Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 42 (2007) 921–933
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

The consequences of scoring docked ligand conformations using free energy correlations

Francesca Spyrakis [a], Alessio Amadasi [a], Micaela Fornabaio [c], Donald J. Abraham [c], Andrea Mozzarelli [a], Glen E. Kellogg [c,*], Pietro Cozzini [b,**]

[a] *Department of Biochemistry and Molecular Biology, University of Parma, 43100 Parma, Italy*
[b] *Department of General and Inorganic Chemistry, University of Parma, 43100 Parma, Italy*
[c] *Department of Medicinal Chemistry, Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond, VA 23298-0540, USA*

Received 18 September 2006; received in revised form 4 December 2006; accepted 29 December 2006
Available online 21 January 2007

Abstract

Ligands from a set of 19 protein–ligand complexes were re-docked with AutoDock, GOLD and FlexX using the scoring algorithms native to these programs supplemented by analysis using the HINT free energy force field. A HINT scoring function was calibrated for this data set using a simple linear regression of total HINT score for crystal-structure complexes vs. measured free energy of binding. This function had an r^2 of 0.84 and a standard error of ± 0.42 kcal mol^{-1}. The free energies of binding were calculated for the best poses using the AutoDock, GOLD and FlexX scoring functions. The AutoDock and GoldScore algorithms estimated more than half of the binding free energies within the reported calibration standard errors for these functions, while that of FlexX did not. In contrast, the calibrated HINT scoring function identified optimized poses with standard errors near ± 0.5 kcal mol^{-1}. When the metric of success is minimum RMSD (vs. crystallographic coordinates) the three docking programs were more successful, with mean RMSDs for the top-ranking poses in the 19 complexes of 3.38, 2.52 and 2.62 Å for AutoDock, GOLD and FlexX, respectively. Two key observations in this study have general relevance for computational medicinal chemistry: first, while optimizing RMSD with docking score functions is clearly of value, these functions may be less well optimized for free energy of binding, which has broader applicability in virtual screening and drug discovery than RMSD; second, scoring functions uniquely calibrated for the data set or sets under study should nearly always be preferable to universal scoring functions. Due to these advantages, the poses selected by the HINT score also required less post-docking structure optimization to produce usable molecular models. Most of these features may be achievable with other scoring functions.

Keywords: HINT; Docking; Free energy scoring; Hydropathic analysis

1. Introduction

The continuing search for therapeutic agents at lower cost and at greater speed (lab bench to bedside) has highlighted the value of computational methods for virtual screening of real or hypothetical libraries to identify new compounds with affinity for the target biomacromolecule. *Docking* of ligands into models of the protein or macromolecule active site and *scoring* of these ligands with respect to "fitness" are the most critical issues in virtual drug screening methods [1–7]. Methods that are both fast and reliable are particularly required for the prediction of binding affinity when screening a large library of compounds. While significant strides in reaching this goal over the past several years have been reported, real application of these computational tools has

* Corresponding author. Tel.: +01 804 828 6452; fax: +01 804 827 3664.
** Corresponding author. Tel.: +39 0521 905669; fax: +39 0521 905556.
E-mail addresses: glen.kellogg@vcu.edu (G.E. Kellogg), pietro.cozzini@unipr.it (P. Cozzini).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2006.12.037

seldom delivered results equal to the promise of the validations and/or runs on test sets.

The docking and scoring paradigm can be thought of as two separate problems: first, to create plausible ''poses'' of the putative ligand within the active site, and second, to identify which of these poses is most likely to be true, i.e., binding most favorably with the target. The first problem is simply geometry, or more broadly informatics, i.e., how can we place a solid object (ligand) within a ''cavity'' of another solid (protein) in well-defined Cartesian space? The second problem is chemical, i.e., analyzing the specific structure and interactions of the docked ligand–protein model with an assigned score or ultimately a binding free energy. Often, however, the pose generation relies on some form of intermediate scoring to increase the plausibility of the poses presented for final evaluation. There are a wide range of algorithms and approaches used to produce docking poses [8–14]. Docking approaches can be classified into three categories: completely rigid body, partial flexibility and complete flexibility. In two-body systems, e.g., docking ligands into proteins, the first approach treats the protein and ligand as two independent but rigid bodies, the second treats only the ligand as flexible, i.e., allowing adjustment of its rotatable bonds, and the third considers both protein and ligand as flexible. While the first approach is too crude to be of much value, of the other two, only the semi-flexible approach is computationally accessible with widely available docking tools.

Docking and scoring programs have been reviewed and benchmarked extensively [5–7,15–19]. To summarize these reports in a few words, most programs are capable of producing viable docking poses, albeit at varying speeds, usually including one or more poses qualitatively similar to the known crystallographic conformation. However, the ability of their associated scoring tools to identify the correct (crystallographic) pose is considerably more problematic. Not unexpectedly, within the training sets for the various docking/scoring codes, the results are significantly better than when these tools are applied to other data sets. The typical validation experiment is to correlate the pose scores with the RMSD (root mean square deviation) of each pose with respect to the crystallographic (experimental) pose [20,21]. In principle the best score will correspond to the lowest RMSD. There are a few realities in computational molecular modeling (and crystallography is, at its core, modeling) that suggest caution in completely relying on this approach. First, except in rare cases of very high resolution or neutron diffraction, there is measurable uncertainty in the ''reference'' experimental pose. There are actually a number of cases where the entire ligand orientation is ambiguous [22,23]. However, the crystallographic pose is a ''frozen'' model of what is actually a very dynamic process at biologically relevant temperatures. The measured free energy of binding, generally obtained at room temperature or higher, is a ''weighted'' composite of these states. In other words, higher quality crystal structures may not completely translate to better free energy predictions as they only represent a few of the many states. Second, in designing and calibrating docking scoring functions, optimization of the correspondence between the best scoring pose and the crystallographic structure (i.e., minimizing RMSD) is a somewhat contrived goal that is perhaps not totally relevant to virtual screening. Last, most dock scoring functions are constructed without explicit consideration of the contribution of entropic terms and hydrophobic interactions distinct from London forces.

We have previously described the properties of the HINT force field and free energy scoring tool for understanding protein–ligand interactions in biological media [24–27]. A number of relevant principles to virtual screening, i.e., the importance of hydrophobic interactions and entropy in modeling free energy, and the energetic influences of active site (bridging) waters and ionization states of acidic and basic residues and ligand functional groups were highlighted. Here we examined 19 protein–ligand complexes for which crystallographic and binding energy data were available. After some limited model cleanup, we calculated an experimental vs. calculated free energy correlation for the set using the HINT free energy scoring tool [24]. The ligands for these complexes were then re-docked into their host proteins using three popular docking codes and the resulting poses were evaluated using several scoring criteria.

Our principal motivation for the present work is to answer the following: does a calibrated free energy scoring tool have distinct advantages over the current collection of docking score functions for evaluating calculated poses? The answer to that question reveals several guidelines governing the usage of docking and scoring studies in modern drug discovery that should be of general interest to medicinal chemists who use or interpret docking results. These results suggest, in part, why virtual screening has yet to deliver results commensurate with expectations [7]. One particular point of emphasis is a comparison of scoring functions with respect to their usefulness in virtual screening as opposed to selecting poses with lowest root mean square deviation (RMSD) between experimental and docked conformations.

2. Materials and methods

2.1. Protein–ligand test set

The structures of the 19 analyzed protein–ligand complexes (Table 1) were retrieved from the Protein Data Bank [28] (www.rcsb.org). The structures were chosen according with the following criteria: non-covalent binding between protein and ligand, crystallographic resolution lower or equal to 3 Å, and inhibition constant values ranging from μM to nM. Also, in order to study a heterogeneous set representative of different existing structural architectures, proteins characterized by multiple structural motifs were selected. In a few cases we chose for the data set multiple complexes of the same receptor bound with different or very similar ligands. This was done in order to test the capability of the docking programs and scoring functions to deal with diverse conformations of the same binding pockets, and to correctly predict the effects of subtle modifications of the ligand structure. All selected complexes passed an additional test for compatibility with

Table 1
Structural and experimental dissociation data for the 19 protein–ligand complexes

PDB code	Protein	Res (Å)	Structure Ref.	Ligand (see Scheme 1)	pK_i (M)	ΔG (kcal mol^{-1})	K_i Ref.
1abe	ABP	1.70	[58]	L-Arabinose (**1**)	7.01	−9.53	[59]
7abp	ABP	1.67	[60]	Fucose (**2**)	6.46	−8.78	[60]
2gbp	GBP	1.90	[61]	Glucose (**3**)	7.40	−10.07	[62]
1cbr	RABP	2.90	[63]	Retinoic acid (**4**)	7.78	−10.58	[64]
1d3q	Human thrombin	2.90	[65]	TRH2 (**5**)	6.54[a]	−8.88	[66]
1d3t	Human thrombin	3.00	[65]	TRH1 (**6**)	6.43	−8.73	[66]
1ett	Bovine thrombin	2.50	[67]	TRB2 (**7**)	5.89	−8.00	[68]
1flr	Immunoglobulin	1.85	[69]	Fluorescein (**8**)	–[b]	−10.98	[70]
2cgr	Immunoglobulin	2.20	[71]	*N*-*N*′*N*″-trisubstituted guanidine compound (**9**)	7.28	−9.89	[71]
1lie	LBP	1.60	[72]	Palmitic acid (**10**)	7.08	−9.63	[73]
1lif	LBP	1.60	[74]	Stearic acid (**11**)	7.10	−9.65	[73]
1pph	Trypsin	1.90	[75]	3-TAPAP (**12**)	5.92	−8.04	[75]
1q9m	PDE4D	2.30	[76]	Rolipram (**13**)	6.00	−8.16	[77]
1y2d	PDE4D	1.70	[78]	Pyrazole-based inhibitor (**14**)	6.00	−8.16	[78]
1soj	PDE3B	2.90	[79]	IBMX (**15**)	6.92	−9.40	[79]
1ulb	Purine nucleoside phosphorylase	2.75	[80]	Guanine (**16**)	5.30	−7.20	[56]
2ack	Acetylcholinesterase	2.40	[81]	Edrophonium ion (**17**)	6.60	−8.97	[82]
2r04	Rhinovirus 14 coat protein	3.00	[83]	Antiviral agent (**18**)	6.52	−8.87	[84]
3ert	ERα	1.90	[85]	4-Hydroxytamoxifen (**19**)	7.97	−10.83	[86]

[a] Derived from K_{ass} app (Lmol$^{-1} \times 10^6$).
[b] No K_i value was available.

the HINT forcefield. Scheme 1 illustrates the chemical structures of the ligands in this study.

2.2. Model building

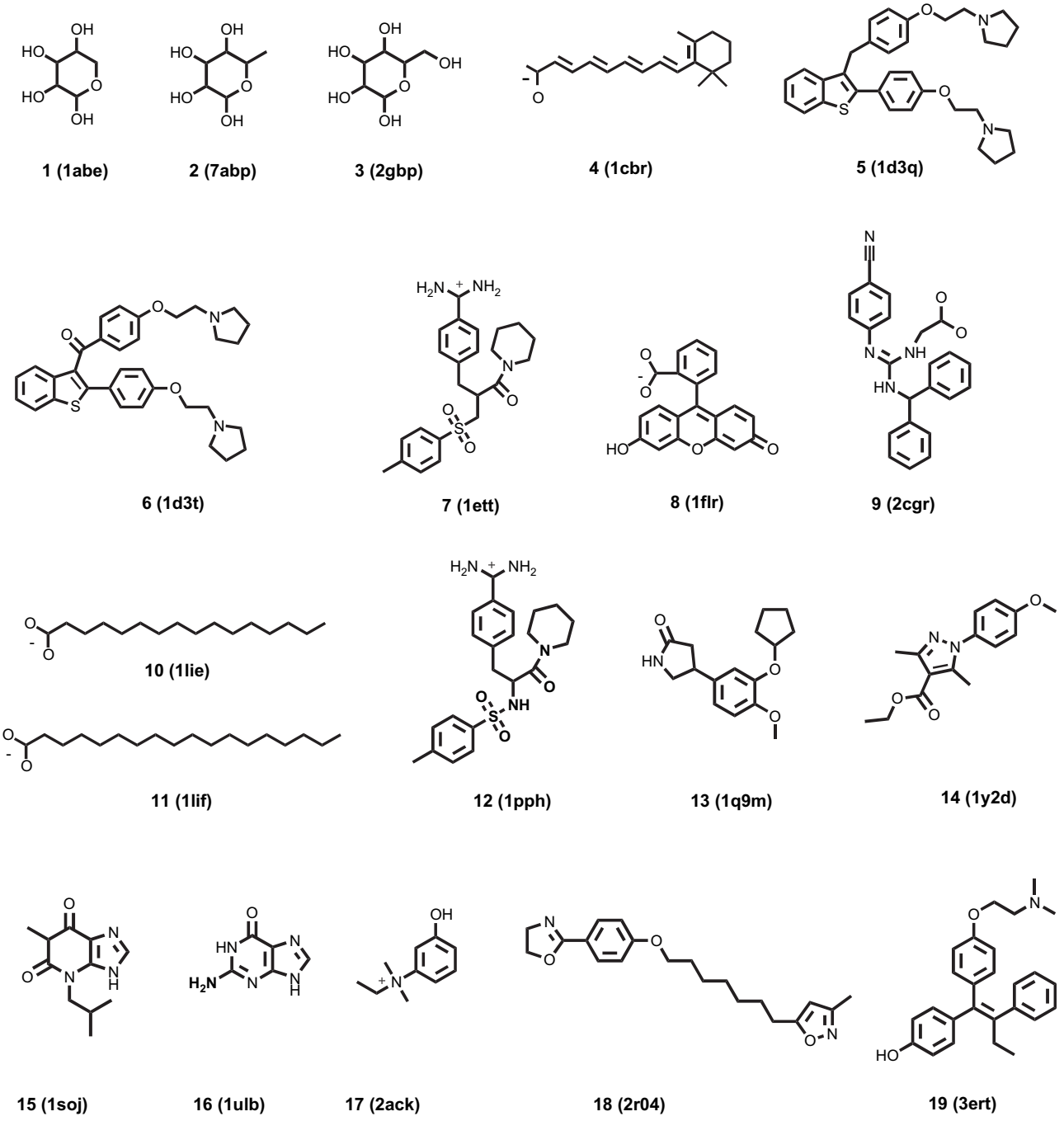
All complexes were modeled with Sybyl version 7.1 (www.tripos.com). The structures were carefully checked and corrected for chemically consistent atom and bond type assignment. Hydrogen atoms, not normally detected with common X-ray diffraction techniques, were computationally added using the Sybyl Biopolymer and Build/Edit menu tools. Water hydrogens were positioned to maximize the number of potential hydrogen bonds. Added hydrogen atoms were then energy minimized, while heavy atoms were fixed, with a convergence gradient of 0.5 kcal (mol Å)$^{-1}$ for 1500 cycles, in order to avoid steric clashes and negative repulsive interactions. Binding site amino acids were carefully inspected to correctly orient the rotatable groups and to properly assign the protonation state of ionizable residues. In these models at physiological pH, aspartate and glutamate were deprotonated, while lysine and arginine were protonated. Analogously, all ligand carboxylic groups were deprotonated, and tertiary amines and guanidiniums were protonated and positively charged.

2.3. Hydropathic analysis of the crystallographic complexes

Before docking, the 19 complexes were hydropathically analyzed and scored with the HINT force field, to determine the correlation between the experimental binding free energies and the HINT score for the protein–ligand interaction (see Fig. 1). The calculations were performed as previously reported [24–27], using a HINT version (3.10Sβ) with local modifications to the commercial version (www.tripos.com). The HINT score is a double sum overall atom–atom pairs of the product of the hydrophobic atom constants (a_i, partial Log $P_{octanol/water}$) and atom solvent-accessible surface areas (S_i) for the two interacting atoms, mediated by a function of the distance between the atoms [24–27]. Partition calculations were performed with the calculate method for the ligands and the dictionary method for the proteins [29]. A new ''semi-essential hydrogens'' partition mode that treats polar hydrogens and hydrogens bonded to unsaturated carbons and carbons α to heteroatoms explicitly was used. In addition, hydrogens bonded to unsaturated carbons were, along with polar hydrogens, allowed to act as hydrogen bond donors. This is in accordance with several recent observations [30,31], suggesting that some C–H···O hydrogen bonds are possible. The HINT option that corrects the S_i terms for backbone amide nitrogens (including their hydrogens) [32] by adding 20 Å^2 was used in this study to improve the relative energetics of inter- and intra-molecular hydrogen bonds involving these nitrogens. The total HINT score considers all possible interactions between the protein and the ligand; in this study of heterogeneous complexes the water contribution was not explicitly considered.

2.4. Docking algorithms and protocols

All docking analyses were carried out on the 19 protein–ligand complexes using the docking algorithms employed in AutoDock 3.05 [13,33–35], GOLD 2.2 [12,36] and FlexX 1.13.5 L [9,37]. While different numbers of docked poses were generated by the three programs due to constraints on this parameter inherent in the programs, documentation for each program asserts that these solutions exhaustively sample ligand conformational space.



Scheme 1.

2.4.1. AutoDock

AutoDock generated the different ligand conformers using a Lamarkian genetic algorithm (LGA), a GA implementation with an adaptive local method search [13]. The simulations started with a predefined number of generation cycles, composed of mapping and fitness evaluation, selection, crossover, mutation and elitist selection steps, and continued with a local search, followed by the scoring of the generated conformers. The energy-based AutoDock scoring function includes terms accounting for short-range van der Waals and electrostatic interactions, loss of entropy upon ligand binding, hydrogen bonding and solvation. The protein and the ligand input structures, prepared as described above, were transformed into the corresponding pdbq format files (containing atom coordinates, partial charges and solvation parameters), with the mol2-topdbqs and AutoTors programs, respectively. The ligand-centered maps were generated by the program AutoGrid with a spacing of 0.375 Å and dimensions of $60 \times 60 \times 60$ points. The ligand rigid roots were automatically set and all possible rotatable bonds and torsions were defined as active. All other parameters were as set by default. For each docking simulation 255 different conformers were generated.

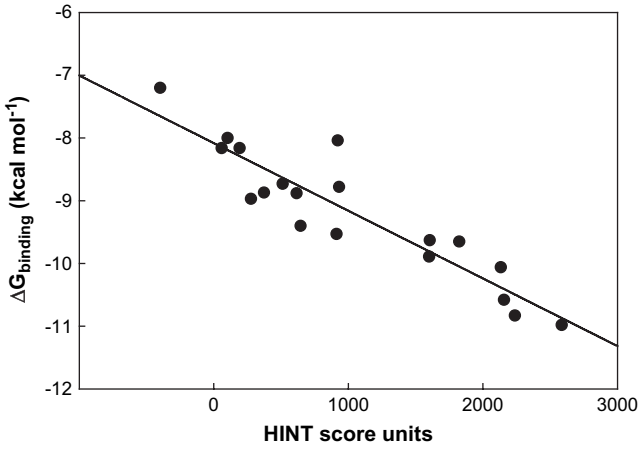


Fig. 1. Correlation between the experimental binding free energies and the calculated HINT score values for the 19 crystallographic protein–ligand complexes.

2.4.2. GOLD

The GOLD docking program is based on the evolutionary development of ligand conformations generated by a genetic algorithm, employing an island model and a niching technique to maintain diversity inside the population [12]. Genetic operations like migration, crossover and mutation were performed on the starting subpopulations, after which new individuals replaced the worst members of the niche. The generated conformers were ranked using the GOLD Fitness scoring function including energy terms for hydrogen bonding, short-range van der Waals interactions and an internal energy term for the ligand that is a sum of ligand steric and torsional energies. Both protein and ligand, including hydrogens, were submitted to the GOLD docking simulations in PDB format. The docking region was defined as a 20 Å sphere centered on the crystallographic ligands. All the parameters, except the number of generated poses that was set to 50, were the defaults. Thus, the number of islands was 5, the population size was 100, the number of genetic operations was 100,000 and the niche size was 2, whereas operator weights for crossover, mutation and migration were set to 95, 95 and 10, respectively. The 50 resulting docked conformations were saved in a multi PDB file.

2.4.3. FlexX

FlexX employs a fragment-based incremental construction algorithm [9]. First, the starting (base) fragment is placed into the binding pocket using a hashing technique and then, the remaining fragments are added step by step, optimizing at each step the new conformation. The ligand interaction centers are mapped on the complementary protein groups and the best scored positions are retained. The scoring function is an adaptation of the Böhm (LUDI) function and includes the following energetic contributions: entropic, hydrogen bonding, electrostatic interactions, aromatic and lipophilic contacts, all of which are scaled by heuristic distance and angle-dependent penalizing functions. The protein structure input data is a heavy atom-only PDB file. The ligand data, prepared as described above, were submitted for docking procedure in mol2 format. The active site and the sub-active site were usually defined by using the reference crystallographic ligand with 10.5 Å and 4.5 Å cutoff distances, respectively. Some particularly shallow binding pockets were analyzed with these parameters set to 6.5 Å and 3.5 Å. Docking runs generating 100 poses each were performed using the default parameters for iterative growing and subsequent scoring of the poses.

2.5. Docking procedure and analysis of results

Each ligand was docked back into its corresponding crystallographic binding site. As required by FlexX, the starting ligand coordinates were energy minimized before docking. Otherwise, all methods used the same receptor and ligand heavy atom coordinates. The accuracy of each predicted pose was evaluated on the basis of the root mean square deviation between the coordinates of the heavy atoms of the ligand in the top docking pose and those in the crystal structure, and of the score calculated by the different native scoring functions in the docking programs. In addition, all the generated poses were submitted for HINT hydropathic analysis. In cases, e.g., AutoDock, where the resulting PDB output file has an incomplete molecular description with respect to atom-typing and bond order, the dock conformation coordinates were superimposed on a fully defined structural prototype for scoring. The energy values assigned by the AutoDock, GOLD, FlexX and HINT scoring functions to the top docking poses were compared and correlated to the experimental binding free energy values. As a further step, the best scored poses were then energy minimized to a convergence gradient of 0.05 kcal (mol Å)$^{-1}$ for 1000 cycles (again the original receptor coordinates were fixed). This calculation step reduced some severe steric clashes between the protein and the ligand. Also, this allowed us to understand the effects of post-docking minimization on the conformations chosen by the various scoring functions.

3. Results and discussion

Identifying new potential lead compounds *in silico*, without actually performing expensive and time-consuming library screening, has become a most intriguing challenge of computational chemistry. Designing and applying docking and scoring tools that are accurate, precise and thus able to identify the correct pose of ligands docked in target binding pockets is a principal goal [8–14]. Many studies have been performed to identify which docking program performs best in reproducing the crystallographic orientation of different compounds in their corresponding receptors [4,5,17,21,38–45]. A low RMSD value between the coordinates of the ligand in the top-scoring docking pose and those in the crystal structure has always been interpreted as an essential requirement of a good docking tool. However, the parallel goal of developing scoring functions that accurately describe the *chemistry* of binding is not necessarily achieved by optimizing scoring functions for low RMSD. Accurate predictions of binding

free energy would allow not only good discrimination between active and inactive molecules, but also among closely related analogs [42]; this latter case being particularly important for drug design. Correctly predicting not just the binding mode but also the binding energy is, in fact, a primary exigency in all docking simulations and, in particular, in virtual screening applications.

Consensus scoring approaches combining multiple scoring functions into a single metric have been created [15,16], because certain combinations of docking algorithms and scoring functions have been shown to work better on some systems than on others [18,19,21,43]. Likewise, simply re-scoring docking poses with independent functions is another valuable adjunct approach [46–50]. In this context we used the HINT scoring function [51] as a post-docking processor tool to evaluate the results of docking simulations performed by three well-known and widely used docking programs: AutoDock, GOLD and FlexX [9,12,13], on a set of 19 crystallographic complexes. We have previously shown that the free energy HINT scoring tool is generally effective at predicting the free energy of binding [24,26,52–54].

The 19 complexes in this study were all characterized by a resolution ≤ 3.0 Å and, as described in Section 2, were chosen to have a set of structurally diverse ligands, complexed with a heterogeneous collection of proteins (see Table 1). Before starting the docking experiments, the interactions between the receptors and the corresponding crystallographic ligand poses were evaluated with the HINT force field [51] and compared to the experimental measurements (Table 1). The correlation of the calculated total HINT scores (H_{TOTAL}) with the experimentally determined free energies of binding is described by the following equation

$$\Delta G_{\mathrm{binding}} = -0.001079 H_{\mathrm{TOTAL}} - 8.08, \quad (1)$$

with a good $r^2 = 0.84$ and standard error of ± 0.42 kcal mol^{-1}. This new calibration of the HINT scoring function applies only to this data set and will be the basis of the following results. The slope of eq. 1 correlation line, indicating that about -925 HINT score units represents 1 kcal mol^{-1} in $\Delta\Delta G$ is different from previous reports [24–26,51], mostly due to recent re-parameterizations of the HINT force field. The *y*-intercept is a composite of numerous other energetic contributions to free energy such as internal energies of the molecules, etc., which can be treated as largely invariant within this data set. Thus, molecules that have overall negative HINT scores can still have favorable (negative) $\Delta G_{\mathrm{binding}}$. As will be seen below, scoring functions with "global" calibrations may not be particularly effective in scoring diverse docked data sets that were not part of their calibration sets.

3.1. Free energy calculated with the native docking score functions

The 19 ligands were extracted and re-docked with AutoDock, GOLD (both based on genetic algorithms) and with FlexX (based on incremental construction). As defined by Kontoyianni [43], docking consists of two parts: (i) accurate prediction of the bioactive ligand conformation in the active site and (ii) estimation of the tightness of target–ligand interactions. We paid particular attention to this second aspect and have analyzed the docking results considering not only the RMSD, but also the predicted binding free energy calculated by the scoring functions of the three programs. In other words, the energy estimation was not just used to identify the top ranked solution, but also to evaluate the affinity of the protein–ligand association for that pose. The best-scored poses generated by each program were considered to be the most probable orientations of those ligands in their binding pockets, and the estimated binding energies were compared to the experimentally determined values (see Fig. 2).

The AutoDock empirical scoring function (Fig. 2a) failed to predict the binding energy of 37% of the analyzed complexes within the reported calibration standard error of ± 2.18 kcal mol^{-1} [13]. In particular, 1ett, 1d3q, 1d3t, 1pph and 2cgr are far removed from the hypothetical correlation and their binding free energies are significantly overestimated. In fact, there is no discernable correlation even ignoring the one proposed by the AutoDock scoring function. We attributed this discrepancy to the different nature of the complexes subjected to this analysis and those in the AutoDock calibration docking study. Clearly, the quality of docking results is highly dependent on the nature of the analyzed systems, but also there is probably just too much diversity of structure and function in protein–ligand complexes for a single scoring function to have universal applicability. Nevertheless, the AutoDock scoring function does perform credibly in identifying the correct orientation of the ligands into their corresponding binding pockets (vide infra).

Most of the same complexes, i.e., 1ett, 1d3q, 1d3t, 1pph, 2r04 and 3ert, were also significantly overestimated by the GoldScore scoring function (Fig. 2b). Fifty conformers for each complex generated by GOLD were ranked using the Fitness algorithm, such that the conformer showing the highest Fitness value was considered to be the most favorable ligand orientation for an optimal protein–ligand association. However, for correlation with experimentally determined binding affinities the Fitness values were transformed to GoldScore by subtracting the intra-molecular term as suggested by Verdonk and co-workers [55]. GoldScore was demonstrated to perform better than either the GOLD Fitness or the Chemscore methods in estimating binding free energies [55]. About 74% of the free energy predictions fall within the standard error range of Verdonk (± 2.51 kcal mol^{-1}), i.e., they are well-predicted by the GoldScore algorithm.

Different, but still not encouraging, results were obtained with FlexX (Fig. 2c), which ranked the generated docking solutions using a scoring function similar to that developed by Böhm [9,56]. Several complexes (1cbr, 1d3q, 1d3t, 1lie, 1lif, 1q9m, 1soj, 1y2d, 2ack, and 2r04) are far removed from the hypothetical correlation and also from the reported standard error interval of the scoring function (± 3.30 kcal mol^{-1}) [9]. In this case the corresponding binding free energies are significantly underestimated. As with AutoDock, there is little

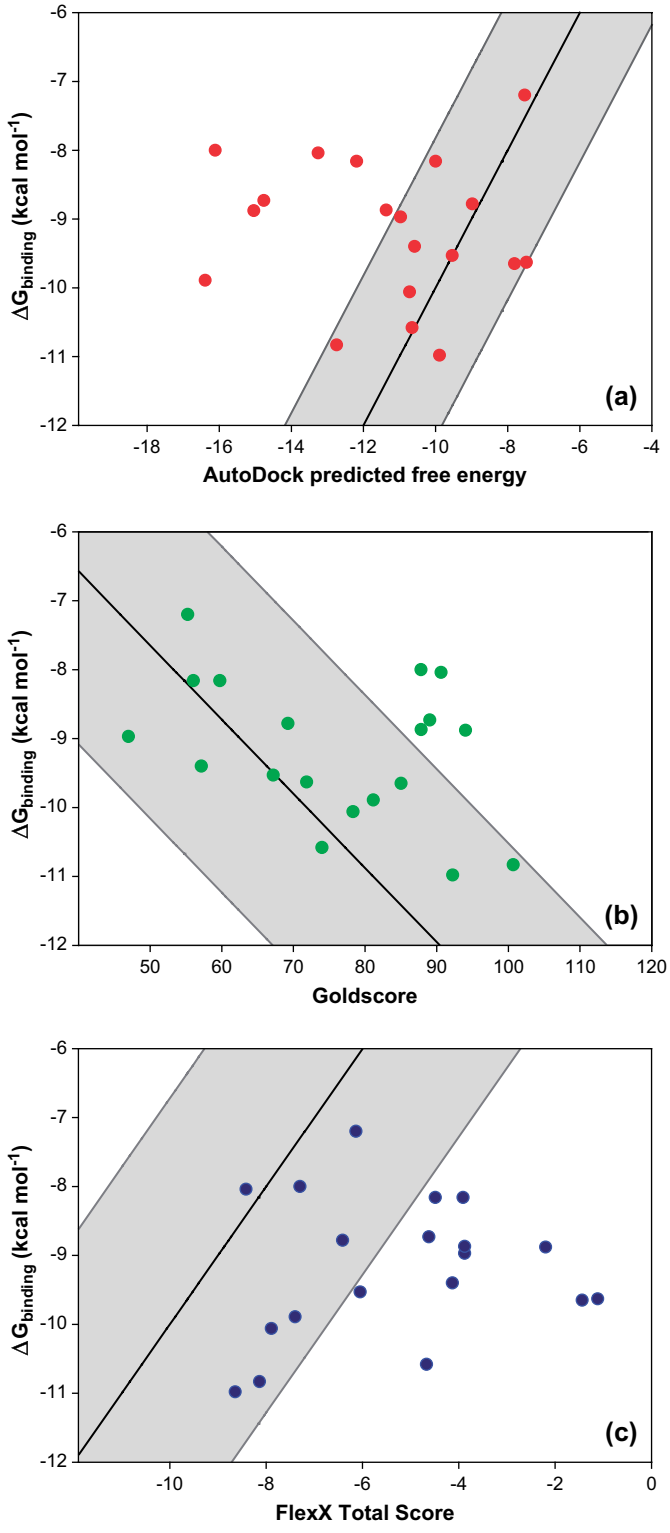


Fig. 2. Correlation between the experimentally determined binding free energies and the binding energies predicted by the docking program's scoring functions for the top scoring poses. (a) Docking simulations using AutoDock 3.05. From the AutoDock calibration [13], the solid line shows a perfect fit while the shaded area illustrates one standard error (±2.18 kcal mol^{-1}). (b) Docking simulations using GOLD 2.2 with binding energies predicted using the GoldScore scoring function. The solid line represents the calibration correlation reported by Verdonk [55], where the standard error is ±2.51 kcal mol^{-1}. (c) Docking simulations using FlexX 1.13.5 L. From the FlexX calibration [9], the solid line shows a perfect fit and the standard error is ±3.28 kcal mol^{-1}.

indication of any correlation between the experimental and calculated free energy values.

Taken together these results suggest that there is, at best, variable reliability in free energy estimates from these scoring functions. We will return to the issues related to identifying the docked conformation closest in RMSD to the crystal structure in a later section. Even if the ligand orientation *is* properly predicted, estimating protein–ligand association energetics remains a very difficult problem crucial for docking *new* compounds into the active site for drug discovery; i.e., not just matching known crystal structures.

3.2. Re-scoring and re-ranking with HINT

Because of our previous successes using the HINT force field in terms of energy prediction [24–26], we used it on these docking results as a post-processor tool: re-scoring and re-ranking all poses generated for each ligand–protein complex by the three docking programs. Thus, we are suggesting that the pose with the highest HINT score corresponds to the most fitting candidate. Usually, the pose identified by the docking program as top ranking did not correspond to the orientation with the highest HINT score; in fact, choosing the "best" pose by the docking program's scoring function or choosing by the HINT force field could lead to very different results and conclusions. The correlations between the HINT score values for the best poses and the experimental binding free energies are shown in Fig. 3. HINT score values for the best poses generated by AutoDock (red symbols and fit line) as selected and scored by HINT led to

$$\Delta G_{\text{binding}} = -0.000725 H_{\text{TOTAL}} - 8.25, \quad (2)$$

with $r^2 = 0.74$ and a standard error of ±0.54 kcal mol^{-1}. GOLD poses (green symbols and line) have a correlation of

$$\Delta G_{\text{binding}} = -0.000931 H_{\text{TOTAL}} - 7.99, \quad (3)$$

with $r^2 = 0.77$ and a standard error of ±0.51 kcal mol^{-1}. Finally, the combination of FlexX docking and HINT (blue symbols and line) scoring resulted in

$$\Delta G_{\text{binding}} = -0.000865 H_{\text{TOTAL}} - 8.20, \quad (4)$$

with $r^2 = 0.78$ and a standard error of ±0.50 kcal mol^{-1}. When the best poses are chosen by the docking programs algorithms but scored by HINT, the correlations are notably worse with r^2 (standard errors) of 0.23 (±0.93 kcal mol^{-1}), 0.46 (±0.80 kcal mol^{-1})[1] and 0.43 (±0.80 kcal mol^{-1}) for AutoDock, GOLD and FlexX, respectively.

Comparing the correlation of Fig. 1 (crystallographic conformation) with those of Fig. 3 suggests that there is little variation between experimental and post-docking free energy correlations for the conformers chosen and scored by HINT.

[1] This correlation was calculated after removing the 2cgr complex from the GOLD data set because of an extremely negative HINT score value arising from steric clashes between the docked pose and protein.

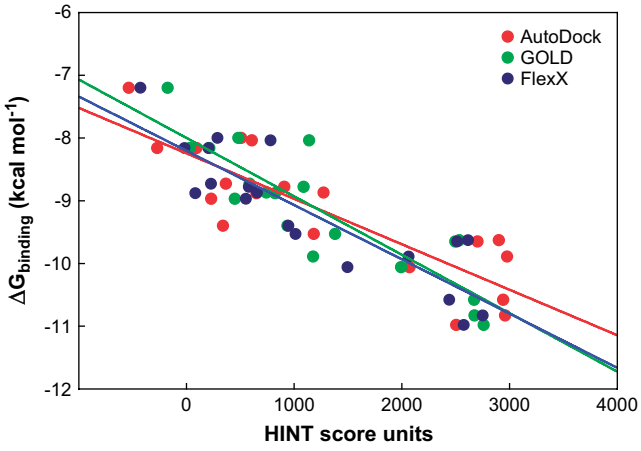


Fig. 3. Correlation between the experimental binding free energies and the HINT score values calculated after extracting and re-docking each ligand into its corresponding binding pocket. Poses selected by best HINT score. Red symbols and line: correlation from the combination of AutoDock and HINT; green symbols and line: correlation from the combination of GOLD and HINT; blue symbols and line: correlation from the combination of FlexX and HINT. (For interpretation of the references to color in text and in figure legends, the reader is referred to the web version of this article.)

The statistical parameters are consistently good, and the slopes of the correlation equations are reasonably similar. Also of significance is that, since the correlations produced by the AutoDock/HINT, GOLD/HINT and FlexX/HINT protocols are similar to each other, HINT's performance as a post-processing tool is independent of the docking algorithm used if the optimal solution is present among the collection of poses calculated by the program.

3.3. The effect of optimizing for lowest RMSD

If the metric of docking success is defined to be identifying the lowest RMSD conformer with the highest score, the docking programs are more successful: the best AutoDock scoring pose has RMSD $\leq$ 1.0 for 5/19 cases, the best scoring GOLD pose has RMSD $\leq$ 1.0 Å for 7/19 cases and the best scoring FlexX pose has RMSD $\leq$ 1.0 for 5/19 cases. Within the top 10% scoring poses, models with RMSD $\leq$ 1.0 Å were identified 7/19, 9/19 and 9/19 times by AutoDock, GOLD and FlexX, respectively. Surprisingly, these low RMSDs do not always correlate to accurate free energy predictions, even with the calibrated HINT score function. This is visually illustrated in Fig. 4, a color map (legend Fig. 4a) summary of the docking results. Fig. 4b illustrates the results when the best conformer is chosen by the docking program that generated the poses. There are three columns for each program's results: first, the RMSD between that pose and the crystal conformation, ranging in shade from white ($\leq$1 Å) to black (>7 Å); second, $\Delta\Delta G_{dock}$: the difference between the free energy calculated by the docking program's score function and the experimental free energy of binding, ranging in color from brown (<-7 kcal mol^{-1}) through light beige (≥-1 kcal mol^{-1}) and light blue ($\leq$1 kcal mol^{-1}) to dark blue (>7 kcal mol^{-1}); and third, $\Delta\Delta G_{HINT}$: the difference between the free energy calculated by the HINT score algorithm and the experimental free energy of binding, ranging in color from dark red (<-7 kcal mol^{-1}) to light pink (≥-1 kcal mol^{-1}) and light green ($\leq$1 kcal mol^{-1}) to dark green (>7 kcal mol^{-1}). As expected the second column, mirroring the data of Fig. 2, shows generally poor free energy predictions, overestimated for AutoDock and GOLD, and underestimated for FlexX. However, $\Delta\Delta G_{HINT}$, where the HINT score was used to predict the binding free energy on the poses selected by the docking programs indicates somewhat improved results.

In Fig. 4c a similar representation is made for poses selected by the HINT scoring function. Again, the first column is the RMSD for the pose with respect to the crystal conformation, while the second column is $\Delta\Delta G_{HINT}$*, where the HINT score was used to predict the binding free energy on the HINT-selected poses. Here, 6/19, 8/19 and 6/19 poses for models built with AutoDock, GOLD and FlexX, respectively, had RMSD $\leq$ 1.0 Å. More interesting is the quality of free energy predictions on the poses selected by HINT. Only a small handful have errors >1.0 kcal mol^{-1}. Of those with larger errors in predicted ΔG, 1lie and 1lif are most problematic, with errors of prediction for the poses calculated by all three docking programs. This is probably due to the highly flexible nature of these ligands (Scheme 1: **10** and **11**) that provide generally indistinct energy landscapes for docking.

There are a few cases where the RMSD results are poor while the free energy predictions are fairly good; for example, in 1d3q and 1d3t (ligands **5** and **6**) neither the native docking functions nor the HINT scoring function are able to select a conformation similar to the crystallographic pose. This is likely due to failures of the search algorithms of the poses generated by FlexX for 1d3q and 1d3t, none showed RMSDs less than 9.14 Å and 7.03 Å, respectively. High RMSD values were found also for many of the high-scoring poses for 2r04 (Fig. 4b and 4c) probably because ligand **18** is pseudo-symmetric, and appears to fit in the binding pocket in two orientations, one inverted with respect to the other (see Fig. 5a). Thus, despite a high RMSD (>7 Å), the energetic predictions for 2r04 are in error by <1 kcal mol^{-1}. This complex might be an interesting case for re-analysis of the diffraction electron density map. Also, while HINT scoring was able to select for 1flr a pose extremely similar to the crystallographic conformation (see Fig. 5b), with RMSD = 0.55 Å and a $\Delta\Delta G_{HINT}$* of only 0.20 kcal mol^{-1}, AutoDock scoring identified an opposite conformation, where the carbonyl and hydroxyl groups on the tricyclic ring are inverted. $\Delta\Delta G_{dock}$ is relatively small for this pose, but RMSD = 4.30 Å and $\Delta\Delta G_{HINT}$ = 6.77 kcal mol^{-1}, which improved to 0.25 kcal mol^{-1}, after structure optimization (vide infra). These cases, where docking free energy predictions track so poorly with RMSD, would appear to be just aberrations, but alternate binding modes are possible with indistinct binding pockets, especially with flexible ligands at biological temperature. Since binding free energy is a weighted composite arising from multiple states, some of these alternate modes are likely contributing to the measured free energy of binding, and are not necessarily "wrong".

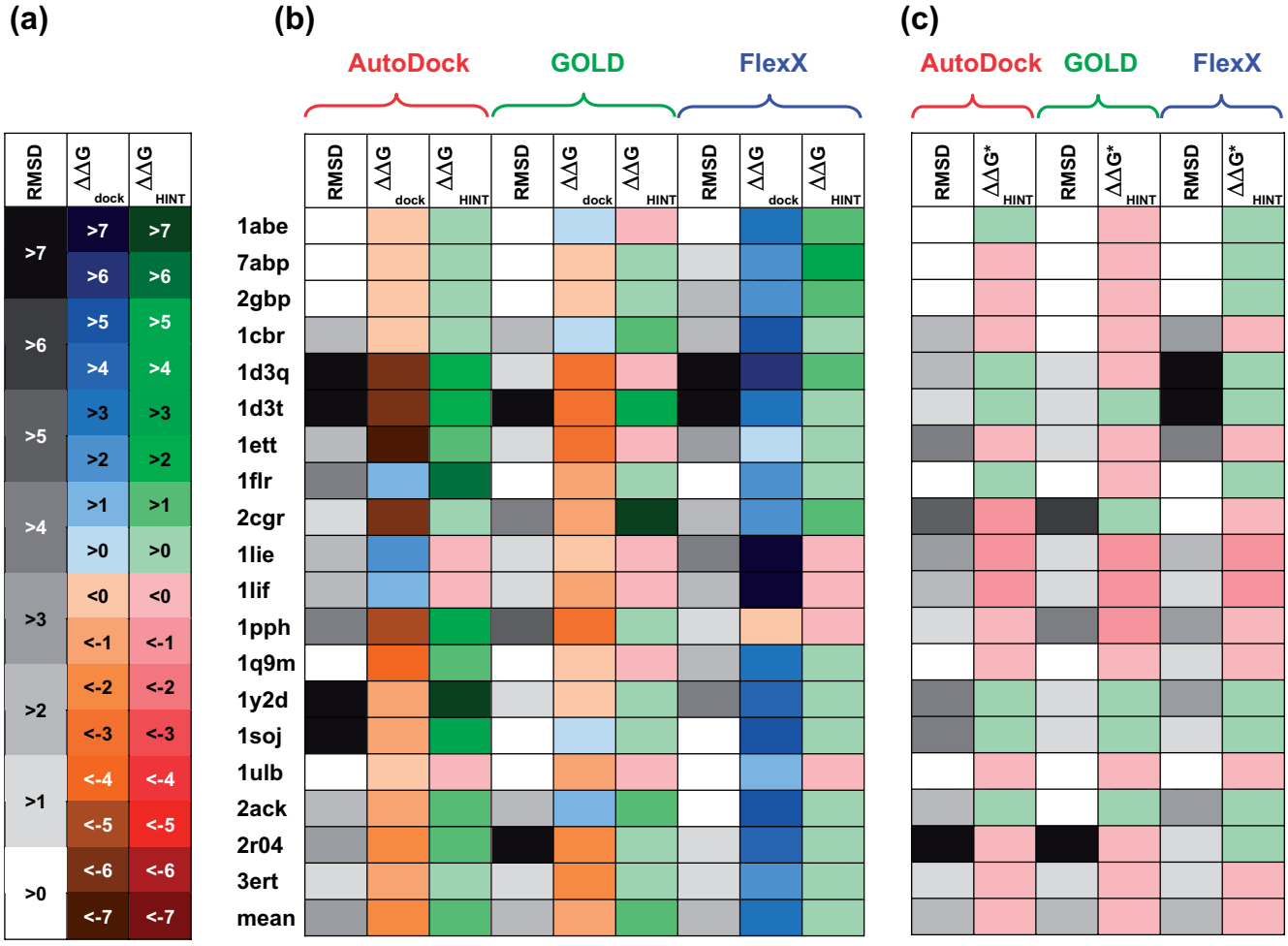


Fig. 4. Overall color–temperature scale representation of the docking results. (a) Temperature codes: first column, RMSD – root mean square deviation between docking pose and crystal conformation in Å; second column, $\Delta\Delta G_{dock}$ – difference between free energy calculated by docking program internal scoring function (and/or published calibration) and measured free energy of binding in kcal mol^{-1}; third column, $\Delta\Delta G_{HINT}$ – difference between free energy calculated by HINT (eq. 1) and measured free energy of binding in kcal mol^{-1}. (b) Results for best scoring conformers chosen by docking programs AutoDock, GOLD and FlexX. The column color scales are as described in (a). (c) Results for best scoring conformers chosen by HINT from poses generated by AutoDock, GOLD and FlexX. The column color scales are as described in (a).

In some cases, high RMSD is indicative of incorrect or inappropriate poses, as can be seen by comparing the top docking poses selected by AutoDock with those selected by HINT scoring of the AutoDock set for the 1pph (Fig. 5c) complex. High RMSD (4.59 Å), $\Delta\Delta G_{dock}$ (−5.22 kcal mol^{-1}) and $\Delta\Delta G_{HINT}$ (5.00 kcal mol^{-1}) values were found for the pose selected by AutoDock for 1pph. Here, the amidinic group occupies a region of the binding pocket that may not actually be sterically accessible to the ligand (see Fig. 5c surface contour), but which has significant functional group complementarity for the amidinic group. The pose selected by HINT has RMSD = 1.39 Å and $\Delta\Delta G_{HINT}{}^{*} = -0.70$ kcal mol^{-1}. This particular case may be due to a parameterization error in the AutoDock algorithm, and even though structure optimization of this local energy minimum decreased the HINT free energy error ($\Delta\Delta G_{HINT}$) to a quite acceptable 0.57 kcal mol^{-1}, the resulting model still may not be meaningful. This example illustrates that, even with the best practices, totally automated docking and virtual screening may yield misleading results.

Regardless of the docking program used, the HINT free energy correlation was able to identify, through highest HINT score, poses that had calculated free energies of binding within 1 kcal mol^{-1} of the experimental values for at least 16/19 of the complexes in the study. We suggest that this method, of basing pose selection on a score function calibrated to free energy of binding, is a superior way to evaluate docking poses as the much more significant application of docking is in virtual screening, where experimental ligand–protein complexes are unavailable and RMSD is incalculable and irrelevant. As noted above, the crystallographic conformer, which can be presumed to be at or near the lowest energy because it is obtained at low temperature, is probably not the *only* biologically relevant pose. While we used the HINT free energy force field to develop the specific energy prediction function in this work, the key point is probably that the specific calibration of score functions for the relevant targets in a docking study is preferable to using a generalized function.

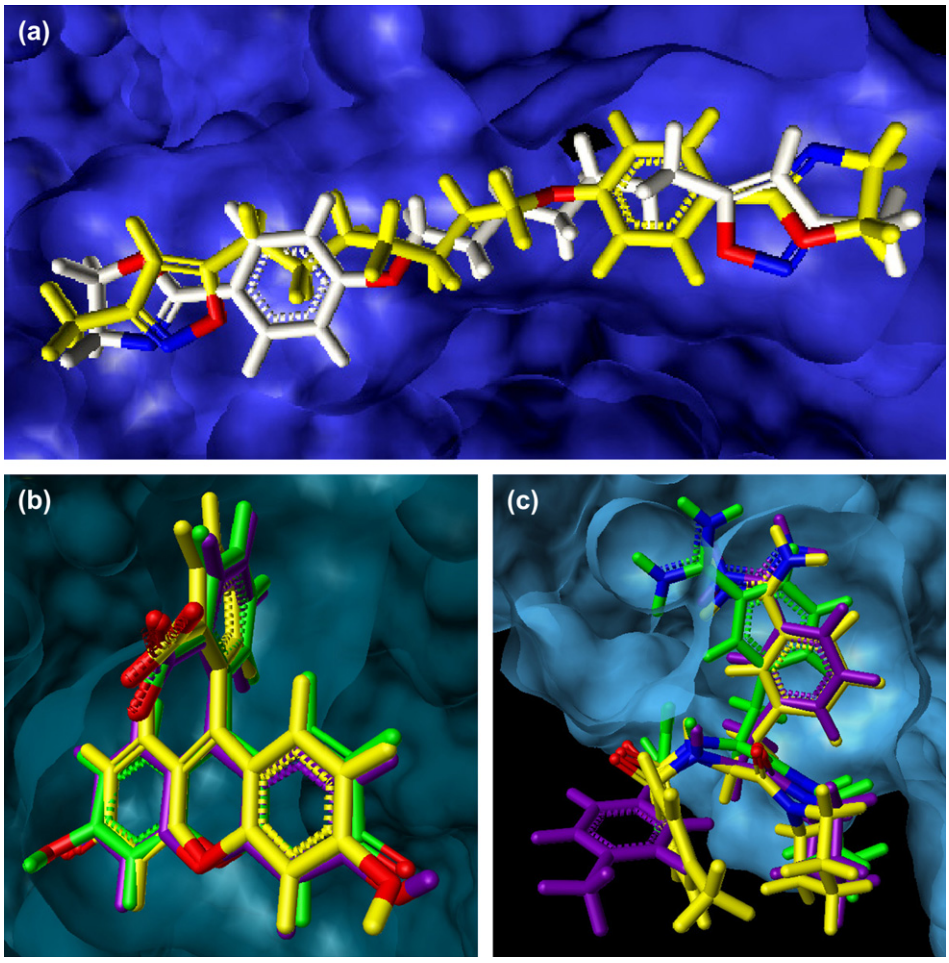


Fig. 5. Overlay of experimental and docked ligands conformations. Ligands are rendered in ball and stick, while the background surfaces represent the corresponding active site cavities. (a) Complex 2r04: the crystallographic ligand conformation and the best scoring GOLD/HINT pose are colored as yellow and white, respectively. Note the inverted conformation. (b) Complex 1flr: superimposition of the crystallographic conformation (yellow) and the structure-optimized top docking poses selected by AutoDock (green) and AutoDock/HINT (purple). Note that the carbonyl and hydroxyl groups on the tricyclic ring are reversed in the AutoDock pose. (c) Complex 1pph: superimposition of the crystallographic conformation (yellow) and the structure-optimized top docking poses selected by AutoDock (green) and AutoDock/HINT (purple). Note the placement of the amidinic group by AutoDock.

3.4. Structure optimization of docked conformations

We were interested to see if the top scoring docked poses, some of which (Fig. 4b) had quite poor calculated free energies of binding, could be ''rescued'' by molecular mechanics structure optimization within the binding pocket [42,57]. Thus, as described in Section 2 the best scoring poses as selected by the docking programs and by HINT were energy minimized. Fig. 6 displays the results of this experiment. In Fig. 6a $\Delta\Delta G_{\mathrm{HINT}}$ for the poses before (black) and after (white) the energy minimization are compared. Clearly about half of the poses are significantly improved by minimization to $|\Delta\Delta G_{\mathrm{HINT}}| \leq 1$ kcal mol^{-1}, but a few (especially among the AutoDock set) are completely unaffected indicating that they are completely wrong. Note, however, that these poses are near local energy minima because the RMSDs are largely unchanged after the minimization procedure; i.e., the main result of the minimization is to ameliorate bad steric contacts and related structural defects of the poses.

The more interesting observation is illustrated in Fig. 6b, where $\Delta\Delta G_{\mathrm{HINT}}$* before and after minimization is shown. Here, minimization had only a very small effect on $\Delta\Delta G_{\mathrm{HINT}}$*, as the starting energies were already good, not only with respect to the HINT force field, which is expected, but *also* with respect to the molecular mechanics force field, which was, at first, completely *unexpected.* In fact, the native docking scoring algorithms tend to de-emphasize van der Waals repulsions with softer functions to both increase volume sampling and partially account for protein side-chain flexibility. Thus, the docked poses selected with the calibrated HINT force field scoring function, which includes a ''hard'' 6-12 Lennard–Jones term, appear to be largely pre-optimized for the binding site and not in need of significant further optimization. By extension, it is not unreasonable to expect that docking other ligands into this set of proteins with any of these (or likely other) docking programs followed by scoring with HINT and eq. 1, would yield well-mannered conformers with reasonable estimates for their free energies of binding.

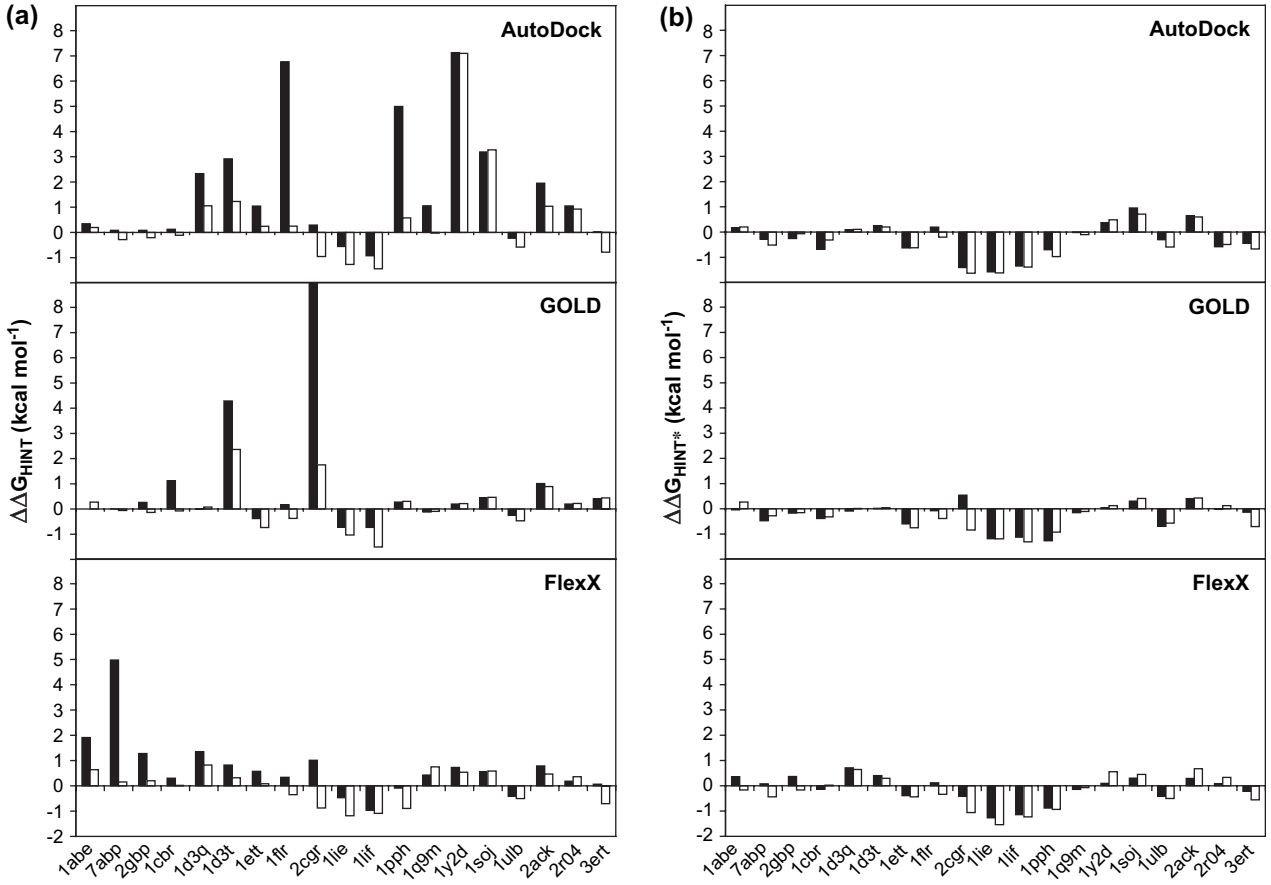


Fig. 6. Analysis of the effects of post-docking minimization. The bars show the difference between the HINT score-derived free energy assigned to the docking pose and the measured free energy of binding before (black) and after (white) energy minimization of the docking pose. Δ*G*s were calculated using eq. 1. (a) ΔΔ*G* for the top scoring docked poses generated and identified by AutoDock, GOLD and FlexX. (b) ΔΔ*G* for the top poses identified by HINT, amongst those generated by AutoDock, GOLD and FlexX.

4. Summary

This work is part of our long-range goal of designing and implementing a reliable and accurate system for scoring and understanding the results from virtual screening. Previous reports were an initial calibration of the HINT scoring function for the free energy of binding in protein–ligand complexes [24], an examination of the role of ionization state in calculating free energy of binding or what we termed ''computational titration'' [25], and a study of the energetic effects of bridging water in the active site of protein–ligand complexes [26]. All of this work is directed at our central hypothesis using a virtual screening score function that is calibrated in terms of free energy, and not just enthalpy, may be a key success factor for discovering new lead compounds computationally. However, the results of this present study indicate that a universal scoring function applicable to *all* biomacromolecular host–guest complexes may be an unrealistic goal. We have no doubts that the AutoDock, GOLD and FlexX scoring functions performed at least as well as reported within their training sets, but real world problems are actually more focused, and the effort to re-calibrate scoring functions for the data set(s) of interest is probably worth the investment of time. While we did not test this assertion, we believe that significantly improved scoring performance would be obtained with the AutoDock, GOLD and FlexX functions if they were re-calibrated for this data set. Perhaps the functional tools for this type of tuning should be included in docking programs. However, it is also apparent from this study that there are consequences in calibrating scoring functions towards minimizing RMSD with respect to known crystal structures rather than optimizing binding free energy. While identifying and scoring the pose identical or most similar to the crystal pose will usually yield a good estimate of binding free energy, the inverse is not as often the case, and some of these alternate poses would likely provide useful information for drug discovery studies. The HINT free energy scoring function, because it is based on pair-wise atom–atom interactions, is inherently easy to apply and calibrate for specific data sets by simple least squares regressions as illustrated by eq. 1.

Acknowledgements

This work was partially supported by funds from the Italian Ministry of Instruction, University and Research within a COFIN2005 and an Internationalization project (A.M.) and U.S.

NIH grant GM71894 (G.E.K.). We acknowledge Dr. Philip D. Mosier for critical and thoughtful suggestions that greatly improved this manuscript.

References

[1] T.P. Lybrand, Curr. Opin. Struct. Biol. 5 (1995) 224–228.
[2] W.P. Walters, M. Stahl, M.A. Murcko, Drug Discov. Today 3 (1998) 160–178.
[3] G.E. Kellogg, Med. Chem. Res. 9 (1999) 439–442.
[4] C. Perez, A.R. Ortiz, J. Med. Chem. 44 (2001) 3768–3785.
[5] C. Bissantz, G. Folkers, D. Rognan, J. Med. Chem. 43 (2000) 4759–4767.
[6] L. Xing, E. Hodgkin, Q. Liu, D. Sedlock, J. Comput. Aided Mol. Des. 18 (2004) 333–344.
[7] G.L. Warren, C.W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, J. Med. Chem 49 (2006) 5912–5931.
[8] R. Abagyan, M. Totrov, R. Kuznetsov, J. Comput. Chem. 15 (1994) 488–506.
[9] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[10] S. Makino, I.D. Kuntz, J. Comput. Chem. 18 (1997) 1812–1825.
[11] C. McMartin, R.S. Bohacek, J. Comput. Aided Mol. Des. 11 (1997) 333–344.
[12] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[13] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639–1662.
[14] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[15] P.S. Charifson, J.J. Corkery, M.A. Murcko, W.P. Walters, J. Med. Chem. 42 (1999) 5100–5109.
[16] R.D. Clark, A. Strizhev, J.M. Leonard, J.F. Blake, J.B. Matthew, J. Mol. Graph. Model. 20 (2002) 281–295.
[17] R. Wang, Y. Lu, S. Wang, J. Med. Chem. 46 (2003) 2287–2303.
[18] B.D. Bursulaya, M. Totrov, R. Abagyan, C.L. Brooks III, J. Comput. Aided Mol. Des. 17 (2003) 755–763.
[19] R. Wang, Y. Lu, X. Fang, S. Wang, J. Chem. Inf. Model. 44 (2004) 2114–2125.
[20] H. Chen, P.D. Lyne, F. Giordanetto, T. Lovell, J. Li, J. Chem. Inf. Model. 46 (2006) 401–415.
[21] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, Proteins 57 (2004) 225–242.
[22] D. Borek, W. Minor, Z. Otwinowski, Acta Crystallogr. D Biol. Crystallogr. 59 (2003) 2031–2038.
[23] A.M. Davis, S.J. Teague, G.J. Kleywegt, Angew. Chem. Int. Ed. Engl. 42 (2003) 2718–2736.
[24] P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, J. Med. Chem. 45 (2002) 2469–2483.
[25] M. Fornabaio, P. Cozzini, A. Mozzarelli, D.J. Abraham, G.E. Kellogg, J. Med. Chem. 46 (2003) 4487–4500.
[26] M. Fornabaio, F. Spyrakis, A. Mozzarelli, P. Cozzini, D.J. Abraham, G.E. Kellogg, J. Med. Chem. 47 (2004) 4507–4516.
[27] A. Amadasi, F. Spyrakis, P. Cozzini, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, J. Mol. Biol. 358 (2006) 289–309.
[28] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[29] G.E. Kellogg, G.S. Joshi, D.J. Abraham, Med. Chem. Res. 1 (1992) 444–453.
[30] Z.S. Derewenda, L. Lee, U. Derewenda, J. Mol. Biol. 252 (1995) 248–262.
[31] R. Vargas, J. Garza, D.A. Dixon, P.H. Benjamin, J. Am. Chem. Soc. 122 (2000) 4750–4755.
[32] M. Porotto, M. Fornabaio, O. Greengard, M.T. Murrell, G.E. Kellogg, A. Moscona, J. Virol. 80 (2006) 1204–1213.
[33] P.J. Goodford, J. Med. Chem. 28 (1985) 849–857.
[34] D.S. Goodsell, A.J. Olson, Proteins 8 (1990) 195–202.
[35] G.M. Morris, D.S. Goodsell, R. Huey, A.J. Olson, J. Comput. Aided Mol. Des. 10 (1996) 293–304.
[36] G. Jones, P. Willett, R.C. Glen, J. Mol. Biol. 245 (1995) 43–53.
[37] B. Kramer, M. Rarey, T. Lengauer, Proteins 37 (1999) 228–241.
[38] J. Wang, P.A. Kollman, I.D. Kuntz, Proteins 36 (1999) 1–19.
[39] S. Ha, R. Andreani, A. Robbins, I. Muegge, J. Comput. Aided Mol. Des. 14 (2000) 435–448.
[40] T. Schulz-Gasch, M. Stahl, J. Mol. Model. (Online) 9 (2003) 47–57.
[41] P. Ferrara, H. Gohlke, D.J. Price, G. Klebe, C.L. Brooks III, J. Med. Chem. 47 (2004) 3032–3047.
[42] E. Perola, W.P. Walters, P.S. Charifson, Proteins 56 (2004) 235–249.
[43] M. Kontoyianni, L.M. McClellan, G.S. Sokol, J. Med. Chem. 47 (2004) 558–565.
[44] R.T. Kroemer, A. Vulpetti, J.J. McDonald, D.C. Rohrer, J.Y. Trosset, F. Giordanetto, S. Cotesta, C. McMartin, M. Kihlen, P.F. Stouten, J. Chem. Inf. Comput. Sci. 44 (2004) 871–881.
[45] J.C. Cole, C.W. Murray, J.W. Nissink, R.D. Taylor, R. Taylor, Proteins 60 (2005) 325–332.
[46] M. Stahl, H.J. Böhm, J. Mol. Graph. Model. 16 (1998) 121–132.
[47] H. Gohlke, M. Hendlich, G. Klebe, J. Mol. Biol. 295 (2000) 337–356.
[48] M. Stahl, M. Rarey, J. Med. Chem. 44 (2001) 1035–1042.
[49] G.E. Terp, B.N. Johansen, I.T. Christensen, F.S. Jorgensen, J. Med. Chem. 44 (2001) 2333–2343.
[50] P.S. Charifson, W.P. Walters, J. Comput. Aided Mol. Des. 16 (2002) 311–323.
[51] G.E. Kellogg, D.J. Abraham, Eur. J. Med. Chem. 35 (2000) 651–661.
[52] J.C. Burnett, G.E. Kellogg, D.J. Abraham, Biochemistry 39 (2000) 1622–1633.
[53] J.C. Burnett, P. Botti, D.J. Abraham, G.E. Kellogg, Proteins 42 (2001) 355–377.
[54] P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, Curr. Med. Chem. 11 (2004) 3093–3118.
[55] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Proteins 52 (2003) 609–623.
[56] H.J. Böhm, J. Comput. Aided Mol. Des. 8 (1994) 243–256.
[57] D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey, T. Lengauer, J. Med. Chem. 42 (1999) 4422–4433.
[58] F.A. Quiocho, N.K. Vyas, Nature 310 (1984) 381–386.
[59] D.M. Miller III, J.S. Olson, J.W. Pflugrath, F.A. Quiocho, J. Biol. Chem. 258 (1983) 13665–13672.
[60] P.S. Vermersch, D.D. Lemon, J.J. Tesmer, F.A. Quiocho, Biochemistry 30 (1991) 6861–6866.
[61] N.K. Vyas, M.N. Vyas, F.A. Quiocho, Science 242 (1988) 1290–1295.
[62] R. Wang, L. Lai, S. Wang, J. Comput. Aided Mol. Des. 16 (2002) 11–26.
[63] G.J. Kleywegt, T. Bergfors, H. Senn, P. Le Motte, B. Gsell, K. Shudo, T.A. Jones, Structure 2 (1994) 1241–1258.
[64] G. Siegenthaler, I. Tomatis, D. Chatellard-Gruaz, S. Jaconi, U. Eriksson, J.H. Saurat, Biochem. J. 287 (Pt 2) (1992) 383–389.
[65] N.Y. Chirgadze, D.J. Sall, S.L. Briggs, D.K. Clawson, M. Zhang, G.F. Smith, R.W. Schevitz, Protein Sci. 9 (2000) 29–36.
[66] D.J. Sall, J.A. Bastian, S.L. Briggs, J.A. Buben, N.Y. Chirgadze, D.K. Clawson, M.L. Denney, D.D. Giera, D.S. Gifford-Moore, R.W. Harper, K.L. Hauser, V.J. Klimkowski, T.J. Kohn, H.S. Lin, J.R. McCowan, A.D. Palkowitz, G.F. Smith, K. Takeuchi, K.J. Thrasher, J.M. Tinsley, B.G. Utterback, S.C. Yan, M. Zhang, J. Med. Chem. 40 (1997) 3489–3493.
[67] H. Brandstetter, D. Turk, H.W. Hoeffken, D. Grosse, J. Sturzebecher, P.D. Martin, B.F. Edwards, W. Bode, J. Mol. Biol. 226 (1992) 1085–1099.
[68] J. Sturzebecher, P. Walsmann, B. Voigt, G. Wagner, Thromb. Res. 36 (1984) 457–465.

[69] M. Whitlow, A.J. Howard, J.F. Wood, E.W. Voss Jr., K.D. Hardman, Protein Eng. 8 (1995) 749–761.
[70] A.L. Gibson, J.N. Herron, X.M. He, V.A. Patrick, M.L. Mason, J.N. Lin, D.M. Kranz, E.W. Voss Jr., A.B. Edmundson, Proteins 3 (1988) 155–160.
[71] L.W. Guddat, L. Shan, J.M. Anchin, D.S. Linthicum, A.B. Edmundson, J. Mol. Biol. 236 (1994) 247–274.
[72] J.M. LaLonde, D.A. Bernlohr, L.J. Banaszak, Biochemistry 33 (1994) 4885–4895.
[73] G.V. Richieri, R.T. Ogata, A.M. Kleinfeld, J. Biol. Chem. 269 (1994) 23918–23930.
[74] Z. Xu, D.A. Bernlohr, L.J. Banaszak, J. Biol. Chem. 268 (1993) 7874–7884.
[75] D. Turk, J. Sturzebecher, W. Bode, FEBS Lett. 287 (1991) 133–138.
[76] Q. Huai, H. Wang, Y. Sun, H.Y. Kim, Y. Liu, H. Ke, Structure 11 (2003) 865–873.
[77] J.E. Souness, D. Aldous, C. Sargent, Immunopharmacology 47 (2000) 127–162.
[78] G.L. Card, L. Blasdel, B.P. England, C. Zhang, Y. Suzuki, S. Gillette, D. Fong, P.N. Ibrahim, D.R. Artis, G. Bollag, M.V. Milburn, S.H. Kim, J. Schlessinger, K.Y. Zhang, Nat. Biotechnol. 23 (2005) 201–207.
[79] G. Scapin, S.B. Patel, C. Chung, J.P. Varnerin, S.D. Edmondson, A. Mastracchio, E.R. Parmee, S.B. Singh, J.W. Becker, L.H. Van der Ploeg, M.R. Tota, Biochemistry 43 (2004) 6091–6100.
[80] S.E. Ealick, Y.S. Babu, C.E. Bugg, M.D. Erion, W.C. Guida, J.A. Montgomery, J.A. Secrist III, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 11540–11544.
[81] R.B. Ravelli, M.L. Raves, Z. Ren, D. Bourgeois, M. Roth, J. Kroon, I. Silman, J.L. Sussman, Acta Crystallogr. D Biol. Crystallogr. 54 (1998) 1359–1366.
[82] Z. Radic, E. Reiner, P. Taylor, Mol. Pharmacol. 39 (1991) 98–104.
[83] J. Badger, I. Minor, M.A. Oliveira, T.J. Smith, M.G. Rossmann, Proteins 6 (1989) 1–19.
[84] J. Badger, I. Minor, M.J. Kremer, M.A. Oliveira, T.J. Smith, J.P. Griffith, D.M. Guerin, S. Krishnaswamy, M. Luo, M.G. Rossmann, et al., Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 3304–3308.
[85] A.K. Shiau, D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard, G.L. Greene, Cell 95 (1998) 927–937.
[86] G. Amari, E. Armani, S. Ghirardi, M. Delcanale, M. Civelli, P.L. Caruso, E. Galbiati, M. Lipreri, S. Rivara, A. Lodola, M. Mor, Bioorg. Med. Chem. 12 (2004) 3763–3782.