J Comput Aided Mol Des (2010) 24:91–105
DOI 10.1007/s10822-009-9317-9

Docking flexible ligands in proteins with a solvent exposure- and distance-dependent dielectric function

Daniel P. Garden · Boris S. Zhorov

Received: 29 October 2009 / Accepted: 26 December 2009 / Published online: 30 January 2010

Abstract Physics-based force fields for ligand–protein docking usually determine electrostatic energy with distance-dependent dielectric (DDD) functions, which do not fully account for the dielectric permittivity variance between ~ 2 in the protein core and ~ 80 in bulk water. Here we propose an atom–atom solvent exposure- and distance-dependent dielectric (SEDDD) function, which accounts for both electrostatic and dehydration energy components. Docking was performed using the ZMM program, the AMBER force field, and precomputed libraries of ligand conformers. At the seeding stage, hundreds of thousands of positions and orientations of conformers from the libraries were sampled within the rigid protein. At the refinement stage, the ten lowest-energy structures from the seeding stage were Monte Carlo-minimized with the flexible ligand and flexible protein. A search was considered a success if the root mean square deviation (RMSD) of the ligand atoms in the apparent global minimum from the x-ray structure was <2 Å. Calculations on an examining set of 60 ligand–protein complexes with different DDD functions and solvent-exclusion energy revealed outliers in most of which the ligand-binding site was located at the protein surface. Using a training set of 16 ligand–protein complexes, which did not overlap with the examining set, we parameterized the SEDDD function to minimize the RMSD of the apparent global minima from the x-ray structures. Recalculation of the examining set with the SEDDD function demonstrated a 20% increase in the success rate versus the best-performing DDD function.

D. P. Garden · B. S. Zhorov (✉)
Department of Biochemistry and Biomedical Sciences,
McMaster University, Hamilton, ON, Canada
e-mail: zhorov@mcmaster.ca

Keywords Monte Carlo-minimization · Generalized coordinates · Force fields · Electrostatic interactions · ZMM program

Abbreviations

AGM	Apparent global minimum
RMSD	Root mean square deviation
PDB	Protein databank
MC	Monte Carlo
MCM	Monte Carlo-minimization
DDD	Distance-dependent dielectric
SEDDD	Solvent exposure- and distance-dependent dielectric

Introduction

In the past decade, growth of high-resolution structures in the protein databank [1], increase of computational power, and new software have boosted applications of in-silico docking of ligands into proteins of known 3D structure and their homology models. Among the aims of these studies is the discovery of new drugs, understanding atomic mechanisms of ligand-receptor interactions, designing new experiments, and improvement of ligand-docking methodology. In-silico docking can predict ligand-binding poses, rank ligands by their interaction energy with the protein, and ideally, predict the ligand affinity. According to a recent assessment, ten popular ligand-docking programs were able to generate ligand-binding poses similar to the crystallographic complexes for some targets, however, were less successful at distinguishing the x-ray structure from the set of docked poses, and were unable to predict

ligand affinities [2]. Most drug candidates are still discovered via experimental high-throughput screening methods, but computational docking is becoming a major source of lead molecules in drug discovery [3].

Scoring functions used in ligand-docking programs can be categorized as knowledge-based [4] and physics-based [5]. Programs GLIDE [6], GOLD [7], and FlexX [8] employ knowledge-based scoring functions developed with training sets of high-resolution structures and search for ligand–protein complexes with the optimal score. Programs AutoDock [9], ICM [10], RosettaLigand [11], and ZMM [12], use physics-based force fields that present the energy as the sum of van der Waals, electrostatic, and solvation components as well as the valence-geometry strain energy of the ligand and protein. The latter three programs search for the apparent global minimum (AGM) and local minima in the space of generalized coordinates using the Monte Carlo-minimization (MCM) method [13]. The local minima may be a few kcal/mol or less from the AGM. A challenge is to tune the force field to ensure that the root mean square deviation (RMSD) of the AGM from the x-ray structure is below 2 Å. Imprecise calculation of electrostatic energy [5] seems to be a major obstacle in meeting this challenge.

The most advanced approach to calculate electrostatic energy treats the protein interior as a low-dielectric medium, the solvent as the high-dielectric medium and numerically solves the Poisson-Boltzmann equation [14–17]. The high computational cost of this method currently prevents its application for high-throughput ligand docking. Besides, available implementations of the method involve inaccuracies that currently are difficult to eliminate [5]. The Generalized Born model [18], which is an approximation of the Poisson-Boltzmann equation, is used e.g. in the CHARMM program for molecular dynamic simulations [19].

A traditional approach to calculate electrostatic energy is based on Coulomb's law, but uncertain dielectric permittivity, unknown location of counterions, and unknown protonation of titrable groups may have a big impact on the electrostatic energy. As an empirical solution, titrable residues can be considered neutral [20, 21] and the linear distance-dependent dielectric (DDD) function $\varepsilon = kr$ is used [22, 23]. In most programs, coefficient k equals 1, 2, or 4 [6–8, 10, 11, 24–26]. However, it is well known that the dielectric permittivity varies from to ~ 2 in the hydrophobic interior of a protein to ~ 80 in bulk water around the protein [4, 27]. Dielectric permittivity at the surface is an order of magnitude smaller than that of the bulk water [28–30] and the permittivity inside the protein may be greater than 2. Besides the linear DDD functions, sigmoidal-shaped DDD functions are used to screen electrostatic interactions at solvent-exposed areas [31–34]. The DDD functions per se do not completely account for large variations of the dielectric permittivity and usually overestimate electrostatic interactions at the protein surface. Computationally efficient solutions for this problem are solvent-exclusion models, which account for solvent screening of electrostatic interactions as well as for hydrophobic interactions between nonpolar groups inside the protein [21, 35].

Here we propose a solvent exposure- and distance-dependent dielectric (SEDDD) function that depends on the distance between a pair of atoms, the degree of their exposure to the aqueous environment, and parameters ε_0 and ε_1 ($\varepsilon_0 \leq \varepsilon_1$). For a water-exposed pair of atoms, a high-range dielectric function $\varepsilon_{ij} = r_{ij}\varepsilon_1$ is used. For a pair of atoms buried inside the protein, a low-range dielectric function $\varepsilon_{ij} = r_{ij}\varepsilon_0$ is used. For a pair of partially exposed atoms, the dielectric function varies between $r_{ij}\varepsilon_0$ and $r_{ij}\varepsilon_1$. Since theoretical derivation of parameters ε_0 and ε_1 is hardly possible, we used an empirical approach. We selected a training set of 16 ligand–protein complexes for which high-resolution x-ray structures are available and systematically varied ε_0 and ε_1 from 1 to 10 with the step of 1. For each combination of ε_0 and ε_1, ligand–protein energy for 500,000 poses of the rigid ligand in the rigid protein was calculated (the seeding stage). The ten lowest-energy poses were refined by MCM with the flexible ligand and flexible protein, and RMSD of the AGM was plotted against ε_0 and ε_1. Considering the ligand flexibility at the seeding stage would have increased the already large computational cost of parameterization by 100 fold. We found that for most of the complexes, the combination $\varepsilon_0 = 2$ and $\varepsilon_1 = 8$ minimizes the RMSD of the AGM from the x-ray structure. Calculations with an examining set of 60 high-resolution ligand–protein complexes were performed with considering the ligands flexibility at the seeding stage. This was achieved by using precomputed libraries of ligand conformers. For each ligand, a library was generated by randomly sampling the ligand torsions, minimizing the repulsive energy, and merging similar conformers. Flexible docking with the SEDDD function demonstrated a 20% increase of the success rate versus the best-performing DDD function.

Methods

Energy components

We use the ZMM program, which is partially described in various applications [36–39]. The website www.zmmsoft.com provides a detailed description of the program, controlling parameters, and tests. Briefly, ZMM minimizes the energy in the space of generalized coordinates [12, 40]

and employs the MCM method [13] for a non-local search of energetically optimal structures. In this study, the van der Waals energy was calculated using the AMBER force field [23, 41]. Standard ZMM computations were performed with the DDD function and the dehydration (solvent-exclusion) energy, which is based on the Gaussian solvent-exclusion model [21]. Ionizable residues were considered in both their neutral states [20, 21] as well as in ionized states. Electrostatic interactions involving metal ions and/or ionized groups of the ligands were calculated at all distances. Other interactions were truncated at a cutoff distance of 10 Å with a 1 Å switching function [42]. The ligands were imported from the PDB files using the ZMM module, which automatically assigns hybridizations of heavy atoms, adds hydrogens, and corrects bond lengths and bond angles that deviate significantly from the standard values. The module then submits ligand coordinates to the MOPAC program and imports atomic charges calculated by the AM1 method [43]. The ZMM-generated chemical structures of the ligands were compared with those at the PDBSum website [44]. Upon importing the x-ray structure of a protein to ZMM, tautomers of histidines and starting orientations of OH and $CONH_2$ groups in amino acids were chosen to minimize steric clashes and maximize H–bonds. The electrostatic energy was calculated with either the DDD or SEDDD function described in "SEDDD function".

Training and examining sets

We optimized our docking protocol (see below) with the DDD function using the first training set of 10 structures (Table 1) and parameterized the SEDDD function using the second training set of the following 16 structures: 1bk0, 1cbx, 1cc8, 1cmd, 1ctq, 1dmp, 1dry, 1ds1, 1g5a, 1g67, 1hb2, 1ky3, 1yds, 2gbp, 2ypi, and 8atc. The SEDDD function accounts for variation in the dielectric permittivity through screening the electrostatic interactions for water-exposed charges and strengthening interactions in buried regions of the protein and ligand-binding pockets. The function is described in "SEDDD function".

The DDD and SEDDD functions were tested with the examining set of 60 high-resolution structures (Fig. 1) chosen to represent diverse ligand-receptor complexes. Fifty of these complexes were used earlier to test ICM, GLIDE, and RosettaLigand [6, 11, 45]. The examining set contains charged and neutral ligands of varying size, shape and number of rotatable bonds (Fig. 2). Structures with resolutions >2.5 Å and complexes involving multi-residue ligands were discarded. Structures with metal ions other than K^+, Na^+, Ca^{2+}, and Zn^{2+} were not considered. Not a single x-ray structure used in this work belongs to more than one set (training or examining).

Libraries of ligand conformers

ZMM docking with flexible ligands and flexible proteins has reproduced various x-ray structures [36, 46], but at a large computational cost. In this study, ligand flexibility was accounted for by generating libraries of ligand conformers (Fig. 3), rigid docking hundreds of thousands of binding poses for each conformer in the protein, and refining low-energy poses by MC minimizing flexible conformers in flexible proteins. First, we sampled a thousand random starting conformers for each ligand and minimized the ligand energy in vacuum. Due to nonbonded attractions, flexible ligands usually adopted compact low-energy conformations that largely differ from the extended conformations seen in the x-ray structures of the ligand–protein complexes. To resolve the problem, we have used an ad hoc force field, which is referred henceforth as AMBERL, to increase the chances of accepting extended conformations of flexible ligands into the libraries. In AMBERL, parameters for nonbonded interactions,

Table 1 Training set of ligand–protein complexes

PDB code	Ligand			Ligand binding site		
	Atoms	Torsions	Heteroatoms	Residues[a]	Ionizable residues[a]	Depth[b]
1b6n	36	17	7	21	6	1.4
1bl7	25	4	7	14	3	1.7
1byg	35	2	7	16	3	1.3
1c2t	35	12	11	18	1	1.3
1dy9	35	19	14	13	5	2.9
1elc	36	15	10	8	3	4.5
1ett	30	8	8	8	4	2.5
1ivc	19	3	6	10	7	1.8
1srh	22	6	7	17	1	2.2
1tnh	12	1	2	13	1	3.0

[a] Within 4 Å from the ligand

[b] Determined as the ratio of the number of flexible residues in the protein double-shell model to the number of the ligand's heavy atoms

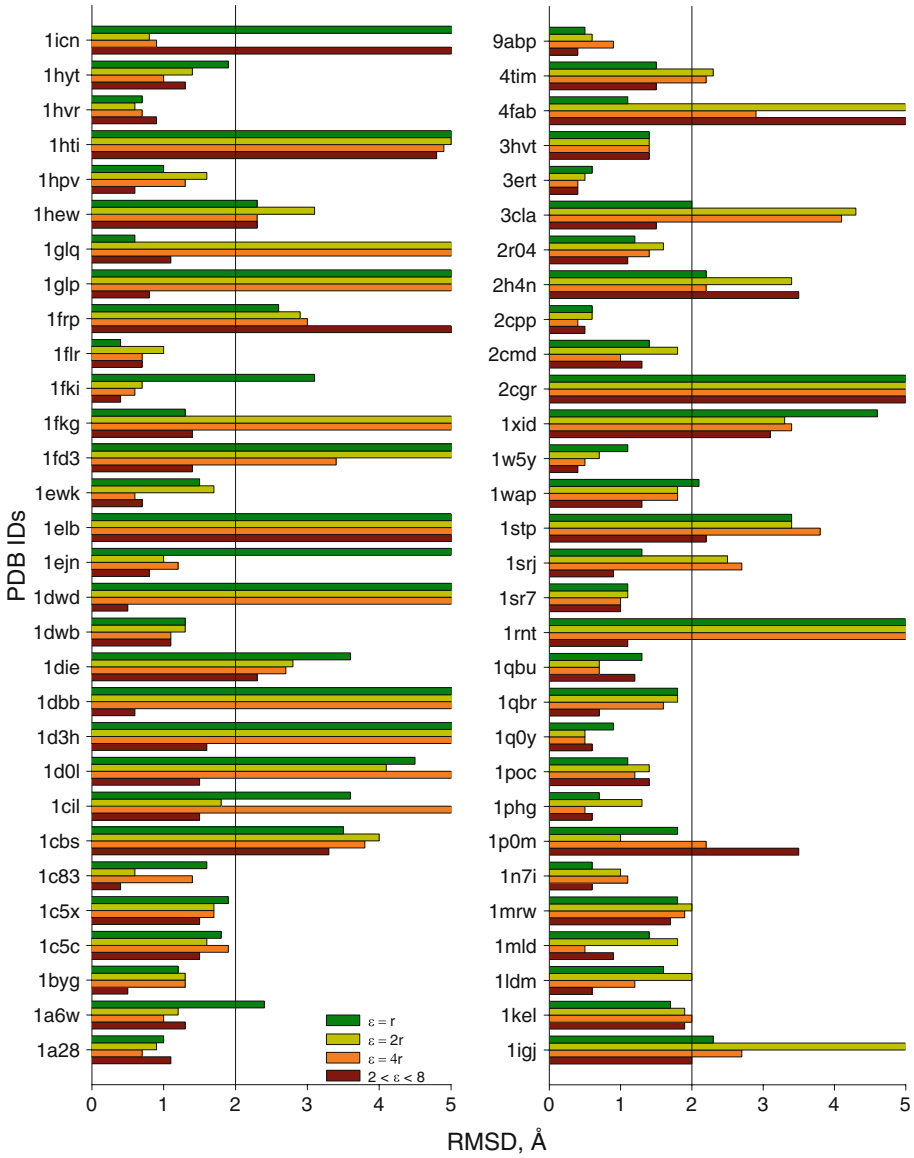

Fig. 1 RMSD of the AGM of the examining-set structures from the corresponding x-ray structures computed with the solvent-exclusion energy and DDD functions $\varepsilon = r$ (*green*), $\varepsilon = 2r$ (*yellow*), and $\varepsilon = 4r$ (*orange*), as well as the SEDDD function without solvent-exclusion energy (*brown*). Flexible-ligand dockings have been performed with the enriched libraries of ligand conformers. The black vertical line shows the 2.0 Å RMSD cutoff. *Bars* crossing the line correspond to false-positive predictions. The best-performing protocol with the DDD function ($\varepsilon = 2r$) and solvent-exclusion energy yields 24 (40.0%) false-positives, while the protocol with the SEDDD function and without solvent yields only 13 (21.7%) false-positives

torsional energy, and bond-angle deformation energy are the same as in AMBER [41], except for the nonbonded-potential well depth ε, which is assigned a value of 0.001 for all atom types. Electrostatic interactions were not included in AMBERL. Most ligands were calculated in vacuum with the exception of highly lipophilic flexible ligands (PDB codes 1icn and 1qbu) that were calculated using the implicit-octanol method [47] and highly hydrophilic ligands (1glp and 1n7i) that were calculated in water. Each unbiased library of ligand conformers was built by sampling a thousand random starting points and minimizing the AMBERL energy during 100 steps. Conformers with energy <10 kcal/mol from the AGM were accumulated in a stack (see below) whose size was limited to 100 conformers. Among the 60 examining-set ligands, for 56 ligands at least one conformer in the stack had RMSD <2 Å from the x-ray conformation in the respective ligand–protein complex and for four ligands the RMSDs was >2 Å (Fig. 4). The latter are highly flexible molecules with >9 "essential" torsions whose variations result in big conformational changes. (Non-essential torsions specify rotations of terminal groups such as $-CH_3$, $-NH_3$, –OH, and –Ph, as well as torsions in 3- to 6-membered rings.) All libraries included conformers with RMSDs up to 7 Å, which served as decoys for flexible ligand docking. In some calculations we used enriched libraries, which were created from the unbiased libraries by replacing the highest-energy conformer with the conformer from the x-ray structure of ligand–protein complex. To tune the docking protocols and parameterize the SEDDD function, we used single-conformer "libraries", which contained just the ligand conformers from the x-ray structures.

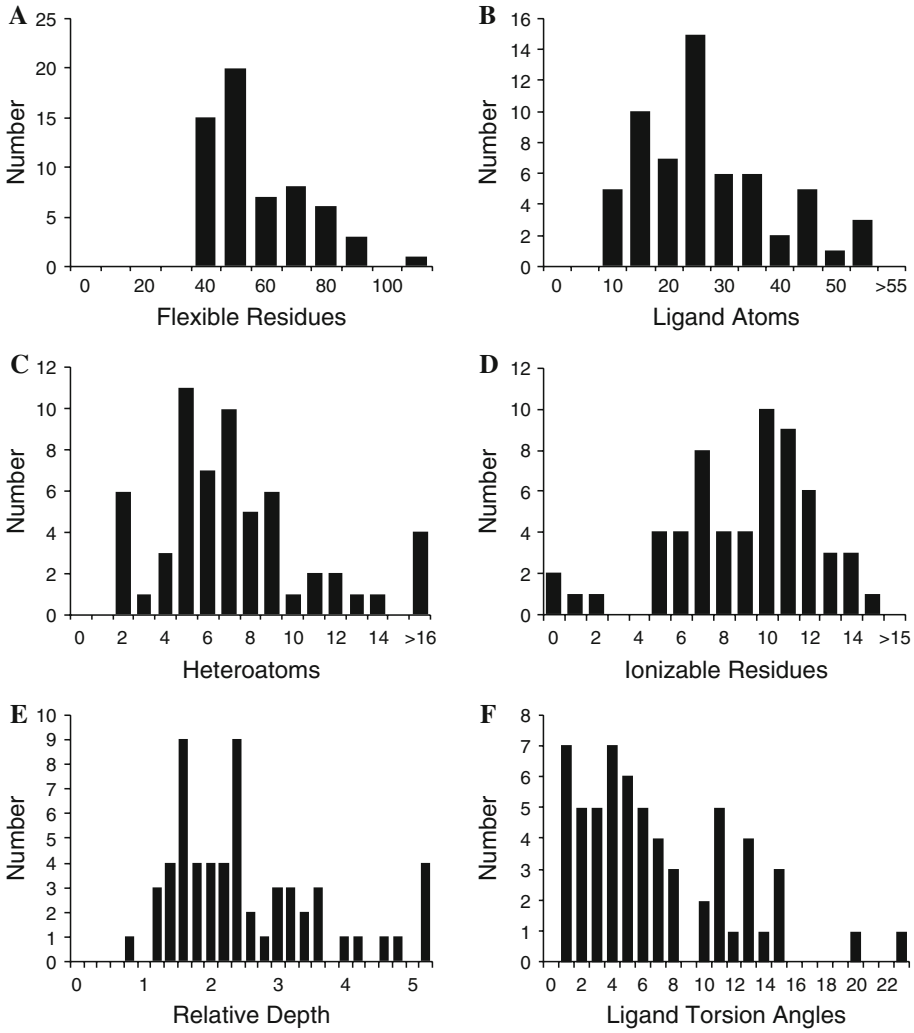

Fig. 2 Characteristics of the examining-set complexes. **a** The number of flexible residues in the double shells whose torsions have been sampled at the refinement stage. **b** The number of heavy atoms in the ligands. **c** The number of heteroatoms in the ligands. **d** The number of ionizable residues in the flexible shells of the proteins. **e** The relative depth of the ligand-binding pockets calculates as the ratio of the number of flexible residues in the double shell to the number of heavy atoms in the ligands. **f** The number of torsion angles in the ligands

Docking protocol

To reduce computational cost, ligands were docked in double-shell models [46]. A double shell contains flexible residues having at least one atom within 8.0 Å from the ligand as seen in the x-ray structure and fixed residues that do not belong to the flexible shell and contain at least one atom within 16.0 Å from the ligand. A protein can adopt different conformations in the presence of different ligands [48], but high throughput docking with flexible ligand and flexible protein remains a challenging problem. Here we employed a two-stage docking protocol. At the first, the seeding stage, each structure from a library of ligand conformers was placed at 500,000 random positions and orientations within the rigid protein (Fig. 5). At the second, the refinement stage, both the ligand and protein were flexible.

The sampling space for the ligand center was a cube with 8 Å edges approximately matching the flexible-shell size. (Other programs use comparable dimensions of the seeding box. In ICM, the ligand mass center is displaced randomly within a sphere of 3 Å radius [49]. GLIDE [6] and RosettaLigand [11] place the ligand center in a cube with edge of 12 and 10 Å, respectively.) Low-energy structures collected at the seeding stage were MC-minimized at the refinement stage, in which position/orientation of the ligand and torsion angles of the ligand and protein were sampled. After each sampling, the energy was minimized in the space of above generalized coordinates as well as bond angles of the ligand. To prevent large deformations of the protein backbones at the refinement stage, α-carbon atoms were constrained (pinned) to their crystallographic position with allowing a penalty-free deviation of up to 1.0 Å from the x-ray position. This two-stage protocol decreased computational cost by rejecting numerous ligand-binding poses that overlapped with the protein at the seeding stage. Parameterization of the SEDDD function was performed with neutral titrable residues; examining-set

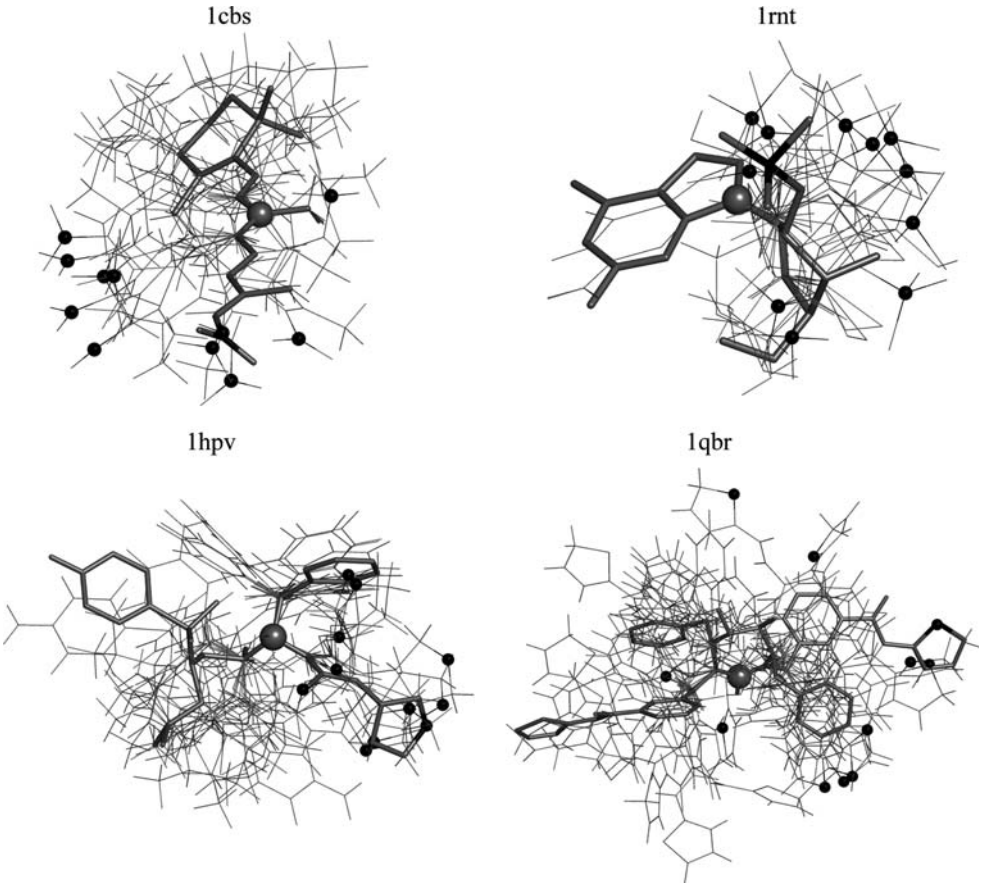

Fig. 3 Four libraries of ligand conformers. The 10 lowest-energy conformers in each library are shown by thin lines. The x-ray conformation is represented by *thick lines*. The atom in the origin of the ligand local system of coordinates is shown as a large *gray sphere*, and a reference atom is marked by a *dot*

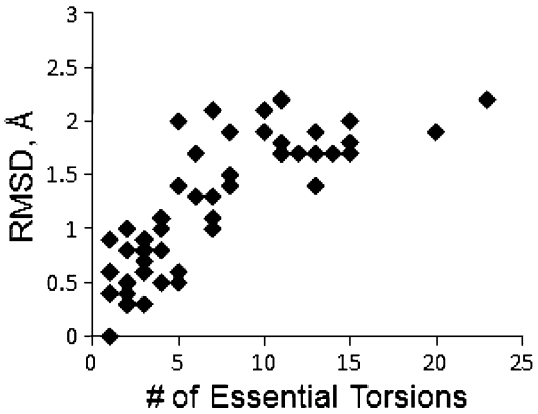

Fig. 4 In the libraries of ligand conformers, RMSD of the best match to the x-ray conformation increases with the number of essential torsions

calculations were performed twice: with all the titrable residues being either neutral or ionized.

Stack control

Multiple predicted 3D structures of a system were collected in a stack and ordered by increasing energy. Each structure in a stack was represented by a record that included its

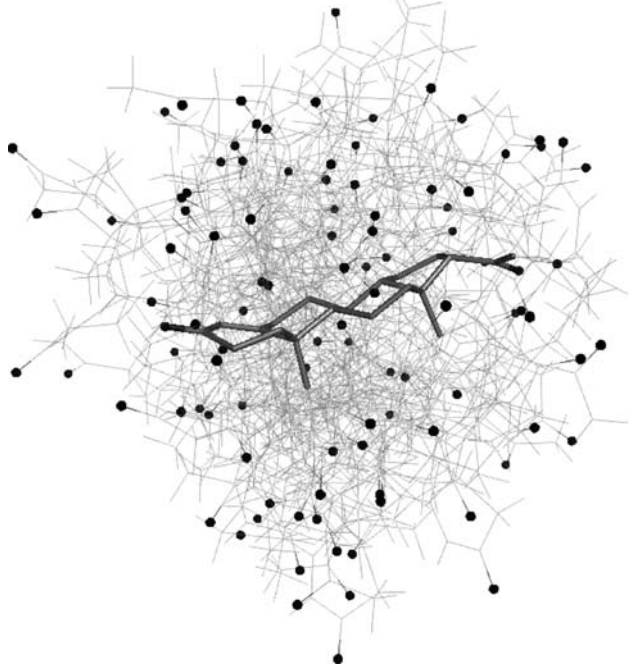

Fig. 5 Sampling progesterone binding poses. The first 100 of 500,000 progesterone poses generated at the seeding stage are shown by *thin lines* with black oxygens. *Thick lines* show the protein-bound progesterone in the x-ray structure (1a28)

energy and generalized coordinates. During sampling a stack was updated as follows. Initially, a large positive value was assigned to the AGM energy E_g. If the energy E_i of a newly generated structure was $<E_g$, the new structure was placed at the top of the stack, E_g was assigned the E_i value, and all structures with $E > E_g + \Delta E$ were removed from the stack. Here ΔE is the threshold to keep a structure in the stack. We used $\Delta E = 10$ kcal/mol to generate libraries of ligand conformers. For docking calculations, we used $\Delta E = 5{,}000$ kcal/mol at the seeding stage and $\Delta E = 100$ kcal/mol at the refinement stage. The high thresholds were necessary to collect at least ten structures at both the seeding and refinement stages. If for a newly-generated structure $E_i < E_g + \Delta E$, the structure was compared with all other structures in the stack. When generating a library of ligand conformers, two conformers were considered similar if none of their matching essential torsions differed by >1 radian. When docking a ligand in a protein, two ligand-binding poses were considered similar if the RMSD of all the matching atoms were <1 Å. If the new structure was dissimilar from any structure in the stack, it was added to the stack. Otherwise, the new structure and its closest match in the stack were compared by energy. If the energy of the new structure was lower than that of the match, the former replaced the latter in the stack. Otherwise the new structure was rejected. Thus, a record in a ligand-conformers stack represented a family of conformers. A record in the seeding-stage stack represented a family of binding poses with similar positions and orientations of the ligand. A record in the refinement-stage stack represented a family of complexes with similar positions, orientations, and conformations of the ligand and similar conformations of side chains in the protein flexible shell.

Success rate estimation

The RMS deviations of the ligand binding poses from the x-ray structure were calculated by comparing Cartesian coordinates of heavy atoms in the predicted and experimental structures. No attempts were made to decrease the RMS deviations by superimposing the structures by the least-squares method. An approximate match to the x-ray structure almost always can be found among the many thousands of ligand poses seeded in the protein. A challenge is to ensure that the refinement stage yields the AGM, which has an RMSD from the x-ray structure below 2.0 Å, the standard success criterion in literature [6, 11, 50, 51]. In some flexible-docking studies the success is measured by calculating RMSD values over both ligand and binding-site atoms. We calculated RMSD only for heavy atoms of the ligand.

Optimizing the docking protocol

Various ligand-docking programs use a two-stage protocol in which many structures are generated at the seeding stage and low-energy structures are refined in the second stage [6, 26, 49]. Computational cost of the global minimization depends, in particular, on the number of starting structures in the seeding stage (N_s), length of the MCM trajectory optimizing each seed (L_s), number of energy-minimizing iterations in each MCM step (N_{im}), number of low-energy seeds submitted for the refinement (N_r), and length of the MCM trajectory refining each seed (L_r). Systematically exploring how the success rate depends on these parameters is hardly possible. Therefore we used a heuristic approach to optimize the docking protocol with the DDD function and the solvent-exclusion energy in the first training set of 10 ligand–protein complexes (Table 1). Earlier $N_{im} = 200$ was found optimal (Zhorov, *unpublished*). To explore how the success rate depends on L_s, we generated 50,000 seeds for each complex and MC-minimized each seed. Brief MC-minimizations decreased the success rate (Fig. 6a) by yielding false-positives. Since optimizing all the seeds in long MCM trajectories is computationally prohibitive, we further ranked the seeds basing on the starting energy. Next we MC-minimized the 100 lowest-energy structures collected for each complex at the seeding stage. The success rate of these calculations increased up to 200 MCM steps then has a plateau at 40% (Fig. 6b). Further increasing the number of MCM steps only increased the computational time without increasing the success rate. Therefore we used $L_r = 200$ for docking experiments.

Next we explored how the success rate of the two-stage protocol depends on the number of low-energy seeds submitted for the refinement and found $N_r = 10$ to be optimal (Fig. 6c). The fact that $N_r > 10$ decreased the success rate indicates that high-energy seeds deviated significantly from the x-ray structures and their MC-minimizations created false-positives. Finally, we explored how the success rate of the two-stage protocol depends on the number of seeds N_s. The success rate fluctuated with $N_s < 50{,}000$, indicating an incomplete sampling and reached a plateau of 60% with $N_s \sim 500{,}000$ (Fig. 6d). The above parameters were used in further calculations.

For a protein–ligand complex, building a library of 100 ligand conformers, generating 500,000 seeds per conformer, determining energies of 50,000,000 seeds and clustering the low-energy seeds requires ~5 h on a 2.2 GHz CPU (AMD Opteron). Refining the top 10 seeds involves a total of 1,000 energy minimizations and takes ~2 h on the same-type CPU. All calculations were performed using a parallel version of ZMM.

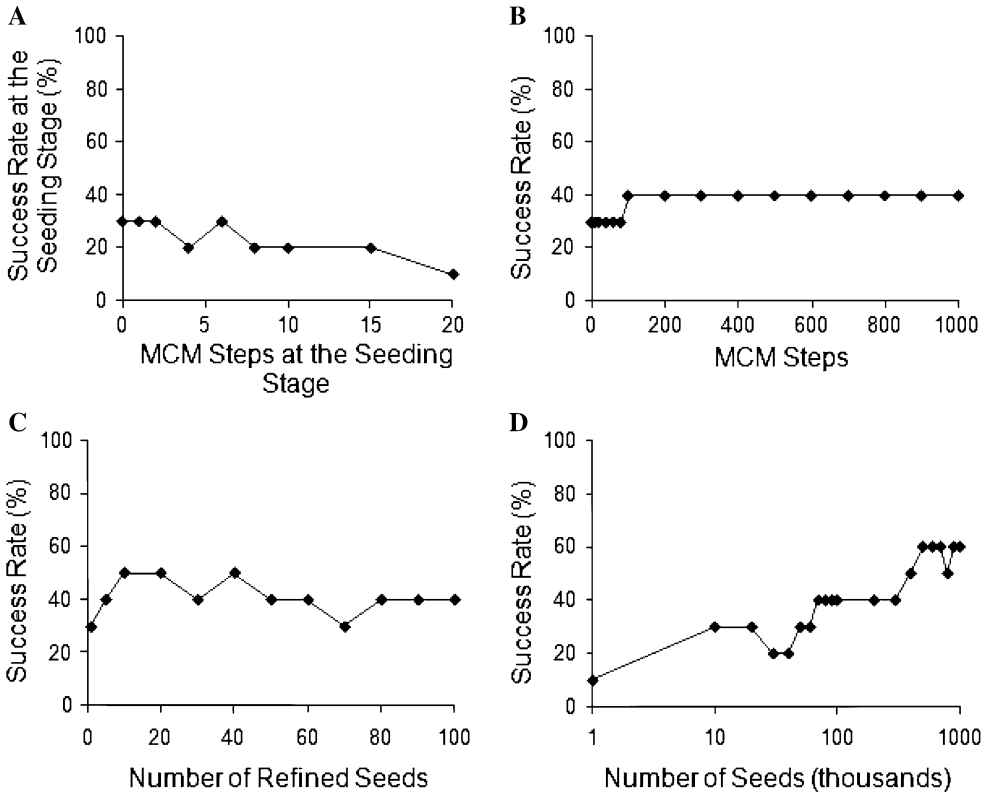

Fig. 6 Optimizing the docking protocol using a training set of ten complexes (Table 1) with the DDD function. **a** Success rate of correct predictions versus length of the seeding-stage MCM trajectories optimizing 50,000 sampled poses of each ligand. The rate decreases with the trajectory length. See text for further explanation. **b** Success rate versus length of the refinement-stage MCM trajectories optimizing 100 lowest-energy poses found at the seeding stage without energy minimizations. The rate plateaus to 40% at 100–200 MCM steps. **c** Success rate of the two-stage docking versus the number of structures taken from the seeding stage to the refinement stage. Note that a better success rate is obtained with small number of seeds. **d** Success rate of the two-stage protocol versus number of sampled seeds

Results

Examining set with the DDD function and the solvent-exclusion energy

Using the enriched libraries of ligand conformers and the DDD function $\varepsilon = kr$ with the solvent-exclusion energy [21] we obtained success rates of 60.0, 60.0 and 58.3, respectively, for $k = 1$, 2, and 4 (Table 2). Most false-positives in these experiments corresponded to ligand-binding sites at the protein surface (Fig. 7). Comparison of the energies of the false-positives with the MC-minimized energies of the corresponding x-ray structures demonstrated large differences in both the desolvation and electrostatic components. The solvent-exclusion energy benefits hydrophobic ligand–protein interactions, which usually occur in deep ligand-binding pockets. Since ligand–protein contacts in such pockets are rather tight, they are accounted for by van der Waals interactions. The solvent-exclusion energy also benefits location of the ligand hydrophilic groups at the protein surface where the dielectric permittivity is high and electrostatic interactions are weak. The DDD function does not completely account for this effect. For example, when calculated with the DDD function $\varepsilon = 2r$, the electrostatic energy $E_{ij} = 332q_iq_j/\varepsilon r$ between atoms i and j bearing equal charges q_i and q_j of 0.5 proton charge units and separated by 6 Å equals 1.15 kcal/mol. This value is an overestimate at the protein surface where electrostatic interactions are very week [27].

SEDDD function

We attempted to resolve the above problem by using a new dielectric function

$$\varepsilon_{ij} = r_{ij} \cdot [\varepsilon_0 + (1 - s_{kl}) \cdot (\varepsilon_1 - \varepsilon_0)], \quad (1)$$

where r_{ij} is the distance between interacting atoms i and j, s_{kl} is the overlap of volumes of solute groups k and l that include atoms i and j, respectively, with the protein and/or ligand, and ε_0 and ε_1 ($\varepsilon_1 > \varepsilon_0$) are parameters. For a buried pair of solute groups that overlap with the protein and/or ligand, $s_{kl} = 1$ and a low-range dielectric $\varepsilon_{ij} = r_{ij}\varepsilon_0$ is used. For a pair of water-exposed atoms, $s_{kl} = 0$ and a high-range dielectric function $\varepsilon_{ij} = r_{ij}\varepsilon_1$ is used. For a pair of partially exposed atoms, $0 < s_{kl} < 1$ and the dielectric function is an intermediate between $r_{ij}\varepsilon_0$ and $r_{ij}\varepsilon_1$. The sigmoid character of the function was chosen to enable smooth transitions between the low- and high-dielectric values. Overlap of the hydration-shell volumes of two solute groups with the protein and/or ligand is given by

$$s_{kl} = c(v_k + v_l), \quad (2)$$

where v_k and v_l are fractions of the hydration-shell volumes of solute groups k and l, respectively, which overlap with the

Table 2 Success rate (SR) and average RMSD (Å) for examining-set docking experiments

Protocol	Solvent-exclusion energy term	Titrable residues	Dielectric function	Match to the x-ray structure			
				AGM		One of top 10 minima[a]	
				SR (%)	RMSD	SR (%)	RMSD
Enriched libraries of ligand conformers							
1	Yes	Neutral	$\varepsilon = r$	58.3	2.65	75.0	1.66
2	Yes	Ionized	$\varepsilon = r$	60.0	2.46	72.7	1.83
3	Yes	Neutral	$\varepsilon = 2r$	58.3	2.94	71.6	1.86
4	Yes	Ionized	$\varepsilon = 2r$	60.0	2.84	75.0	1.71
5	Yes	Neutral	$\varepsilon = 4r$	56.7	2.70	66.7	2.00
6	Yes	Ionized	$\varepsilon = 4r$	58.3	2.64	71.6	1.90
7	No	Neutral	$2r \leq \varepsilon \leq 8r$	71.6	1.82	81.6	1.23
8	No	Ionized	$2r \leq \varepsilon \leq 8r$	78.3	1.70	85.0	1.18
Unbiased libraries of ligand conformers							
9	Yes	Ionized	$\varepsilon = 2r$	45.0	3.24	55.0	2.62
10	No	Ionized	$2r \leq \varepsilon \leq 8r$	58.3	2.51	71.3	2.23

RMSD of the apparent global minima from corresponding x-ray structures averaged over the 60 examining-set complexes

[a] Best-RMSD structure from the 10 lowest energy structures including the AGM

protein and/or ligand. For every solute group in the system, values v are computed upon each update of the interaction list by using Eqs. 6 and 21 derived by Lazaridis and Karplus for their Gaussian solvent-exclusion model [21]:

$$v_m = \sum_{n \neq m} \frac{2V_n}{4\pi\sqrt{\pi}\lambda_m r_{mn}^2} \exp\left[-\frac{(r_{mn} - R_m)}{\lambda_m}\right]^2. \quad (3)$$

Here, λ is the hydration-shell thickness, which equals 6 Å for neutral titrable groups and 3.5 Å for other groups, R is the van der Waals radius of a solute group [52]. V is the hydration-shell volume of a solute group taken from Table 1 of [21].

Summation in (3) is over solute groups, while electrostatic interactions are summed over all pairs of nonbonded atoms including hydrogens. Thus, (3) cannot be used to estimate values v_H for hydrogen atoms. We tested $v_H = 0$ and $v_H = v_s$, where s stands for the hydrogen-containing solute group. Better results were obtained for $v_H = v_s$. Coefficient $c = 0.625$ was empirically calibrated to ensure that $0 < s_{kl} < 1$ for any pair of solute groups at any distance. The SEDDD function is not differentiable, but since the hydration-shell volumes are constant between updates of the interaction list, the electrostatic energy term involving the SEDDD function is differentiable between the updates. Derivatives of energy in the space of generalized coordinates were calculated analytically [12, 40].

Optimizing parameters ε_0 and ε_1

Expressions (1–3) provide a physically meaningful model, but the challenge is to determine the combination of ε_0 and ε_1, which minimizes the RMSD between the predicted and experimental structures. The dielectric permittivity varies from $\sim$2 inside a protein to $\sim$80 in bulk water around the protein [4, 27]. For a pair of solvent-exposed nonbonded atoms at distance of 3 Å, $r_{ij}\varepsilon_1 = 80$ when $\varepsilon_1 \sim 27$. The dielectric permitivity at the protein surface is intermediate between the protein core and bulk water [28–30]. Therefore, we varied ε_0 and ε_1 from 1 to 10 with a step of 1, which gives 55 combinations of ε_0 and ε_1. For each combination, the AGM was searched using the two-stage docking protocol and the single-conformer ligand library. The RMSD of the ligand's heavy atoms between the AGM and the x-ray structure was plotted as the function of ε_0 and ε_1. Trial maps computed with and without the solvent-exclusion energy demonstrated that omitting the solvent-exclusion energy improves RMSD (not shown). This result indicates that calculations with the SEDDD function implicitly account for variable hydration of solute groups at different degrees of exposure to the solvent. We further computed RMSD plots against ε_0 and ε_1 without the solvent-exclusion energy for 16 ligand–protein complexes in the second training set.

Most of the plots have several RMSD minima (Fig. 8). The diagonal corresponds to calculations with the DDD function $\varepsilon = kr$ and without the solvent-exclusion energy. For commonly used DDD functions $\varepsilon = r$ to $\varepsilon = 4r$ the RMSD on the maps is usually above 2 Å. The majority of areas with RMSD < 1.0 Å occurred at regions where $\varepsilon_0 \neq \varepsilon_1$ indicating that when the solvent-exclusion energy term was omitted, calculations with the SEDDD function produced superior results as compared to calculations with the DDD function.

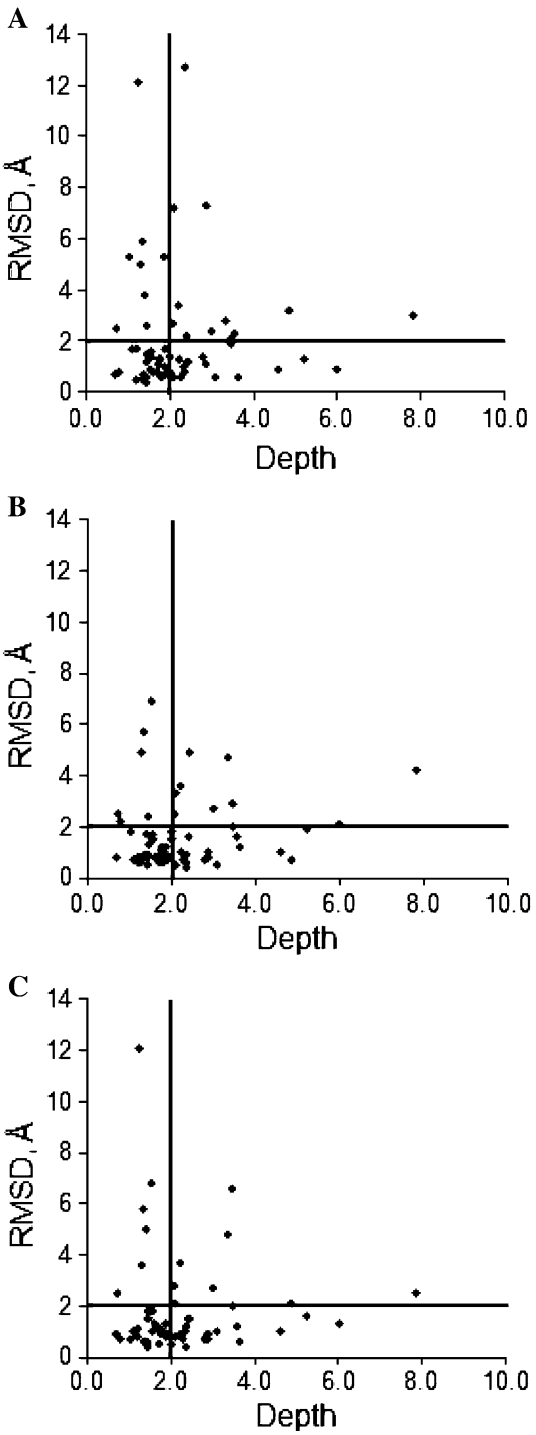

Fig. 7 RMSDs from the x-ray structures of the AGM structures predicted with the DDD function plotted against the depths of the ligand-binding pockets. Each point represents an examining-set ligand-receptor complex. The ligand-binding pocket depth was determined as the ratio of the number of flexible residues in the protein double-shell model to the number of the ligand's heavy atoms. **a**, **b** and **c** show, respectively, data obtained with the DDD functions $\varepsilon = r$, $\varepsilon = 2r$, and $\varepsilon = 4r$. Note high-RMSD points at shallow binding sites

The maps depend, in particular, on the number of the ligand polar groups. In the case of polar ligands, results depend dramatically on the dielectric function as exemplified by complex of chloramphenicol acetyltransferase with the ligand, which has a nitro group and two chlorine atoms (Fig. 8b). Maps for hydrophobic ligands, such as progesterone (1A28) have wide low-RMSD regions indicating the rather low effect of electrostatics (Fig. 9).

Examining set with the SEDDD function

Parameters $\varepsilon_0 = 2$ and $\varepsilon_1 = 8$ minimize RMSD in maps computed for the second training-set complexes (Fig. 8). We used these parameters to test the ability of the two-stage docking protocol to reproduce the examining-set complexes with the unbiased and enriched libraries of ligand conformers. Calculations were performed using titrable residues considered in both ionized and neutral forms (Table 2). In calculations with the enriched libraries and the DDD function, the highest success rate of 60% was obtained with ionized residues and $\varepsilon = r$ or $\varepsilon = 2r$, while calculations with neutral titrable residues yielded slightly lower success rates. Docking using the SEDDD function with ionized residues and enriched libraries resulted in a success rate as high as 78.3%. Besides an increase in the success rate, the SEEDD function also decreased the average RMSD of the AGMs from the x-ray structures from 2.46 to 1.7Å. Ionization of titrable residues had little effect with the DDD functions, but noticeably improved the success rate with the SEDDD function. These results show a high potential of the SEDDD function when the libraries of ligand conformers contain structures close to those in protein-bound complexes.

The final comparison of the DDD and SEDDD functions was performed with ionized residues and unbiased libraries of ligand conformers. The SEDDD function yielded a success rate of 58.3, which is 13.3% better than the best performing DDD function $\varepsilon = 2r$. Thus, flexible ligand docking with both unbiased and enriched libraries of ligand conformers demonstrated that the SEDDD function consistently yields higher success rates than the DDD functions.

Discussion

Electrostatic and solvent-dependent interactions

Various sophisticated, physically rigorous methods have been proposed for calculations of electrostatic interactions between molecules in solvents, e.g., [53] and references therein. Such methods usually involve large computational

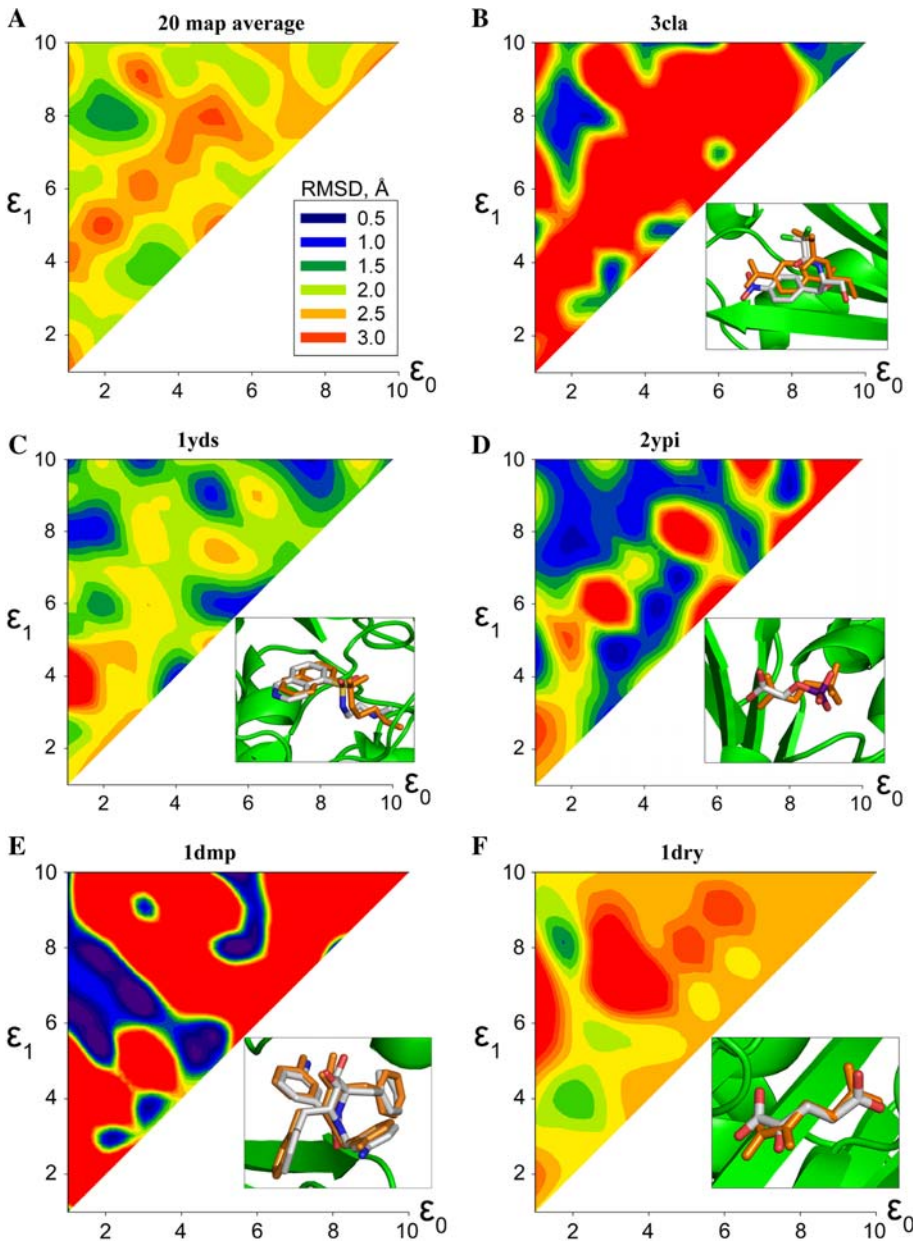

Fig. 8 Parameterizing the SEDDD function with the second training set. **a** The RMSDs of the apparent global minima from the corresponding x-ray structures are averaged over the 20 training-set complexes and plotted against parameters ε_0 and ε_1. Note the global minimum of RMSD at $\varepsilon_0 = 2$ and $\varepsilon_1 = 8$. **b–f** RMSD maps for individual complexes identified by PDB codes. *Insets* show the x-ray structures with CPK-colored ligands superimposed with the *orange-colored* ligands in AGM structures found with $\varepsilon_0 = 2$ and $\varepsilon_1 = 8$

cost and we are not aware of their systematic testing on large examining sets of ligand–protein complexes of various structures. Commonly used ligand-docking programs typically calculate electrostatic interactions with the DDD functions [6–8, 10, 11, 24–26]. In this study we proposed the SEDDD function that accounts for both electrostatic interactions and dehydration effect and increases the success rate of ligand docking.

To parameterize the function, we computed 16 RMSD maps and at each map found a minimum at $\varepsilon_0 = 2$ and $\varepsilon_1 = 8$ (Fig. 8). Other low-RMSD regions at higher values of ε_0 and ε_1 are also seen (Fig. 8) indicating that sometimes underestimation of electrostatic interactions is better than their overestimation. Comprehensive analysis of this observation is hardly possible given an approximate character of classical force fields. However, we found that local minima of RMSD at high values of ε_0 and ε_1 usually appear for complexes with ligand–protein H–bonds. Since H–bonding potential is included in the AMBER force field, the H–bonds were predicted even with underestimated electrostatic energy.

Electrostatic and solvation energy components are interrelated and in many programs the DDD function is used along with the solvent-exclusion energy or other energy terms that account for the dehydration energy. In this study, the highest success rate was obtained in calculations with the SEDDD function without the solvent-exclusion energy. A likely cause of the lower success rate

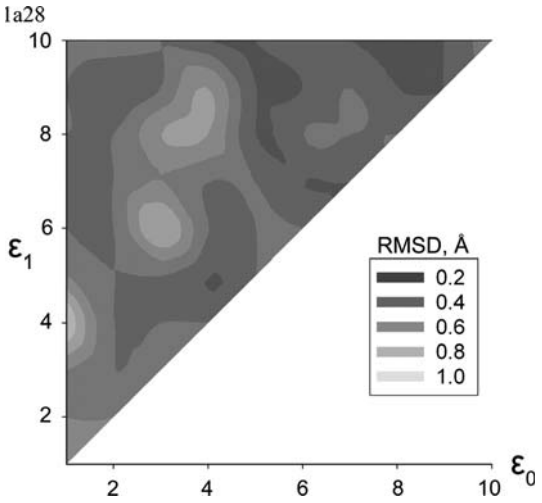

Fig. 9 RMSD map for a complex 1A28 between a progesterone receptor and progesterone, a hydrophobic ligand. Low RMSDs are obtained at all tested combinations of ε_0 and ε_1, indicating a small impact of electrostatic interactions on the results

of calculations with the SEDDD function and the solvent-exclusion energy is double counting of some charge-screening effects. Another cause may be overestimation of electrostatic attractions at solvent-exposed regions where beneficial hydration may overweight electrostatic repulsions. A disadvantage of omitting the solvent-exclusion energy term is that exposure of hydrophobic groups to the solvent is not penalized, but this is partially compensated by favorable van der Waals ligand–protein interactions in the protein buried regions.

Mallik et al. [15] proposed a method to calculate electrostatic energy with taking into account exposure of interacting atoms in the aqueous environment. The method solves the Poisson-Boltzmann equation to predict the electrostatic-potential field and calculates the effective dielectric permittivity as function of distance from the surface. For atoms close to the surface, their exposure to the aqueous environments is also taken into consideration. The method was parameterized using subtilisin, a rather spherical protein. When tested on three small proteins of various shape, the method produced better results as compared with the DDD function and Generalized Born model. However, the method was not tested on a wide variety of ligand–protein complexes. The computational cost of this method seems too high for high throughput ligand docking.

Combining the AMBER force field with the hydration shell model yielded correlation of the experimental and predicted binding energy in a series of 23 glucose analogue inhibitors of glycogen phosphorylase [54]. In our approach, the solvent exposure-dependent component of the SEDDD function weakens interactions at the protein surface thus accounting for the dehydration energy. Indeed, the examining-set calculations with the SEDDD function in vacuum increased the success rate versus analogous calculations with the solvent-exclusion energy (Table 2).

In this study we optimized the ligand-docking protocol and parameterized the SEDDD function using two training sets of 26 ligand–protein complexes, which represent diverse types of ligand–protein interactions. The SEDDD function enables smooth variations of the dielectric permittivity between $2r_{ij}$ and $8r_{ij}$. The lower limit is often used in molecular simulations. The upper limit allows weakening of the electrostatic interactions at the water-exposed surface. All of the 16 RMSD maps, which were computed with the second training set, have a RMSD minimum at $\varepsilon_0 = 2r$ and $\varepsilon_1 = 8r$ and this combination was chosen to parameterize the SEDDD function. Thus the latter apparently does not involve a "knowledge-based" component specific for the training sets and remains as "physics-based" as the DDD functions used with classical force fields including AMBER and CHARMM.

Flexible ligand docking

Flexible ligand docking is commonly performed by two-stage protocols in which precomputed libraries of ligand conformers are used at the seeding stage. Such libraries are easy to build for semirigid ligands and ligands containing a small number of rotatable bonds. However, flexible ligands with six or more rotatable bonds often adopt extended conformations in proteins, which are energetically unfavorable in lone ligands and may be unfavorable in water or lipid environments. An extended ligand conformation in complex with a protein maximizes ligand–protein contacts at the expense of the intra-ligand attractions. The fact that such conformations are observed in experimental structures suggests that the binding of corresponding ligands is an entropy driven process in which water molecules expelled from the ligand-binding site gain entropy. A similar mechanism explains, for example, the binding of antifreeze proteins to ice [55]. Energy-minimized extended conformations of lone ligands calculated with standard force fields have much higher energy than energy-minimized folded conformations. Therefore a flexible-ligand library including both folded and extended energy-minimized conformers accepted with only energy as the criteria (without any bias towards the experimental structure in complex with the protein) may contain many thousands of conformers. We are not aware of published algorithms for generating small libraries of ligand conformers without the protein context that would contain with a high probability a close match to the x-ray conformation. Here we generated

libraries of ligand conformers using an ad hoc force field that ignores intra-ligand attractions. To reduce the number of conformers we used a clustering method in which structures were compared by torsion angles. Matching torsions in a pair of conformers were considered similar if they differed by less than 1 radian. Using this approach, we were able to generate 93% of the examining-set libraries in which at least 1% of conformers had an RMSD < 2.0 Å from the x-ray structures. A disadvantage of this hands-free approach is the size of the 100-conformers libraries compared to libraries of 10 conformers per ligand [11]. Our large libraries contain many decoys. This may explain why ZMM-based flexible docking with the DDD function performed worse than analogous approaches by others with smaller libraries. However, docking with the SEDDD function and large unbiased libraries yielded a success rate comparable with those reported by other groups. The SEDDD function would have performed better with smaller libraries. Indeed, our calculations with the enriched 100-conformers libraries yielded a success rate as high as 78.3% in spite the fact that most of the conformers in the libraries are decoys. Automatic generation of small libraries of ligand conformers having a close match to the protein-bound conformation should involve analysis of the binding-site structure. This challenging problem was not addressed in the current study.

As mentioned above, docking a ligand using ZMM requires hours of CPU time, whereas similar calculations with programs like GLIDE, GOLD, ICM, and FlexX, which are optimized for high-throughput ligand docking, requires minutes. The high docking speed of such programs is due to several features, which are not implemented in ZMM. These include replacing explicit receptor atoms with grid potentials [6, 10] and filtering-out at the screening stage unpromising ligand-binding modes that mismatch the binding-pocket shape [8] or lack chemical complementarity to the binding site [7]. Another cause of large CPU time involved in our study are libraries of ligand conformers, which are much larger than those used by others [11]. In should be emphasized that ZMM was not designed for high-throughput ligand docking. We have employed ZMM and the SEDDD function for mapping drug receptors in homology models a calcium channel [56–58]. Recent simulations of large-scale conformational transitions of ion channels with the SEDDD function has allowed to reach lower RMSDs between the predicted and experimental structures than simulations with the DDD function [59]. In the current work, calculations with the SEDDD function not only demonstrated a higher success rate of predicting correct ligand-binding modes, but they were faster than calculations with the DDD function and the Gaussian solvent-exclusion energy term.

Conclusions

In this work we proposed a new dielectric function SEDDD that depends on the distance between interacting atoms and their exposure to the aqueous environment. The SEDDD function allows quantifying large variations of the dielectric permittivity between the buried and water-exposed areas in the protein. These variations have long been well recognized and different computational approached are proposed to consider them in molecular simulations. A classical example is a combination of the distance-dependent dielectric (DDD) function and the solvent-exclusion energy. However, the DDD function contains a heuristic coefficient whose choice is difficult and may require preliminary computations for a system under consideration. In our test calculations, the SEDDD function with the AMBER force field significantly improved the success rate of correct predictions of ligand–protein complexes versus results achieved with the DDD function and the solvent-exclusion energy. Thus, the simple and physically meaningful SEDDD function may be a method of choice to substitute both the DDD function and the solvent-exclusion energy. The function does not contain parameters specific for particular ligands or proteins and may be adapted to various programs for molecular simulations.

Acknowledgments We thank Denis Tikhonov for helpful discussions. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca). The study was supported by the grant MOP-53229 to BSZ from the Canadian Institutes of Health Research and a Graduate Scholarship to Daniel Garden from SHARCNET.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
2. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931
3. McInnes C (2007) Virtual screening strategies in drug discovery. Curr Opin Chem Biol 11:494–502
4. Gohlke H, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl 41:2644–2676
5. Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
6. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47:1739–1749

7. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
8. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261:470–489
9. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9:1–5
10. Totrov M, Abagyan R (1997) Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins Suppl 1:215–220
11. Meiler J, Baker D (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins 65:538–548
12. Zhorov BS (1981) Vector method for calculating derivatives of energy of atom-atom interactions of complex molecules according to generalized coordinates. J Struct Chem 22:4–8
13. Li Z, Scheraga HA (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci USA 84:6611–6615
14. Fogolari F, Zuccato P, Esposito G, Viglino P (1999) Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. Biophys J 76:1–16
15. Mallik B, Masunov A, Lazaridis T (2002) Distance and exposure dependent effective dielectric function. J Comput Chem 23:1090–1099
16. Gilson MK (1995) Theory of electrostatic interactions in macromolecules. Curr Opin Struct Biol 5:216–223
17. Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149
18. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 112:6127–6129
19. Dominy BN, Brooks CL (1999) Development of a generalized born model parametrization for proteins and nucleic acids. J Phys Chem 103:3765–3773
20. Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials of the naturally occurring amino acids. J Phys Chem 79:2361–2381
21. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35:133–152
22. McCammon JA, Wolynes PG, Karplus M (1979) Picosecond dynamics of tyrosine side chains in proteins. Biochemistry 18:927–942
23. Weiner SJ, Kollman PA, Case DA, Singh UC, Chio C, Alagona G, Profeta S, Weiner PK (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784
24. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15:411–428
25. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 10:293–304
26. Wang J, Kollman PA, Kuntz ID (1999) Flexible ligand docking: a multistep strategy approach. Proteins 36:1–19
27. Finkelstein AV, Ptitsyn O (2002) Protein physics. Academic Press, London
28. Teschke O, Ceotto G, de Souza EF (2001) Interfacial water dielectric-permittivity-profile measurements using atomic force microscopy. Phys Rev E Stat Nonlin Soft Matter Phys 64:011605
29. Rubinstein A, Sherman S (2007) Evaluation of the influence of the internal aqueous solvent structure on electrostatic interactions at the protein-solvent interface by nonlocal continuum electrostatic approach. Biopolymers 87:149–164
30. Bockris JO, Reddy AKN (1977) Modern electrochemistry. Plenum Press, New York
31. Mehler EL, Solmajer T (1991) Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Eng 4:903–910
32. Gelpi JL, Kalko SG, Barril X, Cirera J, de La Cruz X, Luque FJ, Orozco M (2001) Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. Proteins 45:428–437
33. Morreale A, Gil-Redondo R, Ortiz AR (2007) A new implicit solvent model for protein-ligand docking. Proteins 67:606–616
34. Morris GarrettM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
35. Augspurger JD, Scheraga HA (1996) An efficient, differentiable hydration potential for peptides and proteins. J Comput Chem 17:1549–1558
36. Blanchet J, Lin SX, Zhorov BS (2005) Mapping of steroids binding to 17 beta-hydroxysteroid dehydrogenase type 1 using Monte Carlo energy minimization reveals alternative binding modes. Biochemistry 44:7218–7227
37. Zhorov BS, Bregestovski PD (2000) Modeling chloride channels of glycine and GABA receptors with blockers. Biophys J 78:A2092
38. Tikhonov DB, Zhorov BS (2007) Sodium channels: ionic model of slow inactivation and state-dependent drug binding. Biophys J 93:1557–1570
39. Bruhova I, Zhorov BS (2007) Monte Carlo-energy minimization of correolide in the Kv1.3 channel: possible role of potassium ion in ligand-receptor interactions. BMC Struct Biol 7(5):1–13
40. Zhorov BS (1983) Vector method for calculating derivatives of the energy deformation of valence angles and torsion energy of complex molecules according to generalized coordinates. J Struct Chem 23:649–655
41. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force-field for simulations of proteins and nucleic-acids. J Comput Chem 7:230–252
42. Brooks CL, Pettitt BM, Karplus M (1985) Structural and energetic effects of truncating long ranged interactions in ionic polar fluids. J Chem Phys 83:5897–5908
43. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) AM1: a new general purpose quantum mechanical model. J Am Chem Soc 107:3902–3909
44. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM (1997) PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci 22:488–490
45. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 56:235–249
46. Zhorov BS, Lin SX (2000) Monte Carlo-minimized energy profile of estradiol in the ligand-binding tunnel of 17 beta-hydroxysteroid dehydrogenase: atomic mechanisms of steroid recognition. Proteins 38:414–427
47. Hopfinger AJ, Battershell RD (1976) Application of SCAP to drug design. 1. Prediction of octanol-water partition coefficients using solvent-dependent conformational analyses. J Med Chem 19:569–573
48. Teague SJ (2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2:527–541

49. Cavasotto CN, Abagyan RA (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 337:209–225
50. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd (2004) Assessing scoring functions for protein-ligand interactions. J Med Chem 47:3032–3047
51. Cavasotto CN, Orry AJ, Abagyan RA (2003) Structure-based identification of binding sites, native ligands and potential inhibitors for G-protein coupled receptors. Proteins 51:423–433
52. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 4:187–217
53. Hassan SA (2007) Liquid-structure forces and electrostatic modulation of biomolecular interactions in solution. J Phys Chem B 111:227–241
54. Venkatarangan P, Hopfinger AJ (1999) Prediction of ligand-receptor binding thermodynamics by free energy force field three-dimensional quantitative structure-activity relationship analysis: applications to a set of glucose analogue inhibitors of glycogen phosphorylase. J Med Chem 42:2169–2179
55. Jorov A, Zhorov BS, Yang DS (2004) Theoretical study of interaction of winter flounder antifreeze protein with ice. Protein Sci 13:1524–1537
56. Tikhonov DB, Zhorov BS (2008) Molecular modeling of benzothiazepine binding in the L-type calcium channel. J Biol Chem 283:17594–17604
57. Tikhonov DB, Zhorov BS (2009) Structural model for dihydropyridine binding to L-type calcium channels. J Biol Chem 284:19006–19017
58. Cheng RC, Tikhonov DB, Zhorov BS (2009) Structural model for phenylalkylamine binding to L-type calcium channels. J Biol Chem 284:28332–28342
59. Garden DP, Bruhova I, Zhorov BS (2010) In-silico activation and deactivation of the pore domains of voltage- and ligand-gated ion channels. Biophys J Supplement 2687-Pos