JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

ARTICLE

pubs.acs.org/jcim

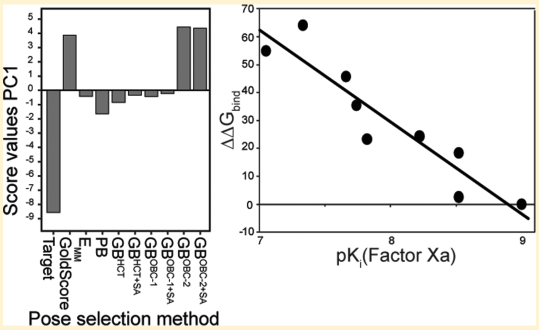
Postprocessing of Docked Protein−Ligand Complexes Using Implicit Solvation Models

Anton Lindström,[†,§] Lotta Edvinsson,[†] Andreas Johansson,[†,⊥] C. David Andersson,[†] Ida E. Andersson,[†] Florian Raubacher,[‡] and Anna Linusson[*,†]

[†]Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

[‡]AstraZeneca R&D Mölndal RA CVGI, Lead Generation, Pepparedsleden 1, SE-431 83 Mölndal, Sweden

S *Supporting Information*



ABSTRACT: Molecular docking plays an important role in drug discovery as a tool for the structure-based design of small organic ligands for macromolecules. Possible applications of docking are identification of the bioactive conformation of a protein−ligand complex and the ranking of different ligands with respect to their strength of binding to a particular target. We have investigated the effect of implicit water on the postprocessing of binding poses generated by molecular docking using MM-PB/GB-SA (molecular mechanics Poisson−Boltzmann and generalized Born surface area) methodology. The investigation was divided into three parts: geometry optimization, pose selection, and estimation of the relative binding energies of docked protein−ligand complexes. Appropriate geometry optimization afforded more accurate binding poses for 20% of the complexes investigated. The time required for this step was greatly reduced by minimizing the energy of the binding site using GB solvation models rather than minimizing the entire complex using the PB model. By optimizing the geometries of docking poses using the GB^{HCT+SA} model then calculating their free energies of binding using the PB implicit solvent model, binding poses similar to those observed in crystal structures were obtained. Rescoring of these poses according to their calculated binding energies resulted in improved correlations with experimental binding data. These correlations could be further improved by applying the postprocessing to several of the most highly ranked poses rather than focusing exclusively on the top-scored pose. The postprocessing protocol was successfully applied to the analysis of a set of Factor Xa inhibitors and a set of glycopeptide ligands for the class II major histocompatibility complex (MHC) A^q protein. These results indicate that the protocol for the postprocessing of docked protein−ligand complexes developed in this paper may be generally useful for structure-based design in drug discovery.

INTRODUCTION

The pharmaceutical industry uses structure-based design to discover and develop potential drug candidates in a cost-effective way.[1−3] A successful molecular docking procedure should be able to identify the bioactive binding pose of a protein−ligand complex and rank that ligand pose with respect to its binding affinity relative to those of other compounds in a series or database of molecules. This is accomplished by sampling the conformations of a complex; when sampling, the ligand is usually treated as being flexible, whereas the protein structure is treated as being nearly or completely rigid. The conformational ensemble is then ranked using a time-efficient scoring function such as knowledge-based equations to select the pose that is most likely to represent the bioactive conformation. In the final step, the binding strength of that pose is estimated to compare and rank different ligands. Several docking programs have been developed that address these steps in different ways. Using the available tools, complexes that are geometrically similar to the experimentally determined binding pose(s) can be generated, although it is not straightforward to select the program that will give the best result for any given target.[2,4] The primary challenge in molecular docking is to obtain accurate estimates of the binding strength of the simulated protein−ligand complexes, which is important when identifying the bioactive conformation and (especially when) seeking to rank-order series of compounds.[2,5,6] Predicting the binding affinity of such systems is difficult due to their complexity; biomolecules have many degrees of freedom, and the process studied occurs in an aqueous environment.[2] One approach to improve the predictions involves postprocessing of the docked protein−ligand complexes using methods that model the effects of water. The MM-PBSA (molecular mechanics Poisson−Boltzmann surface area) method,[7,8] combines molecular mechanics energies (E_{MM}) with continuum solvation models to estimate the binding affinity of a protein−ligand complex. In their original form, molecular dynamics (MD) simulations with explicit water molecules were used to generate sets of representative structures. Solvent molecules and any counterions were then removed from the structures and the average free energy, $\bar{G}$, was calculated as

$$\bar{G} = \bar{E}_{MM} + \bar{G}_{PB} + \bar{G}_{SA} - TS_{solute} \qquad (1)$$

where G_{PB} is the polar solvation energy calculated by numerical solution of the PB equation, G_{SA} is the nonpolar solvation energy estimated by a simple surface area (SA) term, and TS_{solute} is the solute entropy estimated from the MD trajectory. The free energy of

Received: September 8, 2010
Published: February 10, 2011

binding was then calculated as the difference in free energy between the end point and the starting point, that is, between the protein−ligand complex and the sum of the energies of the protein and the ligand in their unbound forms. This method is significantly less computationally demanding than alternatives such as free energy perturbation (FEP) calculations[8] because it relies on the use of implicit water and calculation of the difference in free energy between the two states rather than the relative energy along a mapping coordinate. The MM-PBSA method has been successfully used to predict the free energy of binding of ligands to a number of protein targets, including avidin and streptavidin,[9] HIV-1 RT,[10]and Cathepsin D.[11] Although the MM-PBSA method is much faster than the aforementioned FEP method and thermodynamic integration calculations, it is generally considered to be too slow for use in virtual screening to find new chemical leads, or in lead optimization projects where many compounds are to be evaluated. However, many modifications and simplifications of the MM-PBSA method have been presented to make the method more applicable in docking studies.[12−20] It has been shown that using a single minimized structure to estimate the free energy gives results similar to (and in some cases, better than) those obtained using a set of structures resulting from an MD simulation.[12,16,18] Further, the use of an implicit solvation model—the Generalized Born (GB) approximation—instead of the explicit water used in the MD simulations has been proposed.[13,14] GB methods have also been used to replace PB for energy minimization of single structures[15,18] and for subsequent free energy calculations[13,15,17,18] on docked protein−ligand complexes.

Another common simplification is to neglect the solute entropic term, TS_{solute}, when calculating the free energy. This term is usually calculated via a computationally demanding normal-mode analysis.[7] The exclusion of TS_{solute} is typically justified by assuming that the term will be constant for a set of analogues binding to a mutual protein and will thus not influence their relative binding energies.[8,13,20,21] It has also been shown that it is difficult to produce good estimates of solute entropy and that the use of such estimates can reduce the accuracy of the free energy calculations.[16,21,22]

In the presented study, we investigated the use of implicit solvation models for geometry optimization, pose selection, and estimation of the relative binding affinities of docked protein−ligand complexes based on the MM-PB/GB-SA methodology. The free energy of binding (ΔG_{bind}) was calculated as the difference in free energy between the end point and the starting point without a solute entropy term, applied to multiple docking poses whose geometry was optimized by energy minimization. The first part of the study focused on optimizing the parameters for the energy minimization of the docked complexes. The parameters examined were: the solvation model used, the number of energy minimization cycles, and the number of protein atoms included. Five solvation models implemented in AMBER9[23] were investigated: one PB model and four different GB models (with and without SA terms). The use of molecular mechanics without any solvent term was also examined. In the second part of the study, we sought to determine whether ΔG_{bind} could be used to identify poses that were geometrically similar to those observed in crystal structures. Four different models (one PB and three GB) were investigated for the calculation of the solvation energy when computing ΔG_{bind}. In the third part, the relative binding energies for the binding of a set of analogues to Factor Xa were compared to experimentally determined binding affinities.[24,25] The influence of solvation on ΔG_{bind} was evaluated using four different solvation models (PB and three GB variants), and the binding energies calculated in this way were used for the selection of docking poses and the estimation of relative binding energies. Additionally, these results were compared to those of gas phase calculations. Finally, the newly developed protocol for the postprocessing of docking poses was applied within an ongoing medicinal chemistry project to calculate the relative binding affinities of modified glycopeptides that target the major histocompatibility complex (MHC) class II protein A^q associated with autoimmune arthritis.[26]

■ MATERIALS AND METHODS

Data Sets. Four protein−ligand complex data sets were used to investigate the effects of implicit solvation models on three steps in the postprocessing of docked protein−ligand complexes (geometry optimization, pose selection, and the estimation of relative binding affinities).

Data set 1 was used to investigate implicit solvation models for the geometry optimization of docking poses. The data set consists of 30 complexes whose structures have been determined by X-ray crystallography, 25 obtained from the PDBbind database[24,25] and the other five disclosed in recent publications.[27−30] The set spans 19 different target proteins and a series of ligands with a broad range of physicochemical properties (Table 1). The molecular weights of the ligands range from 150 to 597 au, their logP values from −4.8 to 6.8, and the number of rotatable bonds, from 0 to 21.

Data set 2 was used to identify suitable implicit solvation models for use in rescoring to select docking poses and consists of three macromolecular targets (Protein-tyrosine phosphatase, Beta-trypsin, and Trypsin) and four structurally related ligands with different binding strengths for each of the proteins (Table 2).[24,25] This set of 12 complexes was used to investigate the effect of implicit solvation on pose selection across a range of protein targets, ligand classes, and binding affinities.

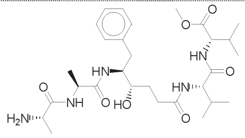
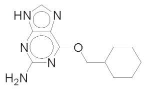
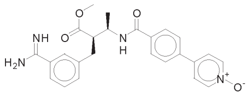
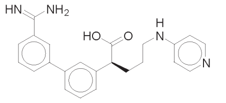
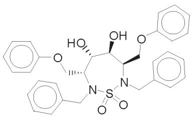
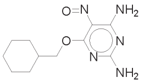
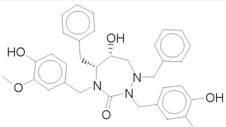
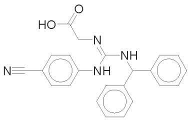
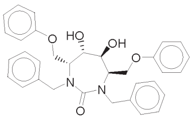
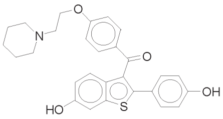
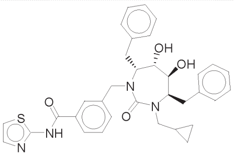
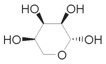
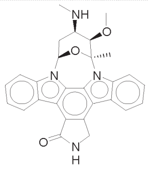
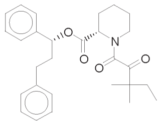
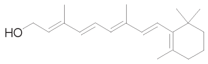
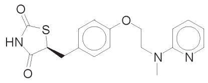
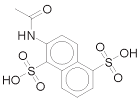
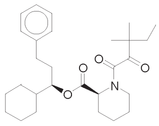
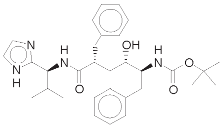
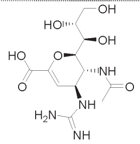
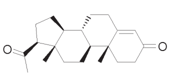
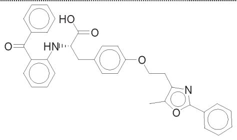
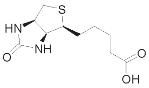
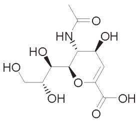
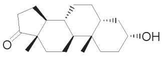
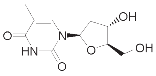
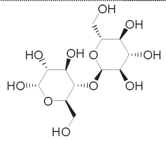
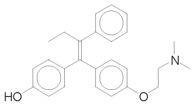
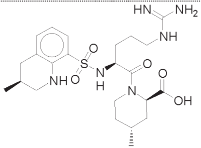
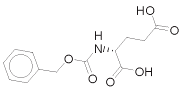
Data set 3 consists of a set of nine Factor Xa ligands with experimentally determined binding affinities (Table 3). The set was used to investigate rescoring using implicit solvation models selected on the basis of the results obtained using data sets 1 and 2. The complexes' structures have been solved at resolutions in excess of 2.2 Å and experimentally reported K_i values were taken from PDBbind.[24,25]

In addition to the three data sets described above, a fourth set of five modified glycopeptides that bind to the class II MHC A^q protein was investigated (Figure 1).[26] This data set was included in the study to investigate and test the general applicability of the new postprocessing protocol. The methods related to this data set are described separately in the last part of the Materials and Methods section.

Protein and Ligand Preparation. The 3D coordinates of the protein−ligand complexes in data sets 1, 2, and 3, determined by X-ray crystallography, were downloaded from the Research Collaborator for Structural Bioinformatics (RCSB) protein data bank.[31] The complexes are named according to their Protein Data Bank (PDB) codes, as entered in the RCSB data bank. The ligand binding poses according to the PDB files are hereafter referred to as "X-ray poses".

Protein structures were extracted from the PDB files and prepared using the Reduce software package[32] to add hydrogens to the structures where required, to suitably orient the hydrogens of SH and OH functional groups, and to specify appropriate tautomers of Asn, Gln, and His. The JACKAL software package[33] was used to check for missing atoms and reconstruct the protein

Table 1. Data Set 1: 30 Protein−Ligand Complexes Used to Investigate Implicit Solvation Models in Geometry Optimization of Docked Poses by Energy Minimization

1AAQ HIV-1 protease 	**1E1V** CDK2 	**1KSN** Factor Xa 	**1XKA** Factor Xa 
1AJV HIV-1 protease 	**1E1X** CDK2 	**1PRO** HIV-1 protease 	**2CGR** Antibody NC6.8 
1AJX HIV-1 protease 	**1ERR** Estrogen receptor 	**1QBU** HIV-1 protease 	**2DRI** D-Ribose binding protein 
1AQ1 CDK2 	**1FKG** FK506 binding protein 	**1RBP** Retinol binding protein 	**2PRG** PPAR-gamma 
1C5C Antibody 21D8 	**1FKH** FK506 binding protein 	**1SBG** HIV-1 protease 	**2QWE** Neuraminidase 
1DBB Antibody DB3 	**1FM9** RXR-alpha 	**1STP** Streptavidin 	**2SIM** Silidase 
1DBJ Antibody DB3 	**1KIM** Thymidin kinase 	**1URG** Maltose binding protein 	**3ERT** Estrogen receptor 
1DWC Thrombin 	**1KR6** Thermolysin 		

structure when needed. SS bonds in the structures were identified using an internuclear distance cutoff of 2.9 Å. Water, other inorganic molecules, and cofactors were removed before calculations were made.

Table 2. Data Set 2: 12 Protein–Ligand Complexes Used to Investigate the Selection of Poses on the Basis of ΔG_{bind}

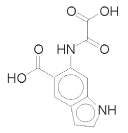
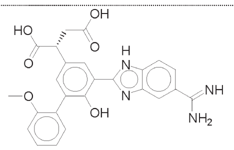
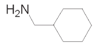
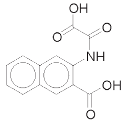
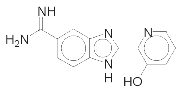
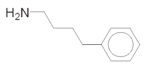
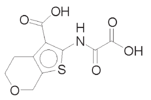
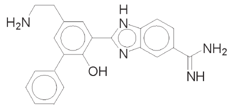
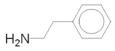
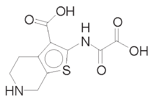
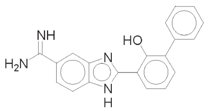
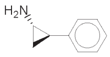
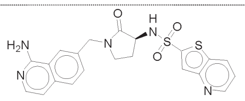
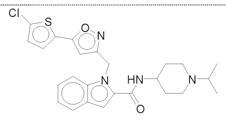
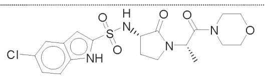
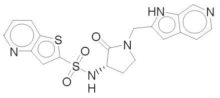
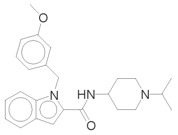
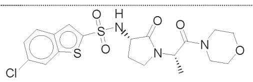
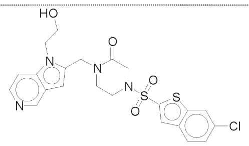
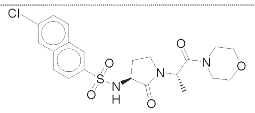
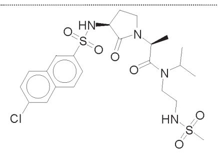
1C83 Protein-tyrosine phosphatase 	1O2Z Beta-trypsin 	1TNG Trypsin 
1C84 Protein-tyrosine phosphatase 	1O32 Beta-trypsin 	1TNI Trypsin 
1C87 Protein-tyrosine phosphatase 	1O37 Beta-trypsin 	1TNJ Trypsin 
1C88 Protein-tyrosine phosphatase 	1O39 Beta-trypsin 	1TNL Trypsin 

Table 3. Data Set 3: Factor Xa–Ligand Complexes Used to Investigate the Rescoring of Docked Poses on the Basis of Their Relative Free Energy of Binding, Calculated Using Various Implicit Solvation Models

1F0R pK$_i$: 7.66 	2BOH pK$_i$: 8.52 	2J2U pK$_i$: 7.33 
1F0S pK$_i$: 7.74 	2BQ7 pK$_i$: 7.05 	2J34 pK$_i$: 7.82 
1NFX pK$_i$: 8.52 	2CJI pK$_i$: 8.22 	2J4I pK$_i$: 9.00 

The chemical structures of the ligands were checked by comparison with information in the HIC-Up database[34] and the articles in which the structures were originally reported. The protonation states of the ligands were adjusted based on visual inspection of the ligands and the protein–ligand complexes and information in original cited articles. The 3D coordinates of the ligands submitted for docking were generated from SMILES using CORINA software.[35]

Generation of Binding Poses. For data sets 1, 2, and 3, the binding poses of the ligands were generated by molecular dock-

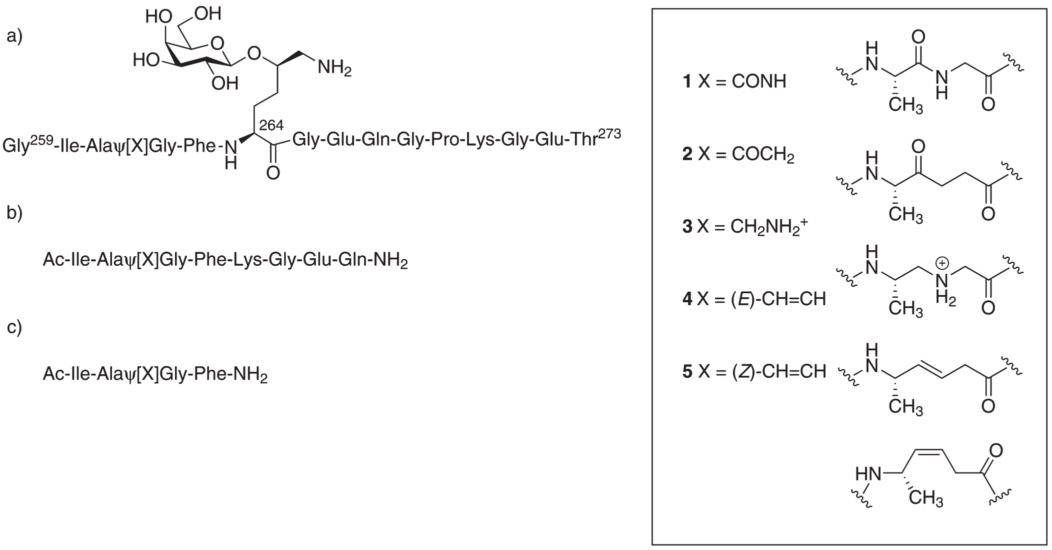

a)
HO
OH
HO
O
O
OH
NH2
264
Gly259-Ile-Alaψ[X]Gly-Phe-N
H
Gly-Glu-Gln-Gly-Pro-Lys-Gly-Glu-Thr273
O
b)
Ac-Ile-Alaψ[X]Gly-Phe-Lys-Gly-Glu-Gln-NH2
c)
Ac-Ile-Alaψ[X]Gly-Phe-NH2
1 X = CONH
2 X = COCH2
3 X = CH2NH2+
4 X = (E)-CH=CH
5 X = (Z)-CH=CH
H
N
CH3
HN
O
H2
⊕

Figure 1. (a) Template structure of modified glycopeptides that bind to the class II MHC A^q protein, including the native ligand (**1**) and the four bioisosteres investigated (**2**−**5**). The glycopeptides were truncated into shorter fragments (b and c) to reduce the number of rotatable bonds in the docking experiments and postprocessing in AMBER9, respectively. X denotes the site of modification in the glycopeptides and the included structural variations (**1**−**5**) are listed to the right.

ing to the prepared proteins using GOLD.[36] The unrefined poses extracted directly from the docking software are hereafter referred to as "docking poses". The atom and bond types of the proteins and ligands were set according to GOLD specifications using SYBYL.[37] The binding poses for data sets 1 and 2 were generated by docking the ligands to the appropriate protein structures using 10 different parameter settings and 20 genetic algorithm (GA) runs, resulting in 200 conformations for each protein−ligand complex. The different parameter settings (Supporting Information) were selected according to the work of Andersson et al.[4] to give a diverse collection of binding poses. For entries in data set 1, the docking pose most similar to the X-ray pose (as judged by the root-mean-square deviation, RMSD) was extracted from the 200 conformations generated by GOLD for each of the 30 protein−ligand complexes. For the 12 protein−ligand complexes in data set 2, several binding poses were extracted for each complex (see Supporting Information). The poses extracted included the one that was most highly ranked by the scoring function GoldScore within GOLD and the one with the lowest RMSD value relative to the crystallographic ligand. Additional poses were selected to give a range of RMSD values with a maximal acceptable RMSD value of 5 Å. Depending on the protein−ligand complexes, this resulted in five to seven selected binding poses per complex.

The binding poses for entries in data set 3 were generated by docking all the ligands to the same Factor Xa protein conformation (PDB code 2J4I, resolution 1.8 Å). Docking was performed using 100 GA runs (except for ligands 1F0R, 2CJI, 2J2U, and 2J4I, for which docking gave binding poses that resembled the X-ray poses using only 10 GA). The five docking poses for each ligand that were ranked the highest by GoldScore were extracted.

All RMSD values presented in this study were based on distances between heavy atoms.

Geometry Optimization. The energies of the docked protein−ligand complexes were minimized using the Sander module in AMBER9;[23] the general AMBER force field (GAFF)[38] was used for the minimization of the ligands, and the AMBER macromolecular force field ff03[39] was used for minimization of the protein structures. The default settings in AMBER9 were used unless stated otherwise. The refined docking poses are hereafter referred to as "energy minimized poses".

File Preparation. The protein and ligand files were prepared using the Antechamber and LEaP modules in AMBER9.[23] Protein and ligand atom types and charges were set according to the ff03[39] and GAFF[38] force fields, respectively. The force field parameters were analyzed, and modifications were made as required where parameters were missing or inappropriate. For entries in data sets 1 and 2, the partial charges of the ligand atoms were based on the restricted electrostatic potential fit (RESP),[40] and for data set 3, the partial charges were based on AM1-BCC.[41,42]

Solvation Models. In implicit water models, the solvent is considered to act as a perturbation of the gas-phase behavior of the system. Thus, water is modeled indirectly through inclusion of an extra term (ΔG_{solv}) in the energy calculations. Generally, the solvation free energy, ΔG_{solv}, can be considered to have three additive components: ΔG_{el}, ΔG_{vdw}, and ΔG_{cav}. ΔG_{el} is the free energy associated with removing all charges from the molecule in vacuum and readding them in an implicit solvent. ΔG_{vdw} is the free energy associated with the van der Waals forces between solvent and solute. ΔG_{cav} is the free energy required to form a cavity within the solvent for the solute to occupy.[43]

In this study, the gas phase energy (E_{MM}) was estimated based on the ff03 and GAFF force fields[38,39] and five different estimations of the electrostatic part (ΔG_{el}) of ΔG_{solv} were investigated.[1,44−46] Four different GB solvation models[47−51] and one PB solvation model[52,53] were used, as implemented in AMBER9.[23] The PB method is generally considered to be the most reliable method for estimating the electrostatic contribution to ΔG_{solv} that is implemented in AMBER.[44,45] The PB electrostatic model represents a molecule as a dielectric body whose shape is defined by atomic coordinates and atomic cavity radii. Using PB, the molecule is treated as a low dielectric cavity with its atomic partial charges mapped onto grid points. The total nonpolar solvation free energy was modeled using a single term that is linearly proportional to the solvent-accessible surface area.[54]

The GB models in AMBER describe each atom in a molecule as a sphere with a charge q_i at its center. The molecule is surrounded by a medium of a high dielectric ε and the interior of each is filled with a medium with a dielectric constant of 1.[55,56] The AMBER GB models estimate the ΔG_{el} ($\sim\Delta G_{\mathrm{GB}}$) according to eq 2:

$$\Delta G_{\mathrm{el}} \sim \Delta G_{\mathrm{GB}} = -\frac{1}{2}\sum \frac{q_i q_j}{f_{\mathrm{GB}}(r_{ij}, R_i, R_j)}\left(1 - \frac{e^{-\kappa f_{\mathrm{GB}ij}}}{\varepsilon_{\mathrm{w}}}\right) \quad (2)$$

where q is the charge of the atom, r_{ij} is the distance between atoms i and j, R is the effective Born radius of the atom, ε_{w} is the dielectric constant of water, and f_{GB} is a smooth function of the model. The electrostatic screening effects of salt ions are incorporated via κ Å^{-1} = 0.316[salt]$^{1/2}$[mol/L].[56]

The most extensively tested GB solvation model in AMBER is the GB$^{\mathrm{HCT}}$ model. However, it has been found that GB$^{\mathrm{HCT}}$ underestimates the effective Born radii of deeply buried atoms.[49] A rescaling function for the effective Born radii that uses adjustable dimensionless parameters has been implemented in the two available GB$^{\mathrm{OBC}}$ models,[50] which are henceforth referred to as solvation models GB$^{\mathrm{OBC\text{-}1}}$ and GB$^{\mathrm{OBC\text{-}2}}$. The fourth GB solvation model is GBn, which adds a pairwise correction term to GB$^{\mathrm{HCT}}$ to eliminate interstitial regions of high dielectric constant that are smaller than a solvent molecule.[51] An additional correction term to the solvation free energy for the atom, SA, was added to the above GB models using the linear combinations of pairwise overlaps (LCPO) approach.[57]

Geometry Optimization Protocols. For all data sets, a preliminary optimization of the positions of the hydrogen atoms in the protein−ligand complexes was conducted based on E_{MM} (i.e., with no solvation model) with the positions of all other atoms frozen. Thereafter, 500 cycles of full geometry optimization through energy minimization was performed on each species. In all cases, the steepest descent method was used for the first 200 cycles, and the conjugate gradient (CG) method in later cycles, as implemented in AMBER9.[23] As part of the study, different geometry optimization protocols were tested on data set 1. The results of that investigation were then further applied to data sets 2 and 3. The different settings used for the different data sets are discussed in detail below.

Data Set 1. Two protocols for geometry optimization were tested on the 30 docked complexes included in data set 1. The first protocol involved optimizing the positions of all protein and ligand atoms. In the second, optimization was only performed on the positions of the atoms of the ligand and those of the protein that featured in a side chain having at least one atom within 6.5 Å of a ligand. Such atoms are henceforth referred to as being within the binding site. A total of nine different implicit water models (four different GB models with or without the SA term, plus the PB model) were used in conjunction with the two optimization protocols to estimate ΔG_{solv}. For comparative purposes, the 30 complexes were also optimized without using any solvation model (E_{MM}).

To investigate the effect of the number of energy minimization cycles used, the binding sites of eight protein−ligand complexes (1C5C, 1AJX, 1FM9, 1KSN, 1QBS, 1URG, 2CGR, and 2SIM) were optimized using 100, 200, 350, 500, 750, and 1000 cycles. This test was performed with the PB solvation model, with the GB$^{\mathrm{HCT}}$ model, and with no solvation model (E_{MM}).

To assess the computational cost of the methods considered, the CPU times were recorded for a 500-cycle optimization of the binding sites of each of the complexes in data set 1, using each of the energy estimation methods (four GB methods ± SA terms, PB, and optimization without implicit solvation) and four Intel Core 2 Duo processors operating in parallel.

Data Sets 2 and 3. Energy minimization of the binding sites of the docked protein−ligand complexes was performed using the GB$^{\mathrm{HCT}}$ solvation model[47,48] with an SA term for 300 cycles.

Estimation of Binding Affinity. ΔG_{bind} is the change in the free energy of the system upon formation of a protein−ligand complex:

$$\Delta G_{\mathrm{bind}} = G_{(\mathrm{complex})} - G_{(\mathrm{protein})} - G_{(\mathrm{ligand})} \quad (3)$$

where $G_{(\mathrm{complex})}$ is the energy of the protein−ligand complex, $G_{(\mathrm{protein})}$ the energy of the unbound protein, and $G_{(\mathrm{ligand})}$ the energy of the unbound ligand. Here, the free energy change was calculated as follows:

$$\Delta G_{\mathrm{bind}} = \Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}} \quad (4)$$

where ΔE_{MM} is the difference between the gas phase energy of the complex and the sum of the gas phase energies of the ligand and the protein in their unbound states, while ΔG_{solv} is the difference between the solvation energy of the complex and the sum of the solvation energies of the ligand and the protein in their unbound states obtained using implicit water models. ΔG_{bind} calculations were performed using the MM-PB/GB-SA module in AMBER9 based on single conformer structures without inclusion of solute entropy terms. ΔG_{solv} calculations were performed using the available methods in AMBER9, i.e. the PB model,[52,53] the GB$^{\mathrm{HCT}}$ model,[47,48] the GB$^{\mathrm{OBC\text{-}1}}$ model,[50] and the GB$^{\mathrm{OBC\text{-}2}}$ model[50] with and without SA terms. The ΔG_{bind} in the gas phase (ΔE_{MM}) was also calculated.

When selecting poses on the basis of ΔG_{bind}, one compares the energies of different poses of the same ligand. ΔG_{bind} is thus directly proportional to $G_{(\mathrm{complex})}$, since $G_{(\mathrm{protein})}$ and $G_{(\mathrm{ligand})}$ are constant for all poses of the same ligand. Hence, the geometries of unbound states are of no significance when ΔG_{bind} is used as a pose selection method, and were therefore not considered when working with data set 2.

For data set 3, the geometry of the unbound state of Factor Xa was taken from the PDB file that was used for docking. The geometries of the unbound states of the ligands were generated using CORINA.[35] The positions of the hydrogen atoms of the molecules in the unbound state were optimized in a preliminary step using gas-phase molecular mechanics followed by energy minimization of the binding site for 500 cycles using the GB$^{\mathrm{HCT}}$ solvation model with an SA term.

Evaluation of Geometry Optimization. The conformations of the energy minimized poses were compared with the crystallographic ligand by calculating the RMSD values between the positions of the heavy atoms of the simulated poses and the corresponding atoms of the X-ray poses as deposited in the RCSB data bank.[31] The results were compared to the RMSD values of the docking poses used as input for the energy minimization according to

$$\mathrm{Diff} = \mathrm{RMSD}_{\mathrm{EM}} - \mathrm{RMSD}_{\mathrm{docked}} \quad (5)$$

where $\mathrm{RMSD}_{\mathrm{EM}}$ is the root-mean-square difference between an energy minimized pose and the crystallographic ligand, and $\mathrm{RMSD}_{\mathrm{docked}}$ is the root-mean-square difference between the docking pose (used as the starting pose for minimization) and the crystallographic ligand. A negative value indicates that the energy minimization has resulted in a pose more similar to the crystal structure than the docking pose. To evaluate differences between results obtained using the different energy minimization methods, the matrix of the Diff values resulting from energy minimizations of the 30 protein−ligand complexes in data set 1 using the 10 different optimization methods was analyzed by principal component analysis (PCA; see below). In addition, the changes in the positions of individual atoms due to energy minimization of docking poses in relation to the crystallographic ligand were calculated. The five maximum individual atom improvements are reported together with the five minimum individual atom improvements. Additionally, all poses were visually inspected to identify parts of the ligands that had adopted a conformation that was more similar to the crystal structure following optimization.

Evaluation of Use of ΔG_{bind} Values for Pose Selection. Three variables were examined to assess the utility of ΔG_{bind} values for selecting ligand binding poses that are geometrically similar to the experimentally determined ligand binding poses, as deposited in the RCSB data bank:[31] the RMSD values for the binding pose with the lowest ΔG_{bind} value, the Euclidean distance between the pose with the lowest ΔG_{bind} value and the pose with the lowest RMSD value, and the correlation between the ΔG_{bind} values and the RMSD values for the different binding poses. The correlations between RMSD and ΔG_{bind} are reported as R^2 values (Pearson's R), and the Euclidean distances (Dist) between two poses were calculated according to

$$\mathrm{Dist} = \sqrt{(\mathrm{RMSD}_x - \mathrm{RMSD}_y)^2 + ((E_x - E_y)\,/\,\mathrm{SD})^2} \qquad (6)$$

where x and y represent two energy minimized poses (the one with the lowest energy value and the one with the lowest RMSD value, respectively), E is the calculated ΔG_{bind}, and SD is the standard deviation of ΔG_{bind} for the set of poses generated with the method used to calculate the ΔG_{bind}. This normalization by SD makes it possible to compare Dist values obtained with different methods. If the pose with the lowest energy value is also the pose with the lowest RMSD value, the distance between these two poses will be zero, while if the pose with the lowest energy value is not the pose with the lowest RMSD value, this metric will give an estimate of the size of the deviation. The impact of solvation on the ΔG_{bind} of the optimized protein−ligand complexes was calculated using the PB,[44,45,52,53] GB$^{\mathrm{HCT}}$,[46−48] and GB$^{\mathrm{OBC\text{-}1}}$,[50] and GB$^{\mathrm{OBC\text{-}2}}$ models[50] with and without SA terms. The ΔG_{bind} in the gas phase (ΔE_{MM}) was also calculated. These results were compared to those for the pose most highly ranked by GOLD using the GoldScore scoring function and with extreme ("Target") values, defined as the lowest RMSD and Dist values, and the highest R^2 values for each complex observed with any of the methods examined. In total, then, nine sets of conditions for evaluating the selection of poses were investigated: eight according to their binding energy (one in the gas phase and seven involving solvation models) and one without geometry optimization (the GoldScore scoring function). These 9 settings were then applied to each of the 12 protein−ligand complexes in data set 2, and the poses so obtained were evaluated with respect to the three metrics described above (RMSD, R^2, and Dist) resulting in a matrix with nine plus one (the "Target" values) experiments and 36 variables (12 × 3). This matrix was subjected to PCA (see details below).

Evaluation of Rescoring Using Relative Free Energy of Binding. Two poses for each ligand in data set 3 were selected for energy minimization and rescoring, the pose most highly ranked by GoldScore and the pose with the lowest ΔG_{bind} values out of the five highest ranked poses. The relative free energy of binding ($\Delta\Delta G_{\mathrm{bind}}$) for the ligands in a data set was calculated as the difference in ΔG_{bind} between each ligand and an arbitrary reference; in the case of data set 3, the reference used was the ligand with the highest affinity (PDB code 2J4I, pK_i = 9.00). The relative binding energies were compared to the experimentally determined pK_i values obtained from PDBbind.[24,25] The following solvation models were employed to calculate the free energy of binding: PB,[52,53] GB$^{\mathrm{HCT}}$,[47,48] GB$^{\mathrm{OBC\text{-}1}}$,[50] and GB$^{\mathrm{OBC\text{-}2}}$ [50] (with and without SA terms). Gas-phase $\Delta\Delta G_{\mathrm{bind}}$ values (ΔE_{MM}) are also reported for comparative purposes.

Principal Component Analysis. PCA is an unsupervised projection method in which systematic variation in a data set is extracted into a few variables, the principal components (PCs).[58] The PCs are linear combinations of the original variables and are uncorrelated to each other, as described by eq 7:

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1' + \mathbf{t}_2\mathbf{p}_2' + \mathbf{t}_3\mathbf{p}_3' + \ldots + \mathbf{t}_A\mathbf{p}_A' + \mathbf{E} = \mathbf{TP}' + \mathbf{E} \qquad (7)$$

where $\mathbf{X}$ is the original data matrix, A is the total number of extracted PCs, and $\mathbf{E}$ is the residual matrix. The new latent variables, the $\mathbf{t}$ scores, show how the experiments (in this case, the different methods) relate to each other, while the $\mathbf{p}$ loadings reveal the importance of the original variables (in this case, the metrics describing the different protein−ligand complexes) for the patterns seen in the scores. Each PC was evaluated based on its eigenvalue, R^2, and Q^2 values.[59] The matrix of Diff values based on data set 1 was mean-centered before the PCA, while the matrix of the three metrics for data set 2 (RMSD, R^2, and Dist) was mean-centered and scaled to unit-variance. The calculations were performed using SIMCA[59] and Evince.[60]

Application of the Postprocessing Protocol to Glycopeptides Binding to A$^{\mathrm{q}}$. A set of five modified glycopeptides that bind to the class II MHC A$^{\mathrm{q}}$ protein[26] was used to test the applicability of the postprocessing protocol for docked protein−ligand complexes to a different protein−ligand system. The 3D coordinates of the glycopeptide/A$^{\mathrm{q}}$ protein complex originated from a previously published comparative model.[26] The docking protocol used, including the technique used to rank the binding poses, has been tailor-made within an ongoing project and thus differs from the protocol used for data sets 1, 2, and 3. Figure 1 shows the glycopeptides that have been synthesized and biologically tested, the simplified docked ligands, and the fragments subjected to postprocessing in AMBER9. The ligands were truncated on the basis of the biological data to reduce the number of rotatable bonds in the docking experiments and the computational cost[26] of the subsequent postprocessing of the docking poses.

The protein−ligand binding poses were generated using FRED.[61] The 3D coordinates of the ligands used in the docking were generated from SMILES using CORINA[35] and a low energy conformational ensemble of each ligand was generated using OMEGA[62] with the parameter settings ewindow = 25, rms = 0.6, and maxconfs = 5000. The resulting conformations were then rigidly docked into the binding site of the A$^{\mathrm{q}}$ protein using parameter settings exhaustive scoring = cgo, clash scale = 0.7, opt = none, maximum number of poses = 1000, and number of alternate poses = 1000. The Lys^{264} side chain extends out into the bulk solvent from the binding pocket, so a positional constraint was imposed during docking to ensure that this would remain true in the docked structures. The 1000 docking poses from FRED were truncated to focus on the modified part of the ligand. To compensate for the rigidity of the ligand and protein during docking, the resulting protein-truncated ligand complexes were subjected to a preliminary geometry optimization step using the MMFF94x force field with an implicit distance dependent dielectric solvation model as implemented in MOE.[63] The RMSD values of the backbone atoms were calculated for all poses using the carbon and nitrogen backbone atoms of the ligand in the comparative model as a reference. The five binding poses with the lowest RMSD values to the reference for each of the five ligands were selected for postprocessing in AMBER9.[23]

The binding poses of the truncated glycopeptides generated in this fashion were fully energy minimized and their relative binding affinities were estimated using the Sander and MM-PB/GB-SA module in AMBER9,[23] using the same postprocessing protocol as for data set 3 (i.e., energy minimization of the ligand binding sites was performed using the $\mathrm{GB^{HCT}}$ solvation model[47,48] with an SA term for 300 cycles). The partial atomic charges on the ligand in the selected docking poses were calculated with AM1-BCC.[41,42] The unbound states of the $\mathrm{A^q}$ protein and the ligands were based on the comparative model,[26] and the docking poses with the lowest RMSD values compared to the reference, respectively. The positions of the hydrogen atoms of the $\mathrm{A^q}$ protein were optimized in a preliminary step using gas-phase molecular mechanics. The $\mathrm{A^q}$ protein and the ligands were energy minimized for 300 cycles using the $\mathrm{GB^{HCT}}$ solvation model with an SA term.

The relative binding affinities were calculated using the native ligand as the reference and were related to experimentally determined affinity data. The data were derived from competitive binding experiments where increasing concentrations of each glycopeptide were incubated with a fixed concentration of biotinylated class II-associated invariant chain peptide (CLIP) and $\mathrm{A^q}$ protein. The amount of $\mathrm{A^q}$-bound biotinylated CLIP was then quantified in a time-resolved fluoroimmunoassay.[26] The experimental data did not support full dose−response curves, so $\mathrm{IC_{50}}$ values could not be determined. The calculated relative binding energies were therefore directly related to the extent of inhibition at a representative ligand concentration of 4 μM.

■ RESULTS AND DISCUSSION

Geometry Optimization of Docking Poses Using Different Implicit Solvation Models. After energy minimization of the docking poses applying any of the four GB methods ± SA terms, the PB method, or in gas phase, a ligand binding pose that was significantly more similar to the crystal structure was obtained for 6 of the 30 docked protein−ligand complexes in data set 1 (PDB codes 1C5C, 1KSN, 1SBG, 1STP, 1XKA, and 2CGR; Figure 2), as judged by improvement in their RMSD values and visual inspection of the poses. In general, these 6 protein−ligand complexes consisted of proteins with relatively open binding pockets and ligands that were more exposed to solvent than in the 24 for which no significant improvement was observed.

The unrefined docking poses geometrically most similar to the X-ray poses (i.e., the poses submitted to geometry optimization) for each of the 30 investigated complexes had RMSD values of 0.27−4.02 Å. The improvements in the RMSD values of the six ligands after energy minimization were similar for all methods examined and ranged from 0.15 to 0.43 Å for when using $\mathrm{GB^{HCT}}$ (see the Supporting Information for data on the other methods). All of the RMSD values for the starting poses of five of these six ligands were between 0.6 and 1.0 Å, while the sixth had a starting pose with an RMSD value below 0.6 Å. This illustrates the point that energy minimization methods optimize toward local minima and thus good starting poses are needed to get a significant effect. It also emphasizes the importance of having a good pose-filtering method to eliminate poor poses.

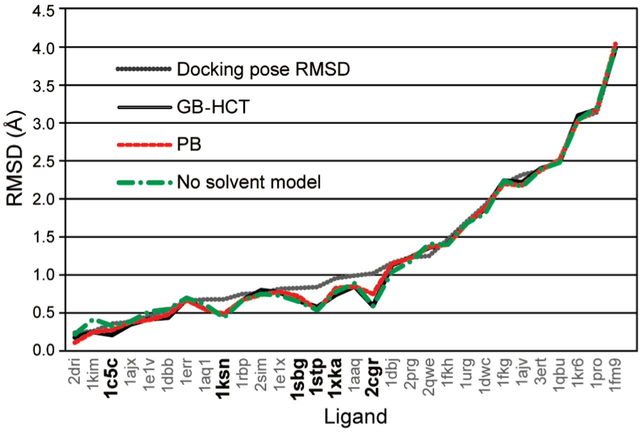


Figure 2. Changes in the RMSD of the docked ligands relative to the crystal structure following energy minimization with the $\mathrm{GB^{HCT}}$ and PB solvent models, and with no solvent model. The RMSD values for the starting poses of the ligands are also shown (docking pose RMSD). The six ligands that after energy minimization were significantly more similar to the crystal structure (discussed in the text) are highlighted. The ligands are denoted by their PDB codes.

Changes in the RMSD values reflect changes in the overall conformational deviation of the molecules, and docking/optimization can displace individual atoms across much greater distances than the RMSD values might suggest. For example, the RMSD values for the six "improved" ligands were between 0.15 and 0.43 Å, but much larger movements of individual atoms were observed (Table 4). Notably, the position of one atom in the docked ligand of complex 1STP was improved by 1.17 Å, and the docked ligands of complexes 1KSN and 2CGR had atoms that moved 0.91 Å toward the positions of the corresponding atoms in the X-ray structure. Inspection of the chemical structure of the six ligands revealed that, in most cases, the improvement in their conformation was due to conformational changes in a ring system and/or aromatic substituents (Figure 3). Although it was possible to identify structural elements whose geometry was often improved in the 6 ligands, no structural features that distinguish the 6 improved ligands from the remaining 24 were identified.

All of the methods for geometry optimization that were examined afforded similar results overall. Nevertheless, PCA of the geometrical differences between the energy-minimized poses generated using the different methods and the docking poses (i.e., their Diff values, see Materials and Methods) revealed interesting differences between the methods. The first three principal components account for 93% of the variation in the data set; inspection of the score plot shows that E_{MM} and PB differ from the GB methods in the first and second components, respectively (Figure 4a), while GBn can be distinguished from the other GB methods along the third component (Figure 4b). The inclusion of SA terms did not have any influence in our study; the position of the GB models with the SA terms on the score plot is similar to that of the GB models without these terms. The loading plot for the first three components (Figure 4c−d) shows the effects of using different optimization methods on

Table 4. Spatial Changes of Individual Ligand Atoms in Relation to the Crystal Structure after Energy Minimization of the Docked Poses

	maximum individual atom improvements (Å)					minimum individual atom improvements (Å)				
1C5C	−0.34	−0.32	−0.25	−0.24	−0.23	0.05	0.04	0.03	−0.06	−0.06
1KSN	−0.91	−0.51	−0.48	−0.48	−0.42	0.23	0.16	0.14	0.12	0.09
1SBG	−0.69	−0.67	−0.57	−0.48	−0.47	0.26	0.25	0.17	0.14	0.12
1STP	−1.17	−0.87	−0.82	−0.71	−0.68	0.43	0.20	0.15	0.05	0.00
1XKA	−0.83	−0.55	−0.54	−0.50	−0.48	0.67	0.55	0.23	0.21	0.08
2CGR	−0.91	−0.81	−0.65	−0.65	−0.61	0.35	−0.21	−0.24	−0.26	−0.29
average	−0.81	−0.62	−0.55	−0.51	−0.48	0.33	0.17	0.08	0.03	−0.01

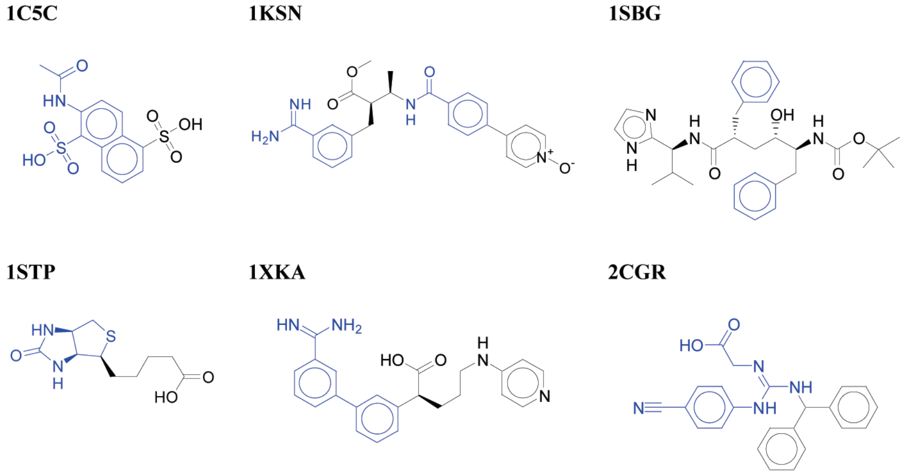


Figure 3. Ligands whose bound conformations became more similar to their crystallographic pose after geometry optimization. Highlighted in blue are the structural elements that underwent significant conformational change. The ligands are denoted by their PDB codes.

specific complexes. No clear general trends were identified from this plot; the relationship between a method and its associated Diff values seems to be dependent on the complex being examined. For example, the binding pose for complex 1KIM obtained using E_{MM} for energy minimization was somewhat worse than that obtained with other methods, whereas the PB method yielded relatively poor results for complex 2CGR. These conclusions were confirmed by visual inspection of the poses. In the case of PB, one might reasonably suggest that the relatively poor performance was due to incomplete convergence of the energy minimization process. However, increasing the number of optimization cycles from 500 to 750 and even to 1000 cycles did not change the result. It should be noted that these differences were small and had no impact on pose prediction in general.

For all of the methods examined, the results obtained by minimizing only the atoms in the binding site were similar to those obtained by minimizing the whole complex, indicating that it is sufficient to minimize only the atoms in the binding site (see Supporting Information). The CPU times used for the various GB methods were all similar and were approximately four times greater than that required for E_{MM}. The CPU time required by the PB implicit solvent model was 12 times greater than that required by the GB methods (see Supporting Information). The inclusion of an SA term did not influence the computational time. Since the PB model does not afford superior results to either E_{MM} or the GB models, the latter methods seem to be preferable in this case.

The geometry and energy of the optimized protein−ligand complexes may be sensitive to the number of optimization cycles performed in the energy minimization process. Thus, the extent of this sensitivity was examined for eight complexes, three of which (1C5C, 1KSN, and 2CGR) had significantly improved poses after minimization. It was found that, for these three complexes, the bulk of the improvement was achieved during the first 200 cycles; some combinations of complex and method did not achieve full convergence even after 1000 cycles. It was also found that for five of the eight complexes (1AJX, 1KSN, 1URG, 2CGR, and 2SIM), increasing the number of cycles to 500 or more resulted in *worse* poses compared to when fewer cycles were used. On the basis of these results, it was decided to standardize to the use of 300 cycles for the energy minimization of docked protein−ligand complexes.

Pose Selection Based on Calculated ΔG_{bind} Energies. If we assume that a crystal structure represents a relevant low-energy conformation of a protein−ligand complex, it follows that a binding pose that does not deviate much from the X-ray structure should have a lower binding energy than one that deviates more substantially. This idea can be reversed: given a set of poses, one could rescore them according to their computed binding energies, and the more negative a pose's binding energy, the more likely it is to match the true pose. Thus, a set of docking poses was generated for specific protein−ligand pairs. Two postprocessing steps were applied to the poses: energy minimization followed by calculation of the binding energy using E_{MM}, PB, or one of the three GB methods ± SA terms. We sought to determine whether the postprocessed pose with the lowest binding energy would also have the lowest RMSD value (and thus be most similar to the X-ray structure), and also whether there is a linear correlation between the binding energies and RMSD values of specific sets of energy-minimized poses.

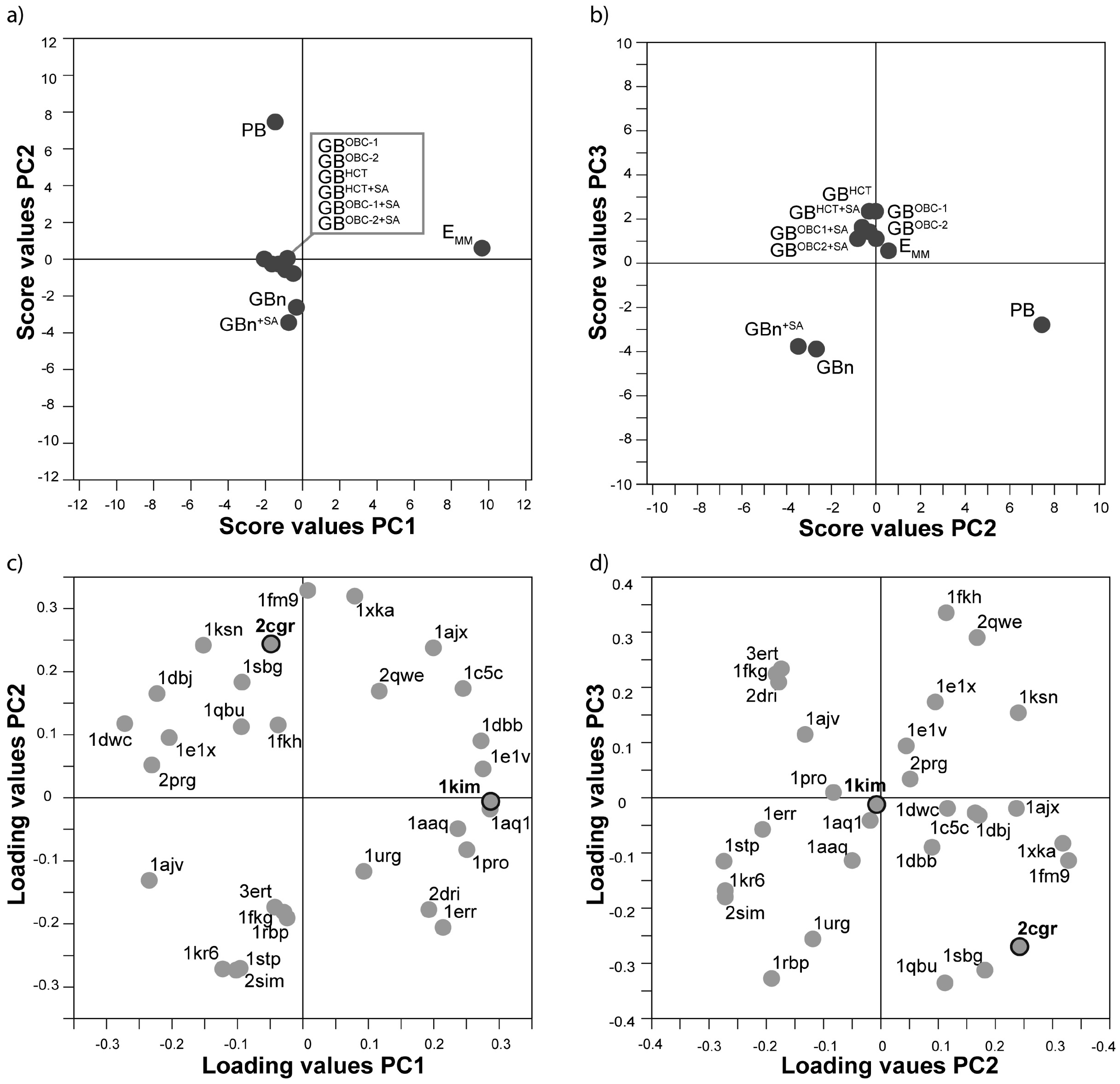


Figure 4. Visualization of the multivariate analysis of the RMSD differences between the crystal structures and poses that have been docked and energy minimized using various techniques. The variation between the methods extracted by PCA ($R^2X = 0.93$, PC's = 3) is displayed in the score plots (a and b) and their relationship with the included protein−ligand complexes is shown in the loading plots (c and d). The methods are denoted by the abbreviations used in the text and the protein−ligand complexes by their PDB codes. Protein−ligand complexes discussed in the text are highlighted.

Four of the methods used for calculating ΔG_{bind} (E_{MM}, PB, GB^{HCT}, and $\text{GB}^{\text{HCT+SA}}$) correctly rescored the pose with the lowest RMSD value for 7 out of the 12 complexes examined (i.e., when one of these methods was used to calculate the ΔG_{bind} of the poses, the pose with the most negative ΔG_{bind} was also that with the lowest RMSD value compared to the X-ray structure). The $\text{GB}^{\text{OBC-1}}$ and $\text{GB}^{\text{OBC-2}}$ (± SA terms) methods correctly rescored the poses of 5 out of the 12 complexes, and the GoldScore function, which ranks ligands without first minimizing their energy, correctly scored the poses with the lowest RMSD values in only 2 cases. There were no clear trends in these results in terms of connecting structural features of the complexes with the likelihood of being correctly scored by any one method.

With all of the methods examined, linear correlations were observed between the calculated values of ΔG_{bind} and the RMSD values of the poses of specific protein−ligand complexes. For example, using the PB method, R^2 values of 0.67, 0.87, and 0.79 were obtained for complexes 1C87, 1O39, and 1TNL, respectively. However, the strength of the correlation was rather variable; the PB method afforded R^2-values of only 0.01, 0.39, and 0.22 for complexes 1C88, 1O2Z, and 1TNI, respectively. The E_{MM}, PB, GB^{HCT}, and $\text{GB}^{\text{OBC-1}}$ (± SA terms) methods used for rescoring, and GoldScore gave similar results, with 5−7 of the 12 investigated protein−ligand complexes having R^2 values greater than 0.45. The $\text{GB}^{\text{OBC-2}}$ and $\text{GB}^{\text{OBC-2+SA}}$ methods afforded results that were clearly worse than those of other methods (only two complexes had R^2 values >0.45). For 5 of the 12 complexes, no correlation between the calculated binding energy and the docked ligand's geometrical similarity to the relevant crystal structure was observed. As before, there were no clear relationships between the structure of the ligands, their binding affinities, or the nature of their target proteins and the likelihood of

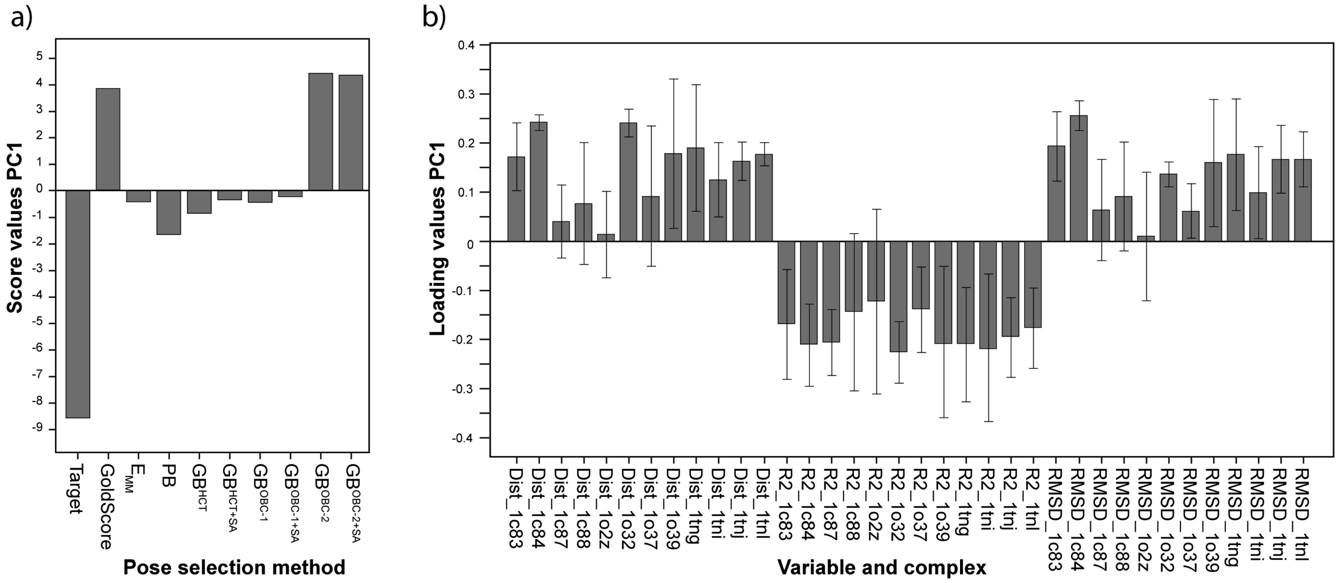


Figure 5. Results of the PCA to investigate the suitability of ranking binding poses according to ΔG_{bind} to identify those whose geometry best matches the crystal structure. The score plot (a) show the nine investigated conditions for selection of docking poses, three GB models ± SA terms, the PB method, in gas phase (E_{MM}), and the GoldScore scoring function. Target illustrates the desired direction of the score values for the investigated methods (see text for details). The loading plot (b) reveals how the three evaluation metrics (RMSD, R^2, and Dist) influence the pattern seen in part a. Conditions with low score values in part a also have relatively lower values of the metrics (variable and complex) seen in part b, and analogously, conditions with high score values also have relatively higher values of the evaluation metrics applied. Parts a and b show, for example, that the use of PB resulted in resulted in lower Dist values, higher R^2 values, and lower RMSD values for all of the investigated complexes compared to GoldScore and $\text{GB}^{\text{OBC-2}}$ ± SA terms (which in turn resulted in higher Dist values, lower R^2 values, and higher RMSD values compared to the other methods).

observing a strong correlation between the calculated RMSD and ΔG_{bind} values. One explanation for the deviation of some highly (re)scored poses from the crystal structure could be that some deviations are not associated with substantial energetic penalties. For instance, significant deviations may be observed in regions of the ligand that do not interact strongly with the protein if they are flexible and/or exposed to solvent. Additionally, the 3D structure of the protein−ligand complex used as a reference arises from an electron density map determined by X-ray crystallography and thus may contain uncertainties and errors that could affect the analysis. Nevertheless, the results show that for seven of the investigated complexes, there is a correlation between a pose's calculated ΔG_{bind} value and its similarity to the reference structure. This supports the use of this methodology for the selection and identification of representative binding poses.

The RMSD values (relative to the X-ray poses) of the poses with the lowest ΔG_{bind} values are noticeably lower than those of the docking poses selected by GoldScore, indicating that postprocessing of docked protein−ligand complexes affords binding poses that more closely match the crystal structures. For example, when using the GB^{HCT} model (± SA term) to compute ΔG_{bind}, the RMSD values of the lowest-energy poses of 7 of the 12 complexes were improved by 0.4 Å or more, and the RMSD values of the remaining 5 complexes were comparable to or slightly better than those of the poses most highly ranked by GoldScore. Similarly, when using the PB model, the RMSD values of six complexes were substantially improved and those of the remaining six were no worse.

The multivariate analysis of the data discussed above (i.e., the Dist, R^2, and RMSD values) by PCA yielded a first component that described the observed general trends ($R^2X = 0.35$) for all of the complexes studied (Figure 5). This can be seen in the loading plot, since all of the complexes had similar loading values for all three metrics (Figure 5b). The score values shown in the column plot indicate that PB was the best of the methods investigated for pose selection, followed by GB^{HCT}, $\text{GB}^{\text{OBC-1}}$ (± SA terms), and E_{MM}. In other words, these are the methods whose scores were most positively correlated with the constructed Target object. GoldScore and $\text{GB}^{\text{OBC-2}}$ (± SA term) are clearly suboptimal methods for the selection of binding poses as their score values are negatively correlated with the Target (Figure 5). Furthermore, the loading plot shows that this conclusion is supported by the results obtained for all three proteins and all three metrics (i.e., Dist, R^2, and RMSD values), confirming the general conclusion.

As the free energy of binding is only calculated from the complexed and unbound forms of the ligand and protein, much less computational time is required for this part of the calculations than for the energy minimization step. Thus, there is relatively little difference in the total CPU time required for postprocessing with the PB model versus the GB models, which supports the use of the PB model for this purpose. However, the GB^{HCT} model may be a useful alternative if larger numbers of poses per ligand are to be investigated.

Estimation of Binding Affinities of Factor Xa Inhibitors Based on Rescoring Using Relative Free Energy of Binding. The geometry optimization and pose selection schemes developed for the postprocessing of docked complexes, that is energy minimization of the binding site of five top-ranked docking poses using $\text{GB}^{\text{HCT+SA}}$ followed by estimation of binding affinity using MM-PB/GB-SA to select the pose with lowest binding energy, were applied to a set of ligands binding to Factor Xa. The aim was to determine whether this MM-PB/GB-SA protocol also could be used to estimate relative binding energies for a set of analogues that would correlate well with their experimentally determined pK_{i} values. The data set used is particularly challenging in that the pK_{i} values span a relatively small range (7.05−9.00). Eight methods (E_{MM}, PB, and three GB methods ± SA terms) were investigated and applied to the

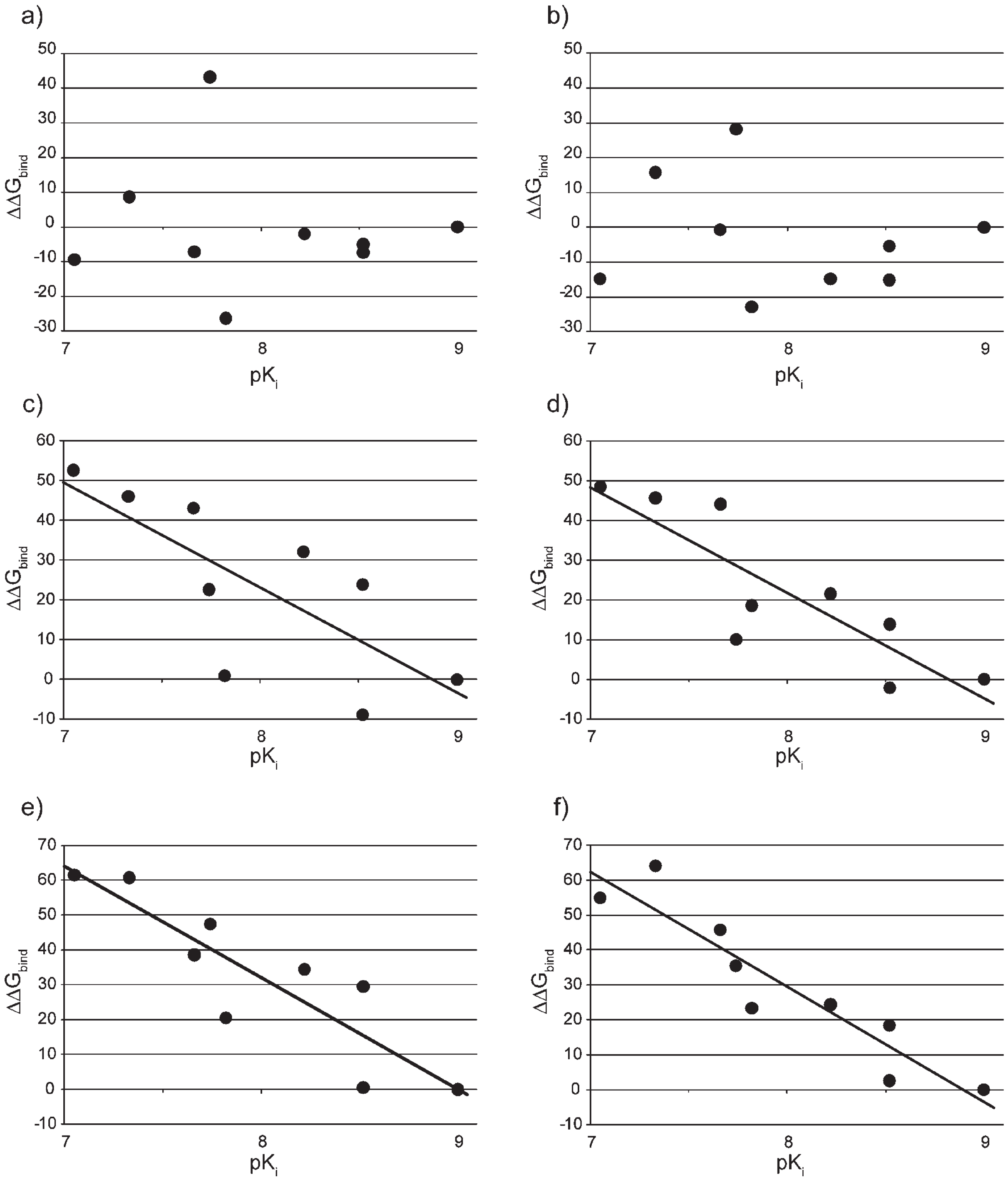


Figure 6. Correlations between calculated relative binding energies of Factor Xa—ligand complexes and their experimental binding data (pK_i values). The relative binding energies shown were calculated for poses that were initially subjected to energy minimization using the $GB^{HCT\text{-}SA}$ model; the poses' binding energies were then computed using molecular mechanics in the gas phase (a,b), with the GB^{HCT} model (c,d) and with the PB model (e,f). Three of the binding energies (a, c, and e) derive from postprocessing of the pose that was top-ranked by GOLD; the other three (b, d, and f) derive from postprocessing of the top five GOLD-ranked poses followed by selection of the optimized pose with the lowest free energy of binding.

five docking poses most highly ranked by GoldScore. The relative binding energies predicted by GoldScore were also included for comparative purposes, although the developers state that the scoring function should only be used for enrichment and not for direct comparisons of binding data.[36]

There was a clear linear correlation between the experimental data and the $\Delta\Delta G_{bind}$ values for the energy-minimized and rescored poses (Table 5 and Figure 6, Supporting Information). The most successful rescoring method was to use the PB solvation model, which afforded an R^2 of 0.78 for the postprocessed pose that had been top-scored by GoldScore. When the postprocessing was applied to the top five poses identified by GoldScore, the lowest-energy pose had an R^2 value of 0.87. Rescoring the poses by calculating their relative binding energies

Table 5. R^2 Values for the Correlations between the Calculated $\Delta\Delta G_{\text{bind}}$ Values and Experimental pK_i Values for Factor Xa−Ligand Complexes

scoring method	correlation[a] for top pose[b]	correlation[a] for lowest energy pose[c]
GoldScore[d]	0.08	
E_{MM}	0.01	0.03
PB	0.78	0.87
GB^{HCT}	0.56	0.73
$\text{GB}^{\text{HCT+SA}}$	0.55	0.75
$\text{GB}^{\text{OBC-1}}$	0.54	0.69
$\text{GB}^{\text{OBC-1+SA}}$	0.53	0.68
$\text{GB}^{\text{OBC-2}}$	0.02	0.04
$\text{GB}^{\text{OBC-2+SA}}$	0.01	0.04

[a] The correlations are expressed in terms of their R^2 values. [b] The docking pose that was highest ranked by GoldScore within GOLD. [c] The docking pose with the lowest ΔG_{bind} out of the top five GoldScore-ranked poses. [d] Force field-based scoring function included in the docking program GOLD, included for comparative purposes.

in the gas phase (E_{MM}) did not show any correlation with experimental data, nor did the scoring values resulting from GoldScore. The relative binding energies calculated using the GB^{HCT} and $\text{GB}^{\text{OBC-1}}$ (± SA terms) models showed linear correlations with experimental data; both models afforded R^2 values of approximately 0.55 for the postprocessed top-scoring pose identified by GoldScore and 0.70 when the lowest energy postprocessed pose was selected from the top five GoldScore poses. The use of the $\text{GB}^{\text{OBC-2}}$ (± SA terms) model to calculate solvation effects did not afford relative binding energies that correlated with the pK_i values.

Thus, the use of an implicit solvation model to calculate relative binding energies has a profound effect on the results obtained using the MM-PB/GB-SA protocol, affording a strong linear correlation with experimental binding data. Furthermore, the results show that for the Factor Xa data set, postprocessing of five highly ranked docking poses resulted in stronger correlations between ΔG_{bind} values and experimental data than simply postprocessing and rescoring the pose that was most highly ranked by GoldScore (Table 5, Figure 6). This effect was seen

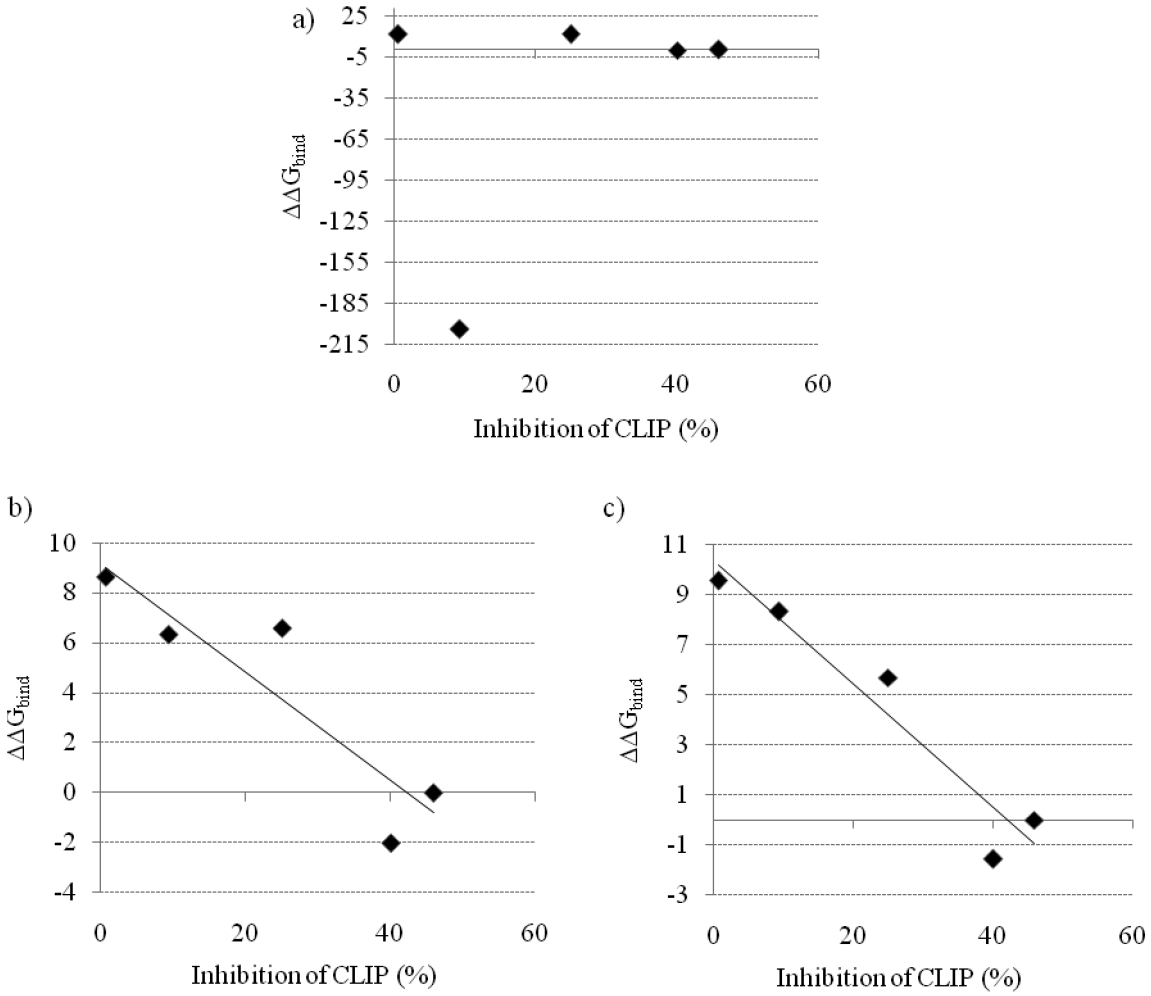


Figure 7. Correlation between the calculated relative binding energies for a series of modified glycopeptide ligands that bind to the MHC class II molecule A^q and experimental binding data determined at 4 μM concentration (see text for details). The structures employed when calculating the relative binding energies were selected as follows: first, the five poses having the lowest backbone RMSD relative to the reference peptide were selected. These were then subjected to energy minimization using the $\text{GB}^{\text{HCT-SA}}$ model and ranked according to their free energy of binding to the protein target. The pose having the lowest binding energy was selected and its binding energy was recalculated with one of the following methods: (a) using molecular mechanics in the gas phase, (b) with the GB^{HCT} solvation model, and (c) with the PB model.

Table 6. Calculated R^2 Correlations between the Calculated $\Delta\Delta G_{\mathrm{bind}}$ Values and Experimental Inhibition Data for a Set of Modified Glycopeptide Ligands That Bind to the Class II MHC A^q Protein

scoring method	correlation[a] for lowest energy pose[b]
E_{MM}	−0.15
PB	0.92
$\mathrm{GB}^{\mathrm{HCT}}$	0.82
$\mathrm{GB}^{\mathrm{HCT+SA}}$	0.81
$\mathrm{GB}^{\mathrm{OBC\text{-}1}}$	0.84
$\mathrm{GB}^{\mathrm{OBC\text{-}1+SA}}$	0.82
$\mathrm{GB}^{\mathrm{OBC\text{-}2}}$	0.36
$\mathrm{GB}^{\mathrm{OBC\text{-}2+SA}}$	0.36

[a] The correlations are expressed in terms of their R^2 values. [b] The docking pose with the lowest ΔG_{bind} out of the top five ranked by lowest backbone RMSD compared to the reference peptide.

for all estimations of relative binding energies using implicit models that gave a linear correlation to the experimental data, i.e. the PB, $\mathrm{GB}^{\mathrm{HCT}}$, and $\mathrm{GB}^{\mathrm{OBC\text{-}1}}$ (± SA terms) solvation models. The inclusion of an SA term in the calculations using the GB models had no significant effect on the quality of the calculated relative binding energies. An investigation of the geometries of the poses showed that the poses selected based on lowest binding energies were in general geometrically more similar to the X-ray poses (based on RMSD and Dist values—see the Supporting Information) compared to the top-ranked poses, which confirm the results seen with data set 2.

Application of the Protocol for the Postprocessing of Docking Poses to Glycopeptidomimetics Targeting the Class II MHC A^q Protein. The five selected docking poses for each of the five truncates corresponding to the glycopeptides binding to the class II MHC A^q protein (Figure 1) were subjected to geometry optimization by energy minimization as previously described (the positions of binding site atoms were optimized using the $\mathrm{GB}^{\mathrm{HCT}}$ solvation model with an SA term, and 300 optimization cycles were performed). The resulting five poses for each of the five ligands were thereafter rescored using PB and three GB models ($\mathrm{GB}^{\mathrm{HCT}}$, $\mathrm{GB}^{\mathrm{OBC\text{-}1}}$, and $\mathrm{GB}^{\mathrm{OBC\text{-}2}}$ ± SA terms). The poses with the lowest relative binding energies for each of the five ligands were then compared to the experimentally determined inhibition data. The calculated relative binding energies of the truncated glycopeptidomimetics correlated well with the ligands' ability to inhibit the binding of CLIP to the A^q protein (Table 6 and Figure 7, Supporting Information). Optimal results were obtained by using the PB model to estimate the free energy of binding; this afforded an R^2 value of 0.92. The $\mathrm{GB}^{\mathrm{HCT}}$ and $\mathrm{GB}^{\mathrm{OBC\text{-}1}}$ models afforded R^2 values of 0.82 and 0.84, respectively. The calculations of relative binding energies without inclusion of a solvation model or using the $\mathrm{GB}^{\mathrm{OBC\text{-}2}}$ model resulted in poor correlations. The poor correlation in the absence of a solvation model is probably due to overestimation of the relative binding energy of the charged aminomethylene isostere fragment in the gas phase (Figure 7a and Figure 1, structure **3**). The addition of SA terms in the calculations had no effect on the relative binding energies.

CONCLUSIONS

We have investigated the effects of applying implicit solvation models on geometry optimization, pose selection, and estimation of the relative binding energies of docked protein−ligand complexes. The time consumed by geometry optimization can be greatly reduced by performing energy minimization with relatively undemanding GB solvation models ($\mathrm{GB}^{\mathrm{HCT}}$ or $\mathrm{GB}^{\mathrm{OBC\text{-}1}}$ ± SA terms) and focusing on optimizing only atoms within the binding sites for 300 cycles rather than using the expensive PB model and optimizing the geometry of the whole complex for 500 cycles. The presented time-efficient postprocessing protocol involves first selecting the five most highly ranked docking poses, and reoptimizing their geometries using the $\mathrm{GB}^{\mathrm{HCT+SA}}$ model. The poses' free energies of binding are then calculated using the PB model, and the poses with lowest binding energy are taken to be representative of the real bound ligand. In general, these poses have geometries similar to those observed in corresponding crystal structures, and the relative free energies of binding obtained in this way using implicit solvation models exhibit strong correlations with experimentally determined binding data. These results indicate that the presented protocol for the postprocessing of docked protein−ligand complexes developed in this paper may be generally useful for structure-based design in drug discovery.

ASSOCIATED CONTENT

Ⓢ **Supporting Information.** Additional information concerning data sets 1 and 2. Tables of parameter settings used when performing docking with GOLD for data sets 1 and 2. RMSD values for data set 1 before and after energy minimization, CPU times for the energy minimizations, and data on the effects of varying the number of cycles on minimization. The full contents of the Dist, R^2, and RMSD value matrices for data set 2 used to investigate pose selection by ranking in terms of free energy of binding. Plots of the relative binding energies versus experimental binding data for Factor Xa and MHC using methods not discussed in the main body of the paper. RMSD and Dist values for data set 3. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anna.linusson@chem.umu.se.

Present Addresses

[§]Department of Radiation Sciences, SE-901 87, Umeå University, Umeå, Sweden.

[⊥]Norwegian Institute for Water Research (NIVA), Gaustadalleen 21, NO-0349, Oslo, Norway.

ACKNOWLEDGMENT

The authors thank the Swedish Research Council and AstraZeneca R&D Mölndal for financial support.

REFERENCES

(1) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.

(2) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.

(3) Talele, T. T.; Khedkar, S. A.; Rigby, A. C. Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Curr. Top. Med. Chem.* **2010**, *10*, 127–141.

(4) Andersson, C. D.; Thysell, E.; Lindström, A.; Bylesjö, M.; Raubacher, F.; Linusson., A. A multivariate approach to investigate docking parameters' effects on docking performance. *J. Chem. Inf. Model.* **2007**, *47*, 1673–1687.

(5) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.

(6) Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomolec. Struct.* **2007**, *36*, 21–42.

(7) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

(8) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(9) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(10) Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.

(11) Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412–1419.

(12) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.

(13) Lee, M. C.; Yang, R.; Duan, Y. Comparison between Generalized-Born and Poisson-Boltzmann methods in physics-based scoring functions for protein structure prediction. *J. Mol. Model.* **2005**, *12*, 101–110.

(14) Brown, S. P.; Muchmore, S. W. High-throughput calculation of protein-ligand binding affinities: Modification and adaptation of the MM-PBSA protocol to enterprise grid computing. *J. Chem. Inf. Model.* **2006**, *46*, 999–1005.

(15) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 243–253.

(16) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–6606.

(17) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.

(18) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865–7877.

(19) Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48*, 1081–1091.

(20) Rastelli, G.; Degliesposti, G.; Del Rio, A.; Sgobba, M. Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chem. Biol. Drug Des.* **2009**, *73*, 283–286.

(21) Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **2008**, *48*, 958–970.

(22) Brown, S. P.; Muchmore, S. W. Rapid estimation of relative protein-ligand binding affinities using a high-throughput version of MM-PBSA. *J. Chem. Inf. Model.* **2007**, *47*, 1493–1503.

(23) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Perlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California, San Francisco, 2006.

(24) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47*, 2977–2980.

(25) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119.

(26) Andersson, I. E.; Dzhambazov, B.; Holmdahl, R.; Linusson, A.; Kihlberg, J. Probing molecular interactions within class II MHC A(q)/Glycopeptide/T-cell receptor complexes associated with collagen-induced arthritis. *J. Med. Chem.* **2007**, *50*, 5627–5643.

(27) Champness, J. N.; Bennett, M. S.; Wien, F.; Visse, R.; Summers, W. C.; Herdewijn, P.; de Clercq, E.; Ostrowski, T.; Jarvest, R. L.; Sanderson, M. R. Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. *Proteins Struct. Funct. Bioinf.* **1998**, *32*, 350–361.

(28) Nolte, R. T.; Wisely, G. B.; Westin, S.; Cobb, J. E.; Lambert, M. H.; Kurokawa, R.; Rosenfeld, M. G.; Willson, T. M.; Glass, C. K.; Milburn, M. V. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature* **1998**, *395*, 137–143.

(29) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927–937.

(30) Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Öhman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753–758.

(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shndyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(32) *Reduce*, version 3.0; The Richardson Laboratory: Duke University, Durham, NC, 2008.

(33) *JACKAL*, version 1.5; The Honig Laboratory: Columbia University, New York, 2008.

(34) Kleywegt, G. J. Crystallographic refinement of ligand complexes. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2007**, *63*, 94–100.

(35) Molecular networks GmbH. CORINA. http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html (accessed January 2008).

(36) *GOLD*, version 3.1.1; The Cambridge Crystallographic Datacenter: Cambridge, U.K., 2008.

(37) *SYBYL*, version 7.2; Tripos Inc.: St. Louis, MO, 2008.

(38) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(39) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, R.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(40) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the Resp model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(41) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(42) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(43) Leach, A. R. *Molecular modelling - Principals and applications*, 2nd ed.; Pearson education limited: Harlow, England, 2001; p 744.

(44) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules. *Annu. Rev. Biophys. Biophysic. Chem.* **1990**, *19*, 301–332.

(45) Davis, M. E.; McCammon, J. A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* **1990**, *90*, 509–521.

(46) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2001**, *56*, 275–291.

(47) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.

(48) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(49) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(50) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes witha modified generalized Born model. *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 383–394.

(51) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born model with simple, robust molecular volume correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.

(52) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2002**, *23*, 1244–1253.

(53) Lu, Q.; Luo, R. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.* **2003**, *119*, 11035–11047.

(54) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.

(55) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(56) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Application of a pairwise generalized Born model to proteins and nucleic acids: Inclusion of salt effects. *Theor. Chem. Acc.* **1999**, *101*, 426–434.

(57) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.

(58) Jackson, J. E. *A users guide to principal components*; John Wiley & sons, Inc.: New York, 1991.

(59) *SIMCA-P+*, version 12; Umetrics: Umeå, Sweden, 2009.

(60) *Evince*, version 2.0; Umbio AB: Umeå, Sweden, 2009.

(61) *FRED*, version 2.2.3; OpenEye Scientific Software: Santa Fe, NM, 2009.

(62) *OMEGA*, version 2.2.0; OpenEye Scientific Software: Santa Fe, NM, 2009.

(63) *MOE*, version 2007.09; Chemical Computing Group: Montreal, Canada, 2009.