexes (1GNG, 1H8F, 1I09, 1J1B, 1J1C, 1O9U, 1PYX, 1Q3D, 1Q3W, 1Q41, 1Q4L, 1Q5K, and 1UV5) are documented in the Protein Data Bank (PDB).[8−14] In addition, one X-ray complex is recorded in the patent literature.[15] An overlap of some of these structures shown in Figure 1 clearly demonstrates two distinct ligand binding sites. One of them corresponds to the so-called active site and the other to the phosphorylation site. A number of different chemical classes of compounds (Figure 2) bind to these two sites and their interactions with the active site residues in their corresponding X-ray illustrated in Figure 3a−f (Supporting Information). In most of the X-ray structures, these interactions are mediated by at least one molecule of water. Interestingly, water molecules mediating protein−ligand interactions are found to have lower temperature (B) factors in the X-ray structures, when compared to those of water molecules located well outside the active sites. In addition to structural elucidation, recently 3D QSAR studies using CoMFA have been presented on a specific class of GSK-3β inhibitors.[16] The results of the CoMFA analyses did not correspond to the active site interactions but claim to characterize the fundamental features of the receptor−ligand interactions.[16]

In light of the biological significance of GSK-3β in various disease areas and in light of the availability X-structure data both in the patent and public domain, we have undertaken a program of molecular modeling studies to determine optimal protocols useful for discovering potential inhibitors with

* Corresponding author e-mail: vnbalaji@jubilantbiosys.com.

10.1021/ci6005036 CCC: $37.00
Published on Web 06/20/2007

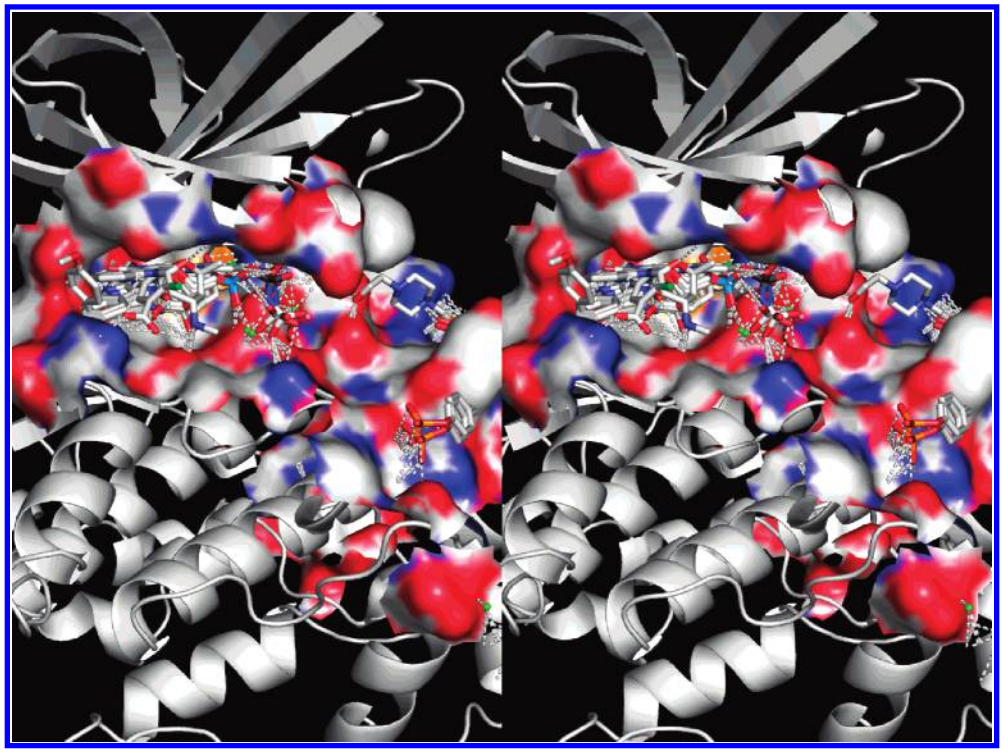

Figure 1. Stereoview of the overlap of X-ray ligands of GSK-3β (1H8F, 1O9U, 1PYX, 1Q4L, 1Q5K, and 1UV5). The ligand in 1H8F (4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid) is bound in the phosphorylation site.

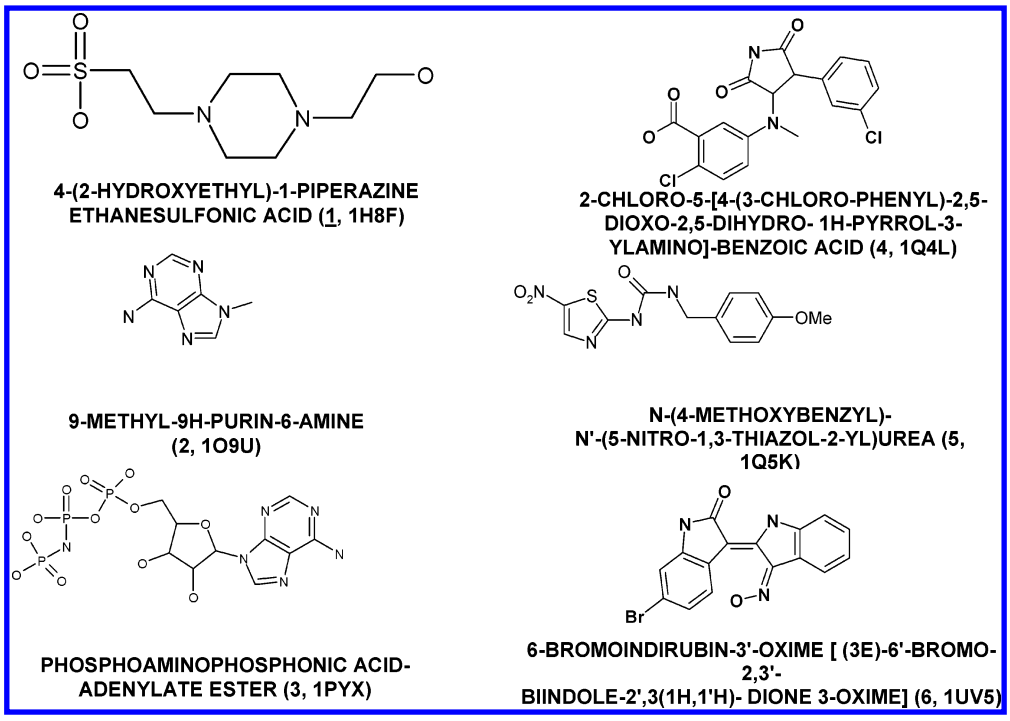

Figure 2. Schematic illustration of ligands in 1H8F, 1O9U, 1PYX, 1Q4L, 1Q5K, and 1UV5.

novel templates. Toward this goal, we have carried out docking studies in the active site of GSK-3β using Schrödinger docking and simulation tools on six of the X-ray structures and on protein models induced to fit noncognate ligands. Specifically, we have focused our attention in this study on two aspects of the docking calculations: (1) the ability to reproduce X-ray pose to within 2 Å root-mean-square deviation (rmsd) and (2) the enrichment of known actives seeded in a collection of decoys. Our studies clearly demonstrate that extraprecision (XP) docking in both the X-ray structures and induced fit models leads to greater pose accuracy relative to X-ray structure pose and greater enrichment of the actives seeded among decoys than SP docking. We also demonstrate that in the event of limited availability of X-ray data, an ensemble of available structure data and induced fit models can be employed to enhance recovery of actives from a database of compounds. Thus, XP docking based protocol is established to be very important in identification of actives that tend to bind to the active site with the right profile of interactions. Furthermore, our studies

provide valuable insights into the role of key active site residues in the interactions of potential inhibitors with the enzyme active site.

METHODS AND NOMENCLATURE

Ligand Preparation. The ligands in the X-ray complexes, 1H8F, 1O9U, 1PYX, 1Q4L, 1Q5K, and 1UV5 (**1**−**6**), were prepared using the LigPrep module of Maestro v7.5.112[17] in the Schrödinger suite of tools. The bond orders of these ligands were manually fixed, and the ligands were "cleaned" through LigPrep specifying a pH value of 7.0. Most probable tautomers of the compounds were chosen based on their interactions with the protein in the X-ray structures. In the case of ligand **3** (present in 1PYX), the negative formal charges were moved to the oxygen atoms which were in closest proximity to either the magnesium ion or a nearby lysine side-chain ammonium. Also, the stereochemistry was fixed as in the crystal geometry. In the case of compound **1**, the SO_3 group carried a formal charge of −1, while two possibilities were considered for the piperazine ring. In the first case, both the piperazines nitrogens carried a formal charge of +1 each. In the second case, both of them were retained in their neutral forms. As will be discussed later, the charge state on the piperazine plays a very important role in producing a pose similar to the X-ray pose. In the final stage of LigPrep, the compounds were subjected to energy minimization with MMFFs.[18]

Protein Preparation. The PDB coordinates of the X-ray complexes 1H8F, 1O9U, 1PYX, 1Q5K and 1UV5 were all aligned to the crystal structure 1Q4L (chosen randomly), using the StructAlign module in Maestro. This allows for the superposition of all the protein chains and ligands in a common framework of the protein active site and the phosphorylation site. We have not chosen the *apo* PDB structures 1GNG and 1I09 for our investigations, since the conformational changes in their active residues relative to those in complexes with ligands could possibly necessitate induced fit docking rather than conventional rigid protein docking. Since we have included induced fit docking studies involving 5 out of the 6 proteins listed above, we have not explicitly considered the proteins in the *apo* structures as a part of our investigations (vide infra).

All the crystallographic water molecules were deleted with the exception of those forming bridge interactions between the protein residues and the ligand atoms. In the case of 1PYX, two active site magnesium ions were also retained as a part of the protein. Once aligned, all the protein−ligand complexes were initially subjected to single point energy calculation in the protein modeling package Prime of the Schrödinger suite of tools. This calculation is carried out using the in-built Schrödinger implementation of OPLS-2001 force field incorporating implicit solvation.[19] This process leads to the addition of missing side-chain atoms and hydrogens to the protein residues. The guanidines and ammonium groups in all the arginine and lysine side chains were made cationic, and the carboxylates of aspartate and glutamate residues were made anionic. A careful examination of the active site indicated that a flip by 180° flip of the Asn185 side-chain amide was needed to optimize the interactions with the ligand in 1PYX. However, such a flip did not impact the interactions in the other 5 crystal structures. Hence, for consistency, all the Asn185 side-chain amides were flipped by 180° as a part of the protein preparation step. In addition to the default setting, the Asp200 side-chain carboxylate was also treated in a neutral form. This is necessary in light of the fact that the binding of the ligand in 1PYX is stabilized by two magnesium ions bound to Asp200 carboxylate. However, there are no Mg^{++} ions present in any of the other five crystal structures. Hence, neutralization of Asp200 is critical to the binding of the 1PYX ligand in cross-docking experiments with the proteins of those five crystal structures.

Following the above steps of preparation, the protein−ligand complexes in the X-ray structures were subjected to ImpRef energy minimization using the Schrödinger implementation of the OPLS2001 force field with implicit solvation[20] in two stages: (a) In the first stage, the positions of water molecule/s and/or active site Mg^{++} ions were optimized keeping the ligand and the protein atoms in their X-ray structure positions. (b) In the second stage, the entire complex was minimized, and the minimization was terminated when the root-mean-square deviation (rmsd) of the heavy atoms in the minimized structure relative to the X-ray structure exceeded 0.3 Å. This helps maintain the integrity of the prepared structures relative to the corresponding experimental structures while eliminating severe bad contacts between heavy atoms.

Docking Studies. Docking studies were carried out in the ligand binding sites of 1H8F, 1O9U, 1PYX, 1Q4L, 1Q5K, and 1UV5 using the Schrödinger docking program Glide.[20,21,23] The protein−ligand complexes prepared as described above were employed to build energy grids using the default value of protein atom scaling (1.0) within a cubic box of dimensions 34 Å × 34 Å × 34 Å centered around the centroid of the X-ray ligand pose. The bounding box (within which the centroid of a docked posed is confined) dimensions were set to 14 Å × 14 Å × 14 Å. No constraints were imposed on any of the active site protein atoms vis-à-vis their interactions with ligand atoms. Both standard precision (SP) and extraprecision (XP) docking of ligands were carried out with default value of ligand atom scaling (0.8). In pose prediction studies, a maximum of 10 poses per ligand was saved, while in enrichment studies one pose was saved per ligand. Both cognate and cross-dockings of ligands were carried out. In the case of 1H8F, the cognate ligand was docked in its binding pocket, while the cross-docking studies were confined to the region deemed to be the active site, as none of the other 5 ligands binds in the phosphorylation site. Thus, a total of six cognate and 25 cross-docking studies were carried out. Root-mean-square deviations of heavy atoms were measured between the docked poses and the X-ray poses obtained in both the cognate and cross-docking experiments. These deviations are represented as histograms in the Discussion section of this manuscript.

For the purposes of docking, random conformations of LigPrep treated structures were created using the conformational analysis module of MacroModel invoked via Maestro v7.5.112. The Monte Carlo method was employed for scanning the accessible conformational space of the molecules, and mainly default options were employed for various parameters (MMFFs force field, constant dielectric of 1.0, no implicit solvation, 500 cycles of TNCG minimization, 1000 Monte Carlo steps). Conformations were deemed to

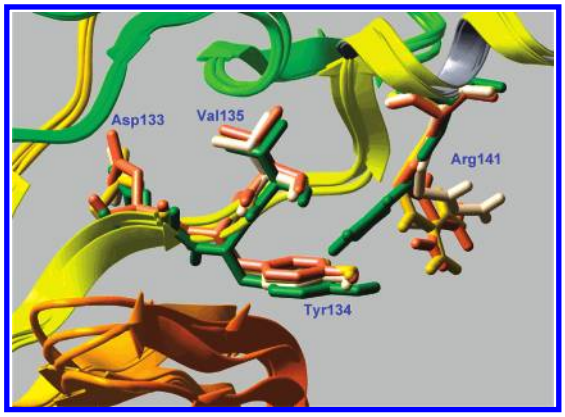

Figure 4. Computer graphic illustration of the overlap of GSK3β crystal structures showing the variations in the side-chain orientations of Arg141. Also shown are the structurally conserved residues Asp133, Tyr134, and Val135.

be different only when the maximum distance between the corresponding pair of atoms in any two structures was greater than 3Å. For each compound, of the conformations generated, one was picked randomly, ensuring that it had an rmsd of at least 2 Å relative to the X-ray structure. Thus, it was ensured that the docking studies were not biased by the starting structure being the X-ray pose.

Induced Fit Docking (IFD) Studies. Following the poor results obtained in some of the initial cross-docking studies (vide infra), it became obvious that the GSK-3β protein structure undergoes differential active site conformational changes (induced fit effects) when binding to different ligands. In this light, we have undertaken IFD studies on protein models of three of the crystal structures, wherein induced fit models have been obtained to fit ligands in noncognate structures. We have used the procedure described by Sherman et al.[22] via a python script executed in the framework of Maestro v7.0.113. We have obtained models of the protein in 1Q5K, 1Q4L, and 1O9U induced to fit the corresponding set of noncognate ligands. Thus for example, the protein in 1Q5K was induced to fit the ligands of 1Q4L, 1PYX, 1O9U, and 1UV5. In the remainder of this manuscript, induced fit protein models are referred to by the PDB code of the protein source and the PDB code of the inducing ligand. For example, 1Q5K_1Q4L refers to the model of the protein in 1Q5K which is induced to fit the ligand of 1Q4L. The ligand in 1H8F was not chosen for any induced fit study since it binds to a site distinct from those of the other ligands.

For each prepared protein model, the centroid of the corresponding X-ray ligand was used to define the energy grids for initial docking with default grid dimensions, derived from the extents of the bound ligand.[22] Twenty poses were chosen to be saved after initial Glide docking which was carried out with van der Waals scaling of 0.4 for both protein and ligand nonpolar atoms. In case of all the proteins, the side chain of Arg141 was mutated to alanine in the initial docking, since this residue undergoes significant movement in various crystal structures as seen from their overlap (Figure 4). It must be emphasized that the mutation of Arg141 to alanine is carried out ***only*** to obtain multiple poses of the ligands (one of which will contain the "experimentally correct" pose) in the initial docking stage. The modified Ala141 is replaced by arginine in the subsequent Prime refinement stage, as described below. It is in this light that the docked poses of the ligands were obtained by scaling the van der Waals radii of nonpolar atoms of the protein and ligand, by 0.4, for reasons cited previously.[22]

Once initial Glide docked poses were obtained, Prime side-chain and backbone refinement together with the minimization of the docked pose were carried out within a sphere of 5 Å from each pose saved. Glide redocking was carried out in Prime refined structures having Prime energy values within 20 kcal/mol of the lowest energy value. The rmsd values of the ligand poses with the best IFD score (as computed in ref 22) relative to the corresponding X-ray poses were determined. In some instances, higher ranked poses (2−5) of the IFD score were also considered for such rmsd calculations.

Enrichment Studies. In order to identify GSK-3β protein models suitable for enrichment of actives, SP and XP docking studies were carried out on a "database" of 2000 ligands-mined from Jubilants' Kinase ChemBiobase. In the present study, these enrichment studies have been confined to the protein model in the X-ray structure of 1Q5K and its corresponding induced fit models 1Q5K_1O9U, 1Q5K_1Q4L, and 1Q5K_1UV5. The database consisting of 10 known active inhibitors (Figure 5), seeded in a decoy set of 1990 molecules, has a random hit rate of 0.5% (=10*100/2000). As stated earlier, the top-ranked pose for each docked ligand (based on GlideScore) was saved. Percentages of actives retrieved in the top 5% (100) and 10% (200) of the database ranks were counted. In addition to looking at the retrieval of actives in individual structures, we have also considered such retrieval from the ensemble of protein structures. Towards this end, all the hits obtained by docking into each of the members of the ensemble are pooled and reranked by GlideScore, using the utility called "glide_sort" available at the distribution of Schrödinger products. Redundancy in the hits obtained is eliminated; for each hit, only the model/pose with the lowest GlideScore is retained. This is followed by the counting of actives retrieved from the top 100 (5%) and 200 (10%) ranks.

The active inhibitors belong to a variety of classes both in the space of core structures and substituents on them and span the molecular weight range of 300−500. The decoy set of molecules also spans the same range of molecular weight and has at least 200 different chemotypes represented in it. These molecules were chosen randomly from a larger collection of 20 000 NCI compounds which were treated with LigPrep as described earlier (vide supra). Figure 6a−d (Supporting Information) shows the histograms of distributions of molecular weight, LogP, and the number of hydrogen bond donors and acceptors for both the active and decoy sets.

RESULTS AND DISCUSSIONS

Binding Site Analyses in X-ray Structures. In the following text, we discuss the intermolecular protein−ligand interactions that are observed in various X-ray structures. In addition to rmsd values relative to X-ray poses, conservation of a majority of these interactions is an important criterion in judging the accuracy of pose prediction by various docking algorithms.

Figure 3a−f (Supporting Information) shows computer graphic illustrations of environments present in the active sites of 5 of the 6 crystal structures present and the phosphorylation site in 1H8F. The ligand **1** in 1H8F (Figure

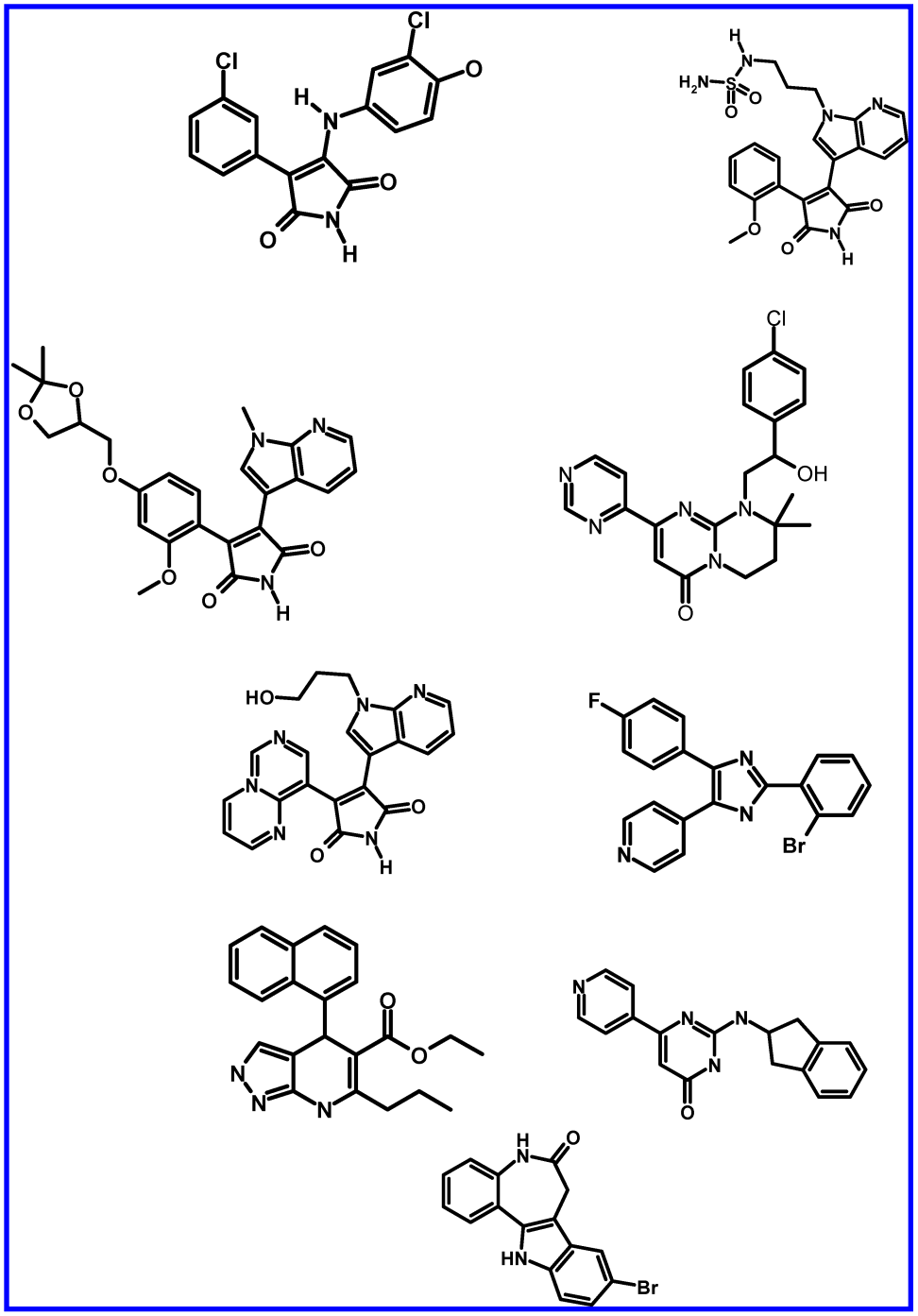

Figure 5. Schematic illustration of the representative chemotypes found in the set of 21 GSK-3β actives used in enrichment studies.

3a, Supporting Information) is bound in the phosphorylation site with ionic interactions between the SO_3^{2-} moiety on the ligand and the side chains of Arg96, Arg180, and Lys205 on the protein. In addition, the aromatic side chain of Tyr216 is in close proximity to the SO_3^{2-} moiety, suggesting the possibility of electrostatic stabilization with the phenolic hydroxyl group. The piperazine nitrogens are not involved in any specific interactions with nearby protein atoms suggesting that at least one of them and possibly both are uncharged in the environment of the binding site. The terminal CH_2OH moiety is hydrogen bonded to a nearby water molecule that in turn forms bridge interactions with nearby protein backbone atoms.

In the case of ligands bound in the active site of the GSK-3β crystal structures, 4 out of 5 form hydrogen bond interactions with the backbone carbonyl oxygen of Asp133 (Figure 3b−d,f, Supporting Information). The only exception to this bonding is the ligand in 1Q5K **5** which forms hydrogens bonds both the at N−H and C═O moieties of Val135 (Figure 3e, Supporting Information). However, all five ligands (**2**−**6**) form hydrogen bonds with the backbone N−H or C═O of Val135. In addition to these hydrogen bonds, ligands **2**−**5** are also hydrogen bonded to active site waters which hydrogen bond with the backbone NH moiety of Asp200 or the side-chain amide of Gln186. Furthermore, the aromatic phenol in Tyr134 provides hydrophobic stabilization to most of the ligands (**3**−**6**) (Figure 3c−f, Supporting Information) in the active site.

Accuracy of SP and XP Docked Poses. All the poses obtained in SP and XP docking for every ligand were rank

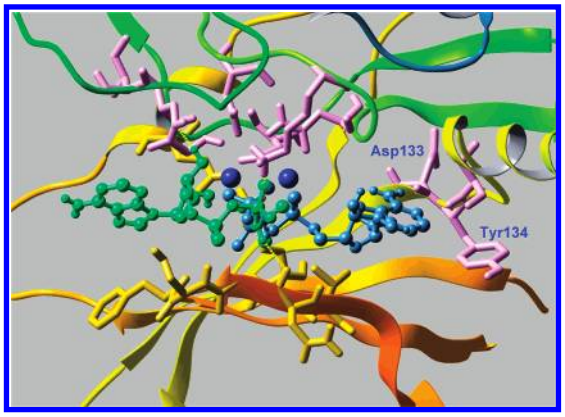

Figure 7. Computer graphic illustration of the top-ranked SP docked pose of **3** in the active site of 1PYX (aqua-blue). The X-ray pose is shown in green. Note that both the SP docked pose and the X-ray pose have their phosphate moieties coordinated with the Mg++ ions.

Table 1. Root-Mean-Square Deviations of Top-GlideScore Ranked Poses Obtained by SP Docking in the Active Sites of GSK-3β Crystal Structures

	protein					
ligand	1H8F	1O9U	1PYX	1Q4L	1Q5K	1UV5
1	7.66					
2	3.58	3.63	9.30	3.01	3.84	3.75
3	8.36	5.08	11.8	6.29	4.33	9.01
4	3.51	3.31	5.43	0.43	6.69	3.63
5	5.38	5.19	12.4	8.69	0.55	0.69
6	2.19	6.11	10.3	1.14	6.57	0.28

ordered according to GlideScore. It is interesting to note that the ranking of these poses by GlideScore parallels the ranking by E-model score, which includes contributions from van der Waals and electrostatic components of interactions between the protein and ligand.[22] Thus, a pose top-ranked by GlideScore is also top-ranked by E-model score in these docking studies.

SP Docked Poses. Table 1 shows root-mean-square deviations of SP docked poses ranked #1 in both the cognate and cross-docking experiments, relative to the corresponding X-ray poses. The columns represent the protein models in the six X-ray complexes, while the rows correspond to ligands. From this table, it is evident that the top-ranked SP docked poses of three ligands (**1**−**3**) are very dissimilar to X-ray poses (rmsd > 3 Å), while the corresponding poses of the remaining 3 X-ray ligands are very similar to their X-ray poses (rmsd < 1 Å). The deviation is highest for ligand **3**, with a value of ∼12 Å in the cognate docking. In its top-ranked pose, the phosphate groups are coordinated with the two magnesium ions in the binding site. However, the purine is swung away from the protein backbone comprising the residues Asp133 to Val135, as observed in the crystal structure (Figure 7). Thus, a combination of strong ionic interactions and a strong penalty for the solvent exposed adenine ring seems to be responsible for the large rmsd relative to the X-ray pose. Also, given the somewhat significant role of electrostatic interactions between the ligand and the enzyme active site, the lack of explicit polarization effects[24] in the scoring function may also cause the wrong orientation of the adenine ring to be favored in the top-ranked pose. Investigations employing QPLD (quantum polarized ligand docking) methods[24] on this enzyme−ligand system

Table 2. Lowest Root-Mean-Square Deviations of Any of the Poses Obtained by SP Docking in the Active Sites of GSK-3β Crystal Structures

	protein					
ligand	1H8F	1O9U	1PYX	1Q4L	1Q5K	1UV5
1H8F	2.22 (4)					
1O9U	0.63 (3)	0.73 (2)	7.88 (3)	0.59 (2)	1.13 (5)	1.12 (4)
1PYX	6.54 (4)	5.08 (1)	2.74 (3)	6.18 (4)	4.33 (1)	8.72 (2)
1Q4L	3.25 (7)	3.02 (8)	5.43 (1)	0.43 (1)	4.77 (4)	2.36 (6)
1Q5K	5.38 (1)	4.66 (2)	7.85 (4)	2.81 (9)	0.55 (1)	0.69 (1)
1UV5	2.19 (1)	2.78 (2)	8.77 (6)	1.14 (1)	6.57 (1)	0.28 (1)

are in progress and will be reported elsewhere. One of the furanose ring hydroxyls hydrogen bonds with the Ser203 side-chain hydroxyl group. Interestingly, a significantly lower rmsd value of 2.74 Å is obtained for the second ranked pose of **3** (Table 2).

The top-ranked pose of the rigid ligand **2** is flipped relative to the X-ray pose and yet occupies roughly the same volume of space as the latter (Figure 8a, Supporting Information), leading to an rmsd of 3.63 Å. This pose is characterized by hydrogen bonding interactions with the backbone NH and C═O of Val135. It lacks any interaction with either the active site water or the backbone C═O of Asp133, observed in the X-ray pose (Figure 7b). As in the case of **3**, the second ranked pose of **2** has a significantly lower rmsd of 0.73 Å and has interactions characteristic of the X-ray pose (Table 2; Figure 8b, Supporting Information). In the case of **1**, the top-ranked docked pose is significantly displaced relative to the X-ray pose (Figure 9a, Supporting Information), forming much weaker hydrogen bonding interactions with the backbone NH of Arg96 and side-chain carboxylate of Glu97, compared to the strong ionic interactions seen in the case of the X-ray pose. Interestingly, the fourth ranked pose is the only one in the vicinity of the X-ray pose (rmsd = 2.2 Å; Figure 9b, Supporting Information) and is characterized by the same ionic interactions involving the SO_3^- moiety. However, the hydroxyl group at the end opposite to the SO_3^- moiety interacts with the backbone carbonyl oxygen of Gly202 rather than the crystallographic water molecule (HOH82).

While conformational flexibility of **1** and **3** could have potentially played a role in the sampling of non-X-ray conformations in the SP docking leading to low energy conformations that differently than in the X-ray structures, the rmsd value for the top-ranked pose of a very rigid ligand **2** is somewhat puzzling. Perhaps the very small size of the ligand causes binding site placement degeneracy leading to the identification of a non-X-ray pose with favorable interactions. Interestingly, the interaction energy terms and GlideScore values (−6.0 and −5.6) associated with the first and second ranked poses of **2** are not significantly different. The only exception to this comes from the Coulombic term, which has values of −3.2 and −7.2 kcal/mol for the first and second ranked poses, respectively. Unlike in the case of ligands **1**−**3**, the top-ranked poses of ligands **4**−**6** are practically identical to the corresponding X-ray poses in their respective cognate dockings, as evident from the rmsd values of less than 0.6 Å. As is to be expected, these docked poses have the same profile of interactions with the protein residues as in the corresponding X-ray structures, and hence they are not shown exclusively.

Table 1 clearly indicates that with two exceptions, the top-ranked poses obtained in all cross-docking experiments are

Table 3. Root-Mean-Square Deviations of Top-GlideScore Ranked Poses Obtained by XP Docking in the Active Sites of GSK-3β Crystal Structures

	protein					
ligand	1H8F	1O9U	1PYX	1Q4L	1Q5K	1UV5
1	7.62					
2	3.51	3.52	3.69	3.63	3.78	3.76
3	7.07	5.27	1.79	6.07	4.94	9.65
4	3.49	3.47	5.52	0.77	5.21	3.43
5	8.07	4.77	8.69	3.41	1.05	0.67
6	4.17	5.81	4.14	1.19	6.52	0.24

Table 4. Lowest Root-Mean-Square Deviations of Any of the Poses Obtained by XP Docking in the Active Sites of GSK-3β Crystal Structures

	protein					
ligand	1H8F	1O9U	1PYX	1Q4L	1Q5K	1UV5
1	0.22(4)					
2	3.51 (1)	0.82 (2)	0.87 (2)	3.01 (2)	2.69 (4)	3.76 (1)
3	3.01 (5)	5.27 (1)	1.79 (1)	6.07 (1)	2.87 (2)	7.71 (2)
4	3.49 (1)	3.41 (3)	5.52 (1)	0.77 (1)	5.15 (2)	3.19 (2)
5	4.59 (3)	4.77 (1)	8.69 (1)	3.36 (3)	1.05 (1)	0.67 (1)
6	2.19 (3)	5.81 (1)	4.14 (1)	1.19 (1)	6.52 (1)	0.24 (1)

significantly deviated from the corresponding cognate X-ray poses. Ligand **6** is able to cross-dock into 1Q4L protein with a rmsd value of 1.14 Å relative to the X-ray pose in 1UV5. Similarly, ligand **5** is able to dock into the active site of 1UV5 with a rmsd value of 0.69 Å relative to the X-ray pose in 1Q5K. As in the case of cognate docking, some of the lower ranked poses in cross-docking studies have lower rmsd values relative to the X-ray poses. For example, the top-ranked poses of ligand **2** cross-docked into 1H8F, 1Q4L, 1Q5K, and 1UV5 have rmsd values of greater than 3 Å (Table 1). On the other hand, the corresponding third, second, fifth, and fourth ranked poses have rmsd values of less than 1 Å (Table 2). The reduction in rmsd values for lower ranked poses is not as dramatic as in the case **3**−**6**. Figure 10 (Supporting Information) shows a collection of histograms on the differences in the rmsd values of top-ranked and lower ranked poses obtained in SP docking studies. Ligands **4**−**6** have at least one lower ranked pose in cross-docking studies that are within 3 Å of their X-ray poses. However, in the case of ligand **3**, all the cross-docking studies yield poses that are considerably displaced from its X-ray pose 1PYX. This is primarily due to the fact that the other protein models lack the active site Mg^{++} ions present in 1PYX. Their absence causes the phosphate moiety of **3** to swing away toward the side chain of Arg141 leading to large rmsd values.

XP Docked Poses. Table 3 shows rmsd values of XP docked poses ranked #1 in both the cognate and cross-docking experiments, relative to the corresponding X-ray poses. In cognate cases, the rmsd values are comparable to those obtained in corresponding SP docking with the exception of 1PYX (ligand **3**), where the docked pose is within 2 Å of the X-ray pose. Thus **1** and **2** have large rmsd values (Table 3). However, when second ranked poses are considered in the case of these two ligands, the corresponding rmsd values are less than 1 Å (Table 4). As in the case of SP docking most top-ranked poses in cross-docking studies are significantly deviated from the respective X-ray poses. As earlier, the cross-docking of **6** in 1Q4L leads to a top-ranked pose with an rmsd of 1.19 Å. Also, the top-ranked pose obtained in the cross-docking of **5** in the active site of 1UV5 is very similar to the X-ray pose as was also observed in the SP docking (rmsd = 0.67 Å).

The impact of GlideScore rank on the rmsd with respect to the X-ray poses is illustrated in Table 4. In the case of **2**, the second ranked pose obtained in cross-docking into 1PYX has an rmsd of less than 1 Å, while in other cross-docking studies there is no significant improvement in the rmsd values relative to its X-ray pose. This is in contrast to the corresponding results obtained via SP docking, where consideration of lower ranked structures led to considerable improvement of rmsd values in cross-docking studies. For the remaining ligands, the rmsd values for lower ranked poses are comparable to those obtained with SP docking as discussed earlier (Table 2).

How different are the top-ranked XP and SP docked poses in cognate docking studies? Figure 11a−e shows the overlaps of these XP and SP docked poses with the lowest rmsd relative to the respective X-ray poses. It is obvious that with the exception of ligand **3**, the best XP and SP poses are practically identical with rmsd values relative to each other being less than or equal to 1 Å. In the case of ligand **3**, the corresponding rmsd value is 2.14 Å, attributable mainly to its conformational flexibility, which allows for a more exhaustive search of the conformational space to optimize the interactions with the protein, as measured by the E-model score value. This score includes contributions from van der Waals and electrostatic interactions between the protein and the ligand, in addition to the GlideScore.[13] However, all the interactions of the top-GlideScore ranked SP and XP poses with the protein and the active site Mg^{++} are conserved as in the crystal structure of 1PYX.

As with cognate docking, the top-ranked poses obtained in the cross-docking studies either via SP or XP are very similar to each other. As noted in Tables 1−4, their rmsd values relative to X-ray poses are similar, but the rmsd values relative to themselves are typically less than 1 Å. Two of the exceptions to this general observation come from the cross-docking of ligands **3** and **5** in the active site region of 1H8F. Here the typical rmsd values between the top-ranked SP and XP poses are in the range of 3−4 Å with significantly higher values for lower ranked poses. Thus, it is gratifying to note that generally the docked poses produced by standard and extraprecision docking methods are qualitatively consistent, although many of the SP poses are rejected quantitatively by XP docking.

Impact of Asp200 Protonation. As discussed earlier in this manuscript, the interactions of ligand **3** in its cognate X-ray complex are mediated by two magnesium ions which are absent in other crystal structures. Hence, it is not surprising that any cross-docking of this ligand will potentially result in poses that are significantly different from the X-ray poses. In this light, we have prepared additional protein models of 1O9U, 1Q4L, 1Q5K, and 1UV5 in which the side chain of Asp200 has been protonated. The proton location was determined based on optimum interactions of **3** in the absence of magnesium ions. These protein models were subjected to preparation, as stated in the Methods section (vide supra). Since the other ligands do not have any direct interactions with Asp200 in X-ray structures, protonation of this residue has no influence on their poses or their ranks. However, the docking of **3** is somewhat influenced by the

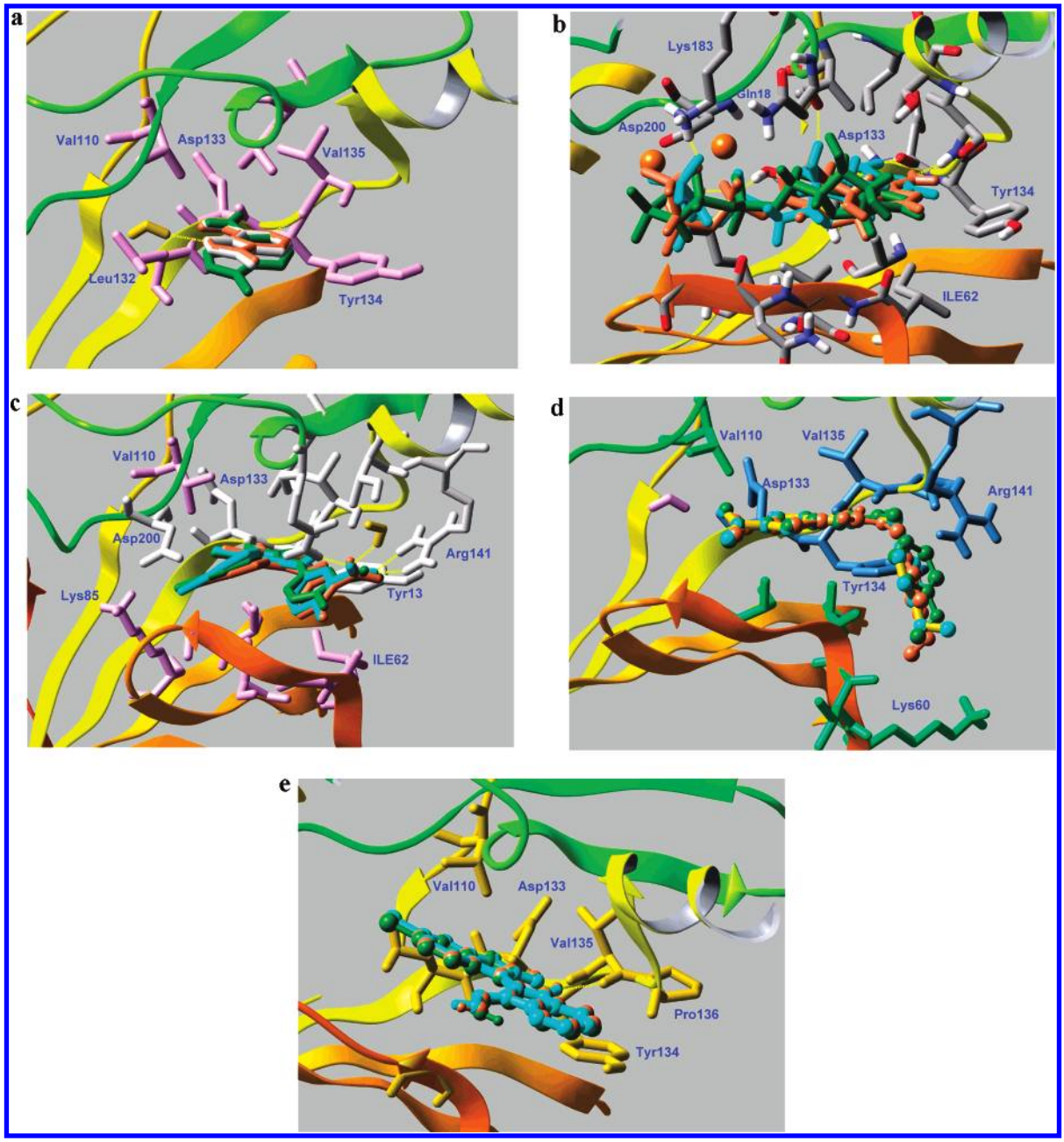

Figure 11. **a:** Computer graphic illustration of SP (brown) and XP (green) docked poses with the lowest rmsd relative to the X-ray pose (white) of **2** in the active site of 1O9U. **b:** Computer graphic illustration of SP (brown) and XP (green) docked poses with the lowest rmsd relative to the X-ray pose (cyan) of **3** in the active site of 1PYX. Also shown are the two magnesium ions in the active site of 1PYX. **c**: Computer graphic illustration of SP (brown) and XP (green) docked poses with the lowest rmsd relative to the X-ray pose (cyan) of **4** in the active site of 1Q4L. **d:** Computer graphic illustration of SP (brown) and XP (green) docked poses with the lowest rmsd relative to the X-ray pose (cyan) of **5** in the active site of 1Q5K. **e:** Computer graphic illustration of SP (brown) and XP (green) docked poses with the lowest rmsd relative to the X-ray pose (cyan) of **6** in the active site of 1UV5.

Asp200 protonation state. Figure 12 (Supporting Information) shows a comparison of the rmsd values obtained for top-ranked and any ranked pose with the lowest rmsd in XP docking simulations involving the two states of Asp200. In the case of docking with anionic Asp200, only cognate docking produces a pose with rmsd below 3 Å. All cross-docking studies have poses that are deviated by at least 4.3 Å. On the other hand, docking with a neutral Asp200 leads to at least one pose with an rmsd below 3 Å in four of the five protein models (2.77 Å in 1O9U [#3]; 2.0 Å in 1PYX [#2]; 2.17 Å in 1Q5K [#3]; 2.55 Å in 1UV5 [#6]). The corresponding rmsd is 6.4 Å in 1Q4L. By contrast, the top-ranked poses in these cases have rmsd values of 5.31 Å in 1O9U; 3.17 Å in 1PYX; 2.49 Å in 1Q5K; and 8.28 Å in 1UV5.

An overview of the results in the cross-docking studies indicates that the significant deviations obtained for the docked poses relative to X-ray poses are typically larger than the rms deviations of the backbone atoms of the protein structures. In view of the limitations of resolution in X-ray data, docked poses within 2 Å of the X-ray poses are deemed to be practically identical to the X-ray pose so long as key protein−ligand interactions are conserved. The larger deviations in cross-docking studies emphasize the significance of induced fit effects caused by conformational changes in the side-chain atoms of active site residues possibly together with

Table 5. Comparison of RMSD Values of Top GlideScore Ranked Docked Poses and Lowest RMSD Values in Any of the Docked Poses Obtained in the Active Site of 1Q5K by SP and XP Docking Protocols with the Corresponding Values Obtained for the Corresponding Poses Ranked by IFDScore in Induced Fit Models of 1Q5K

ligand	SP docking	XP docking	induced fit docking
2	3.84 (1), 1.13 (5)	3.78 (1), 2.69 (4)	3.48 (1), 1.11 (9)
3	4.33 (1)	2.38 (1), 2.17 (3)	2.84 (1), 1.74 (4)
4	6.69 (1), 4.77 (4)	5.21 (1), 5.15 (2)	1.56 (1), 1.62 (2)
5	0.55 (1)	1.05 (1)	
6	6.57 (1)	6.52 (1)	1.02 (1), 0.72 (2)

Table 6. Comparison of RMSD Values of Top GlideScore Ranked Docked Poses and Lowest RMSD Values in Any of the Docked Poses Obtained in the Active Site of 1O9U by SP and XP Docking Protocols with the Corresponding Values Obtained for the Corresponding Poses Ranked by IFDScore in Induced Fit Models of 1O9U

ligand	SP docking	XP docking	induced fit docking
2	3.63 (1), 0.73 (2)	3.52 (1), 0.82 (2)	
4	3.31 (1), 3.02 (8)	3.47 (1), 3.41 (3)	2.14 (1)
5	5.19 (1), 4.66 (2)	4.77 (1)	3.71 (1), 3.69 (2), 1.86(3)
6	6.11 (1), 2.78 (2)	5.81 (1)	1.31 (1)

Table 7. Comparison of RMSD Values of Top GlideScore Ranked Docked Poses and Lowest RMSD Values in Any of the Docked Poses Obtained in the Active Site of 1Q4L by SP and XP Docking Protocols with the Corresponding Values Obtained for the Corresponding Poses Ranked by IFDScore in Induced Fit Models of 1Q4L

ligand	SP docking	XP docking	induced fit docking
2	3.01 (1), 0.59 (2)	3.63 (1), 3.01 (2)	3.01 (1), 0.42 (8)
4	0.43 (1)	0.77 (1)	
5	8.69 (1), 2.81 (9)	3.41(1), 3.36 (3)	**1.64 (1), 0.86 (2)**
6	1.14 (1)	1.19 (1)	1.06 (1), 0.97 (2)

relatively minor backbone movements. The results of induced fit docking investigations are detailed in the following paragraphs.

Induced Fit Models. The results of the above-discussed docking studies have clearly emphasized the role of the subtle active site conformational flexibility in the ability to produce experimentally consistent binding modes of the ligands. In this light, we have carried out induced fit docking (IFD) studies in which we have developed protein models induced by noncognate ligands (as described in the Methods section − vide supra). The results of these studies are described below.

Table 5 compares the rmsd values obtained for docked poses in induced fit models of 1Q5K (i.e., 1Q5K_1O9U, 1Q5K_1PYX, 1Q5K_1Q4L, and 1Q5K_1UV5) with the corresponding values obtained in cross-docking studies. These values correspond to the top-ranked model based on the IFD score. As described earlier,[22] this score is a sum of the SP docking score during the redocking stage (vide supra) and 5% of the Prime score computed using the OPLS-AA force field with implicit solvation. Similar data for protein models of 1Q4L and 1O9U are described in Tables 6 and 7, respectively.

Table 5 clearly demonstrates significant improvement in the prediction of a docking pose similar to the X-ray pose in an induced fit model when compared to XP or SP docking in the protein model from the crystal structure. The improvement is most dramatic in the case of ligand **6**, where the rmsd changes by an order of magnitude—from ~6 Å in SP/XP docking to ~0.2 Å in induced fit docking. The development of the induced fit model (1Q5K_1UV5) allows for adequate protein flexibility in the active site residues such that the X-ray pose is readily accessible to the conformationally constrained ligand. Figure 13a illustrates the computer graphic overlap of the X-ray and top-ranked XP pose of **6** in the active site of the X-ray structure 1Q5K and the top IFD-score ranked pose in the induced fit model 1Q5K_1UV5. Residues within 3 Å of the bound ligand are also shown in this illustration. Figure 13b also shows the overlap of just the ligand poses in the X-ray (from 1UV5) and top 2 IFD-score ranked poses, which have rmsd values of 1.02 Å and 0.72 Å, respectively. The corresponding rmsd values are greater than 2 Å in lower ranked induced fit models. Thus, in this case, the second ranked induced fit model has the ligand pose closest to the X-ray pose. However, it should be noted that differences in the two rmsd values are smaller compared to the corresponding rmsd differences in the protein models, calculated relative to the X-ray structure of 1UV9.

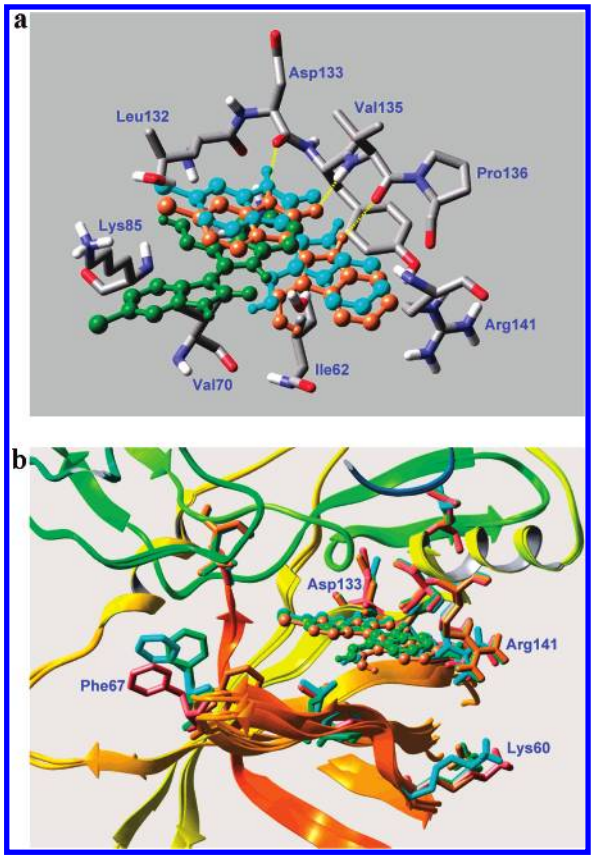

Figure 13. a: Computer graphic illustration of the overlap of ligand **6** in X-ray pose (**cyan**), top-ranked XP docked pose in 1Q5K X-ray model (**green**), and top IFD-ranked pose in the induced fit model 1Q5K_1UV9 (**brown**). Note that the rmsd value (relative to the X-ray pose) for the induced-fit model pose is significantly lower than in the top-ranked pose obtained via XP docking in the X-ray structure of 1Q5K. **b:** Computer graphic illustration of the overlap of ligand **6** in X-ray pose (**cyan**) and top two IFD score-ranked poses in the induced fit model 1Q5K_1UV9 (**green** and **brown**). Also shown is the X-ray crystal structure of 1Q5K aligned to the three models (**pink**). Note that the side chain of Phe67 in the top-ranked induced model (**brown**) has a significantly different orientation from that of the side chain in the second ranked induced fit model and the X-ray structures of 1UV5 (**cyan**) and 1Q5K (**pink**).

Tables 5a and 5b (Supporting Information) list the χ^1 and χ^2 angles, respectively, for those residues within 10 Å of the active site, where these angles vary by more than 30°. Comparing the active site residues in the top 2 IFD-score ranked induced fit models, we find that most of the side variations are confined to a few side chains on the lowered side of ligand pose relative to the conserved stretch Asp133-Val135. Of specific interest is the variation in the side-chain conformation of Phe67 (present on a loop between two β-strands). In the top-ranked IFD model, this side chain swings into the active site ($\chi^1 = 52°$), while in the second ranked IFD model this side chain has a conformation similar to that in the X-ray structures of 1Q5K and 1UV5 ($\chi^1 = -46°$).

The two induced fit models 1Q5K_1Q4L have ligand **4** poses with rmsd values of 1.55 Å and 1.62 Å, relative to the X-ray pose of this ligand in the PDB structure 1Q4L (Figure 14b,c). The reductions in these rmsd values relative to those obtained with the top two ranked XP poses in 1Q5K crystal structure are less dramatic than in the case of **6**. Nevertheless, here the induced fit model is very helpful in docking to a pose very similar to the X-ray pose of **4** and eliminates the short contacts of the latter with the side-chain atoms of Val70 and Lys85 observed in the active site of the X-ray structure of 1Q5K protein. These contacts are mainly responsible for the larger values of rmsd with respect to the X-ray pose of **4** in 1Q4L.

The results obtained for the induced fit models 1Q5K_1PYX are not as impressive as in the case of above 2 sets of models. In the case of **3**, the rmsd relative to the X-ray pose of the docked pose in the top-ranked induced fit model (2.82 Å) is higher than the lowest value of 2.0 Å obtained for the docking of this ligand in the active site of 1Q5K with protonated Asp200. Interestingly, the fourth and fifth IFD-ranked models have poses with lower rmsd (1.76 Å and 1.72 Å, respectively). These observations can be attributed partly to the somewhat extended conformational preference of the phosphate moiety in the ligand instead of the more compact form observed in the X-ray data.

Top-ranked induced fit models of 1Q5K_1O9U have ligand poses that are significantly deviated from the X-ray pose of **2** in the X-ray structure of 1O9U (Figure 15a). Only the ninth ranked model has an rmsd of less than 2 Å. The main reasons for larger rmsd values for the ligand poses in the top-ranked models seem to be (a) lack of necessary hydrophobic collapse in the protein structure surrounding a small ligand, (b) lack of inclusion of explicit water molecules in the active site which are present in the crystal structure of 1O9U, and (c) potential for significant hydrogen bonding by the ligand atoms over a small surface area. The latter causes a number of different poses to be located in the active site where stabilizing hydrogen bonds with the protein backbone can be formed.

The induced fit models with top 9 IFD score ranks, obtained by the flexible docking of **6** in the active site of 1O9U, have ligand poses with rmsd values less than 2 Å relative to the X-ray pose. The top-ranked model has a ligand pose with rmsd of 1.31 Å, which is significantly smaller than the rmsd values obtained for the top-ranked poses in the corresponding XP and SP docking studies (Table 6). The top-ranked induced fit docked model of 1O9U_1Q4L has a ligand pose within 2.14 Å of the X-ray structure. All the higher ranked models obtained with the flexible docking of **4** have poses with significantly higher rmsd values. Thus, the induced fit model leads to only marginal improvement in pose prediction relative to XP and SP docking. However, all the key interactions characteristic of the X-ray structure are preserved in the induced fit model. Specifically, the cis-amide group in the five-membered ring of the ligand hydrogen bonds with the backbone C═O of Asp133 and N−H of Val135. In addition, the carboxylic acid group on one of the phenyl rings forms an ionic interaction with the guanidine moiety of Arg141 and hydrogen bonding with the hydroxyl group of Thr138.

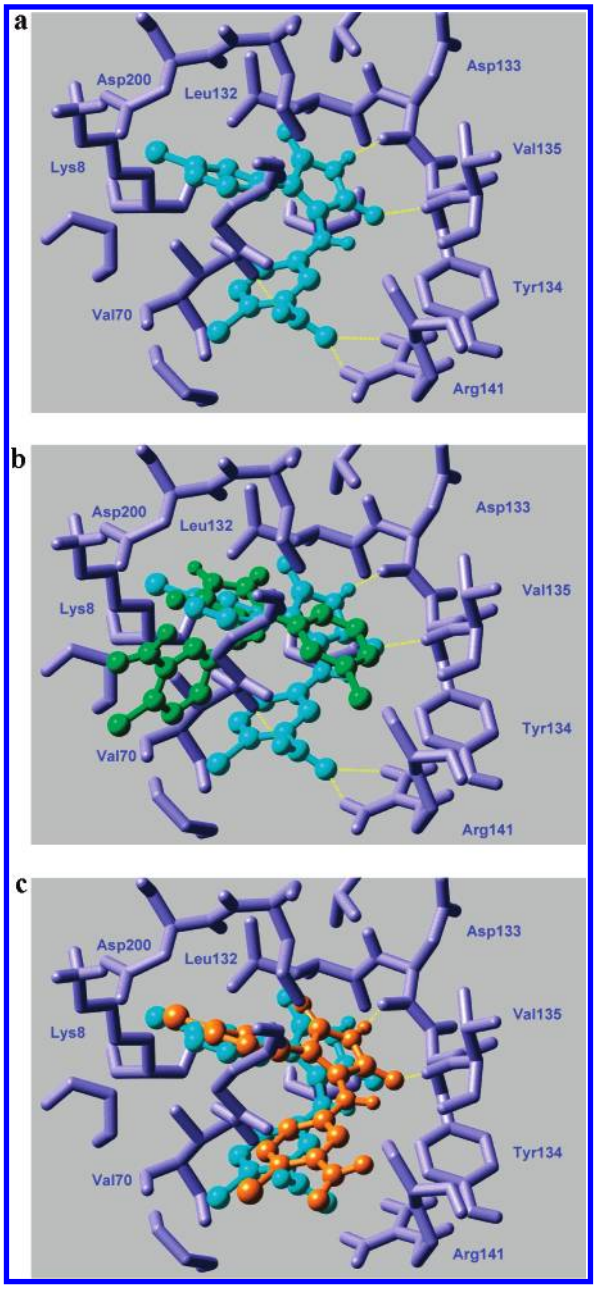

Figure 14. (a). Computer graphic illustrations of ligand **4** in X-ray structure (**cyan**) (a), top-ranked pose obtained in XP docking (**green**) (b), and top IFD score-ranked pose in the induced fit model 1Q5K_1Q4L (**brown**) (c).

The top-ranked (based on IFD score) induced fit model obtained with ligand **5** (IF_1O9U_1Q5K) has a pose with a rmsd value of 3.72 Å, relative to the X-ray pose, while the second and third ranked poses have corresponding rmsd

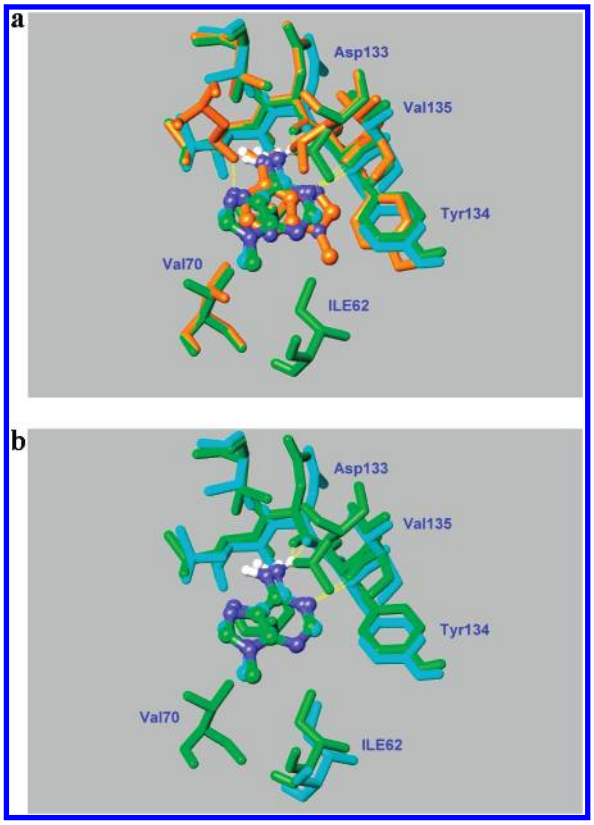

Figure 15. (a). Computer graphic illustrations of the overlap of ligand **2** in X-ray structure (**cyan**) and top-ranked pose obtained in induced fit docking (**brown**). (b). The corresponding overlap for the ligand pose (green) with second ranking in IFDScore is shown in (b).

values of 3.69 Å and 1.86 Å, respectively. These rmsd values are lower than those obtained either by SP or XP cross-docking (Table 6). Interestingly, the poses in the three top-ranked models form the same set of hydrogen bonds with the active site residues as in the X-ray complex of 1Q5K (Asp133 [C═O], and N−H and C═O groups of Val135). The somewhat larger values of the rmsd for the poses in the first two ranked IF structures are due to the reorientation of the phenyl methoxy moiety in the ligand. In the X-ray pose and in the third ranked IF model, this moiety is projected toward the pocket formed by the side-chain atoms Ile62, Lys60, and Gln72. On the other hand, in the first two ranked IF structures, this moiety is projected toward the side chain of Thr138. In this orientation, the methoxy oxygen is poised for a weak hydrogen bonding interaction with the Thr138 side-chain hydroxyl. This variation in the phenyl methoxy orientation corresponds to a change of ∼170° in the dihedral angle connecting it with the rest of the ligand and results in the large values of the rmsd. Nevertheless, it is gratifying to note that these values are smaller than those obtained in cross-docking studies, emphasizing the significance of active site conformational flexibility as captured by induced fit docking.

The induced fit models with the top 10 IFD score ranks, obtained by the flexible docking of **6** in the active site of 1Q4L, have ligand poses with rmsd values less than 1.5 Å relative to the X-ray pose. The top 2 ranked models have ligand poses with rmsd values of 1.06 Å and 0.97 Å, which are very similar to the rmsd values obtained for the top-ranked poses in the corresponding XP and SP docking studies (Table 7). The root-mean-square deviations for poses of ligand **5** in the top 2 IFD score ranked models of 1Q4L induced to fit the 1Q5K ligand are less than 2 Å (Table 7). The top six ranked induced fit models of the 1Q4L protein using ligand **2** have poses that are considerably deviated from the X-ray pose (rmsd > 3 Å) as seen in Figure 15a. However, the seventh IFDScore-ranked structure has a ligand pose practically identical to the X-ray pose (rmsd = 0.42 Å) as illustrated in Figure 15b. This rmsd value is close to the lowest value of 0.59 Å obtained for the second ranked pose in the SP docking. Although the ligand pose in the top-ranked IFD structure has a rmsd of 3.1 Å relative to the X-ray pose, it forms two hydrogen bonds with the backbone carbonyl oxygen of Asp133 (via the exocyclic amino group of the ligand) and N−H of Val135 (via the unsubstituted nitrogen in the five-membered ring). In the X-ray crystal structure, the latter is involved in hydrogen bonding with the active site water molecule. Since the corresponding site in the 1Q4L protein does not have a water molecule, it appears that the ligand is flipped about the bond connecting the exocyclic amino group to the six-membered ring. Interestingly, in the top-ranked induced fit model, the six-membered ring nitrogen (normally involved in hydrogen bonding with Val135 N−H moiety in X-ray structure) hydrogen bonds with the S−H group of the Cys199 side chain.

For reasons discussed above, we have not carried out induced fit docking of the protein models in 1O9U and 1Q4L against the ligand in 1PYX. To recap, the binding of this ligand in its X-ray pose requires an appropriate environment including two magnesium ions and a collection of well placed water molecules in the active site. The phosphate moiety in ligand **3** binds to a region of the active site that is dominated by an intricate balance of cationic and anionic residues together with key water molecules, which are not included explicitly during the docking simulations. Hence, as noted in the IFD models of 1Q5K_1PYX, the rmsd values of this ligand in the corresponding induced fit models of 1Q4L and 1O9U are expected to be larger than 3 Å.

The above discussions have dwelt mainly on the accuracy of ligand poses and the role played by protein flexibility in predicting them. In light of the results obtained, we find that the induced fit protein models compare very favorably with the known experimental answers. In order to make such comparisons we superimposed amino acid residues within shells of 8 Å around the ligand poses in the induced fit models and the corresponding X-ray structures (experimental answers). rmsd values were measured for all the protein backbone atoms and the side-chain carbons (C^{β}, C^{γ}, C^{δ}, and C^{ϵ}) in the superimposed models. These ranged from 0.6 to 2 Å in case of various induced fit models, including those where the ligand pose rmsd was larger than 3 Å. For example, in the case of induced fit models of 1Q5K induced to fit ligands **2**, **4**, and **6**, the typical rmsd between the induced fit models with the corresponding crystal structures (1O9U, 1Q4L, and 1UV5, respectively) vary between 0.6 and 1.0 Å. However, when all the atoms including the hydrogens are considered for the computation of the rmsd, the corresponding values are between 3 and 5 Å. This is mainly due to significant deviations in the side-chain conformations (χ^1 and χ^2) in a few amino acid residues, as listed in Tables 5a and 5b (Supporting Information). The induced fit models of 1Q4L and 1O9U are also similarly

Table 8. Percentages of Actives Recovered by the Top 5% and 10% of the Database Ranked by GlideScores Obtained in SP and XP Docking Protocols in the Active Sites of GSK-3β X-ray Structures[a]

protein	SP (5%)	SP (10%)	XP (5%)	XP (10%)
1O9U	40	60	60	80
1PYX	0	0	40	40
1Q4L	80	80	90	90
1Q5K	0	10	0	10
1UV5	50	60	60	60
ensemble	70	90	90	90

[a] Also shown are the retrieval percentages obtained in the ensemble of protein structures.

Table 9. Percentages of Actives Recovered by the Top 5% and 10% of the Database Ranked by GlideScores Obtained in SP and XP Docking Protocols in the Active Sites of GSK-3β X-ray Structure 1Q5K and Its Models Induced to Fit Ligands **2**, **4**, and **6**[a]

protein	top 5% (SP)	top 10% (SP)	top 5% (XP)	top 10% (XP)
1Q5K (X-ray)	0	10	0	10
IF_1O9U	10	10	60	70
IF_1Q4L	30	50	80	90
IF_1UV5	10	20	30	40
ensemble	30	50	50	50

[a] Also shown are the retrieval percentages obtained in the ensemble of protein structures from X-ray structure and induced fit models.

displaced relative to the corresponding experimental answers, and hence those details are not discussed.

Enrichment Studies. We have utilized both the X-ray crystal structures and induced fit models to carry out molecular docking studies to retrieve known GSK3β actives seeded in a decoy set. These studies help us identify the optimal protocol for the highest enrichment of actives in the top 5% and 10% of the database GlideScore ranks.

Table 8 shows the retrieval of GSK-3β actives seeded into the decoy database as described in the methods section, from the SP and XP docking into the X-ray structures of 1O9U, 1PYX, 1Q4L, 1Q5K, and 1UV5. Specifically, percentages of active compounds recovered in the top 5% and top 10% of the GlideScore ranks are listed. The results of ensemble docking into these 5 crystal structures are also illustrated in this table.

The results indicate that at least half the actives are retrieved in the top 5% (100) of the SP docking ranks in two out of the 5 X-ray structures, while 3 out of 5 structures do so in the top 5% of XP docking ranks. As is to be expected, larger percentages of actives are retrieved in the top 10% (200) of the database ranks in all the X-ray structures. Of particular interest are the cases of docking the database of actives and decoys into the active sites of 1Q5K and 1PYX proteins. In the case of 1Q5K no actives are retrieved in the top 5% of either SP or XP docking, while only one out 10 actives is retrieved in the top 10%. No actives are retrieved by the SP docking into the 1PYX active site within the 10% of the ranks, while 40% of the actives are obtained in the top 100 and 200 ranks in XP docking. Ensemble docking into the five proteins retrieves at least 70% of the actives within the top 5% of the GlideScore ranks, while the corresponding percentage is 90 when the top 10% of the ranks are considered. Thus, it is clear that upon SP docking of the database, screening the top 10% of the ranks will recover 90% of the actives yielding a 9-fold enrichment of actives. In the case of XP docking into the five-protein ensemble, an 18-fold enrichment is obtained since 90% of the compounds are recovered in the top 5% of the GlideScore ranks (Table 8).

In addition to crystal structures, we have also examined the enrichment of actives by SP and XP docking in the active sites of induced fit models. As an example, we have chosen the models of 1Q5K induced to fit ligands **2**, **4**, and **6**. The results of these docking studies are outlined in Table 9.

As stated earlier, only one out of 10 actives is retrieved by either the SP or XP docking in the active site of the 1Q5K crystal structure (row 1 Table 9; row 5 Table 8). However, there is significant improvement in the retrieval of actives by the three induced fit models (Table 9), with the exception of SP docking into the active site of 1Q5K_1O9U. Here only one compound is returned in the top 5 and 10% of the ranks. Interestingly, retrieval of actives by SP docking in an ensemble of the 1Q5K crystal structure and three induced fit models is better than that of the crystal structure and two of the induced fit models (1Q5K_1O9U, 1Q5K_1UV5). However, it is the same as that obtained in 1Q5K_1Q4L. On the other hand, the XP docking in the ensemble retrieves fewer actives than by the induced fit models 1Q5K_1O9U and 1Q5K_1Q4L. However, these retrieval percentages are higher than those obtained from docking to the binding pockets of the X-ray structure and that of the induced fit model IF_1UV5.

The above results clearly demonstrate the utility of induced fit models in the enrichment actives, particularly when limited X-ray data are available. Thus, the lack of retrieval of most actives in the top 10% of the database ranks by XP or SP docking in the active site is compensated by corresponding docking in the active sites of induced fit models which enrich a considerably larger number of actives. The ensemble docking is also helpful, although not to the same extent as in the case of the ensemble of X-ray structures.

One of the referees has suggested that given the differences in properties such as molecular weight, LogP, and number of hydrogen bond donors and acceptors, between some of the actives and decoys, it may be possible to correlate the enrichment of actives to these properties. A recent study on kinase inhibitors has derived QSPR models to enrich actives using essentially 2-D properties such as cited above.[25] However, preliminary investigations along these lines clearly indicate that enrichment of actives is obtained only upon taking into account explicit active site interactions, via the GlideScore. The details of these investigations are beyond the scope of this manuscript and will be published elsewhere. In addition, the question of sensitivity of enrichment results to the composition of the decoy set of molecules has been raised. As an example, we have carried out supplemental studies using the decoy sets employed in the Glide validation studies[20,21,23] and find the results reported in Table 8 are not altered to any significant extent, so long as the retrieval of actives in the top 5% and 10% of the database ranks are considered. It must be emphasized that the decoy set chosen in this study represents a practical situation in a drug discovery process pursued in the pharmaceutical industry. The details of the decoy set sensitivity studies will be discussed elsewhere.

CONCLUSIONS

We have presented Glide docking and enrichment studies in the active sites of six GSK-3β kinase X-ray structures and some of their variants induced by ligands that cross-dock poorly in them. Cognate SP dockings in X-ray models lead to top GlideScore ranked poses similar to corresponding poses in three of the six cases. On the other hand, poses similar to X-ray poses are found within top 5 ranks in the remaining three cases. In the case of XP docking studies, top GlideScore ranked poses are very similar to X-ray poses in four out of six cases, with the rankings being within the top 5 in the other two cases. Thus, cognate XP docking leads to higher accuracy of pose prediction compared to cognate SP docking. The protonation state of a side-chain carboxylate group in a key active site aspartate residue is shown to play a key role in predicting the binding mode of a ligand containing an anionic phosphonate moiety while not significantly affecting the docked poses of neutral ligands. Most of the cross-docking studies generally yield poses significantly different from the X-ray poses, indicative of induced fit effects in the binding of various ligands. This observation is generally supported by the markedly reduced root-mean-square deviations of docked poses relative to X-ray poses in a number of models obtained by the induced fit docking of ligands that dock poorly in cross-docking studies. The induced fit models are clearly very useful in the enhanced recovery of actives in enrichment studies carried out with SP and XP docking methods. Generally, higher enrichment is obtained by XP docking compared to SP docking. Larger enrichments are obtained by docking into an ensemble of X-ray structures versus the corresponding docking into an ensemble of an X-ray structure and its induced fit models.

ACKNOWLEDGMENT

The authors would like to thank Dr. Shashidhar Rao from Schrodinger Inc. for training/guidance provided on Schrodinger Tools.

Supporting Information Available: Tables 5a and 5b and Figures 3, 6, 8−10, and 12. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Frame, S.; Cohen, P.; Biondi, R. M. A common phosphate-binding site explains the unique substrate specificity of GSK3 and its inactivation by phosphorylation. *Mol. Cell* **2001**, *7*, 1321−1327.

(2) Mukai, F.; Ishiguro, K.; Sano, Y.; Fujita, S. C. Alternative splicing isoform of tau protein kinase I/glycogen synthase kinase 3 beta. *J. Neurochem.* **2002**, *81* (5), 1073−1083.

(3) Doble, B. W.; Woodgett, J. R. GSK-3, tricks of the trade for a multi-tasking kinase. *J. Cell. Sci.* **2003**, *116*, 1175−1186.

(4) McManus, E. J.; Sakamoto, K.; Armit, L. J.; Ronaldson, L.; Shapiro, N.; Marquez, R.; Alessi, D. R. Role that phosphorylation of GSK3 plays in insulin and Wnt-signalling defined by knockin analysis. *EMBO J.* **2005**, *24* (8), 1571−1583.

(5) Wagman, A. S.; Johnson, K. W.; Bussiere, D. E. Discovery and development of GSK3 inhibitors for the treatment of type 2 diabetes. *Curr. Pharm. Des.* **2004**, *10* (10), 1105−1137.

(6) Li, L.; Yuan, H.; Weaver, C. D.; Mao, J.; Farr, G. H.; Sussman, D. J., III; Jonkers, J.; Kimelman, D.; Wu, D. Axin and Frat1 interact with dvl and GSK, bridging Dvl to GSK in Wnt-mediated regulation of LEF-1. *EMBO J.* **1999**, *18*, 4233−4240.

(7) Bhat, R. V.; Budd, S. L. GSK3 beta signalling casting a wide net in Alzheimer's disease. *Neurosignals* **2002**, *11* (5), 251−261.

(8) Bax, B.; Carter, P. S.; Lewis, C.; Guy, A. R.; Bridges, A.; Tanner, R. et. al. The Structure of Phosphorylated GSK-3β Complexed with a Peptide, FRATtide, that Inhibits β-Catenin Phosphorylation. *Structure* **2001**, *9* (12), 1143−1152.

(9) Dajani, R.; Fraser, E.; Roe, S. M.; Young, N.; Good, V.; Dale, T. C.; Pearl, L. H. Crystal structure of glycogen synthase kinase 3β: structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* **2001**, *105*, 721−732.

(10) Ter Haar, E.; Coll, J. T.; Austen, D. A.; Hsiao, H. M.; Swenson, L.; Jain, J. Structure of GSK3β reveals a primed phosphorylation mechanism. *Nat. Struct. Biol.* **2001**, *8*, 593−596.

(11) Dajani, R.; Fraser, E.; Roe, S. M.; Yeo, M.; Good, V. M.; Thompson, V. et al. Structural basis for recruitment of glycogen synthase kinase 3β to the axin-APC scaffold complex. *EMBO J.* **2003**, *22* (3), 494−501.

(12) Bertrand, J. A.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H. M.; Flocco, M. Structural Characterization of the GSK-3β Active Site Using Selective and Non-Selective ATP-Mimetic Inhibitors. *J. Mol. Biol.* **2003**, *333*, 393−407.

(13) Bhat, R.; Xue, Y.; Berg, S.; Hellberg, S.; Ormo, M.; Nilsson, Y.; Radesater, A. C.; Jerning, E.; Markgren, P. O.; Borgegard, T.; Nylof, M.; Gimenez-Cassina, A.; Hernandez, F.; Lucas, J. J.; Diaz-Nido, J.; Avila, J. Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J. Biol. Chem.* **2003**, *278* (46), 45937−45945.

(14) Meijer, L.; Skaltsounis, A. L.; Magiatis, P.; Polychronopoulous, P.; Knockaert, M.; Leost, M.; Ryan, X. P.; Vonica, C. A.; Brivanlou, A.; Dajani, R.; Crovace, C.; Tarricone, C.; Musacchio, A.; Roe, S. M.; Pearl, L. H.; Greengard, P. Gsk-3-Selective Inhibitors Derived from Tyrian Purple Indurubins. *Chem. Biol.* **2003**, *10* (12), 1255−1266.

(15) Bussiere, D. E.; He, M.; Le, V. P.; Jansen. J. M.; Chin, S. M.; Martin, E. (Chiron Corporation, U.S.A.). Crystallization and crystal structure of human glycogen synthase kinase 3β protein and methods of use thereof. WO 0224893, A2 20020328, 2002, 200.

(16) Lescot, E.; Bureau, R.; de Oliveira Santos, J. S.; Rochais, C.; Lisowski, V.; Lancelot, J. C.; Rault, S. 3D-QSAR and Docking Studies of Selective GSK-3β Inhibitors. Comparison with a Thieno[2,3-*b*]-pyrrolizinone Derivative, a New Potential Lead for GSK-3β Ligands. *J. Chem. Inf. Model.* **2005**, *45* (3), 708−715.

(17) *Maestro (v7.0.113) − A unified interface for all Schrödinger products*; developed and marketed by Schrödinger, LLC (http://www.Schrödinger.com), 120 W. 45th St., 32nd Floor, New York, NY 10036. Copyright 2005.

(18) (a) Halgren, T. A Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519. (b) Halgren, T. A Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520−552. (c) Halgren, T. A Merck Molecular Force Field. III. Molecular Geometrics and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586. (d) Halgren, T.; A Nachbar, R. B. Merck Molecular Force Field. IV. Conformational Energies and Geometries. *J. Comput. Chem.* **1996**, *17*, 587−615. (e) Halgren, T. A Merck Molecular Force Field. V. Extension of MMFF94 using Experimental Data, Additional Computational Data and Empirical Rules. *J. Comput. Chem.* **1996**, *17*, 616−641. (f) Halgren, T. A MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720−729. (g) Halgren, T. A MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.

(19) (a) Rizzo, R. C.; Jorgensen, W. L. OPLS all-atom model for amines: resolution of the amine hydration problem. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836. (b) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. Force field validation. using protein side chain prediction. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(20) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Sander, L. C.; Shenkin, P. S. Glide: A new method for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(21) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new method for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(22) Sherman, B.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49* (2), 534−553.

(23) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P.; Mainz, D. T. Extra

Precision Glide - docking and scoring based on a new theory of molecular recognition. *J. Med. Chem.* **2006**, *49*, 6177−6196.
(24) Cho, A. E.; Guellar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26* (9), 915−931.
(25) Subramanian, J.; Sharma, S.; B-Rao, C. A novel computational analysis of ligand-induced conformational changes in the ATP binding sites of cyclin dependent kinases. *J. Med. Chem.* **2006**, *49* (18), 5434−5441.

CI6005036