Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 1259–1268
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis

Tommy Sander, Tommy Liljefors, Thomas Balle *

Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark

Received 9 July 2007; received in revised form 22 November 2007; accepted 23 November 2007
Available online 4 December 2007

Abstract

Highly flexible proteins constitute a significant challenge in molecular docking within the field of drug design. Depending on the efficacy of the bound ligand, the ligand-binding domain (LBD) of the ionotropic glutamate receptor iGluR2 adopts markedly different degrees of domain closure due to large-scale domain movements. With the purpose of predicting the induced domain closure of five known iGluR2 partial to full agonists we performed a validation study in which normal mode analysis (NMA) was employed to generate a 25-membered ensemble of iGluR2 LBD structures with gradually changing domain closures, followed by accurate QM/MM docking to the ensemble. Based on the docking scores we were able to predict the correct optimal degree of closure for each ligand within 1–3° deviation from the experimental structures. We demonstrate that NMA is a useful tool for reliable ensemble generation and that we are able to predict the ligand induced conformational change of the receptor through docking to such an ensemble. The described protocol expands and improves the information that can be obtained from computational docking when dealing with a flexible receptor.

Keywords: Protein flexibility; Molecular docking; Normal mode analysis; Elastic network model; Ensemble generation; iGluR2 receptor; Domain closure

1. Introduction

Computational docking of small molecules to a macromolecular target is a widely used tool in drug discovery and development projects ranging from high-throughput virtual screening (HTVS) of billions of compounds to the more focused optimization of one or a few candidates [1,2]. Currently, most standard docking protocols incorporate ligand flexibility into the docking process while considering the protein as a rigid structure. Bearing in mind the often significantly different conformations observed for many protein structures deposited in the protein data bank (PDB) [3], this simplification is not easily justified. On the other hand, due to the exponential nature of the problem, sampling the many degrees of freedom of a protein quickly becomes a daunting task, both practically and computationally. Using only one rigid target structure, however, not only reduces the chance of identifying a ligand that would bind to the receptor in another conformation [4], it also eliminates the possibility to predict the conformational change of the receptor induced by the ligand upon binding.

Docking methods taking receptor flexibility into account have been developed, and they can roughly be divided into two groups: induced fit docking focusing on the ligand–receptor interface, and docking to an ensemble of receptor conformations [5]. Induced fit docking can be done indirectly by allowing closer ligand–receptor contacts through softer van der Waals (vdW) repulsion terms, or more directly by sampling and refining the structure of the binding pocket allowing it to adjust itself to the docked ligand [1,6]. While these methods account for local changes in side chain conformations, large-scale domain movements in the receptor are not explored. Ensemble docking better allows for such structural changes to be included in the docking protocol. However, this requires that an adequate number of conformationally different structures are available for inclusion in the ensemble. Despite significant advances in

* Corresponding author. Tel.: +45 35 33 64 09; fax: +45 35 33 60 40.
E-mail address: tb@farma.ku.dk (T. Balle).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.11.006

X-ray crystallography [7] giving access to more and more protein structures, comprehensive experimentally determined conformational ensembles are still relatively rare. Hence, computational methods such as molecular dynamics (MD) [8–10] and normal mode analysis (NMA) [11–13] have become popular methods for generating such ensembles. Especially NMA seems suitable for describing large-scale domain movements of a flexible protein at low computational cost.

An example of a highly flexible receptor displaying large-scale rigid body domain movements is found among the ligand-gated ionotropic glutamate receptors (iGluRs), which are the primary mediators of fast excitatory neurotransmission in the central nervous system [14]. For the iGluR2 subtype, of which more than 50 crystal structures are available in the PDB [3], it has been observed that the ligand-binding domain (LBD) adopts significantly different conformations depending on the co-crystallized ligand. In the inactivated apo state or with bound antagonists the two subdomains, D1 and D2, of the LBD (Fig. 1) are separated in an open conformation, whereas a full agonist such as glutamate (Glu) induces a rapid, full closure where D1 and D2 move together like a clamshell [15]. Partial agonists induce intermediate domain closures, and it has been shown that the degree of domain closure induced by a ligand is proportional to its efficacy [16]. Hence, it is of high relevance to be able to predict the ligand induced receptor conformation through ensemble docking. With the availability of an experimental series of conformationally different structures, the iGluR2 LBD is an ideal system for testing the principles of docking to a computationally derived ensemble.

Therefore, we performed a validation study where NMA, in the simplified elastic network model (ENM) form [11,17,18], was applied as a tool for the generation of an iGluR2 ensemble to which five partial to full agonists inducing a known degree of LBD closure were docked. First, to ensure that the employed docking program was able to discriminate between the different receptor conformations, a cross-docking experiment was carried out where each of the five ligands was docked to each of the five receptor structures. Then a perturbation of the semi-closed iGluR2 LBD structure in complex with kainate (KA), a weak partial agonist, along the 100 lowest-frequency normal modes was performed to yield 25 model structures with a varying degree of domain closure. The five ligands were docked to all 25 models, and the obtained scores were evaluated to establish whether this ensemble docking protocol makes it possible to predict the experimentally observed degree of closure for each ligand.

With the described method we present an initial evaluation of a new approach to overcome some of the difficulties in docking to a highly flexible protein. It should be emphasized that the purpose of this study was to establish whether it was possible, with the proper setup, to predict the receptor closure induced by a ligand. Thus, we do not attempt to solve the non-trivial problems of e.g. side chain flexibility and water molecule interactions which are often important parts of the system setup in a study such as the present. However, the basic principles followed could serve as inspiration in cases where only a few conformationally different protein structures are available, and where further exploration of the conformational space with the purpose of docking is desired.

2. Methodology

All calculations were performed with 64-bit AMD Opteron processors. Molecular modeling tasks were performed with

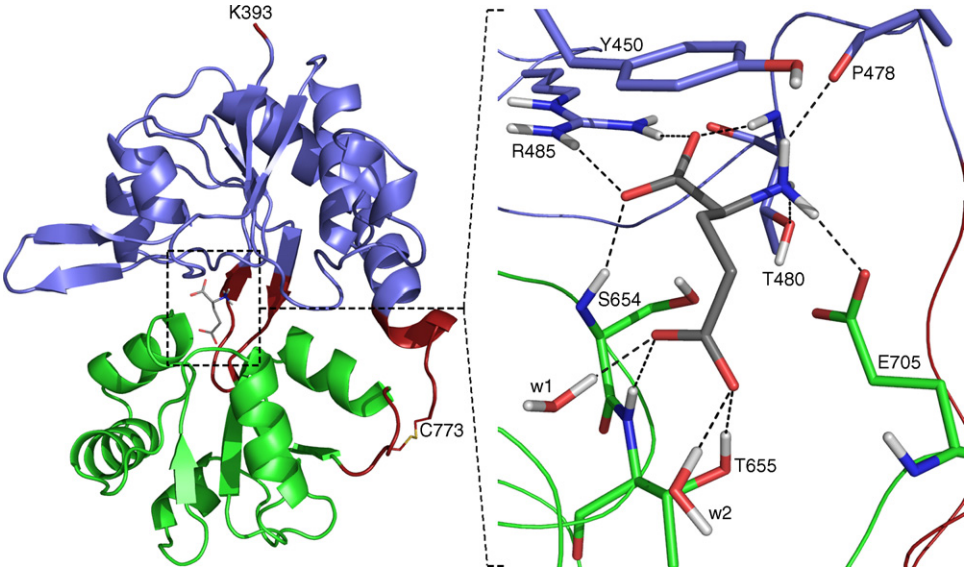


Fig. 1. (Left) Domain definitions of the LBD exemplified with the iGluR2–Glu structure in cartoon representation. D1 backbone is slate blue, D2 is green, hinge and terminal regions are ruby. Terminal residues are indicated, and the C718–C773 disulfide bridge is shown. Glu (gray carbons) is shown inside the binding pocket. (Right) Close-up view of iGluR2–Glu with a ribbon backbone showing Glu, the pocket residues, water molecules w1 and w2, and important hydrogen bond interactions (black dashed lines). The structure shown is that obtained after geometry optimization with heavy atoms frozen, as described in Section 2.1. Non-polar hydrogens are left out for clarity. Carbon coloring in pocket residues corresponds to the domain definitions in the left hand picture. Non-carbon coloring: H, white; N, blue; O, red; S, yellow.

Maestro [19], and polynomial regressions were calculated in Microsoft Office Excel 2003. If not otherwise stated, default settings were used during all calculations. Amino acids are designated with their one-letter code.

2.1. Selection and preparation of crystal structures

Five structures [15,20,21] of the iGluR2 LBD in complex with different partial or full agonists (Scheme 1) were downloaded from the PDB along with the apo structure (PDB code 1FTO) [15]. The five ligands and corresponding structures were selected based on the following considerations: (1) the ligands should have similar binding modes while being structurally different and (2) the LBD structures should have different degrees of domain closures. Since this can be considered a pilot study on the applicability of docking to ensembles generated with NMA, similar binding modes were sought to ensure that the induced degree of domain closure was the only significant difference in the ligand–receptor interactions. This criterion limited the set of appropriate structures to the LBDs co-crystallized with agonists. With KA and ACPA (Scheme 1) representing the extremes within this group in terms of induced domain closure, these ligands were obvious choices. Likewise, Glu was natural to include as the endogenous iGluR2 agonist. Finally, the Br-HIBO and Hwd bound structures completed a wide distribution of domain closures while introducing ligands with different ring structures (Scheme 1).

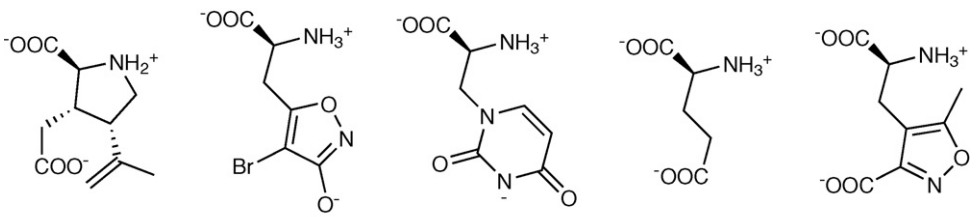
In the following, the experimentally determined protein structures (excluding the bound ligand) are designated according to their co-crystallized ligand, i.e. iGluR2–KA, iGluR2–Br-HIBO etc. The protein sequence is numbered according to the mature iGluR2 subunit as adopted in Ref. [15]. The LBD starts with the N-terminal N392, K393, or T394, depending on the X-ray data, and ends with C773 (Fig. 1).

The following definitions are applied to the LBD (Fig. 1): subdomain D1 is comprised of residues T394-M496 and Y732-W766, and correspondingly D2 is defined by residues S501-Q714 and T720-N726. The remaining residues make up hinge [15] or terminal regions belonging to neither subdomain. The pocket residues Y450, P478, T480, R485, S654, T655, and E705 are those having significant electrostatic, hydrogen bonding, or vdW interactions with all agonists.

The agonist bound structures were treated as follows: all ions, co-crystallized ligands, chains other than chain A, and water molecules except the two shown in Fig. 1, designated w1 and w2, were removed. Water molecules w1 and w2 are present in all published iGluR2 crystal structures [22]. N- and C-terminals were modeled as ammonium and carboxylate groups, respectively, and the sulfur atoms of C718 and C773 were connected to form a disulfide bridge. The docking software Glide [23–25] was used in this study, and prior to the actual docking this program was also used for protein preparation (pprep), which attaches hydrogens at the correct positions and performs a constrained geometry optimization. To avoid biasing towards a specific ligand, the binding pockets were filled with water molecules during this step. With the solvation tool *genbox* from the MD package GROMACS [26], water molecules were placed in and around the protein. All water molecules more than 10 Å away from the one judged to be the most centrally placed one inside the binding pocket were removed. Pprep was applied to this system with the most centrally placed water molecule chosen as the ligand. Subsequently, all water molecules except w1 and w2 were deleted.

M708 is a long and flexible residue, which extends into the binding pocket and displays markedly different conformations in the various crystal structures. This interferes with the underlying presumption of this study that the degree of domain closure is the only significant determinant for the preferred receptor conformation for each ligand. Therefore, this residue was mutated to an alanine after the pprep step. Potentially, instead of the M708A mutation one could generate different rotameric states of M708 or use an induced fit docking protocol to account for the flexibility of this particular residue. However, we chose not to introduce further complications into the study. A brief examination of the effect of having M708 present in a conformation giving maximum space for ligand binding was actually carried out for the cross-docking test set. For this step the $\chi_1 - \chi_3$ torsions of M708 were adjusted after pprep to equal those in the iGluR2–(*S*)-ATPA crystal structure (PDB code 1NNK). This was followed by a geometry optimization with MacroModel [27] keeping everything but the side chain atoms of M708 frozen, using the OPLS 2005 force field, the GB/SA solvation model, and the PRCG method converging on gradient with a 0.1 kJ/(mol Å) threshold. These structures were treated like those with the M708A mutation, including the ligand docking step.

Since all ligands are tightly bound to the receptor through many hydrogen bonds, correct orientations of these are pivotal for obtaining reliable docking results. Therefore, hydrogen bonds and water molecules had to be oriented optimally. For



Scheme 1. Ligand structures (with the applied protonation states shown), PDB codes, and degrees of domain closure of the five iGluR2 LBDs constituting the cross-docking test set. Domain closures are measured relative to the iGluR2 apo structure (PDB code 1FTO) as described in the text.

this step, Glu was temporarily inserted into each binding pocket in the orientation corresponding to that observed in the X-ray structure by superimposing the five structures after pprep onto the pocket residue Cα atoms of the original iGluR2–Glu. Subsequently, a geometry optimization with Glu inside the pocket and with all heavy atoms except w1 and w2 frozen was performed on each of the crystal structures with MacroModel, using the OPLS 2005 force field, the GB/SA solvation model, and the PRCG method converging on gradient with a threshold of 0.1 kJ/(mol Å). To avoid distortion of the two hydrogen atoms in R485 binding directly to the α-carboxylate group of the Glu (Fig. 1), these were also held frozen during the optimization. In this way the shape of the binding pockets were not biasing towards Glu.

Finally the receptor grid, which is the way Glide represents a protein cavity, was calculated with the previously inserted Glu as the ligand defining the binding pocket centre.

2.2. Ensemble generation with NMA

A modified version of the program *nexe*, which runs at the AD-ENM web server [11,28], was provided by Dr. Wenjun Zheng from the NIH, Bethesda, MD. The modifications allowed the calculation and use of an arbitrary number of normal modes as well as reporting for each normal mode the sign and weight necessary to perform a certain perturbation.

The program was used to calculate the normal modes and to do the actual perturbations with the ENM approximation connecting all Cα atoms within a cut-off distance of 10 Å. First the normal modes of the semi-closed input structure, iGluR2–KA, were calculated. Then the weight and corresponding sign, i.e. direction, was identified for each normal mode responsible for changing the conformation of iGluR2–KA to that of the closed target structure, iGluR2–Glu. This target structure was chosen due to Glu being the endogenous agonist as well as the degree of closure being relatively far from that of iGluR2–KA compared to the Br-HIBO and Hwd bound structures. Finally, these weights and an amplitude varying from −0.4 to 2.0 in steps of 0.1 were applied to the input structure to perturb iGluR2–KA along the lowest 100 non-trivial modes. This yielded an ensemble of 25 conformations, designated model 1 through 25, with increasing degrees of closure. Note that with an amplitude of 0.0, model 5 has a conformation corresponding to that of iGluR2–KA.

The generated conformational ensemble included only backbone and, for non-glycine residues, Cβ atoms. Thus, employing a python script in Maestro all side chain atoms were reinserted and all non-backbone torsional angles were adjusted to match those of the iGluR2–Glu crystal structure. Again this was chosen as template structure so that only two experimental structures (iGluR2–KA and iGluR2–Glu) were needed for the ensemble generation. No atoms were added to glycine and alanine residues during this step. No changes were made to torsional angles in proline residues due to their cyclic nature, which was not easily incorporated into the script handling the torsion adjustments. Besides, only P480 is placed close to the ligand-binding site and in all models this was given the correct conformation directly upon reinsertion, so that no adjustment was necessary for this residue. After adjusting all non-backbone torsion angles in the models, Y450 and E705 were still oriented relatively far from those in iGluR2–Glu. Since the side chain atoms of these residues interact closely with all ligands it was deemed necessary to also adjust some of their bond angles to match those of iGluR2–Glu. Specifically, the H-Cα-Cβ angles in Y450 and E705, and the Cα-Cβ-Cγ angle in Y450, were adjusted to equal those in the template after the default addition of hydrogens to the Cα atoms of these residues in Maestro. The two water molecules, w1 and w2, were inserted into each model by superimposing it onto the D2 Cα atoms of iGluR2–Glu. From this point, each model was treated as described in Section 2.1 for the five crystal structures.

2.3. Docking and pose evaluation

The five ligands were built as shown in Scheme 1, i.e. with a zwitter-ionic α-amino acid group and a deprotonated distal part, followed by a geometry optimization with MacroModel using the MMFFs force field, the GB/SA solvation model, and the PRCG method converging on gradient with a threshold of 0.05 kJ/(mol Å).

Ligands were docked to all receptor grids using Schrödinger's QM-Polarized Ligand Docking (QPLD) protocol [29]. This QM/MM approach is a computationally rather expensive method and is therefore not the first choice in a docking experiment. However, agonist binding to iGluR2 is based almost solely on hydrogen bonds between charged or highly polar groups, interactions which are not accurately described with classical static charge force fields of regular docking protocols such as Glide itself. We settled on the QPLD approach since initial cross-dockings with Glide in both standard and extra precision modes failed to reflect the experimentally observed properties of the ligands in terms of both scoring pattern and correct pose identification (data not shown). In QPLD, an initial Glide Standard Precision (SP) docking is performed, followed by a QM derived charge calculation for the top 5 poses of the ligand in the field of the receptor, and finally re-docking and re-scoring with the five new sets of charges, keeping the 10 highest ranked poses. Standard settings were kept, except for the QM Level, which was set to Medium (3-21G basis set, BLYP functional, Ultrafine SCF accuracy level).

For each docking pose, Glide reports two different scoring values, namely Gscore and Emodel [24]. Whereas the purpose of Gscore is to predict differences in binding affinity between different ligands, Emodel is better for ranking different poses of the same ligand. In addition, the algorithm for calculating the Gscore value for a docking pose reduces the net ionic charge in formally charged groups by ca. 50%. Five such groups are present in each ligand–receptor pair of this study (three in the ligand, and R485 and E705 in the receptor; cf. Fig. 1). By contrast, the Coulomb terms of Emodel is based on the full charges. Since each ligand was evaluated separately and because Coulomb interactions involving charged groups play a crucial role in the current receptor system, we based our docking results on the Emodel scores.

The best (most negative) score for a correct pose was taken as the result for each docking. The criteria for a correct pose were (1) a heavy atoms root mean square deviation (RMSD) below 1.5 Å compared to the experimental ligand structure and (2) a correct hydrogen bonding pattern of especially the α-amino acid part of the ligand (Fig. 1). For measuring ligand RMSD values (docking poses vs. crystal structures) the five experimentally determined receptor bound ligand structures were placed as correctly as possible in all binding pockets. This was done by superimposing all structures/models onto the pocket residue Cα atoms of iGluR2–Glu. In this way the five experimentally determined ligands were placed such that approximate RMSD values could be measured after docking to both native and non-native structures. No distinction was made between the two possible orientations of the isopropenyl group in a correct KA pose, i.e. either direction of the double bond was regarded equally correct, provided that the substituent was in a pseudo-equatorial position.

2.4. GRID calculation

The GRID program [30,31] was used to determine energetically favorable binding sites for water molecules in the iGluR2–Glu structure. This was done to verify that the two water molecules w1 and w2 included in all dockings could have been identified without prior experimental knowledge of their presence as mentioned in Section 2.1. The grid spacing was set to 0.5 Å (NPLA = 2), and the box dimensions were set automatically. The OH2 probe was used, and the protein structure was chain A of the iGluR2–Glu crystal structure (PDB code 1FTJ) after deletion of all water molecules, ions, and ligands.

2.5. Measurements

RMSD values were calculated with Maestro for ligands and with PyMOL [32] for protein structures. Domain closures relative to chain A of the iGluR2 apo structure were calculated in PyMOL with a python script [33]. In short, this script measures the rotational component of the rotation-translation matrix in degrees required to move the target protein from an iterative least squares Cα superposition on D1 to that on D2 of a reference structure (here the iGluR2 apo structure). Distances between the centers of mass of D1 and D2 Cα atoms were measured with the *g_dist* program from GROMACS. All Cα RMSD values, domain closures, and distances reported were measured after pprep as described in Section 2.1.

3. Results and discussion

3.1. Cross-docking to experimentally determined receptor structures

The cross-docking study was performed to test the underlying assumption that the docking protocol is able to identify the correct receptor conformation as the preferred structure for each ligand.

Before analyzing the docking scores we briefly describe the ability of Glide to dock the ligands in the correct orientations and conformations, according to the criteria for correct and incorrect poses described in Section 2.3. For Br-HIBO, Hwd, Glu, and ACPA the best scoring pose of the top 10 poses reported by Glide was correct in all five receptor structures. For KA, this was the case in its own receptor only; in its non-native structures no correct pose was found among the top 10. In iGluR2–Glu no KA pose was reported at all due to the Emodel score being above zero. RMSD values for the top ranked pose for each ligand in its native receptor were below 0.6 Å (see Supplementary Data).

Fig. 2 shows the Emodel scores for the top ranked pose for each ligand in each receptor (see Supplementary Data for values in tabulated form). KA, Hwd, Glu, and ACPA all obtained the best score in their native receptor, with a margin in Emodel score of at least 2.4 kcal/mol. Especially KA was highly favored in its own receptor by the scoring function, since no correct pose was reported in the others. Br-HIBO obtained the second best score in its native structure, the top ranked pose in iGluR2–Hwd being 7.0 kcal/mol better. Although this blurs the otherwise perfect scoring pattern it should be noted that iGluR2–Hwd is the structure which most closely resembles iGluR2–Br-HIBO, with a difference of ca. 1° in domain closure. Except for the small differences in backbone conformation, no significant differences between the binding pockets are visible. Taking the relatively good score in its native receptor into account as well, the result for Br-HIBO can thus be considered in good accordance with the expected outcome.

Based on these cross-dockings we find that the test set is valid in terms of the ability to predict the correct degree of domain closure for each ligand with the employed docking procedure.

As mentioned previously, M708 is found to occupy multiple conformational states depending on the co-crystallized ligand. To avoid moving the focus of this study from the large-scale domain movements to the challenge of flexible side chains in the binding pocket, we chose to neglect this residue. Of course, the M708A mutation allows for a larger ligand substituent in that region of the cavity than would normally be allowed by the

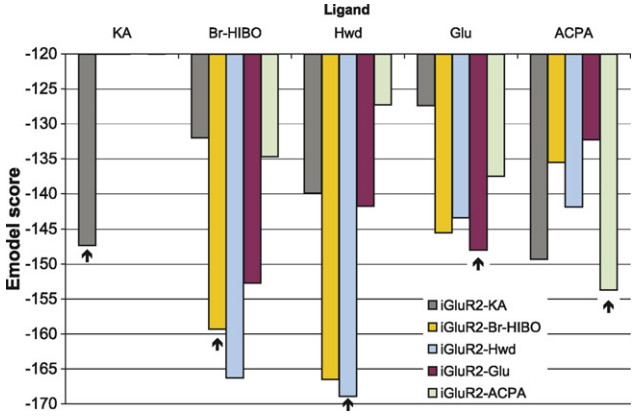


Fig. 2. Results of the cross-docking. Emodel scores are shown for the top scoring pose of each ligand in each of the experimentally determined receptor structures. Arrows indicate the score of the ligand in its native receptor.

methionine. A solution to this, without having to explicitly include side chain flexibility, could be to set the rotameric state of M708 to that seen in e.g. the iGluR2–(*S*)-ATPA structure (PDB code 1NNK). This conformation seems to give rise to the largest empty space in the pocket. Cross-docking to the five test set structures with the rotameric state of M708 set to that in iGluR2–(*S*)-ATPA resulted in the exact same pattern of ranking the ligands (results not shown), although the Emodel scores for Br-HIBO were consistently higher. Thus, keeping in mind the need to keep the system setup as simple as possible in this study, we find the M708A mutation employed in this study well justified.

3.2. Ensemble generation

The described NMA protocol resulted in a smooth rigid-body perturbation of the iGluR2–KA structure towards iGluR2–Glu, as evident from plotting the degree of domain closure and the D1–D2 centre-of-mass distance against the model number (Fig. 3). The models were subjected to pprep before these measurements were performed, hence the small fluctuations. The lowest overall Cα RMSDs of the generated models compared to each of the five experimental structures are in the range of 0.15–0.48 Å as depicted in Fig. 4. The close resemblance to iGluR2–KA is due to the fact that this was the input structure of the perturbation. The observed hinge-bending domain movement is illustrated in Fig. 5, where the open model 1 and the intermediately closed model 13 are superimposed onto D1 Cα of model 25 which has the highest degree of closure.

3.3. Docking to the generated ensemble

Analysis of the binding mode of the reported docking poses, according to our definition of correct and incorrect poses in Section 2.3, revealed that the docking of Hwd, Glu, and ACPA resulted in a correct pose as the highest ranked in all 25 models. A correct Br-HIBO pose was top ranked in models 1–20, 22, and 23; only incorrect poses were reported in the remaining three models. For KA a correct pose was ranked first in models 1–13 and 17, and third and second in models 14 and 15, respectively. In the remaining models 16 and 18–25 no correct

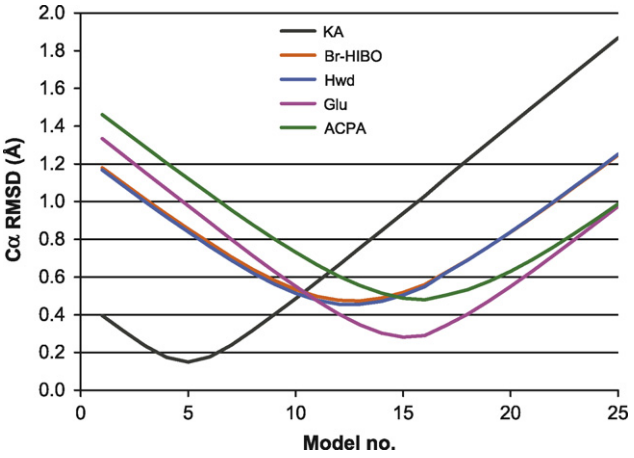


Fig. 4. Overall Cα RMSD for the generated models after pprep compared to each of the experimental structures (designated by their co-crystallized ligand).

KA pose was reported. The highest heavy atoms RMSD values for a correct pose compared to the crystal structure was 1.4 Å (ACPA in model 1, and Hwd in model 25), and 85% of the 113 correct poses reported had RMSD < 1.0 Å.

Scoring results from the ensemble docking is shown in Fig. 6, where the Emodel scores for each ligand is plotted against the domain closure of the corresponding ensemble member. Only values for the highest ranked correct poses are included; the top ranked, incorrect poses for KA and Br-HIBO had Emodel values in the ranges of −123.2 to −95.2 and −125.8 to −114.8, respectively. See Supplementary Data for the Emodel scores of all top ranked poses.

The predicted optimal domain closure for a ligand can be determined from these dockings in two ways: (1) by taking the closure of the model in which the single best score is obtained or (2) by fitting a mathematical expression to the scoring values and then take the minimum of this to define the optimal degree of closure. For the latter solution we fitted 2nd order polynomial regression curves to the correct pose Emodel values (Fig. 6), with the corresponding squared correlation coefficients (R^2) also indicated. Fig. 7 compares the results obtained from both methods with the experimentally determined domain closures for each ligand. The overall tendency of the more efficacious ligands preferring the more closed structures, and vice versa, is

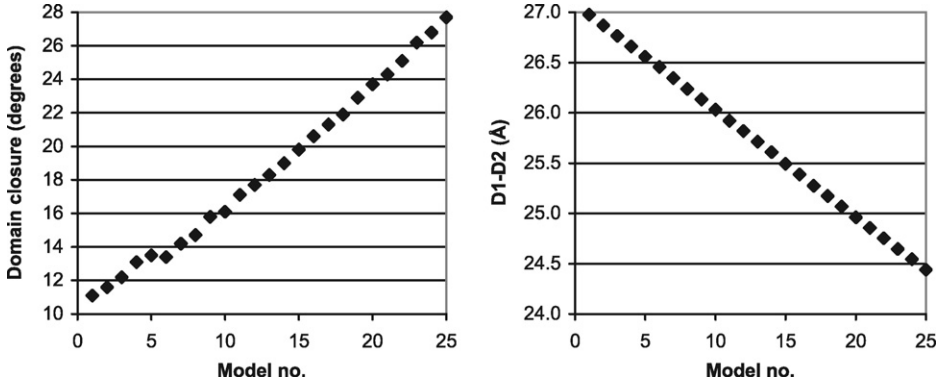


Fig. 3. Changes in degree of domain closure (left) and the D1–D2 centre-of-mass distance (right) for the 25 generated models.

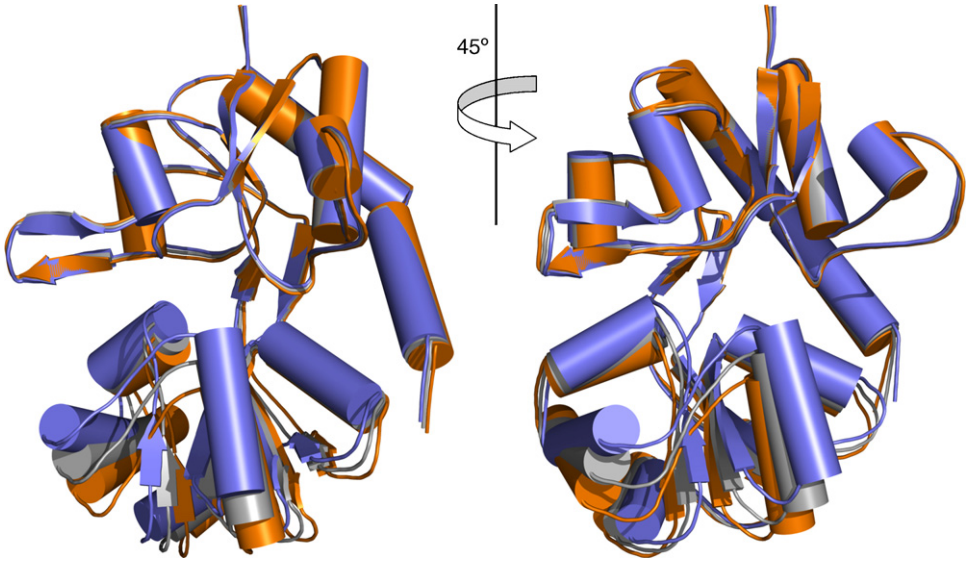


Fig. 5. The backbone conformational change of D2 relative to D1 in the NMA perturbation, shown in cartoon representation with cylindrical helices. Model 1 (orange), representing the most open conformation of the ensemble, and model 13 (light gray) are aligned to D1 Cα atoms in model 25 (slate blue) having the highest degree of closure. The left image is oriented as in Fig. 1; the right image is after a 45° rotation left to front.

observed regardless of which method is chosen. The deviations lie in the range of 1–3°, with Glu representing the largest deviation when taking the minimum of the 2nd order polynomial as the optimal degree of closure.

One should keep in mind the constant fluctuations of e.g. a protein–ligand complex, in the sense that a given optimal closure of e.g. 18° represents a certain interval (e.g. 17–19°) within which the receptor closure is found most of the time, rather than a completely static structure. From MD simulations of the iGluR2 LBD, Arinaminpathy et al. [34,35] have shown a reduction in domain mobility which is more pronounced for full agonists compared to partial agonists. This is in agreement with the similar scores seen in Fig. 6 for the partial agonists KA, Br-HIBO and Hwd in the models with closures close to the curve minimum. Thus, deviations such as those seen in Fig. 7 are expected.

Although both methods for determining the optimal degree of closure result in good accordance with experimental data, we find that a distinction between the two approaches should be made in the general case when evaluating such docking studies. When dealing with distinct and different states of a receptor, such as the five experimentally determined crystal structures in the cross-docking experiment, the single best score should be taken to

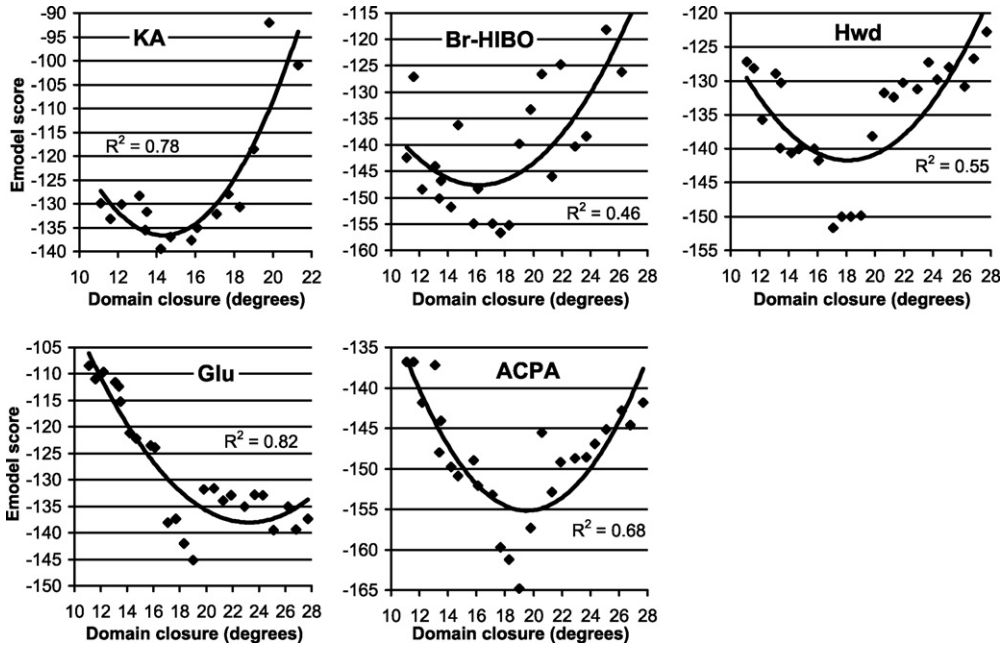


Fig. 6. Results of the QM/MM docking to the generated ensemble of iGluR2 LBD structures. Emodel scores of the highest ranked correct pose for each ligand in each model are shown as a function of the degree of domain closure. The 2nd order polynomial regression lines and corresponding R^2 are also shown.

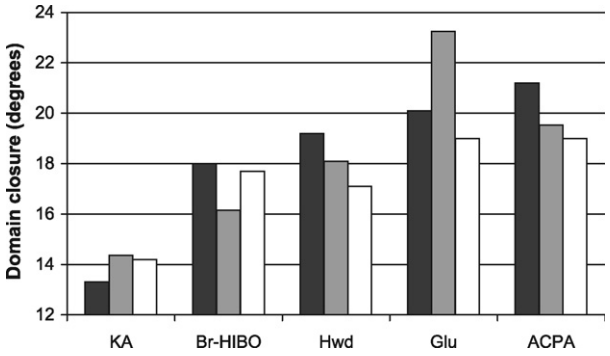


Fig. 7. Experimentally observed domain closures (dark gray) compared to the degrees of closure as predicted from the ensemble docking using the minima of the regression curves of the Emodel scores (light gray) and the single best Emodel scores (white).

indicate the optimal receptor structure. By contrast, the ensemble generated through NMA consists of 25 perturbed models of the same starting structure (iGluR2–KA). Accordingly, results from docking to such an ensemble would be expected to follow a pattern of improving (i.e. more negative) scoring values until reaching the optimal degree of closure, followed by increasingly poorer (i.e. less negative) scores. However, as evident from the results (Fig. 6), the docking scores fluctuate along this general trend due to variabilities inherent to the pprep step and the docking algorithm itself. Considering this in combination with the relatively large number of ensemble members, it seems inadequate to merely determine the optimal degree of closure from the single best score. Therefore, in the general case of ensemble docking we find it more reasonable to fit a mathematical expression to the scoring values and then identify the preferred receptor conformation by locating the minimum of this function. Although the exact equation (if any) describing the distribution of such data is probably not the 2nd order polynomial in particular, this function was used here as a trend line to locate the minima. R^2 values in the range 0.46–0.82 (Fig. 6) support the use of this method. It is still important, however, to look at the single best score. If this does not lie reasonably close to the minimum of the regression line the validity of the docking experiment should be questioned.

3.4. Correct and incorrect docking poses

As described in Section 2.3, only correct poses were considered when evaluating the docking scores. In principle, if this docking procedure were to be carried out with ligands of unknown binding modes, this rule would not be immediately applicable. Unless clearly violating common physical–chemical sense, every pose would have to be taken into account. However, as for the present study a few points should be made. First of all, as an evaluation of the principle of predicting the preferred receptor conformation through ensemble docking, the results of this study should be based on poses reflecting reality. And secondly, the vast majority of the highest ranked docking poses were correct. Only in the case of KA the lack of correct poses reported in the more closed crystal structures and ensemble models could present a problem. However, even if the binding mode for KA were unknown, all reported incorrect KA poses would undoubtedly be critically evaluated in terms of common physical–chemical sense. As exemplified in Fig. 8, the pyrrolidinium ring in KA was inverted in all incorrect poses, resulting in a pseudo-axial position of the isopropenyl group. Compared to the true binding conformation, the lowest conformational energy of a pseudo-axial conformation with the distal carboxymethyl group oriented as in Fig. 8 is approximately 6 kcal/mol higher (as calculated with MacroModel, MMFFs, and GB/SA solvation). Although not unrealistically high, such an energy difference should cause suspicion, especially when high scoring, low energy poses were found in the more open models.

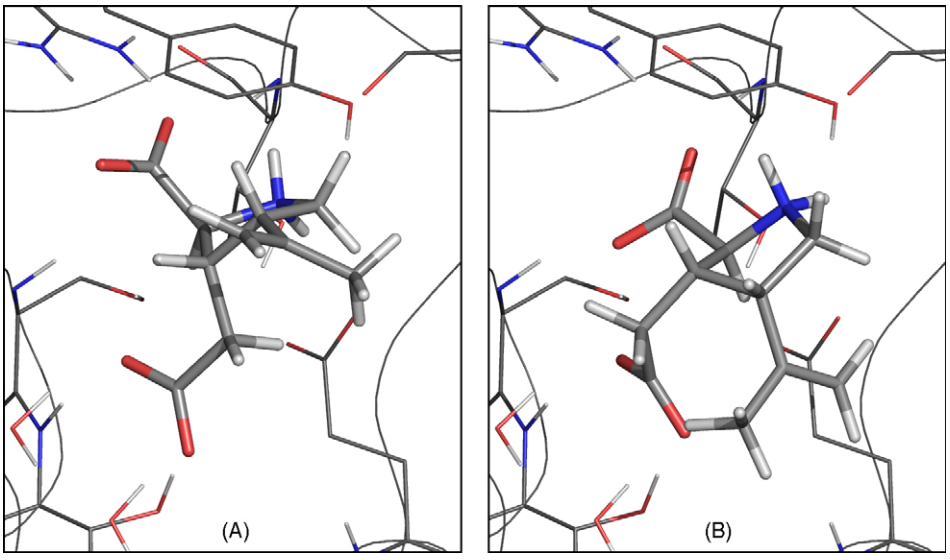


Fig. 8. (A) The correct KA pose ranked third in model 14, with the isopropenyl group in pseudo-equatorial position, compared to (B) the top ranked, incorrect, and high-energy pose in the same model, with a pseudo-axial isopropenyl group. The backbone of model 14 is drawn in ribbon representation; atoms are shown for pocket residues. For clarity, non-polar hydrogens are shown for KA only. Non-carbon coloring as in Fig. 1.

The argument of poorer scoring values accompanied by a change in ligand conformation and/or orientation also applies to Br-HIBO, for which only incorrect poses were reported in models 21, 24, and 25. Seen as isolated cases, these poses were in fact reasonable in terms of both conformation and orientation, if the true binding mode was unknown. However, the sudden change of pose coupled with poorer scoring values would probably lead to the conclusion of these poses being incorrect. With the high accuracy docking protocol followed in this study we find the exclusion of incorrect poses justified, as it would have been possible with critical assessments to identify the incorrect poses, even without prior knowledge of the true binding mode.

3.5. The inclusion of w1 and w2

The two water molecules, w1 and w2, have been identified to be present at the same location in all reported iGluR2 crystal structures [22]. Thus, it was straightforward in this study to decide upon the inclusion of both due to the prior structural knowledge of all ligand–receptor interactions. Preliminary cross-dockings without w1 and/or w2 inside the pocket were also carried out, but results from these (not shown) clearly indicated the necessity of their presence in order to obtain correct results.

In real cases with no such prior experimental knowledge this would not be immediately possible, thus raising the important question whether the methods of this study will be applicable at all in these circumstances. Three points should be stressed in this regard. Firstly, similar to the correct/incorrect pose discussion above, keeping the method validating nature of this study in mind we prioritized setting up a reliable reference system. Secondly, this ligand–receptor system represents a hydrophilic extreme in which water molecules, and hydrogen bonds in general, play a much more important role than would be the case in many other systems. Finally, with a program such as GRID it is possible to accurately predict energetically favorable positions for water molecules within a protein, so that prior experimental knowledge may not be necessary.

We briefly investigated the applicability of GRID for determining the energy of a water probe within the pocket of the iGluR2–Glu structure. Excluding the part of the pocket filled by Glu, four separate positions with interaction energies below ca. −8 kcal/mol were found. Two of those corresponded exactly to w1 and w2 (Fig. 1) and were found by GRID to have interaction energies of −12.2 and −13.2 kcal/mol, respectively. The third favorable position (−9.8 kcal/mol) was found where a water molecule is seen in iGluR2–Glu binding to the side chain oxygen atoms of Y702 and T686 but not directly to Glu. With an interaction energy of −8.7 kcal/mol the final position corresponded to a water molecule forming hydrogen bonds to the backbone N of E705 and the distal carboxylate of Glu in the iGluR2–Glu structure.

The predicted energy differences agree well with the fact that w1 and w2 have never been observed to be displaced as opposed to the other two water molecules. Although the exact energy values cannot provide definite answers to whether a water molecule could be displaced under certain circumstances, calculations like these would provide a strong indication as to whether or not it would be realistic to include a water molecule in the binding pocket during docking.

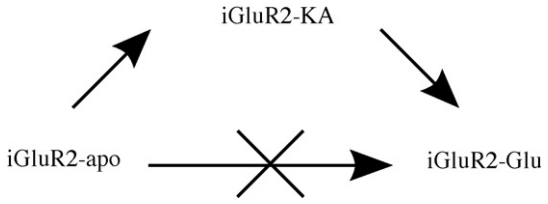


Fig. 9. The partially agonized iGluR2–KA structure does not represent an intermediate in the probably unrealistic direct perturbation of the iGluR2 LBD in the apo (or antagonized) state towards the fully agonized state represented by iGluR2–Glu.

3.6. Rigid-body flexibility varies in different conformations

A potential pitfall in a study like this is the assumption of a direct perturbation with one specific combination of normal modes describing the actual conformational transition, in this case from the semi-open iGluR2–KA to the closed iGluR2–Glu. As reported by Bjerrum and Biggin [33], the overall rigid-body motions of the iGluR2 LBD should be distinguished in terms of which combination of eigenvectors is the main descriptor of a certain motion. Interestingly, they found that the transition between the open apo or antagonist bound structures and the semi-open partial agonist bound structures (e.g. iGluR2–KA) is described by a different combination of eigenvectors than that of iGluR2–KA towards the closed full agonist bound structures. Correspondingly, the partial agonist bound structures are not found as intermediates in a direct NMA perturbation from a fully open to a fully closed conformation (Fig. 9) similar to the one employed here for iGluR2–KA towards iGluR2–Glu. Preliminary attempts to perform a perturbation of iGluR2–apo towards iGluR2–Glu resulted in models with greatly increased RMSD values compared to those shown in Fig. 4 and with significant structural distortions (results not shown). Similar to the observations of Bjerrum and Biggin, this indicated that the transition could not be described realistically by a direct perturbation with one specific combination of normal modes. Therefore, we focused on the iGluR2 structures representing the partially to fully agonized states, excluding the open antagonized or apo structures.

In general, identifying such patterns in the rigid-body flexibility of a protein is not trivial, and it presents a limitation to the methods described here when applying them to a system where only the extremes are known.

4. Conclusions

We have performed a study to establish whether it is possible to predict the ligand induced conformational change of a receptor through high accuracy QM/MM docking of five known agonists to an extensive structural ensemble of the highly flexible iGluR2 receptor generated with normal mode analysis (NMA). Our results show that the docking scores follow a distribution from which such predictions are possible and in good agreement with

the experimentally determined agonist bound iGluR2 structures. Additionally, the computationally inexpensive NMA protocol employed was shown to be a useful method for reliable ensemble generation, needing only two conformationally different input structures. Since large-scale domain movements in a receptor are often coupled to biological function, the ability of the method to facilitate such predictions of ligand induced receptor conformation may be of great value in the early stages of drug discovery projects in cases where only a few experimentally determined receptor structures are available. Although our results are based on a rather detailed system setup, to some extent taking advantage of prior experimental knowledge, we believe that the fundamental ideas of this study can serve as inspiration when dealing with docking to flexible receptors.

Acknowledgements

The authors would like to thank Dr. W. Zheng from the NIH, Bethesda, MD, for providing the *nexe* program as well as for general help with the NMA method. The Carlsberg Foundation and the Lundbeck Foundation are gratefully acknowledged for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.11.006.

References

[1] V. Mohan, A.C. Gibbs, M.D. Cummings, E.P. Jaeger, R.L. DesJarlais, Docking: successes and challenges, Curr. Pharm. Des. 11 (2005) 323–333.
[2] G. Schneider, H.J. Böhm, Virtual screening and fast automated docking methods, Drug Discov. Today 7 (2002) 64–70.
[3] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[4] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy, J. Med. Chem. 47 (2004) 45–55.
[5] X. Barril, S.D. Morley, Unveiling the full potential of flexible receptor docking using multiple crystallographic structures, J. Med. Chem. 48 (2005) 4432–4443.
[6] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, Novel procedure for modeling ligand/receptor induced fit effects, J. Med. Chem. 49 (2006) 534–553.
[7] T.L. Blundell, S. Patel, High-throughput X-ray crystallography for drug discovery, Curr. Opin. Pharmacol. 4 (2004) 490–496.
[8] H.B. Broughton, A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening, J. Mol. Graphics Modell. 18 (2000) 247–257.
[9] H.A. Carlson, K.M. Masukawa, K. Rubins, F.D. Bushman, W.L. Jorgensen, R.D. Lins, J.M. Briggs, J.A. McCammon, Developing a dynamic pharmacophore model for HIV-1 integrase, J. Med. Chem. 43 (2000) 2100–2114.
[10] K.L. Meagher, H.A. Carlson, Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case, J. Am. Chem. Soc. 126 (2004) 13276–13281.
[11] W. Zheng, S. Doniach, A comparative study of motor-protein motions by using a simple elastic-network model, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13253–13258.
[12] C.N. Cavasotto, J.A. Kovacs, R.A. Abagyan, Representing receptor flexibility in ligand docking through relevant normal modes, J. Am. Chem. Soc. 127 (2005) 9632–9640.
[13] N. Floquet, J.D. Marechal, M.A. Badet-Denisot, C.H. Robert, M. Dauchez, D. Perahia, Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors, FEBS Lett. 580 (2006) 5130–5136.
[14] M.L. Mayer, Glutamate receptor ion channels, Curr. Opin. Neurobiol. 15 (2005) 282–288.
[15] N. Armstrong, E. Gouaux, Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core, Neuron 28 (2000) 165–181.
[16] A. Frandsen, D.S. Pickering, B. Vestergaard, C. Kasper, B.B. Nielsen, J.R. Greenwood, G. Campiani, C. Fattorusso, M. Gajhede, A. Schousboe, J.S. Kastrup, Tyr702 is an important determinant of agonist binding and domain closure of the ligand-binding core of GluR2, Mol. Pharmacol. 67 (2005) 703–713.
[17] M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis, Phys. Rev. Lett. 77 (1996) 1905–1908.
[18] K. Hinsen, Analysis of domain motions by approximate normal mode calculations, Proteins 33 (1998) 417–429.
[19] Maestro, Version 7.5, Schrödinger, LLC, New York, NY, 2006.
[20] A. Hogner, J.S. Kastrup, R. Jin, T. Liljefors, M.L. Mayer, J. Egebjerg, I.K. Larsen, E. Gouaux, Structural basis for AMPA receptor activation and ligand selectivity: crystal structures of five agonist complexes with the GluR2 ligand-binding core, J. Mol. Biol. 322 (2002) 93–109.
[21] R. Jin, T.G. Banke, M.L. Mayer, S.F. Traynelis, E. Gouaux, Structural basis for partial agonist action at ionotropic glutamate receptors, Nat. Neurosci. 6 (2003) 803–810.
[22] U. Pentikainen, L. Settimo, M.S. Johnson, O.T. Pentikainen, Subtype selectivity and flexibility of ionotropic glutamate receptors upon antagonist ligand binding, Org. Biomol. Chem. 4 (2006) 1058–1070.
[23] Glide, Version 4.0, Schrödinger, LLC, New York, NY, 2005.
[24] 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, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
[25] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, J. Med. Chem. 47 (2004) 1750–1759.
[26] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[27] MacroModel, Version 9.1, Schrödinger, LLC, New York, NY, 2005.
[28] AD-ENM: The Elastic Network Model Server. http://enm.lobos.nih.gov/ (accessed April 23, 2007).
[29] Schrödinger Suite 2006 QM-Polarized Ligand Docking Protocol; Glide Version 4.0, Schrödinger, LLC, New York, NY, 2005; Jaguar Version 6.5, Schrödinger, LLC, New York, NY, 2005; QSite Version 4.0, Schrödinger, LLC, New York, NY, 2005, 2006.
[30] GRID, Version 22a, Molecular Discovery, Ltd., Pinner, Middlesex, 2005.
[31] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[32] The PyMOL Molecular Graphics System, Version 0.99, DeLano Scientific, Palo Alto, CA, USA, 2002.
[33] E.J. Bjerrum, P.C. Biggin, Rigid-body essential X-ray crystallography: distinguishing the bend and twist of glutamate receptor ligand-binding domains, Proteins, in press.
[34] Y. Arinaminpathy, M.S. Sansom, P.C. Biggin, Molecular dynamics simulations of the ligand-binding domain of the ionotropic glutamate receptor GluR2, Biophys. J. 82 (2002) 676–683.
[35] Y. Arinaminpathy, M.S. Sansom, P.C. Biggin, Binding site flexibility: molecular simulation of partial and full agonists within a glutamate receptor, Mol. Pharmacol. 69 (2006) 11–18.