

Predicting the Binding Mode of Known NCp7 Inhibitors To Facilitate the Design of Novel Modulators

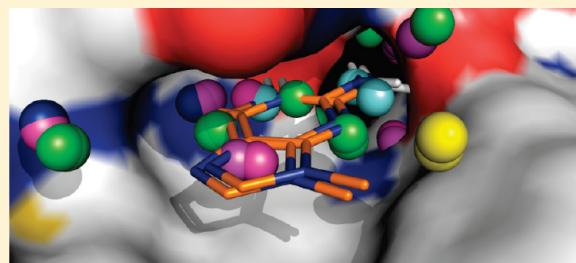
Mattia Mori,[†] Fabrizio Manetti,[†] and Maurizio Botta^{*,†,‡}

[†]Dipartimento Farmaco Chimico Tecnologico, University of Siena, via A. Moro 2, I-53100 Siena, Italy

[‡]Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, BioLife Science Building, Suite 333, 1900 N. 12th Street, Philadelphia, Pennsylvania 19122, United States

 Supporting Information

ABSTRACT: The HIV-1 nucleocapsid protein (NCp7) is an emerging target for antiretroviral therapy. Five hits have been reported to inhibit the NCp7–viral nucleic acids interaction at micromolar concentrations. We used two computationally refined structures of NCp7 as receptors to propose a reliable binding pose for these compounds, by means of computational methods. Theoretical binding modes are in agreement with available experimental data. Results lay the foundations for a rationale development of more effective NCp7 inhibitors.



■ INTRODUCTION

The HIV-1 nucleocapsid protein-7 (NCp7) is a small and highly basic zinc-binding protein that is strongly involved in many processes of the viral life cycle,^{1,2} operating the strand transfer that promotes the activity of the reverse transcriptase (RT),³ as well as participating in the specific genome encapsidation by binding to the different stem loops of the RNA packaging signal ψ (PSI-RNA) and enhancing the final release of new mature virus particles.⁴ For these reasons, NCp7 is being considered a promising target for the identification of new anti-HIV-1 therapeutic agents. A recent high throughput screening (HTS) of a small compound library led to the identification of five molecules, able to interact with NCp7 and to compete with the binding of the viral genome without promoting zinc ejection.⁵ These hit compounds have a low molecular weight and do not share a common chemical scaffold, and, although showing a very similar micromolar inhibitory activity toward NCp7, at the state of the art they represent the only reliable and valuable starting point for further studies.

Recently, we applied a computational protocol based on density functional theory (DFT) and molecular dynamics (MD) simulations to the complexes between NCp7 and PSI-RNA or PBS-DNA,⁶ derived from NMR experiments and deposited in the protein data bank (PDB).⁷ The resulting theoretical 3D structures, differently from the original NMR models, are in very good agreement with a plethora of biological data and, therefore, could be considered as reliable tools for *in silico* structure-based ligand design approaches. Here, these computationally derived NCp7 structures were used as receptors to model the possible binding conformations of the five hit compounds (namely **1–5**, whose molecular structures are reported in Figure 1, and main features are described in Table 1)⁵ and guanine, taken as representative of the

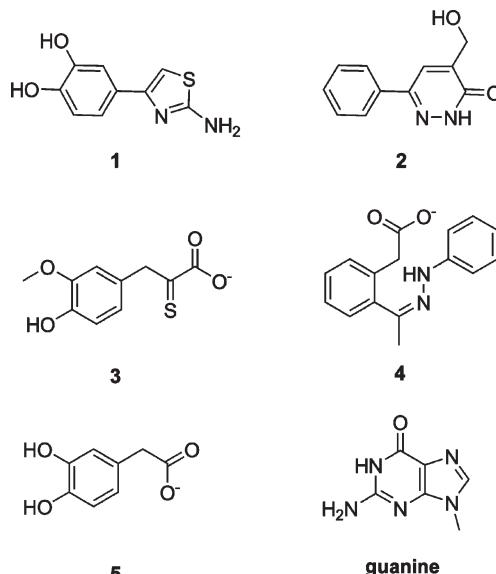


Figure 1. Chemical structures of **1–5** and guanine. Ionization states reported here were selected for docking simulations and were predicted for being representative of physiological pH values ($\text{pH } 7.0 \pm 1.0$).

nucleic acids. The main aim of the work is to propose a reliable binding mode for these compounds toward NCp7, able to account for the available experimental data (i.e., SAR) and usable to facilitate the identification of novel NCp7 modulators by means of virtual screening or rational design.

Received: October 8, 2010

Published: December 20, 2010

Table 1. Main Features of Compounds 1–5 and Guanine

compound	name in ref 5	molecular weight	number of rotatable bonds	number of heavy atoms
1	A10	208.24	4	14
2	C07	202.07	3	15
3	E03	226.20	5	15
4	H02	268.32	5	20
5	H04	190.02	4	12
guanine		165.06	0	12

Molecular interaction fields (MIFs) were first calculated within the putative ligand binding site of NCp7. Next, three docking programs were applied to explore the possible binding conformations of the small molecules toward NCp7, and to identify the molecular determinants mostly involved in protein–ligand interaction. Finally, the calculated docking energies, and scoring and rescore values as well, were compared with experimental affinity values, to strengthen the reliability of the applied computational protocol and of the binding modes predicted for 1–5.

COMPUTATIONAL METHODS

Selection of NCp7 Structures. PDB 1A1T⁸ and PDB 2JZW⁹ were selected from the RCSB-PDB⁷ as representative for the complex between NCp7 and RNA (5'-GGACUAGCGGAGG-CUAGUCC-3') and DNA (5'-GTCCCTGTCGGGC-3'), respectively, as described elsewhere.⁶ Coordinates of the lowest energy model with the least restraint violation of each NMR structure were selected for this study and were considered as representative for all NMR models (see Supporting Information). MD-based structures of the same complexes were retrieved from a previous computational work.⁶ The frame with the lowest root-mean-square deviation (rmsd) with respect to the average structure of each MD production was selected and used in this study, upon energy minimization. Nucleic acids, water molecules, and counterions were removed from each complex, where present.

Relaxation of the PDB Structures. The two NMR-based structures 1A1T and 2JZW were pretreated with the Protein Preparation Wizard included into the Maestro suite,¹⁰ with the aim to relax atom coordinates until the geometric convergence (0.01 Å rmsd). The OPLS2005 force field was used.¹¹

Calculation of Molecular Interaction Fields (MIFs). The GRID suite was applied to compute energy minima for the interaction of representative probe atoms on the accessible surface of the NCp7.^{12–14} MIFs were computed for all the four protein structures by using the following parameter set: number of planes for Angstrom was set to 8 (NPLA=8); the grid center was set on the W37 residue, and the box dimensions were 20 × 20 × 20 Å along the x, y, z axes. GRID single atom probes O (sp2 carbonyl oxygen), O:: (sp2 carboxy oxygen atom), OH (phenol or carboxy OH), DRY (the hydrophobic probe), C1=(sp2 CH aromatic or vinyl), N1 (neutral flat NH), and N2 (neutral flat NH₂) were used. For each probe atom, coordinates of GRID points showing a probe interaction energy lower than −2.00 kcal · mol^{−1} were selected and converted into ASCII-compatible energy maps by the k2f program included in the GRID suite. These maps were used to describe molecular interaction fields calculated within the putative ligand binding site of both NMR- and MD-based NCp7 structures.

Docking with AutoDock. The AutoDock 4.2.3 program¹⁵ was used to predict the binding conformation and the free energy of interaction for 1–5 and guanine. Molecular structures of 1–5 were retrieved from the literature.⁵ Each compound was energy minimized by means of the MacroModel program¹⁶ implemented in the MAESTRO Schrödinger software suite, using the OPLS2005 force field.¹⁷ 3D ligand structures were then prepared for docking calculations by means of the AutoDockTools program (version 1.5.6)¹⁵ through the addition of atomic Gasteiger partial charges and by the automatic selection of the active rotatable bonds.

AutoDockTools was also used for the calculation of potential interaction grid maps for all NCp7 structures used as rigid receptors during docking simulations. A grid box of 70 × 70 × 70 points with a 0.375 Å point spacing was centered in the mass-center of residues K33, K34, G35, W37, M46, and Q45. This box covers most of the accessible surface of each NCp7 structure used in this work. Potential grid maps for the interaction of ligand atom-types were then calculated with the auxiliary program AutoGrid. Docking calculations were performed for 250 hybrid GA-LS runs, performing a maximum of 2.5 M energy evaluations for a maximum of 27 000 generations. Initial ligand coordinates and degrees of freedom were set as random, whereas a rmsd tolerance of 2 Å was set for cluster grouping. All other parameters were kept as default, and the protein rotatable bonds were kept frozen during the flexible docking of ligands.

Docking with GOLD. The GOLD program version 4.1.2,¹⁸ distributed by the Cambridge Crystallographic Data Center, was used. Compounds 1–5 and guanine were docked toward the surface of the four NCp7 structures. The ligand binding site was defined as the space region having a radius of 15 Å, centered on the CD2 atom of the W37 side chain. The flexibility of side chains was turned off, whereas the automatic detection of the cavity was enabled. The search efficiency of the Genetic Algorithm was set at 200%. Both the ChemScore^{19,20} and the GoldScore^{21,22} functions were employed for the docking and the rescore processes. In detail, two different protocols were followed: (i) docking with ChemScore and rescore with GoldScore and (ii) docking with GoldScore and rescore with ChemScore. A consensus “score by score” protocol was also used for the analysis of docking results.²³

Docking with Glide. The Glide program²⁴ is currently implemented as a module of the Maestro Schrödinger software suite, released by the Schrödinger LLC. The program version 5.5 (revision 1.76) was used in this work, both in the Standard Precision mode (SP) and in the Extra Precision mode (XP). The enclosing box, where ligands 1–5 and the guanine moiety have been confined, was centered on the CD2 atom of the W37 side chain. Parameters for receptor grid generation and docking calculations were kept at their default value for both the SP and XP modes.

Enrichment Rate of the DUD (Directory of Useful Decoys). The “All DUD Decoy” benchmarking data set was enriched with compounds 1–5.²⁵ The resulting library was then docked toward both the NCp7 conformations, by means of the GOLD program. The binding site was defined as the space region centered on the CD2 atom of the W37 residue, having a radius of 15 Å. Docking accuracy was set to the 50%, in order to reduce the computational cost. Both the ChemScore and the GoldScore functions were employed for the docking process.

Rescoring Docking Results with the MM-PBSA Module of AMBER10. The theoretical binding poses selected for compound

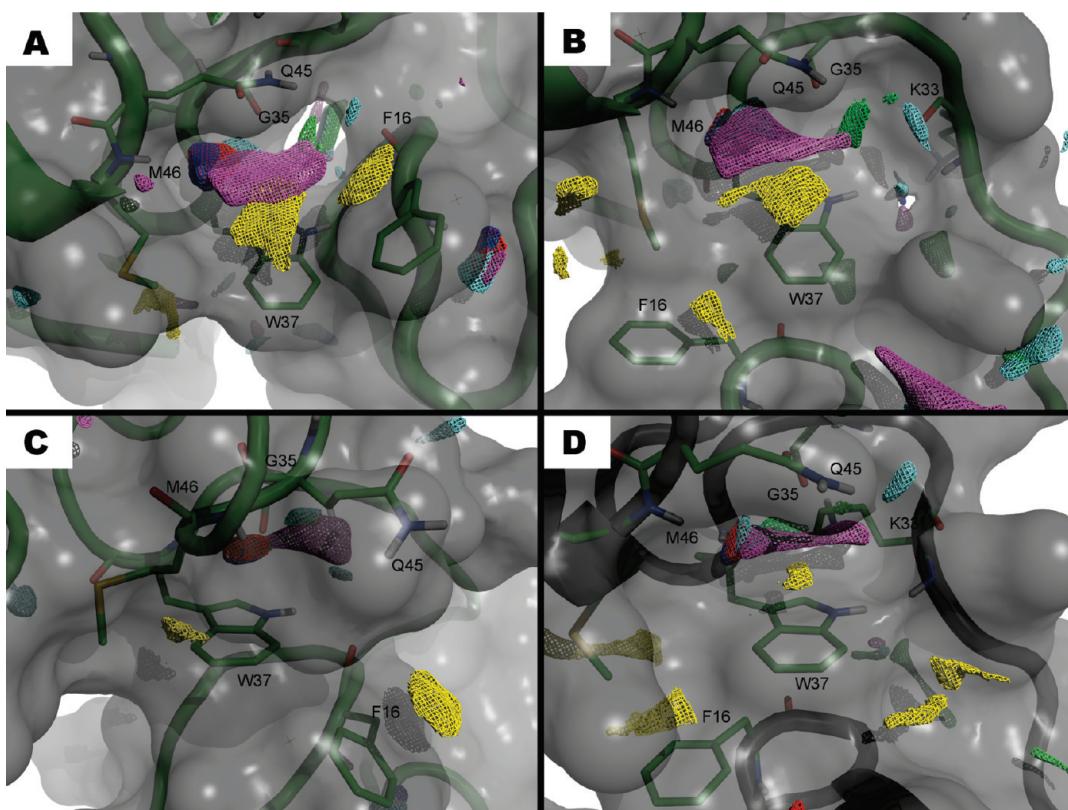


Figure 2. Molecular interaction fields calculated for the four NCp7 structures. Residues bounding the lipophilic pocket are labeled. MIFs are color coded: yellow = hydrophobic; magenta = hydrophobic-aromatic; green and cyan = H-bond donor; blue and red = H-bond acceptor. (A) MD-based structure of (RNA)NCp7, and (B) (DNA)NCp7. C structure of the NCp7 protein in PDB 1A1T and D) in PDB 2JZW.

1–5 were rescored with an external scoring function. The MM-PBSA perl script^{26,27} implemented in the AMBER10 program²⁸ was used to solve the Poisson–Boltzmann equation and to calculate the delta energy of binding of each ligand toward NCp7. Coordinates of each protein–ligand complex were retrieved from docking calculations. The protein was parametrized with the ff03 AMBER force field.²⁹ Additional parameters for the two zinc ions and for zinc-binding residues were retrieved from a previous computational work.⁶ Ligands were treated with the General Amber Force Field (GAFF) for organic molecules. Systems were energy minimized in implicit water solvent for 200 steps by using the steepest descendent (SD) algorithm and further for 300 steps by using the conjugate gradient (CG) method. These short relaxations are principally aimed at resolving bad contacts and possible steric clashes and to drive each complex to the closest local energy minimum. After local optimization, input coordinates for the MM-PBSA calculation were generated by a second energy minimization process lasting two steps SD and three steps CG. In this way, the delta energy of binding can be considered as a current energy value for the docked conformation, relaxed until the local minimum. For resoring purposes, the delta-PBTOT values were taken into account.

RESULTS AND DISCUSSION

Description of Receptor Structures. Structures of the complexes between NCp7 and PSI-RNA or PBS-DNA obtained by MD simulations,⁶ and the corresponding NMR-based models (PDBs 1A1T⁸ and 2JZW⁹) were used, upon removing coordinates of nucleic acids, counterions, and water molecules, where present. The MD-based conformation of NCp7 taken from the

complex with DNA is thereafter referred to as (DNA)NCp7, while the MD-based conformation of NCp7 taken from the complex with RNA is thereafter referred to as (RNA)NCp7. The putative ligand binding site is constituted by a small lipophilic pocket and a flanking groove on the protein surface, mainly composed of hydrophobic and basic residues, and it is placed in correspondence to the W37 residue that is crucial for nucleic acids binding, as suggested by biological evidence and MD calculations.⁶

Analysis of Molecular Interaction Fields. To identify the structural determinants mostly involved in ligand–protein interactions, GRID was applied to calculate MIFs of NCp7 extracted from each complex. Representative single atom probes were selected (see Computational Methods) to calculate the interaction energy of the most common chemical moieties of small organic compounds. Results clearly showed that MIFs are mainly accommodated in proximity to W37 of the MD-based models of NCp7 (Figure 2A, 2B). In particular, an extended and flat region accounting for the interactions of aromatic rings was found over the indole ring of W37 (Figure 2, magenta), and it is flanked by a hydrophobic region that is able to reach F16. Energy minima of the polar moieties, such as hydrogen bond donors (Figure 2, green and cyan) and acceptors (Figure 2, blue and red), were found in the inner part of the lipophilic cavity. In detail, the most profitable regions for hydrogen bond donors were found in close proximity to the backbone atoms of F16, K34, and G35 for (RNA)NCp7, and of K33 and G35 for (DNA)NCp7. On the other hand, MIFs for hydrogen bond acceptor groups are mostly accommodated close to the backbone atoms of W37 and M46. It is important to note that, despite the significantly different 3D

organization of the MD-derived NCp7 conformation in complex with RNA or DNA, MIFs found within the lipophilic pocket are very similar because of the highly similar folding of this protein region.⁶ Moreover, MIFs orientation finely reflects the binding mode of the guanine nucleotide of both PSI-RNA and PBS-DNA, as described in the complexes derived from MD calculations.⁶ On the contrary, MIFs calculated on the NMR-based structures of NCp7 are in disagreement with the binding mode of the guanine moiety of RNA and DNA. In the binding conformation to DNA, the best MIF regions showed a disposition similar to that found for the MD models, although characterized by a lower energy of interaction (Figure 2D versus 2A and 2B). Differently, the 3D arrangement of the ligand binding site of the complex with RNA completely precludes the accessibility of the W37 aromatic ring to the solvent, resulting in a displacement of MIFs slightly distant from the W37 side chain (Figure 2C).

This preliminary analysis of MIFs suggests that the MD-based structures of NCp7 could represent a reliable tool for the design of ligands that interact in correspondence to the W37 side chain of the lipophilic pocket, better than the corresponding NMR structures. Moreover, further docking calculations provided additional details on the higher reliability of computational NCp7 structures in comparison to the corresponding NMR models for receptor-based ligand design purposes.

Docking Calculations. Docking calculations were performed with AutoDock, GOLD, and Glide to identify the possible binding mode of **1–5** and guanine toward both NMR- and MD-based NCp7 structures. ChemScore and GoldScore were both applied as docking and scoring functions in GOLD, whereas ChemScore and GlideScore were used in Glide. The default force field was used in AutoDock.

Since no experimental structure of the complex between NCp7 and small ligands are available, the reliability of these software was preliminarily evaluated, by checking for their capability to retrieve a binding mode of guanine comparable to that previously found in the MD-derived NCp7–RNA and –DNA complexes. This approach is supported by the evidence that computational models were reported to have a significantly higher agreement with experimental data than the corresponding NMR-based structures.⁶ Satisfactory results were obtained by docking guanine toward MD-based NCp7 receptor structures (rmsd between MD-based and docked conformation of guanine was lower than 1.0 Å). On the contrary, poor results were obtained by all docking software toward the NMR-based NCp7 structure taken from PDB 1A1T (rmsd between MD-based and docked conformation of guanine is higher than 4.0 Å), while only Glide with the extra precision setting (XP) and AutoDock were able to identify the correct binding pose of guanine on the NCp7 structure taken from PDB 2JZW. Since these NMR models suffered from several structural limitations, these structures were not further used for docking **1–5**.

Docking of guanine and **1–5** toward both the MD-based NCp7 structures showed that ligands take place preferentially within the lipophilic pocket of the binding site. All docking programs provided a unique binding mode for the guanine moiety on the two MD-based conformations of NCp7 (Figures 3A, 4A), characterized by a hydrophobic interaction with the W37 aromatic ring. Parallelism, reciprocal orientation of the aromatic rings, and the distance between centroids suggest a possible π – π interaction.^{30,31} Polar contacts are found between guanine and the backbone atoms of G35, W37, and M46. An additional hydrogen bond is made with the F16 backbone in the

(RNA)NCp7 conformation, and with the K33 backbone in the (DNA)NCp7 conformation. All these interactions were already described for guanine and are observed here also for **1–5**.⁶ Moreover, this interaction pattern is in very good agreement with MIFs, thus providing further support to the reliability of the binding mode predicted for these compounds.

The best docked pose of the 4-(2-amino-1,3-thiazol-4-yl)-benzene-1,2-diol **1** ($K_i = 11.0 \pm 1.0 \mu\text{M}$)⁵ showed the aminothiazole ring accommodated within the inner side of the lipophilic pocket and making two hydrogen bonds with the G35 and F16 carbonyl groups of (RNA)NCp7 (Figure 3B), or G35 and K33 of (DNA)NCp7 (Figure 4B). Moreover, hydrophobic interactions occurred with the aromatic rings of W37 and F16. This binding mode accounts for the evidence that alkylation of the thiazole ring induces in vitro a complete loss of inhibitory activity toward the NCp7–genome interaction, because of the steric hindrance of the alkylated thiazole nucleus within the lipophilic pocket. The most active compound currently available to inhibit the binding of NCp7 toward the viral genome is the 4-(hydroxymethyl)-6-phenyl-2,3-dihydropyridazin-3-one **2** ($K_i = 8.5 \pm 0.9 \mu\text{M}$).⁵ The unique binding mode of this compound within the lipophilic pocket showed hydrogen bond interactions between the pyridazinone ring and F16, and between the hydroxy group and M46 of the (RNA)NCp7 (Figure 3C). In a similar way, the pyridazinone moiety made hydrogen bond contacts with G35 and M46 of the (DNA)NCp7, whereas the hydroxy group pointed toward the solvent (Figure 4C). The key role of the hydroxy group found by docking simulations is also supported by the fact that its replacement with a β -phenylethyl chain decreased the inhibitory activity, probably due to steric hindrance of this group that could strain the compound. Finally, in both conformations of NCp7, the side chain of W37 made hydrophobic interactions with the pyridazinone nucleus, in agreement with MIFs. Docking calculations on 3-(4-hydroxy-3-methoxyphenyl)-2-sulfanylidene propanoic acid **3** ($K_i = 10.5 \pm 0.9 \mu\text{M}$)⁵ identified a unique binding mode toward (RNA)NCp7. The usual contact with the aromatic portion of W37 is found. The hydroxy group made hydrogen bonds with G35 and M46 (Figure 3D), while the carboxylate moiety interacted with the side chain of K47. This interaction was also found for the sole docking pose predicted by Glide for **3** toward (DNA)NCp7. Differently, AutoDock and GOLD identified a binding conformation that allowed the carboxylate group of the ligand to interact with the K34 side chain of the protein groove flanking the lipophilic pocket (Figure 4D). The most large and flexible compound is the 2-[2-[1-(2-phenylhydrazin-1-ylidene)ethyl]phenyl]acetic acid **4** ($K_i = 13.0 \pm 1.0 \mu\text{M}$).⁵ Docking calculations provided the same binding mode toward both (RNA)NCp7 and (DNA)NCp7. The central aromatic ring interacted with W37, while the NH and the carboxylate terminus contacted the carbonyl and the NH moiety of the Q45 backbone. The monosubstituted phenyl ring was accommodated within the protein groove and pointed toward the solvent. All docking programs found very similar orientations of **4**, with the exception of AutoDock that replaced the Q45 side chain with R32 in the interaction with (DNA)NCp7 (Figures 3E, 4E). Finally, the hydroxy groups of the 2-(3,4-dihydroxyphenyl)acetic acid **5** ($K_i = 15.0 \pm 1.0 \mu\text{M}$)⁵ interacted with the backbone of F16, G35, W37, and M46 of the inner part of the lipophilic pocket. Hydrophobic contacts were found between the phenyl ring and W37 (Figures 3F, 4F). Binding conformations of **1–5** described here were retrieved from the highest populated clusters of

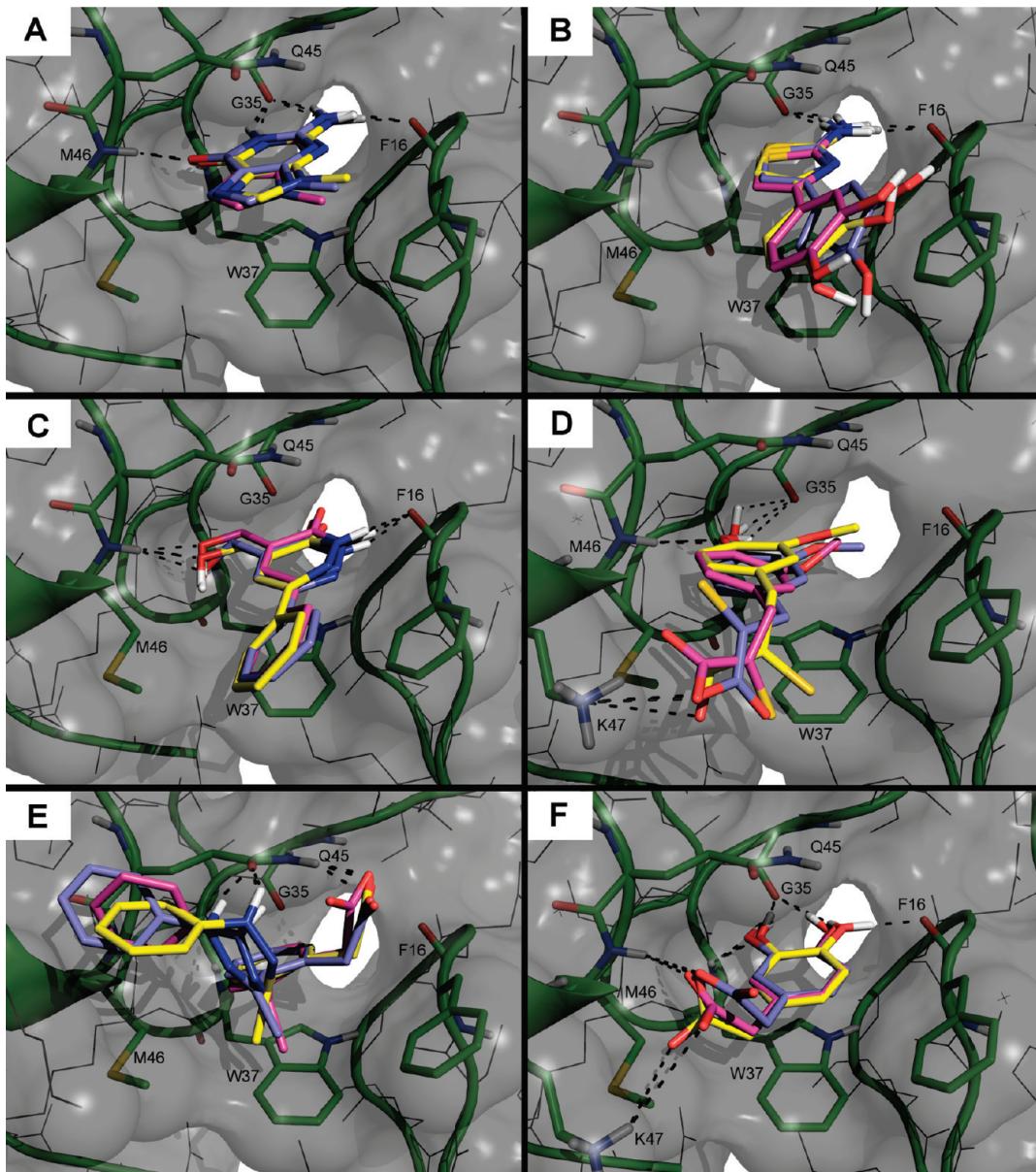


Figure 3. Predicted binding modes for the guanine moiety (A) and compounds 1–5 (B–F) toward the MD-based (RNA)NCp7 structure. Protein cartoon and surfaces are colored green and gray, respectively. Ligands docked by AutoDock are colored magenta, those docked by GOLD are blue, and those docked by Glide are yellow. Polar contacts are showed as dashed lines.

docking poses and were generally characterized by the lowest theoretical binding energy (or top-ranking score).

Docking results are in agreement with MIFs (superimposition between docking poses and MIFs is showed in Supporting Information). In fact, aromatic rings of 1–5 are placed over the W37 indole ring, in correspondence of the magenta region of Figure 2, whereas the amino group and hydrogen bond donors are placed in the inner part of the cavity, in the regions where MIFs for the interaction of H-bond donor groups were found (Figure 2, green and cyan). Hydrogen bond acceptor groups are mostly docked in proximity to the backbone of M46 and W37, in correspondence to the blue and red areas of Figure 2. Hydrophobic groups interact with the F16 phenyl ring, thus occupying the outer part of the binding site in the yellow regions of Figure 2. In summary, all the currently available NCp7 inhibitors bind preferentially within the same lipophilic pocket

that accommodates the guanine nucleotide. Compounds 1–5 make a common hydrophobic interaction with the side chain of W37, that is considered as the key residue for ligand binding to NCp7 (it is also crucial for nucleic acid interaction). Most ligands show multiple hydrogen bond interactions with backbone of M46 and G35, while additional contacts are made with W37 backbone, F16, and Q45 in both MD-based receptor conformations. Ligand interactions are also found with residues located at the boundary of the lipophilic pocket, such as K47, and in the protein groove mainly composed of hydrophobic and basic residues, such as R32, K33, and K34. Chemical modifications of 1–5 could be designed to optimize the interactions with these residues. Moreover, they could be targeted during the *de novo* design of novel anti-HIV agents able to inhibit the NCp7-induced nucleic acid destabilization, and to compete with the binding of nucleic acids to the NCp7 surface.

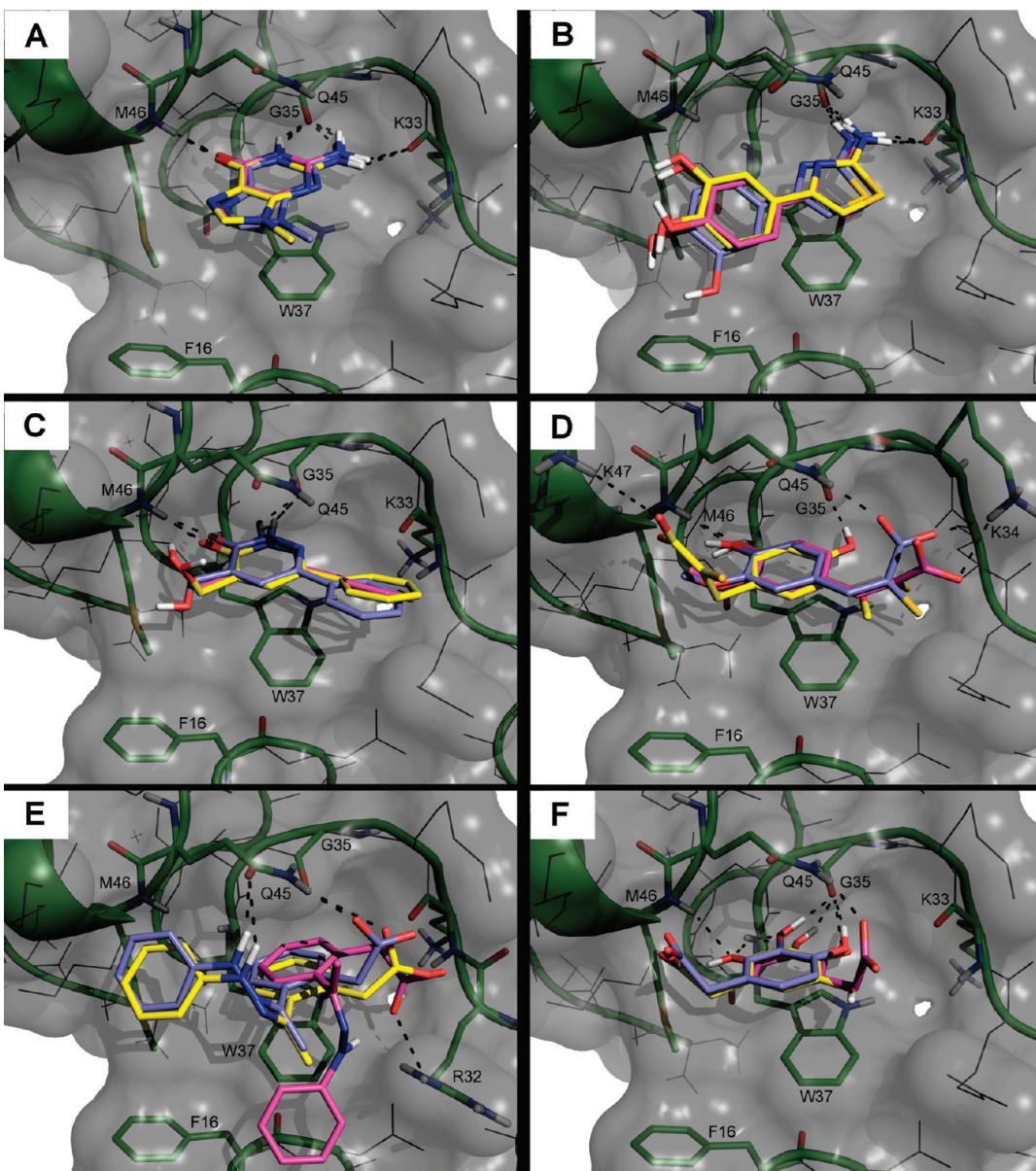


Figure 4. Predicted binding modes for the guanine moiety (A) and compounds **1–5** (B–F) toward the MD-based (DNA)NCp7 structure. Color codes are the same as in Figure 3.

Theoretical energies of binding and scoring values, calculated for the binding conformations of **1–5** described in Figure 3 and Figure 4, were then compared with the experimental K_i values (Table 2). The very similar binding affinity of these compounds (K_i values range from 8.5 to 15.0 μM) prompted us to consider **1–5** as equipotent toward NCp7, with the sole exception of **2** that is almost 2-fold more active than **5**. The significant difference between **2** and **5** is taken into account during the comparison of the experimentally observed affinity trend with the theoretical binding energies.

The scoring trend provided by the GoldScore function was not satisfactory at all. On the other hand, the ChemScore function seemed to be more accurate in the prediction and in the scoring of the binding pose,²³ and its use in the GOLD program provided the most precise results. In addition, the whole ligands+DUD (directory of useful decoys) benchmarking data set,²⁵ enriched with **1–5**, was docked with GOLD toward both the protein

conformations. ChemScore ranked **1–5** within the top 1.5% of the results, whereas GoldScore ranked **1–5** within the top 25%, thus confirming the higher reliability of the ChemScore function for docking small molecules toward NCp7. The ChemScore function was also used during the preliminary steps of the Glide docking, whose results satisfactorily match with experimental data. In the case of AutoDock, an overestimation of the binding energy was observed for **3** and **4**, which was attributed to the excessive contribution of electrostatic interactions between the ligands carboxy group and the side chain bearing a positive charge (data not shown). In fact, through the smoothing of the electrostatic contributions of a factor 10, a higher agreement with the experimental trend was observed. Since electrostatic interactions occur mainly in the solvent-accessible region, we thought that a smoothed value could be accurate anyway, especially in the case of a highly basic protein such as NCp7. Finally, in agreement with previous findings,³² results of Glide in the SP and

Table 2. Docking Results for Compounds 1–5 and the Guanine Ring

protein structure	compound and $K_i(\mu\text{M})^{\text{s}}$	AutoDock final intermolecular energy ^a	GOLD fitness (CS) ^b	Glide energy (XP) ^c
(RNA)NCp7	1 (11.0 ± 1.0)	-5.99	21.01	-25.42
	2 (8.5 ± 0.9)	-6.09	22.46	-26.55
	3 (10.5 ± 0.9)	-5.72	23.10	-26.56
	4 (13.0 ± 1.0)	-5.88	21.57	-25.84
	5 (15.0 ± 1.0)	-5.46	18.22	-24.39
(DNA)NCp7	1 (11 ± 1.0)	-5.57	22.50	-24.70
	2 (8.5 ± 0.9)	-6.33	21.69	-30.66
	3 (10.5 ± 0.9)	-6.02	20.24	-26.92
	4 (13 ± 1.0)	-5.92	18.95	-23.02
	5 (15 ± 1.0)	-5.02	18.48	-24.13
consensus ^d	1 (11 ± 1.0)	-5.78	21.75	-25.06
	2 (8.5 ± 0.9)	-6.21	22.07	-28.60
	3 (10.5 ± 0.9)	-5.87	21.67	-26.74
	4 (13 ± 1.0)	-5.90	20.26	-24.43
	5 (15 ± 1.0)	-5.24	18.35	-24.26

^a Sum of vdW + Hbond + desolv energy + (electrostatic energy/10).

^b ChemScore function. ^c Extra precision settings. ^d score by score.

XP settings are very similar (results of Glide XP are reported in Table 2). The main interaction that occurs between 1–5 and NCp7 is of hydrophobic nature. W37 constitutes, in fact, the key residue of the hydrophobic binding pocket for ligands and nucleic acids. Docking results show that, in this work, functions providing the highest accuracy in both docking and scoring of 1–5 are ChemScore and GlideScore. Interestingly, these empirical functions are reported to be more adequate for docking compounds into hydrophobic binding sites,^{33,34} while GoldScore, that here provided a poor accuracy, is described as more precise when the protein binding site is hydrophilic and when hydrogen bond interactions represent the most important anchor points to the receptor. Our results are in line with these findings, and it is not surprising that, for studying ligand interaction toward the hydrophobic pocket of NCp7, ChemScore and GlideScore outperformed the force field-based GoldScore and AutoDock functions. However, it is worth noting that AutoDock provided satisfactory results, in agreement with available biological data, only after smoothing the electrostatic contribution to the binding energy, as reported above.

Consensus Analysis of Docking Results. In order to account for the possible protein flexibility and to reduce the number of possible false-positive and false-negative data, a “score by score” consensus analysis was performed by averaging scoring values obtained by docking toward both NCp7 structures (Table 2). The use of multiple receptor conformations is known to improve the quality of the docking results, especially in the cases of a very flexible protein as NCp7.³⁵ Here we report that the consensus score of the docking results fits better with the experimental trend than the results obtained from a single receptor structure. The use of multiple NCp7 conformations in further computational studies, such as virtual screening protocols, would probably account for the accurate identification of new putative NCp7 ligands with inhibitory activity better than that of 1–5.

Rescoring Docking Results. It is a common finding that rescoring docking results could provide a more accurate description of the ligand binding energy toward a receptor.^{23,36} Here we applied the rescoring protocol implemented in GOLD and Glide, as well as we used an external scoring function (Poisson–Boltzmann surface

Table 3. Rescoring Results with the PBSA Approach

protein structure	compound and $K_i(\mu\text{M})^{\text{s}}$	AutoDock poses	GOLD poses	Glide poses
		delta-PBTOT ^a	delta-PBTOT ^a	delta-PBTOT ^a
(RNA)NCp7	1 (11.0 ± 1.0)	-37.74	-37.04	-38.20
	2 (8.5 ± 0.9)	-42.13	-40.80	-41.17
	3 (10.5 ± 0.9)	-41.05	-35.45	-43.79
	4 (13.0 ± 1.0)	-40.91	-34.20	-36.15
	5 (15.0 ± 1.0)	-38.61	-34.01	-31.80
(DNA)NCp7	1 (11.0 ± 1.0)	-34.91	-30.40	-30.55
	2 (8.5 ± 0.9)	-34.71	-34.43	-35.86
	3 (10.5 ± 0.9)	-34.54	-36.00	-23.36
	4 (13.0 ± 1.0)	-29.64	-25.96	-19.43
	5 (15.0 ± 1.0)	-27.30	-22.39	-18.76
consensus ^b	1 (11.0 ± 1.0)	-36.32	-33.72	-34.37
	2 (8.5 ± 0.9)	-38.42	-37.61	-38.51
	3 (10.5 ± 0.9)	-37.79	-35.72	-33.57
	4 (13.0 ± 1.0)	-35.27	-30.08	-27.79
	5 (15.0 ± 1.0)	-32.95	-28.20	-25.28

^a Delta-PBTOT is in $\text{kcal} \cdot \text{mol}^{-1}$. ^b score by score.

area approach, PBSA). Although the rescoring of the ChemScore binding pose with the GoldScore function in the GOLD program lowered the matching with experimental data, the use of the default GlideScore function in the Glide program did not produce significant benefits. The PBSA was hence employed to provide a higher accuracy in energy calculation (Table 3). In fact, the quality of this method for the estimation of ligand–protein delta energy of binding is widely appreciated especially to rank a set of molecules, and it is being routinely applied in MD and drug design protocols.³⁷ Moreover, rescoring docking poses with PBSA could overcome some source of errors that are intrinsic of docking programs, generally providing highly reliable results. The binding poses of 1–5 described in Figure 3 and Figure 4 were selected, and their respective docking-based complexes with NCp7 were energy minimized by means of AMBER10. As expected, no significant conformational changes were found by energy minimization, and rms differences between coordinates of docking-based complexes and their corresponding conformation at the local minimum were lower than 1.0 Å (calculated on all heavy atoms). The current delta energy of binding of each ligand–protein complex at the state of local energy minimum was calculated with the MM-PBSA module of AMBER10. Delta-PBTOT values are reported in Table 3.

When the previously described “score by score” consensus approach was applied to the PBSA-based rescoring data, very similar delta energy values for compounds 1, 3, and 4 were found, whereas different energy values were still obtained for 2 and 5, in agreement with experimental K_i values.

The application of this protocol to the very flexible NCp7 provided satisfactory results in this work, even though this could not be generally extended to all target proteins. Moreover, from an informatics standpoint, rescoring docking poses by means of the PBSA does not require significant CPU effort (45 s for each protein–ligand adducts running on a single core, processor Intel E5420 2.5 GHz, including the local optimization) and could be easily automated and implemented in scripts for virtual screening.

CONCLUSIONS

In conclusion, we used two computationally refined conformations of NCp7 to predict, by means of docking simulations,

the possible binding modes of five active compounds recently identified by in vitro high throughput screening. Docking results were in agreement with the calculated molecular interaction fields and with available experimental data. The highest matching was obtained when results obtained by docking toward two receptor structures were rescored and then merged by means of a consensus approach. Rescoring with the PBSA improved the quality of docking results and the matching with available experimental data.

Finally, NCp7 residues that constitute the putative binding site on the surface of NCp7, and that are mostly involved in ligand binding, were described. These residues, along with hydrophobic and basic residues of the protein groove, could be targeted during the rational optimization of compounds **1–5** by chemical modifications. In addition, ligand interaction motifs herein identified could be incorporated into a pharmacophoric model (or split into multiple pharmacophores) that could be used as a query to filter commercial databases, for further extensive searches of structurally diverse active ligands toward the very flexible NCp7 protein by means of virtual screening methodologies.

■ ASSOCIATED CONTENT

S Supporting Information. Superimposition between MIFs and docking poses into the protein binding site, and structural alignment of all NMR models of NCp7 structures. This information is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +39 0577 234306. Fax: +39 0577 234333. E-mail: botta.maurizio@gmail.com.

■ ACKNOWLEDGMENT

This work was supported by the European Union: “THINC”, Collaborative Project, grant number HEALTH-2007-2.3.2-1; “CHAARM”, Collaborative Project, grant number HEALTH-F3-2009-242135. It was also supported by Italian Ministero dell’Istruzione, dell’Università e della Ricerca - Prin 2008 research project, grant number 2008CE7SSA_004.

■ REFERENCES

- (1) Cen, S.; Khorchid, A.; Gabor, J.; Rong, L.; Wainberg, M. A.; Kleiman, L. Roles of Pr55(gag) and NCp7 in tRNA(3)(Lys) genomic placement and the initiation step of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* **2000**, *74* (22), 10796–10800.
- (2) Poon, D. T.; Li, G.; Aldovini, A. Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. *J. Virol.* **1998**, *72* (3), 1983–1993.
- (3) Cruceanu, M.; Stephen, A. G.; Beuning, P. J.; Gorelick, R. J.; Fisher, R. J.; Williams, M. C. Single DNA molecule stretching measures the activity of chemicals that target the HIV-1 nucleocapsid protein. *Anal. Biochem.* **2006**, *358* (2), 159–170.
- (4) Dietz, J.; Koch, J.; Kaur, A.; Raja, C.; Stein, S.; Grez, M.; Pustowka, A.; Mensch, S.; Ferner, J.; Moller, L.; Bannert, N.; Tampe, R.; Divita, G.; Mely, Y.; Schwalbe, H.; Dietrich, U. Inhibition of HIV-1 by a peptide ligand of the genomic RNA packaging signal Psi. *ChemMed-Chem.* **2008**, *3* (5), 749–755.
- (5) Shvadchak, V.; Sanglier, S.; Rocle, S.; Villa, P.; Haiech, J.; Hibert, M.; Van, D. A.; Mely, Y.; de Rocquigny, H. Identification by high throughput screening of small compounds inhibiting the nucleic acid destabilization activity of the HIV-1 nucleocapsid protein. *Biochimie* **2009**, *91* (7), 916–923.
- (6) Mori, M.; Dietrich, U.; Manetti, F.; Botta, M. Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome. *J. Chem. Inf. Model.* **2010**, *50* (4), 638–650.
- (7) Protein Data Base; Brookhaven National Laboratory, Upton, NY ; <http://www.rcsb.org/pdb>, accessed Nov 25, 2010.
- (8) De Guzman, R. N.; Wu, Z. R.; Stallings, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998**, *279* (5349), 384–388.
- (9) Bourbigot, S.; Ramalanjaona, N.; Boudier, C.; Salgado, G. F.; Roques, B. P.; Mely, Y.; Bouaziz, S.; Morellet, N. How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *J. Mol. Biol.* **2008**, *383* (5), 1112–1128.
- (10) Maestro, version 9.0, Schrödinger, LLC, New York, NY, 2009.
- (11) Jorgensen, W.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
- (12) GRID version 22 for LINUX; Molecular Discovery Ltd, Pinner Middlesex HASSNE (UK), <http://www.moldiscovery.com>, accessed April, 2010.
- (13) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849–857.
- (14) Reynolds, C. A.; Wade, R. C.; Goodford, P. J. Identifying targets for bioreductive agents: using GRID to predict selective binding regions of proteins. *J. Mol. Graph.* **1989**, *7* (2), 103–8, 100.
- (15) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
- (16) MacroModel, version 9.7, Schrödinger, LLC, New York, NY, 2009.
- (17) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energies and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (18) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245* (1), 43–53.
- (19) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33* (3), 367–382.
- (20) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11* (5), 425–445.
- (21) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
- (22) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein–ligand docking using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504–6515.
- (23) O’Boyle, N. M.; Liebeschuetz, J. W.; Cole, J. C. Testing assumptions and hypotheses for rescoring success in protein–ligand docking. *J. Chem. Inf. Model.* **2009**, *49* (8), 1871–1878.
- (24) Glide, version 5.5, Schrödinger, LLC, New York, NY, 2009.
- (25) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.
- (26) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897.

- (27) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **1998**, *16* (3), 671–682.
- (28) AMBER 10, University of California, San Francisco, CA, 2008.
- (29) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012.
- (30) McGaughey, G. B.; Gagne, M.; Rappe, A. K. pi-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **1998**, *273* (25), 15458–15463.
- (31) Wang, W.; Hobza, P. Theoretical study on the complexes of benzene with isoelectronic nitrogen-containing heterocycles. *Chemphyschem* **2008**, *9* (7), 1003–1009.
- (32) Li, X.; Li, Y.; Cheng, T.; Liu, Z.; Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J. Comput. Chem.* **2010**, *31* (11), 2109–2125.
- (33) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49* (4), 1079–1093.
- (34) Corbeil, C. R.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 3. Impact of input ligand conformation, protein flexibility, and water molecules on the accuracy of docking programs. *J. Chem. Inf. Model.* **2009**, *49* (4), 997–1009.
- (35) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 178–184.
- (36) Cole, J. C.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Comparing protein-ligand docking programs is difficult. *Proteins* **2005**, *60* (3), 325–332.
- (37) 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* (4), 1493–1503.