

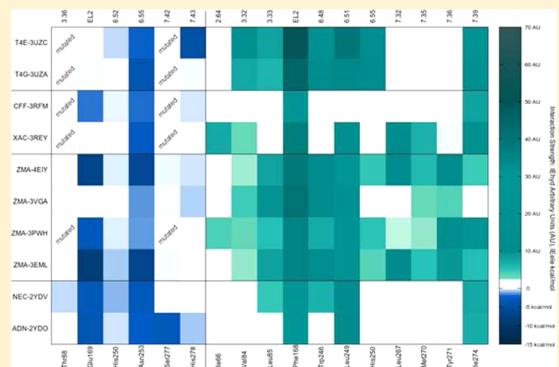
Alternative Quality Assessment Strategy to Compare Performances of GPCR-Ligand Docking Protocols: The Human Adenosine A_{2A} Receptor as a Case Study

Antonella Ciancetta, Alberto Cuzzolin, and Stefano Moro*

Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, via Marzolo 5, 35131 Padova, Italy

Supporting Information

ABSTRACT: The progress made in the field of G protein-coupled receptors (GPCRs) structural determination has increased the adoption of docking-driven approaches for the identification or optimization of novel potent and selective ligands. In this work, we compared the performances of the 16 different docking/scoring combinations using the recently released crystal structures of the human A_{2A} AR (hA_{2A} AR) in complex with both agonists and antagonists. The proposed evaluation strategy encompasses the use of three complementary “quality descriptors”: (a) the number of conformations generated by a docking algorithm having a RMSD value lower than the crystal structure resolution (R); (b) a novel consensus-based function defined as “protocol score”; and (c) the interaction energy maps (IEMs) analysis, based on the identification of key ligand–receptor interactions observed in the crystal structures.



INTRODUCTION

The progress made in the field of G protein-coupled receptors (GPCRs) structural determination has increased the adoption of docking-driven approaches for the identification or the optimization of novel potent and selective ligands.^{1–8} As routinely demonstrated, docking programs are usually successful in generating multiple poses that include binding modes similar to the crystallographically determined bound structure, whereas scoring functions are much less successful at correctly identify the corresponding “bioactive” binding mode.⁹ This intrinsic limitation generally implies the need for the calibration of the docking protocol through benchmark studies prior to applying it. Traditionally, these benchmarks have focused on redocking the cognate ligand of a crystallographic receptor–ligand complex to measure geometric pose prediction accuracy.^{10–12}

In this work, we propose an alternative quality assessment strategy to compare the performances of the 16 different docking/scoring combinations using the recently released crystal structures of the human A_{2A} AR (hA_{2A} AR) in complex with both agonists and antagonists, as summarized in Table 1.^{13–19} Among them, one is cocrystallized with the endogenous agonist adenosine (PDB ID: 2YDO¹⁵), one with its synthetic analogue NECA (N-ethyl-5'-carboxamido adenosine, PDB ID, 2YDV¹⁵), and the remaining eight with five antagonists, namely, ZM 241385 (4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo-(2,3-*a*)(1,3,5)triazin-5-yl-amino)ethyl)phenol, PDB IDs 3EML,¹³ 3PW¹⁶, 3VGA,¹⁷ 4EIY¹⁹); T4G (6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, PDB ID 3UZA¹⁸);

T4E (4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, PDB ID 3UZC¹⁸); caffeine (PDB ID 3RFM²⁷); and XAC (N-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl-2,3,6,7-tetrahydro-1*H*-purin-8-yl)phenoxy]acetamide, PDB ID 3REY¹⁶). The structures of the cocrystallized ligands are shown in Figure 1. The pharmacology of adenosine receptors and the potential applications of their agonists and antagonists have already been extensively described and recently reviewed.²⁰

The performances of all docking/scoring combinations were evaluated using an alternative assessment strategy defined by three complementary “quality descriptors”: (a) the number of conformations generated by the docking algorithm having a root mean square deviation (RMSD) value lower than the crystal structure resolution (R); (b) a novel consensus-based function defined as “protocol score”; and (c) the interaction energy maps (IEMs) analysis, based on the identification of key ligand–receptor interactions observed in the crystal structures. The comparison of IEMs enables a fast and graphical selection of the conformations based on the quality of the interactions (in terms of the number of established interactions and their relative strength) occurring between each docking pose and selected key residues. This type of analysis was proposed in the past by several authors and in different forms but rarely applied due to the lack of automation and the subjectivity of interactions selection.^{21–24} Both limitations have been overcome by developing an in house python script.

Received: May 12, 2014

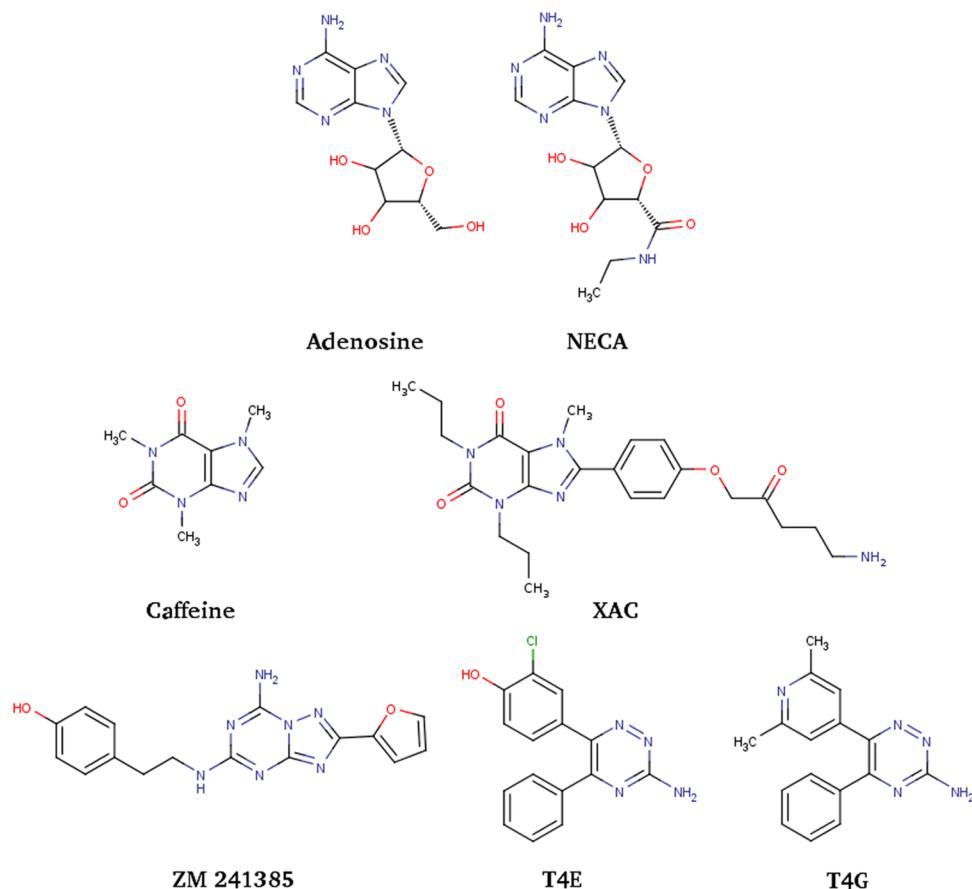
Published: July 21, 2014



Table 1. hA_{2A} AR Crystal Structures Available to Date

PDB ID	release date	R (Å)	ligand name	ligand name abbreviation ^b	crystallization strategy	ligand type
3EML	08/10/14	2.60	ZM 241385	ZMA	T4 lysozyme fusion ^c	antagonist
3QAK ^a	11/03/09	2.71	UK-432,097	UKA	T4 lysozyme fusion ^c	agonist
2YDO	11/05/18	3.00	adenosine	ADO	StaR ^d	agonist
2YDV	11/05/18	2.60	NECA	NEC	StaR ^d	agonist
3PWH	11/09/07	3.30	ZM 241385	ZMA	StaR ^e	antagonist
3REY	11/09/07	3.31	XAC	XAC	StaR ^e	antagonist
3RFM	11/09/07	3.60	caffeine	CFF	StaR ^e	antagonist
3VG9 ^a	12/02/01	2.70	ZM 241385	ZMA	Fab2838 complex ^f	antagonist
3VGA	12/02/01	3.10	ZM 241385	ZMA	Fab2838 complex ^f	antagonist
3UZA	12/03/21	3.27	T4G	T4G	StaR ^e	antagonist
3UZC	12/03/21	3.34	T4E	T4E	StaR ^e	antagonist
4IEY	12/07/25	1.80	ZM 241385	ZMA	ApoCytochrome b562RIL chimera ^g	antagonist

^aStructures not considered in this study. ^bThree letter code assigned in the PDB file. ^chA_{2A} AR-T4L-ΔC: IL3 replaced by T4 and C-Term deleted. ^dhA_{2A} AR-GL31-ΔC: thermostabilizing mutations (L48A, A54L, T65A, and Q89A), N154A mutation, and C-Term deleted. ^ehA_{2A} AR-StaR2-ΔC: thermostabilizing mutations (A54L, T88A, K122A, V239A, R107A, L202A, L235A, and S277A), N154A mutation, and C-Term deleted. ^fhA_{2A} AR-Fab2838-ΔC: complex with mouse monoclonal-antibody Fab fragment (Fab2838), N154A mutation, and C-Term deleted. ^ghA_{2A} AR-BRIL-ΔC: IL3 replaced by apocytochrome b562RIL and C-Term deleted.

Figure 1. Structures of hA_{2A} AR cocrystallized ligands used to perform the docking benchmark study.

MATERIALS AND METHODS

Numbering and Name Conventions. The residues of the hA_{2A} AR are indicated according to the following scheme: the three letter residue name is followed by the residue number and the Ballesteros and Weinstein notation reported in brackets.²⁵ The latter nomenclature scheme, generally indicated by “TM ± 50”, identifies the residues by the helix number (TM) followed by the position relative (±) to a reference residue among the most conserved amino acids in that helix, to which the number

50 is arbitrarily assigned. Compound names correspond to the ligands three letter codes assigned in the PDB file, whereas protein identifiers are the corresponding PDB entries. Docking protocols are named according to the following scheme: “program name abbreviation-scoring function/search algorithm”.

Computational Facilities. Energy calculations and analyses of docking poses were performed with the Molecular Operating Environment (MOE, version 2012.10) suite.²⁶ Extraction and analysis of docking results were performed by

using in-house bash and python scripts. Maps and graphs were created with Gnuplot, version 4.4.²⁷

Protein Structures Selection and Preparation. Of the 12 hA_{2A} AR crystal structures available, listed in Table 1, the following structures were used to perform the benchmark study (PDB IDs): 2YDO,¹⁵ 2YDV,¹⁵ 3EML,¹³ 3PWH,¹⁶ 3REY,¹⁶ 3RFM,¹⁶ 3UZA,¹⁸ 3UZC,¹⁸ 3VGA,¹⁷ and 4EIY.¹⁹ The structures identified by the 3QAK¹⁴ and 3VGA¹⁷ PDB IDs were not considered: We excluded from our analysis the ligand cocrystallized in the 3QAK structure because it has a high number of heavy atoms and a number of rotatable bonds exceeding the cutoff allowed for some of the in-house available docking programs; the inverse agonist cocrystallized in the 3VGA structure was not considered because it has low occupancy in the PDB structure.

The selected structures were retrieved from the RCSB PDB database (<http://www.rcsb.org>).²⁸ Before the preparation procedure, all the proteins were aligned and superimposed to a selected reference structure (3EML¹³). Antibody portions, ions, and crystallization solvents were removed, whereas water molecules and cocrystallized ligands were retained for the hydrogen atoms assignment step and then removed. Fused proteins (lysozyme and apocytochrome in the specific cases) as well as point mutations were retained, and the structures were not subjected to any conformational changes. Missing loop domains, N-terminal and C-terminal, were not modeled. Ionization states and hydrogen positions were assigned with the "Protonate-3D" tool,²⁹ as implemented in the MOE suite. Then, to minimize contacts among hydrogen atoms, the structures were subjected to energy minimization with Amber99 force field³⁰ until the root mean square (RMS) of the conjugate gradient was <0.05 kcal·mol⁻¹·Å⁻¹, by keeping the heavy atoms fixed at their positions. After the protonation step, for each structure, the coordinates of the binding site center (barycenter of cocrystallized ligand) were determined and saved, then ligand and water molecules were removed and protein atoms partial charges computed with the Amber99 force field.³⁰

Ligand Structures. Co-crystallized ligands were extracted from the corresponding crystallographic complex and checked for errors. Hydrogen atoms were added, and the protonation state (pH 7.4) was assigned. The structures were not subjected to energy minimization, so that for each ligand the starting conformation is the same one observed in the crystal structure. Partial charges on ligand atoms were computed on the basis of the PM3/ESP semiempirical Hamiltonian.^{31,32}

Docking Settings. The performances of the following six docking programs were assessed: Autodock,³³ GOLD,³⁴ Glide,³⁵ PLANTS,³⁶ Molegro Virtual Docker,³⁷ and MOE-dock.³⁶ The versions of the programs were the most up to date available at the time we performed the calculations. Among all the scoring functions and search algorithms available in the considered programs, we discarded those that did not allow us to return a user-defined number of output conformations without duplicates and postdocking refinement. In the end, a total amount of 16 different docking algorithm/scoring function combinations were assessed, as detailed in Table 2. To make the results obtained with different protocols as homogeneous as possible, we set the common settings reported in Table 3.

Docking Stages. Each ligand structure was first docked into the corresponding crystal structure with the different docking protocols (cognate ligand docking). Then, for the

Table 2. List of Docking Programs along with Search Algorithm (or Placing Method) and Scoring Functions Used to Perform the Docking Benchmark Study

program	search algorithm (+ placing method)	scoring function	abbreviation
Autodock 4.2	genetic algorithm	AutoDock SF	AD-GA
	Lamarkian GA	AutoDock SF	AD-LGA
	<i>local search</i>	AutoDock SF	AD-LS ^a
Glide 5.8	glide algorithm	Standard Precision	Glide-SP
	genetic algorithm	Goldscore	Gold-Gold
	genetic algorithm	Chemscore	Gold-Chem
	genetic algorithm	ASP	Gold-ASP
GOLD 5.1	genetic algorithm	PLP	Gold-PLP
	ACO algorithm	ChemPLP	Plants-CPLP
	ACO algorithm	PLP95	Plants-PLP95
PLANTS 1.2	ACO algorithm	PLP	Plants-PLP
	systematic search + alpha triangle	London dG	MOE-AT
	systematic search + alpha PMI	London dG	MOE-APMI
MOE 2012.10	systematic search + triangle matcher	London dG	MOE-TM
	iterated simplex	MolDock SF	MVD-IS
	MolDock optimizer	MolDock SF	MVD-MDO
Molegro Virtual Docker 5.5	<i>MolDock simplex evolution</i>	<i>MolDock SF</i>	<i>MVD-MDSE^a</i>

^aThe protocols have been excluded from the study, and the corresponding results will be not discussed.

Table 3. Common Docking Settings for the Evaluated Protocols

parameter	value/setting
ligand input conformation	X-ray binding mode
ligand initial partial charges	PM3/ESP
water molecules	excluded
output	20 conformations
RMSD threshold	1.0 Å
binding cavity center	ligand barycenter in X-ray structure
binding cavity radius	20 Å
grid spacing (for grid-based calculations)	0.3 Å
refinement and rescore	turned off

protocols giving the best performances in the cognate ligand docking stage, we performed ensemble docking runs to assess whether the protocol is able to assign to each ligand (or ligand conformation in case of ZMA) the corresponding cocrystallized structure by selecting it among all the tested proteins. For the ensemble docking step, three different strategies for the definition of the binding site center have been evaluated. In particular, the binding site center was set as (i) the centroid of the barycenters of the ligands; (ii) the ZMA barycenter in 3EML structure; and (iii) the barycenter of each ligand in its corresponding crystal structure. Finally, we evaluated the effect of two additional parameters on the docking outcomes: the reconstruction of the second extracellular loop (EL2) and the starting conformation.

Analysis of Docking Results. To judge the performances of the different tested protocols, the RMSD values between predicted and crystallographic poses were calculated. In the

case of the XAC ligand, which has a highly flexible solvent-exposed tail, the RMSD values were computed only for the heavy atoms of the aromatic cores.

The performances of the docking protocols were evaluated on the basis of the lowest, highest, and average RMSD values (RMSD_{\min} , RMSD_{\max} , and RMSD_{ave} , respectively) as well as the highest number of conformations with a RMSD value lower than the corresponding X-ray resolution (R), $N^{(\text{RMSD} < R)}$.

Protocol Score. To compare at a glance the performances of the different protocols tested, we merged the above-discussed parameters and in particular the RMSD_{ave} and the $N^{(\text{RMSD} < R)}$ in a unique statistical value, that we called protocol score, defined as follows: (a) one point is assigned to each protocol that has the RMSD_{ave} value lower than R ; (b) one point is assigned to each protocol that generates at least 10 conformations having RMSD values with respect to the X-ray binding mode lower than R ; and (c) two points are assigned to the protocols that satisfy both of the above-mentioned requirements. Moreover, to discern the best protocols among the good ones, three points are assigned to the protocols that give the lowest RMSD_{ave} value and, at the same time, returns the highest number of conformers with a RMSD value lower than R . The protocol score assignment criteria are summarized in Table 4.

Table 4. Protocol Score Assignment Criteria

condition	score
$\text{RMSD}_{\text{ave}} < R$	1
$N^{(\text{RMSD} < R)} > 10$	1
$\text{RMSD}_{\text{ave}} < R$ and $N^{(\text{RMSD} < R)} > 10$	2
protocols returns: $\text{Min}(\text{RMSD}_{\text{ave}})$ and $\text{Max}(N^{(\text{RMSD} < R)})$	3

Analysis of Ligand–Receptor Interactions. To analyze the ligand–receptor interactions, we calculated the individual electrostatic and hydrophobic contributions to the interaction energy (hereby denoted as IEele and IEhyd, respectively) of key residues involved in the binding with the ligands, as emerged from detailed analyses and comparisons among the different crystallographic binding modes. In particular, the electrostatic contribution is computed on the basis of the nonbonded electrostatic interaction energy term of the force field,³⁸ whereas the hydrophobic contributions is calculated by using the directional hydrophobic interaction term based on contact surfaces as implemented in the MOE scoring function.²⁶ As a consequence, energy (expressed in kcal/mol) is associated with the electrostatic contribution, whereas a score (the higher the better) is related to the hydrophobic contribution. The analysis of these contributions have been reported as “interaction energy maps” (hereby indicated as IEMs), graphically displayed as heat-like maps reporting the key residues involved in the binding with the considered ligands along with a quantitative estimate of the occurring interactions.

RESULTS AND DISCUSSION

Overview of X-ray Binding Modes. Prior to discussing the results of our docking benchmark, we briefly report an overview of the binding modes observed in the ligand-hA_{2A} AR complexes under study. It has to be pointed out that the analysis herein reported lacks water mediated interactions, as we intentionally did not include water molecules in our docking simulations. We instead briefly addressed this topic in a recent study,³⁹ and a deeper investigation of the role of water

molecules in hA_{2A} AR ligand binding is the focus of a study being currently conducted in our research group.⁴⁰

Figure 2 depicts the IEM (for more details, see the Materials and Methods section) of the ligands under study. The map has been derived stepwise (see Figure S1, Supporting Information). We first computed for all the ligands the individual contribution of each residue to the interaction energy (per residue analysis). From a comparison of the results of the per residue analyses, we then identified the key residues involved in the binding with all ligands and reported the occurring interaction in the IEM in Figure 2. The common interaction pattern for all ligands involves an aromatic π–π stacking with the conserved Phe168, located in the second extracellular loop (EL2), and additional hydrophobic contacts with Leu249 (6.51) and Ile274 (7.39) side chains. Strong polar interactions are established with the side chain of the conserved Asn253 (6.55).⁴¹ The IEM along with the three-dimensional representation of the corresponding binding modes (Figure S2, Supporting Information) helps in appreciating the different extent (and type) of interaction networks and the different sizes of the ligands by allowing a direct comparison of all ligands to each other, between pairs of structurally related compounds and different conformations of the same molecule. With respect to the antagonists structures, agonists (NECA and adenosine) interact through fewer hydrophobic interactions while establishing hydrogen bonds with two additional residues, namely, Thr88 (3.36) and Ser277 (7.42), mediated by the ribose moiety. With respect to the latter interaction, it has to be pointed out that Thr88 (3.36) and Ser277 (7.42) are mutated to alanine in the so-called “StaR2” structures (3PWH, 3REY, 3RFM, 3UZC, and 3UZA), and therefore for those constructs, the interaction with these residues cannot be detected. Among the antagonist structures, the differences in the binding patterns between ZMA and caffeine reflect their binding affinities to the hA_{2A} AR: ZMA shows an extended pattern with strong polar and hydrophobic interactions, whereas caffeine establishes fewer and less intense interactions.

All of the above-described interactions are consistent with the available mutagenesis data: in particular, the far back known role of Asn253 (6.55) and the more recent mutagenesis data highlighting the roles of Phe168 (EL2) and Leu249 (6.51) for both agonist and antagonist binding and that of Thr88 (3.36) for agonist binding.⁴²

Workflow. The workflow of the computational protocol is shown in Figure 3. Starting from the protein–ligand complexes (PDB files, see Table 1), the protein and ligand structures were prepared and docking simulations run with the selected protocols, listed in Table 2. Then, for each protein structure, the RMSD values of the conformations generated by the different docking protocols with respect to the ligand cocrystallized binding modes were calculated and statistical analyses performed. As detailed in the Materials and Methods section, during the protein preparation step, water molecules and cocrystallized ligands were retained. Once the hydrogen atoms were added and minimized, water molecules were removed and no longer considered. Complexed antibody portions, ions, and crystallization solvents were removed, whereas fused proteins as well as point mutations were retained. Unsolved protein regions (loops, N-terminal, and C-terminal) were not remodeled, and the structures were not energy minimized. Ligand structures were extracted from the original PDB files and checked for errors, and hydrogen atoms were added. The structures were not energy minimized in order

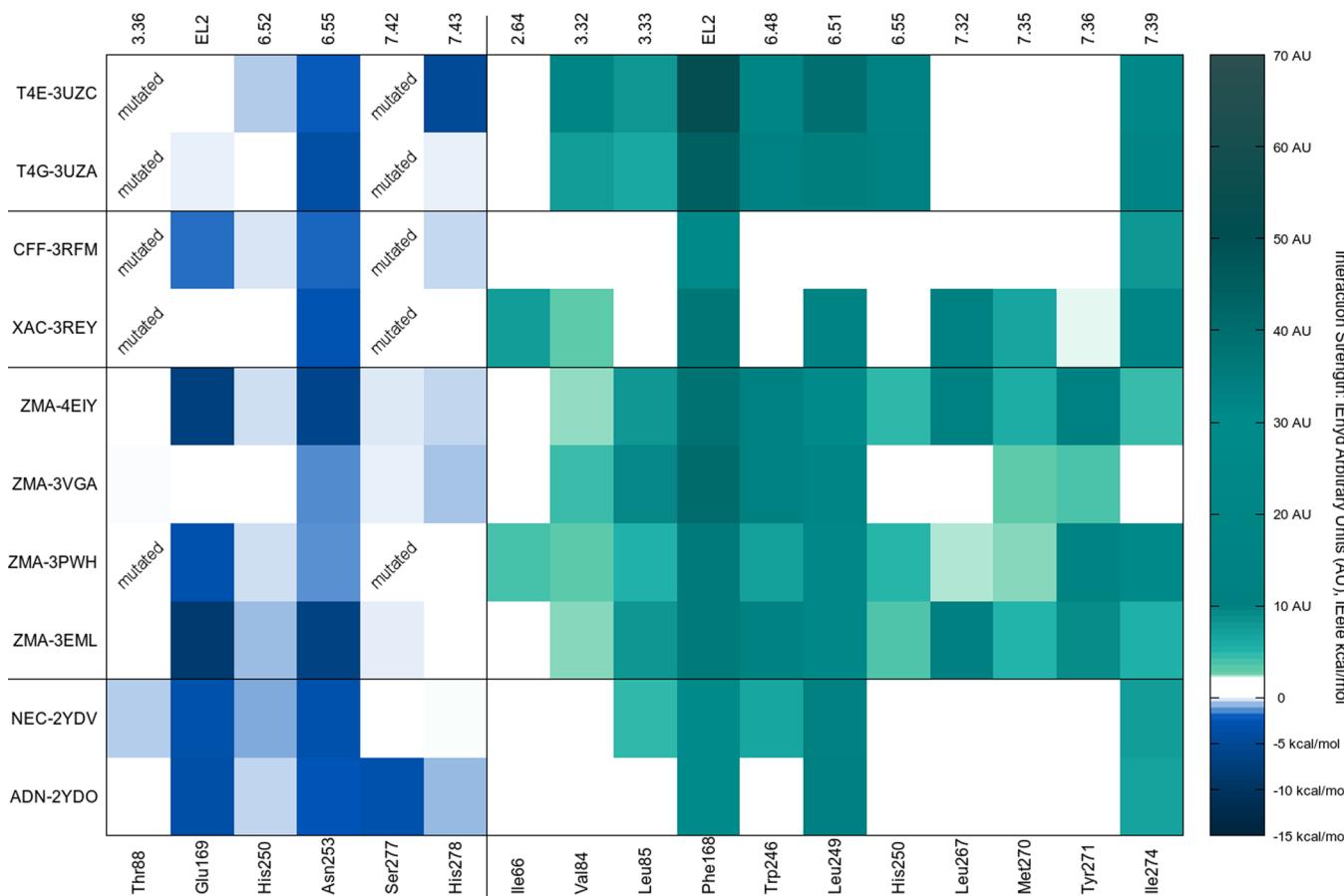


Figure 2. Interaction energy map (IEM) for the hA_{2A} AR cocrystallized ligands under study. Ligands are identified by the three letter codes assigned in the PDB file followed by the PDB IDs. The IEM has been computed for all the considered crystal structures after protein and ligand preparation and removal of water molecules. Electrostatic energy values (left) are expressed in kcal mol⁻¹, whereas hydrophobic scores (right) are expressed in arbitrary hydrophobic units.

to retain for each ligand the X-ray observed conformation. As it is generally accepted that the performances of docking protocols, especially those relying upon genetic algorithms, can be affected by the starting conformation, we also run test calculations on selected cases by supplying different random generated conformations as input and evaluated their effects on the performances of the docking protocols. The results are discussed in the following text.

After the protein and ligand preparation steps, we run the different docking protocols listed in Table 2: the assessed scoring functions include knowledge-based (Gold-ASP), force-field-based (AutodockSF and GoldScore), and empirical scoring functions, whereas tested algorithms comprise both deterministic (MOE and Glide) and stochastic search method approaches. Moreover, the variety of tested protocols also encompasses different types of protein representation, such as grid (Autodock and Glide) and all atom. We defined common settings (summarized in Table 3) for the different programs in order to ascribe the differences in the performances to the selected combination of search algorithms (or placing method) and scoring functions. In particular, we chose the same settings for the binding cavity (center, radius, and grid spacing for grid-based calculations), the ligand input (conformation and partial charges), and the program output (number of saved conformations, RMSD threshold value, refinement, and rescoring).

The results of the docking calculation were collected, and RMSD values with respect the cocrystallized ligand were computed for all conformations generated by the protocols. The performances were evaluated on the basis of lowest, highest, and mean RMSD values as well as the highest number of conformations with a RMSD value lower than the corresponding X-ray resolution, $N^{(RMSD < R)}$. As the considered crystal structures range from high (4EIY, 1.8 Å) to low (3RFM, 3.6 Å) resolutions, to evaluate the latter statistical parameter ($N^{(RMSD < R)}$) we decided to compare each structure with its own R value rather than setting a fixed threshold. With this procedure, we intrinsically take into account the quality of the experimental data to be reproduced: the less confidence we have in the atomic coordinates experimentally determined, the less strictly we judge the performances of a protocol in reproducing them.

Cognate Ligand Docking. The results of the cognate ligand docking step are reported in Table S1 (Supporting Information), and the most relevant statistical parameters are graphically summarized in Figure 4. In the map in Figure 4A, the minimum RMSD value ($RMSD_{min}$) of all the tested docking protocols are reported for each considered crystal structure: At first glance, it can be noted that there are protocols, such as Gold-ASP; Gold-Gold; Gold-PLP; and Plants-PLP, able to generate at least one pose that reproduces the X-ray binding mode with satisfying accuracy regardless of the specific structure under consideration. From the map, it is

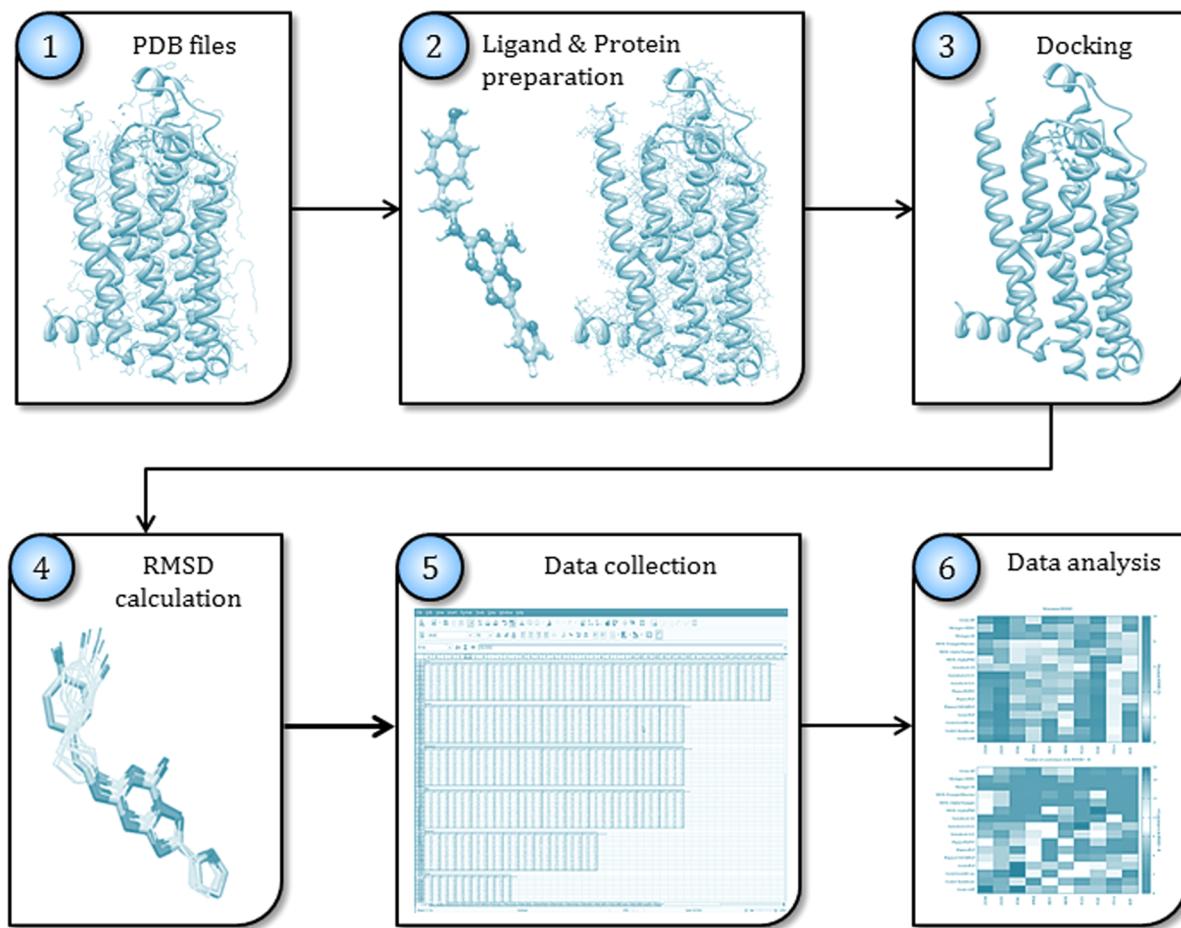


Figure 3. Workflow of the benchmark study.

also straightforward that agonist binding poses are predicted better than antagonist ones and in particular that the caffeine binding mode is the most challenging to be reproduced. As we also highlighted in our recent membrane molecular dynamics simulations of caffeine docking poses,³⁹ the dynamical evolution of different starting binding modes involves a consistent number of water molecules in rapid exchange around the ligand structure as a consequence of the weak interactions established with the receptor. This feature makes it challenging to reproduce the crystallographic binding mode without taking into account the dynamical behavior of surrounding water molecules. The difficulty is even increased if water molecules are not considered at all, as in this specific docking exercise.

The ability to reproduce at least once the X-ray observed binding mode is not a sufficient criterion to judge the quality of the docking protocols. Indeed, when extending the analysis by computing the average RMSD value (RMSD_{ave} , Figure 4B), some of the above-mentioned protocols worsen their performances. Moreover, other protocols that accurately reproduced at least in one conformer the X-ray binding mode (Gold-Chem, MVD-IS) show RMSD_{ave} values well over the structure resolutions. These data are consistent with the well-known limit of the docking procedure. Although search algorithms are usually successful at generating poses that include binding modes crystallographically observed, the scoring functions developed to date are much less successful at identifying and ranking the correct binding pose.⁹ The best protocol would be

the one that is able to accurately reproduce the X-ray binding mode and rank the conformations with the lowest RMSD values at the top of the list. This is also in view of the application of the docking protocol in a virtual screening framework, where only a selected percentage of the top generated conformations is considered and further processed. With this in mind, we defined an additional parameter by computing for each protocol the number of conformers having a RMSD value lower than the X-ray structure resolution ($N^{(\text{RMSD} < R)}$), which represents somehow the best accuracy one might expect from a computational procedure based on experimental data. From the map in Figure 4C, it can be evinced that there are only a few protocols able to generate a high number of conformations close to the crystallographic binding mode and that only in a few cases (Gold-ASP/2YDO; Gold-ASP/3EML; Gold-Gold/3EML; Gold-PLP/3EML; AD-LGA/3UZA; MOE-APMI/3UZC) all of the conformations generated by the protocol have RMSD values below the structure resolution.

We therefore tried to merge all of the above-discussed parameters in a unique statistical value that we called protocol score, defined as detailed in the Materials and Methods section and summarized in Table 4. The map in Figure 4D graphically displays the scores assigned to all of the tested protocols. The obtained results suggest that it is not possible to identify the best protocol for low resolution structures ($R > 3.00$) and that the protocols that perform better for most structures are

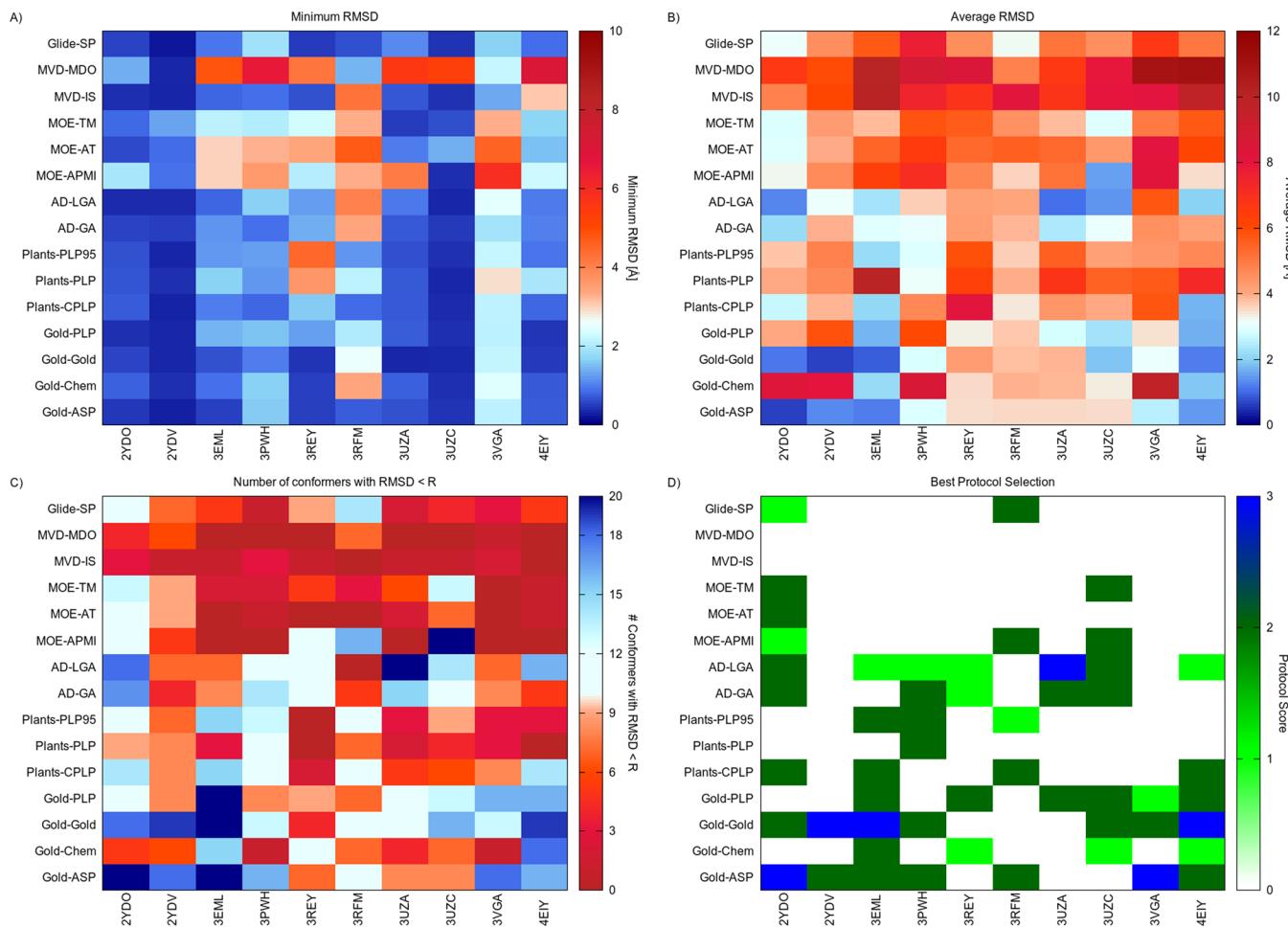


Figure 4. Results of the cognate ligand docking procedure.

represented by the docking program Gold combined with the GoldScore and ASP scoring functions.

Consensus Scoring. A well-known technique to improve the performances of the scoring functions is to combine the results of different functions into a consensus score.⁴³ Moreover, it has been demonstrated that combining different types of scoring functions increases the accuracy, as each scoring function compensates for the weaknesses of the other one. We therefore combined the results of the two scoring functions performing at best for the majority of the considered structures (ASP and GoldScore) and evaluate the performances of the thus obtained consensus scoring function (Table 5). As can be noted, the consensus score improves the overall prediction accuracy and gives good performances for all the structures (protocol score = 2). Moreover, for several protein structures, such as 2YDV, 3EML, 3PWH, 3UZC, and 4EIY, the consensus score represents the best performing protocol (protocol score = 3).

Ensemble Docking. For the above-described consensus protocol, we run ensemble docking calculations to assess whether the protocol is able to assign to each ligand (or ligand conformation in the case of ZMA) the corresponding cocrystallized structure by selecting it among all the considered proteins. For ensemble docking, we evaluated three different strategies for the definition of the binding site center by setting it as (i) the centroid of the barycenters of the ligands; (ii) the ZMA barycenter in the 3EML structure (upon which all of the

Table 5. Consensus Score Results^a

PDB ID	consensus protocol			best single protocol		
	RMSD _{ave} [Å]	N ^(RMSD < R)	score	RMSD _{ave} [Å]	N ^(RMSD < R)	score
2YDO	0.694	20	2	0.593	20	3
2YDV	0.415	20	3	0.600	19	3
3EML	0.783	20	3	0.853	20	3
3PWH	2.661	17	3	2.861	16	2
3REY	2.719	11	2	3.266	9	2
3RFM	3.661	12	2	3.163	14	2
3UZA	2.969	14	2	1.033	20	3
3UZC	1.081	19	2	1.489	20	2
3VGA	2.766	16	2	2.470	18	3
4EIY	1.127	19	3	1.156	19	3

^aFor each protein structure, the best protocols among all the tested ones in this study are reported in bold face text.

other structures were aligned at the beginning of the docking procedure); and (iii) the barycenter of each ligand in its corresponding crystal structure. The results collected in Table 6 report the percentage of protein selection for the generation of 20 conformers for each ligand. As can be noted, the definition of the binding site center does not significantly affect the outcomes. In all of the considered cases, the protocol preferentially selects three proteins, namely, the 4EIY, 2YDO, and 2YDV structures. These structures are characterized by high to low resolution (1.8, 2.6, and 3.0 Å, respectively) and the

Table 6. Percentage of Protein Selection of the Ensemble Docking Runs

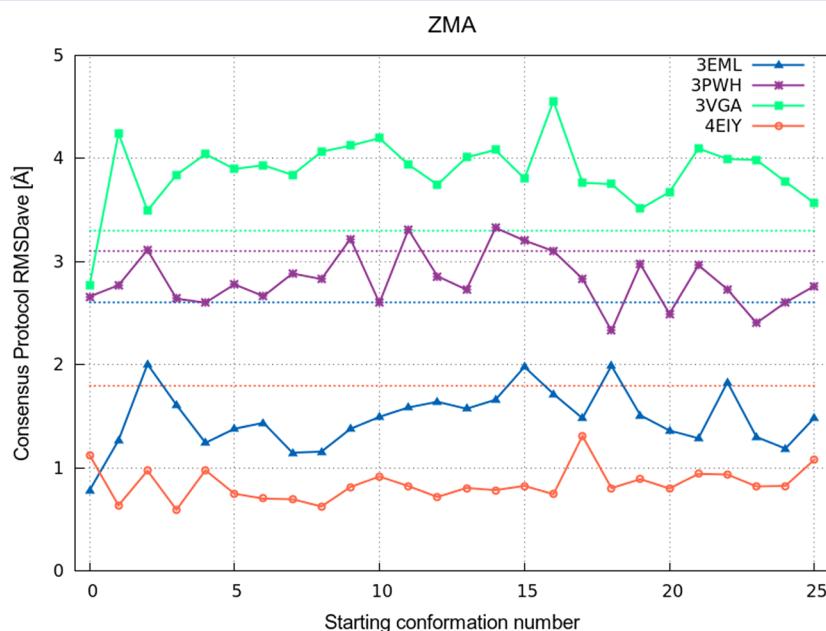
PDB ID	strategy 1	strategy 2	strategy 3
2YDV	28.00	26.00	27.50
2YDO	24.50	25.50	22.50
3EML	1.00	2.00	1.50
3PWH	2.00	1.50	1.00
3REY	0.00	0.00	0.00
3RFM	1.50	1.50	1.00
3UZA	1.00	0.00	0.50
3UZC	1.00	0.00	0.00
3VGA	0.50	0.00	0.00
4EIY	40.50	43.50	46.00

complete resolution of the second extracellular loop (EL2). We therefore ascribe the preference of the docking protocol for these structures to their completeness rather than to their resolution.

Evaluation of Additional Parameters. IEMs Inspection. As mentioned in the Introduction section, we employed a complementary metrics to evaluate the protocol performance, that we called IEMs. In our implementation, the IEMs are based on the analysis of key ligand-binding interactions observed in the crystal structures (Figure 2) and enable a fast and graphical selection of the conformations generated by the docking algorithms. The selection is guided by the quality, in terms of the number of established interactions and their relative strength, of the interactions occurring between each docking pose and selected key residues. The identification of key residues can be either knowledge-based, such as a comparison with available protein–ligand crystal structures (in our case) or mutagenesis data, or blind. In the latter case, the IEMs are computed for all residues surrounding the binding site within a user-defined radius and are devoid of any subjectivity. Moreover, to solve the automation issue that has been previously pointed out by several authors,^{21–24} we developed a python script that automatically generates the

IEMs from the computed interactions. Figure 6A depicts the IEMs of the output conformations generated by the consensus protocol for the 3UZA structure. A first glance at the map highlights the presence of two clusters of conformations. The comparison with the interaction energy computed for the cocrystallized ligand (representing the lowest row of the map) suggests that poses 2, 6, 8, 9, 12, and 13 considerably differ from the binding mode observed in the crystal structure. A superimposition of the poses (Figure 6B, Video S7, Supporting Information) reveals that those structures are placed far away from the binding site by the docking protocol and therefore have a higher RMSD with respect to the cocrystallized ligand. In this specific case, the IEM inspection helped in identifying pose clusters according to the types of interaction they establish with the receptor. The comparison of IEMs relative to the conformation generated by the same protocols for different structures (Figures S4–S8 and Videos S1–S10, Supporting Information) helps in evaluating protocol performances as well as the reproducibility of a crystal binding mode. By comparing panels A and B of Figure S5 (Supporting Information), it is straightforward that the binding mode of ZMA in the 3PWH structure (Video S4, Supporting Information) is more challenging to be reproduced as compared to the one observed in the 3EML structure (Video S3, Supporting Information). However, the irregularity of the interactions in the IEMs of the 3PWH, 3REY, and 3VGA structures (Figures S5B; S6A; and S8A; and Videos S4; S5; and S9, respectively, Supporting Information) mirrors the higher average RMSD obtained for those structures. The same conclusions can be drawn by comparing the performances of different protocols on the same structure (data not shown). We therefore believe that the IEMs can represent a complementary metrics that can help in evaluating the results of docking calculations in a fast graphic way.

Reconstructed EL2. The results obtained in the ensemble docking procedure show that the protocol preferentially selects complete structures. We therefore decided to evaluate the effect of the reconstruction of the second extracellular loop (EL2) for

**Figure 5.** Effect of starting conformation on consensus RMSD_{ave} . Structure resolution is depicted as a dotted line.

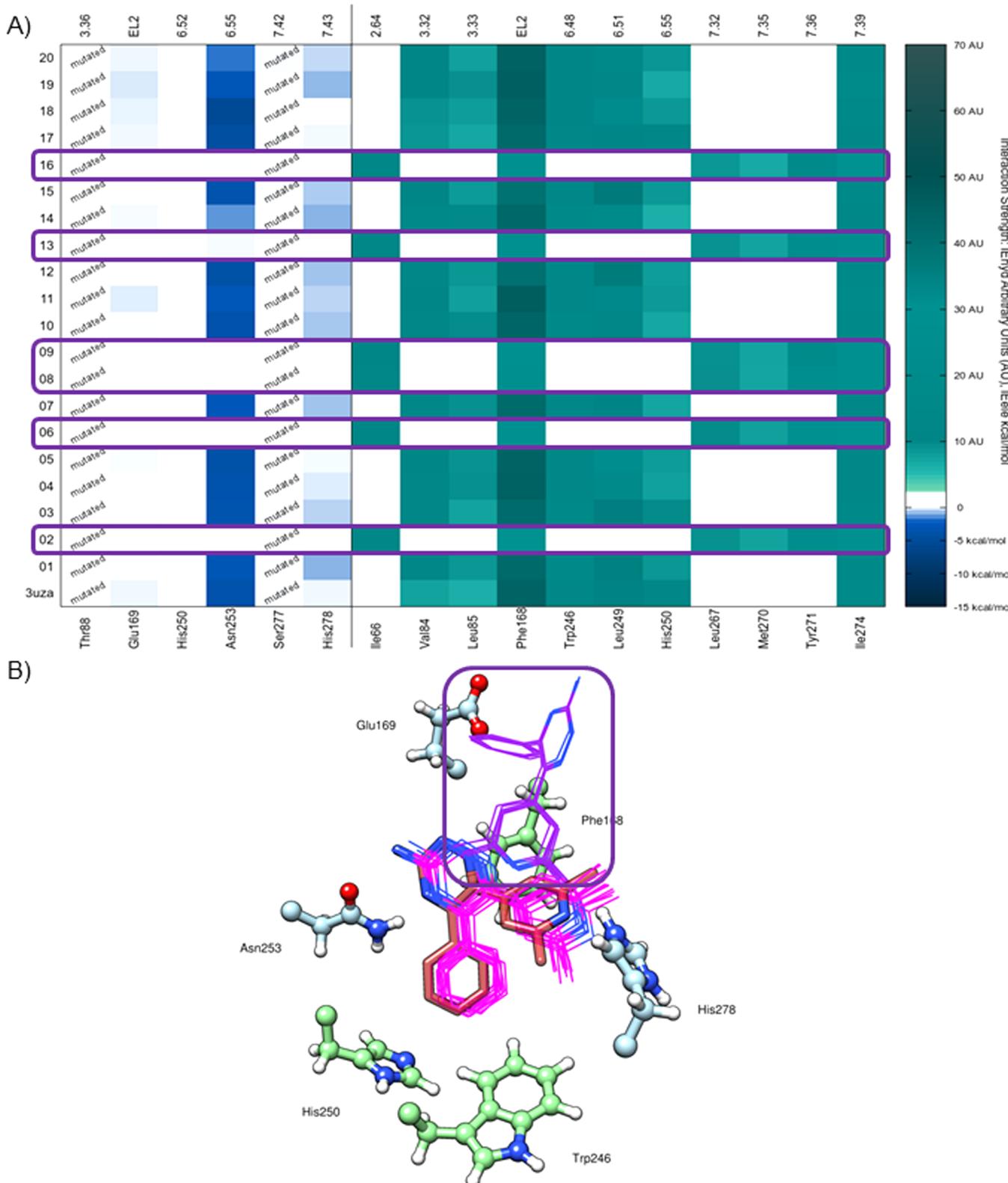


Figure 6. (A) IEM of the conformations generated by the consensus protocol for the 3UZA structure. (B) The two clusters of conformations are superimposed to the crystal structure (red sticks): the poses with the lower RMSD values are represented in magenta wires, whereas poses with higher RMSD are depicted as purple wires.

the structures that have this portion not completely solved. The loop has been reconstructed by using the “Paste Fragment” tool implemented in MOE. The missing portion of the EL2 (see Table S2, Supporting Information) has been copied from the

4EIY structure and pasted into the others, by using the “graft” options, that fits the fragment by superposing the flanking residues of the missing sequence and performs a short minimization. This method represents the least computa-

ally expensive choice and resulted in the most meaningful option in our specific case, as three of the available crystal structures have been solved with complete EL2 and structure superposition highlights very high structural similarity among them (see Figure S3, Supporting Information).

We repeated the cognate ligand docking calculations for the thus-obtained structures by comparing the performances of the consensus protocol prior to and after loop reconstruction (Table 7). As can be noted, the reconstruction of the EL2

Table 7. Results of Consensus Protocol for Structures with Reconstructed EL2

PDB ID	original structure			reconstructed EL2		
	RMSD _{ave} [Å]	N ^(RMSD<R)	score	RMSD _{ave} [Å]	N ^(RMSD<R)	score
3EML	0.783	20	3	1.680	20	2
3PWH	2.661	17	3	2.998	15	2
3REY	2.719	11	2	3.490	8	0
3RFM	3.661	12	2	3.684	0	0
3UZA	2.969	14	2	2.376	15	2
3UZC	1.081	19	3	0.811	20	3
3VGA	2.766	16	2	2.939	14	2

worsens the performance of the consensus protocol for the 3REY and 3RFM structures. The 3UZC structure, on the contrary, benefits from loop reconstruction, whereas the performances of other structures are only slightly affected. These results suggest that it is not possible to draw a general conclusion about the benefits of loop reconstruction. We therefore recommend when selecting a crystal structure for docking studies to pay attention to the completeness of the structure in the surroundings of the binding cavity. In case none of the available structures is completely solved, different loop reconstruction methods should be tested.

Starting Ligand Conformation. To evaluate the effects of the starting structure on the docking outcomes, we carried out test calculations by supplying different conformations as input. We performed the analysis by considering 25 different conformations of ZMA as initial structures: 24 conformations were obtained through a stochastic conformational search by leaving the default setting in the MOE conformational search (apart from the RMSD threshold that was increased to 1.0 Å), and another conformation was generated by simply performing an energy minimization. In both cases, the starting point was a structure designed with the MOE builder tool. We evaluated the performances of the consensus protocol for the 3EML, 3PWH, 3VGA, and 4EIY structures (Figure 5). As can be noted, the performances of the consensus protocol are less affected by the starting conformation, and the RMSD_{ave} values remain well below the crystal structure resolution for the 4EIY and 3EML structures, whereas for the 3PWH and to a greater extent for the 3VGA structures the performances are considerably worse. In the case of the 3VGA structure, all of the RMSD_{ave} values are all above the structure resolution. In all cases, no consistent trend is evidenced.

CONCLUSIONS

We have presented here an alternative assessment strategy to compare the performances of GPCR-ligand docking protocols based on complementary “quality descriptors”: (a) the number of conformations generated by a docking algorithm having a RMSD value lower than the crystal structure resolution

(N^(RMSD<R)); (b) a novel consensus-based function defined as protocol score; and (c) the IEM analysis, based on the identification of key ligand–receptor interactions observed in the crystal structures. We have selected as test case the hA_{2A} AR in complex with different ligands and evaluated the performances of 16 different docking/scoring combinations in generating poses close to the conformations observed in the X-ray structures. Common settings among the different selected docking programs have been defined, and two issues potentially affecting the docking outcomes, such as the input conformation and the reconstruction of protein missing portions around the binding site, have been tested and discussed.

The conclusions of our analysis can be summarized into a few points: (i) as expected, no universal docking protocol exists that can reproduce with satisfying accuracy all of the observed X-ray binding modes even when relative to the same receptor subtype cocrystallized with structurally related ligands; (ii) in the analyzed test case, the overall performances of the docking protocols benefit from the use of a consensus scoring function and are not considerably affected by the input conformation as well as by EL2 reconstruction. Exceptions to those general considerations have been observed for low resolution structures and for structures cocrystallized with low affinity ligands.

In view of the obtained results, we suggest using complementary metrics and additional statistical parameters to the traditional RMSD value to judge and compare the performances of docking protocols.

ASSOCIATED CONTENT

S Supporting Information

Procedure to compute IEMs; overview of crystal binding modes; cognate ligand docking results; comparison of EL2 completeness in the crystal structures; superimposition of structures having solved EL2; computed IEMs for the consensus protocol and all considered crystal structures; and video files. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel: +39 049 8275704. Fax: +39 049 827 5366. E-mail: stefano.moro@unipd.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was carried out with financial support from Italian Ministry for University and Research (MIUR), Rome, Italy. The molecular modelling work coordinated by S.M. has been carried out with financial support from the University of Padova, Italy, and the Italian Ministry for University and Research, Rome, Italy. S.M. is also very grateful to Chemical Computing Group for the scientific and technical partnership. S.M. participates in the European COST Action CM1207 (GLISTEN).

ABBREVIATIONS

ARs, adenosine receptors; EL2, second extracellular loop; GPCRs, G protein-coupled receptors; NECA, N-ethyl-5'-carboxamido adenosine; T4E, 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol; T4G, 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine; TM, transmembrane; ZM

241385, 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)-triazin-5-yl-amino)ethyl)phenol; XAC, *N*-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl- 2,3,6,7-tetrahydro-1*H*-purin-8-yl)-phenoxy]acetamide

■ REFERENCES

- (1) Carlsson, J.; Yoo, L.; Gao, G. Z.; Irwin, J. I.; Shoichet, B. K.; Jacobson, K. A. Structure-based discovery of A2A adenosine receptor ligands. *J. Med. Chem.* **2010**, *53*, 3748–3755.
- (2) Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. Structure-based discovery of beta2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6843–6848.
- (3) Kufareva, I.; Rueda, M.; Katritch, V.; Stevens, R. C.; Abagyan, R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **2011**, *19*, 1108–1126.
- (4) Carlsson, J.; Coleman, R. G.; Setola, V.; Irwin, J. J.; Fan, H.; Schlessinger, A.; Sali, A.; Roth, B. L.; Shoichet, B. K. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat. Chem. Biol.* **2011**, *7*, 769–778.
- (5) Katritch, V.; Jaakola, V. K.; Lane, J. R.; Lin, J.; Ijzerman, A. P.; Yeager, M.; Kufareva, I.; Stevens, R. C.; Abagyan, R. Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J. Med. Chem.* **2010**, *53*, 1799–1809.
- (6) Reynolds, K. A.; Katritch, V.; Abagyan, R. Identifying conformational changes of the beta(2) adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 273–288.
- (7) de Graaf, C.; Rognan, D. Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J. Med. Chem.* **2008**, *51*, 4978–4985.
- (8) de Graaf, C.; Kooistra, A. J.; Vischer, H. F.; Katritch, V.; Kuijer, M.; Shiroishi, M.; Iwata, S.; Shimamura, T.; Stevens, R. C.; de Esch, I. J.; Leurs, R. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J. Med. Chem.* **2011**, *54*, 8195–8206.
- (9) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (10) Katritch, V.; Abagyan, R. GPCR agonist binding revealed by modeling and crystallography. *Trends Pharmacol. Sci.* **2011**, *32*, 637–643.
- (11) Vilar, S.; Karpiak, J.; Berk, B.; Costanzi, S. In silico analysis of the binding of agonists and blockers to the beta2-adrenergic receptor. *J. Mol. Graphics Modell.* **2011**, *29*, 809–817.
- (12) Beuming, T.; Sherman, W. Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *J. Chem. Inf. Model.* **2012**, *52*, 3263–3277.
- (13) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.
- (14) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Cherezov, V.; Stevens, R. Structure of an agonist-bound human A2A adenosine receptor. *Science* **2011**, *332*, 322–327.
- (15) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, *474*, 521–525.
- (16) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the adenosine A2A receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **2011**, *19*, 1283–1293.
- (17) Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; Cameron, A. D.; Kobayashi, T.; Hamakubo, T.; Iwata, S.; Murata, T. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **2012**, *482*, 237–240.
- (18) Congreve, M.; Andrews, S. P.; Dore, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-triazine derivatives as adenosine A2A antagonists using structure based drug design. *J. Med. Chem.* **2012**, *55*, 1898–1903.
- (19) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Heitman, L. H.; Ijzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **2012**, *337*, 232–236.
- (20) Müller, C. E.; Jakobson, K. A. Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochim. Biophys. Acta* **2011**, *1808*, 1290–1308.
- (21) Perez-Nueno, V. I.; Rabal, O.; Borrell, J. I.; Teixido, J. APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening. *J. Chem. Inf. Model.* **2009**, *49*, 1245–1260.
- (22) Novikov, F. N.; Stroylov, V. S.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. Developing novel approaches to improve binding energy estimation and virtual screening: a PARP case study. *J. Mol. Model.* **2009**, *15*, 1337–1347.
- (23) Agostino, M.; Sandrin, M. S.; Thompson, P. E.; Yuriev, E.; Ramsland, P. In silico analysis of antibody-carbohydrate interactions and its application to xenoreactive antibodies. *Mol. Immunol.* **2009**, *47*, 233–246.
- (24) Agostino, M.; Sandrin, M. S.; Thompson, P. E.; Yuriev, E.; Ramsland, P. A. Identification of preferred carbohydrate binding modes in xenoreactive antibodies by combining conformational filters and binding site maps. *Glycobiology* **2010**, *20*, 724–735.
- (25) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relationships in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
- (26) MOE (Molecular Operating Environment), version 2012.10; Chemical Computing Group Inc.: Montreal, Canada, 2012.
- (27) Williams, T.; Kelley, C. Gnuplot: An Interactive Plotting Program; available at <http://www.gnuplot.info> (accessed Apr 18, 2014).
- (28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Protein Data Bank. *Nucl. Acids. Res.* **2000**, *28*, 235–242.
- (29) Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187–205.
- (30) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (31) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.
- (32) Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.
- (33) 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**, *16*, 2785–2791.
- (34) Glide, version 5.8; Schrödinger, LLC: New York, NY, 2012.
- (35) GOLD Suite, version 5.0; Cambridge Crystallographic Data Centre: Cambridge, UK. <http://www.ccdc.cam.ac.uk>.
- (36) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: application of ant colony optimization to structure-based drug design. *Lec. Notes Comput. Sci.* **2006**, *4150*, 247–258.
- (37) Thomsen, R.; Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49*, 3315–3321.
- (38) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

- (39) Sabbadin, D.; Ciancetta, A.; Moro, S. Bridging molecular docking to membrane molecular dynamics to investigate GPCR–ligand recognition: The human A_{2A} adenosine receptor as a key study. *J. Chem. Inf. Model.* **2014**, *54*, 169–183.
- (40) Sabbadin, D.; Ciancetta, A.; Moro, S. Perturbation of fluid dynamics properties of water molecules during GPCR-ligand recognition: the human A_{2A} adenosine receptor as a key study. *J. Chem. Inf. Model.* **2014**, Submitted.
- (41) Jaakola, V.-P.; Lane, J. R.; Lin, J. Y.; Katritch, V.; Ijzerman, A. P.; Stevens, R. C. Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: Identification and characterization of essential amino acid residues. *J. Biol. Chem.* **2010**, *285*, 13032–13044.
- (42) Cristalli, G.; Lambertucci, C.; Marucci, G.; Volpini, R.; Dal Ben, D. A2A adenosine receptor and its modulators: Overview on a druggable GPCR and on structure-activity relationship analysis and binding requirements of agonists and antagonists. *Curr. Pharm. Des.* **2008**, *14*, 1525–1552.
- (43) 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*, 1079–1093.