# Pushing the Limits of Computational Structure-Based Drug Design with a Cryo-EM Structure: the Ca$^{+2}$ Channel α2δ-1 Subunit as a Test Case

*Martin Kotev, ξ Rosalia Pascual, ζ Carmen Almansa, ζ Victor Guallar, ⊥,Π,, Robert Soliva ξ**

ξ Nostrum Biodiscovery, Jordi Girona 29, Nexus II D128, 08034 Barcelona, Spain.

ζ ESTEVE, Drug Discovery and Preclinical Development, Carrer Baldiri Reixac, 4-8. Parc Científic de Barcelona, 08028 Barcelona , Spain.

⊥ Barcelona Supercomputing Center (BSC), Jordi Girona 29, E-08034, Barcelona, Spain.

Π ICREA, Passeig Lluís Companys 23, E-08010, Barcelona, Spain.

1

ABSTRACT. Cryo electron microscopy (cryo-EM) is emerging as a real alternative for structural elucidation. In spite of this, very few cryo-EM structures have been described so far as successful platforms for *in silico* drug design. Gabapentin and pregabalin are some of the most successful drugs in the treatment of epilepsy and neuropathic pain. Although both are in clinical use and are known to exert their effects by binding to the regulatory α2δ subunit of voltage gated calcium channels, their binding modes have never been characterized. We describe here the successful use of an exhaustive protein-ligand sampling algorithm on the α2δ-1 subunit of the recently published cryo-EM structure, with the goal of characterizing the ligand entry path and binding mode for gabapentin, pregabalin and several other aminoacidic α2δ-1 ligands. Our studies indicate that: i) all simulated drugs explore the same path for accessing the occluded binding site on the interior of the α2δ-1 subunit; ii) they all roughly occupy the same pocket; iii), the plasticity of the binding site allows the accommodation of a variety of aminoacidic modulators, driven by the flexible "capping loop" delineated by residues Tyr426-Val435 and the floppy nature of Arg217; iv) the predicted binding modes are in line with previously available mutagenesis data, confirming Arg217 as key for binding, with Asp428 and Asp467 highlighted as additional anchoring points for all amino acidic drugs. The study is one of the first proofs that latest-generation cryo-EM structures combined with exhaustive computational methods can be exploited in early drug discovery.

**Introduction**

The Nobel Prize in Chemistry 2017 was recently awarded to Jacques Dubochet, Joachim Frank and Richard Henderson for the development of technologies that were key in propelling cryo-electron microscopy (cryo-EM).[1] This technique has been evolving over decades, moving from low-resolution structural determination of huge macromolecular complexes such as viral capsids[2] or the ribosome,[3] to the near-atomic elucidation of small enzymes such as β-galactosidase.[4] The latest successful applications unveil the structures of proteins with molecular weights as low as 64 KDa[5] and, in extreme cases, with resolutions as low as a remarkable 1.8 angstroms. The first cryo-EM structures of receptor-ligand complexes have started to emerge. Some notorious examples of complexes of pharmaceutical relevance recently published are: the structure of substrate-bound human telomerase[6], the ATPase P97 in complex with a small-molecule inhibitor solved at 2.4 Å resolution,[7] γ-secretase in complex with a dipeptide inhibitor solved at 4.2 Å,[8] the TRPV1 ion channel in complex with resinferatoxin at 2.9 Å resolution[9] and isocitrate dehydrogenase bound to ML309 at 3.8 Å.[10] Cryo-EM structures have also been used to guide the design of selective *Plasmodium falciparum* proteasome inhibitors[11] and to determine structural differences between *Saccharomyces cerevisae* and *Arabidopsis thaliana* IGPD enzyme to drive inhibitor selectivity[12].

Since many complexes have resolutions between 2.5 and 5 Å, the technique is clearly approaching an accuracy that enables rational drug design. It is reasonable to argue that cryo-EM will soon become a complement to X-ray and NMR in the structural elucidation of drug targets.[13,14,15] However, the latest advances have taken place so recently that there are very few examples where cryo-EM structures have been used as a platform for computer-assisted molecular design.

The voltage-gated calcium channel Cav1.1 is one of the latest multiprotein complexes solved by cryo-EM at a near-atomic resolution (3.6 Å).[16] Voltage-gated calcium channels comprise several subunits: the α1 pore-forming subunit which is encoded by 10 different genes giving rise to different physiological classes based on their voltages of activation; the intracellular β subunit with four different subtypes, the transmembrane γ subunit encoded by eight different genes and the α2δ subunit for which four genes have been found. The α2δ subunit is located extracellularly and its structure comprises four cache domains and one VWA domain, and consists of two proteolytically cleaved peptides (α2 and δ), which are stabilized by multiple disulfide bridges.

Pregabalin (**1**) and gabapentin (**2**) are two approved drugs currently in clinical practice for treating a range of conditions such as epilepsy, neuropathic pain, fibromyalgia or restless leg syndrome (Figure 1). Despite being alkylated derivatives of the neurotransmitter GABA (gamma amino-butyric acid), the lack of interaction with GABAA and GABAB receptors has been demonstrated by receptor-binding and electrophysiological assays.[17] Instead, they have proven to be high affinity selective ligands for the α2δ-1 and α2δ-2 subunits of the calcium channels.[18,19,20] Further, in vivo experiments with genetically modified mice containing one aminoacid mutation in the α2δ-1 subunit, which significantly diminishes binding of gabapentin, have confirmed the

implication of this subunit in the therapeutic effect of the drugs.[21,22] However, the binding mode of any of the reported α2δ-1 ligands has not been properly characterized and in view of the recent publication of the cryo-EM structure of Cav1.1 including the α2δ-1 subunit, we decided to approach the in silico characterization of its binding site and the possible ligand entry path.

Computational structure-based binding mode prediction usually relies on small molecule docking to experimentally solved receptor structures,[23,24] in the majority of cases by X-ray crystallography or NMR. Often, receptor flexibility is neglected in docking calculations, leading in some cases to incorrect *in silico* generated pose predictions. In low-throughput regimes, receptor flexibility may be taken into account. Most methods achieve this by allowing conformational changes of a defined set of aminoacids in the active site, by truncating them to alanine, or by using a softened potential in the docking step, followed in all cases by a refinement, simulated annealing and energy minimization once the compound has been docked.[25] An alternative to emulate the flexibility is docking into an ensemble of protein conformations that can be obtained from a previously run molecular dynamics (MD) simulation of the receptor or from a set of crystallized structures. Mixed methods including both approaches have also been developed.[26]

However, complex formation from its initial to its final phase is rarely explored. In fact, the complete simulation of a binding event has been accessed only recently and thanks to the design of specialized supercomputers,[27] which have allowed to successfully reproduce a whole binding event in explicit solvent MD simulations.[28] However, these calculations cannot be routinely applied in most drug discovery settings. PELE[29] is an alternative method of exhaustive sampling that allows simulating the whole binding event. It has been used in a wide range of recognition problems such as the exploration of binding mechanisms for nuclear hormone receptor ligands,[30]

the computational prediction of resistance to HIV-1 protease inhibitors[31] or pose prediction for low-molecular-weight fragments,[32] and has been tested as an induced-fit docking algorithm for prospective binding mode prediction on a variety of targets with remarkable results, both in academic and industrial settings.[33,34]

In this paper, we systematically investigate the recognition properties of the α2δ-1 subunit. We fully explore with a variety of computational techniques the binding pathways and binding modes of three α2δ-1 modulators pregabalin (**1**), gabapentin (**2**) and the proline derivative **3**, which reached phase I for the treatment of neuropathic pain.[35] We characterize the plasticity of this target and highlight the motions needed for its ligands to access the binding site. Additionally, we use the structures generated by the simulations to rationalize SAR changes in other disclosed α2δ-1 modulators. Overall, this study is one of the first proofs that cryo-EM structures of near-atomic resolution can be utilized, when coupled to "state of the art" computational techniques, in a computer-assisted medicinal chemistry program.

## Results and discussion

### Overall flexibility of the putative binding site

Firstly, the overall flexibility of the α2δ-1 subunit was studied by means of MD simulations in explicit solvent. The homology-built model of α2δ-1 was simulated for 0.6 μs in explicit solvent in the absence of any ligands (apo form). Particular attention was given to the flexibility of three segments whose deletion caused a significant loss in binding of gabapentin: Thr205–Gly223 (TPNKIDLYDVRR**R**PWYIQG, noted here as segment I, with Arg217 in bold), Thr515–Gly538 (TLDFLDAELENDIKVEIRNKMIDG, segment II) and Thr582-Gly604 (TYSFYYIKAKIEETITQARSKKG, segment III). Attention was focused as well on the first

segment residue Arg217, which was the only point mutation out of the six tested from the three different fragments that caused gabapentin to substantially lose its affinity.[19,20]

Once the simulation reached a geometric equilibrium in terms of Root Mean Square deviation (RMSd), a series of conformations were clustered and a representative of each cluster was visualized. As can be seen in Figure 2, segments I, II and III are far apart from one another. Segment I is linear, mostly unstructured and highly flexible, not forming any stable secondary structure with the only exception of its N-terminal end, where the key Arg217 residue is found at the beginning of an alpha helix. The flexibility of this segment is somewhat surprising as it runs towards the interior of the protein. Segment II forms two alpha helices on the back of segment I and seems to be rather fixed, and segment III forms a series of beta sheets located far from Arg217, on the opposite site of the structure from segment I. Most probably these two segments are important for gabapentin binding because they guarantee the structural integrity of the whole domain. A detailed inspection of the flexibility of Arg217 (Figure 2, top right), located between the Cache1 and the VWA domains, reveals it is highly mobile. Its side chain is dynamically H-bonding to amino acids located in segment I, in an extension of segment II that runs parallel to segment I, and also in the "capping loop" made up of residues Tyr426-V435; by doing so, it radically changes the overall topology of the observed adjacent cavity. Detailed analysis also reveals that there are two negatively charged residues in the vicinity of Arg217, namely Asp467 and Asp428. The former seems to be quite rigid, as it is found in the middle of a beta sheet running through the core of the protein. The latter, in contrast, is located in the capping loop, which is a fairly mobile structure acting as a cap on top of the cavity next to Arg217. It seems to be able to engage Arg217, as well as adopting other conformations by which it H-bonds to Trp181 and also intramolecular H-bonds to the backbone NH of other loop residues. The

choreography of these three charged residues was early on taken as a potential 3-point anchoring site that could be exploited by ligands, considering many inhibitors bear positively and negatively charged groups (Figure 1) and the key role of Arg217 for gabapentin binding as identified by the mutagenesis experiments. Accordingly, molecular interaction potentials were calculated with the original cryo-EM configuration, to check for the volume and reactivity of this cavity. As expected, Arg217 generates a strong potential for a negative probe. Importantly, Asp467 and Asp428 generate a strong potential energy well for a positive probe next to the latter (Figure 2, bottom). Overall, the cavity seems rather small but highly flexible, capable of binding a variety of modulators.

**Molecular recognition of gabapentin, pregabalin and compound 3 by the α2δ-1 subunit**

As a second step, and taking into account the results from the apo MD simulations and interaction potentials, we simulated how the two approved drugs, gabapentin and pregabalin as well as compound **3** bind their target.

Initially, gabapentin and compound **3** were docked into the original cryo-EM structure of the α2δ-1 subunit and into the human homology-built model. For gabapentin, a series of docking poses were generated both next to Arg217 and also in a variety of other pockets. Compound **3** could be docked next to Arg217 only when the van der Waals potentials were significantly softened and the search area was restricted to the small cavity detected.

In order not to bias the binding modes generated by restricting rigid docking to some particular area or anchoring point, it was decided to run full exhaustive induced-fit docking simulations

with PELE with the three studied compounds starting in all cases from an unbound initial pose, on the surface of the subunit.

Gabapentin was initially placed on different random positions at the surface of the α2δ-1 subunit, with its center of mass approximately 25 Å away from the reference point, which was set next to Arg217, due to its proven importance for ligand binding. These geometries were used as initial structures for an unbiased dynamic exploration of ligand-receptor interactions with PELE. As seen in Figure 3, although many independent simulations of gabapentin end up locating the ligand in a variety of binding cavities on the surface of the α2δ-1 subunit, one of them finds its way into the expected binding site, ending at distances ca. 2 Å to the reference point. Interestingly, this trajectory reaches also in this location the most favorable protein-ligand interaction energies (around -80 kcal/mol). Monitoring how this copy accesses the binding site suggests the ligand entry path for gabapentin. The access seems to take place primarily by traversing a tunnel formed by segment I, the N-terminal extension of segment II and the capping loop, together with a gating mechanism enabled by the movement of two residues, namely Tyr212 and Arg217. Gabapentin travels the length of segment I from its most solvent-exposed side through to its interior until it reaches Tyr212, which must first move away to enable the drug to approach Arg217 (Figure 4). The latter must in turn move away from the entry path to allow gabapentin's entry, but closes in again on the drug, engaging its carboxylic acid in a double H-bond.

The same process was repeated for pregabalin and in this case the maximum distance to the reference point was set to ca. 23 Å. As depicted in Figure 3, the dynamic exploration of α2δ-1 for pregabalin discovered several poses, but in contrast to Gabapentin, the lowest one is not located next to Arg217. Our simulations find 5 different low interaction energy poses at different

distances to the reference point. This can be explained by the more flexible nature of pregabalin, which facilitates visiting several superficial pockets. A close inspection of all these low interaction poses reveals that none of them coincides with any of the low-energy interaction poses found for gabapentin with the exception of the pose found at the putative binding site, where it adopts a binding mode highly similar to the one found for gabapentin (Figure 4). The entry of pregabalin is again facilitated by the movement of Tyr212 and Arg217 but in inverted order. The binding mode found for the lowest-energy PELE trajectory features the pregabalin carboxylic acid engaging Arg217 and the backbone NH of Ala429 located in the capping loop. The ammonium group of pregabalin establishes two direct salt bridges with the carboxylates of Asp428 and Asp467 by a double salt bridge interaction and the isopropyl group is located in the same environment as the cyclohexyl of gabapentin. However, the H-bonds seen with indole N of Trp219 and the hydroxyl oxygen of Tyr212 in gabapentin are missing for pregabalin, explaining the less favorable interaction energy predicted (ca. 70kcal/mol). These two missing interactions could be a result of an incomplete sampling in our simulations, which could be overcome by running them in thousands of cores instead of hundreds.

Additionally, the same protocol was followed for compound **3**, a proline derivative that has a bulkier hydrophobic substituent than gabapentin or pregabalin (Figure 3). As in the previous cases, 3 finds its way into the deep cavity in the interior of the subunit only in a few simulations, but the few copies that reach the binding site are estimated to have the lowest interaction energies with the receptor. In fact, the profile for **3** is even sharper than that of gabapentin, being the pose inside the putative binding site next to Arg217 clearly the one with the most favorable interaction energies (below -100 kcal/mol). The entry path again involves a route along segment

I, and the final engagement requiring movement of Tyr212 and Arg217 in a similar dynamic movement as the one seen for gabapentin.

The poses for gabapentin and the proline derivative **3** were subsequently optimized by an explicitly solvated 0.6 μs MD simulation, from which an average structure was derived and energy-minimized (Figures 4 and 5). In the MD average structure, the gabapentin pose is defined by 8 electrostatic interaction points with the α2δ-1 subunit. Its carboxylic acid moiety is forming a double salt bridge with Arg217 as well as a H-bond with the indole nitrogen of Trp219 and the hydroxyl oxygen of the gating Tyr212, all of them located in segment I. An additional H-bond is formed with backbone NH of Ala429, located in the capping loop, which seems to be important for keeping it in a closed conformation. On the other side of the molecule, its amine group is engaging the protein with a double salt-bridge interaction with Asp428 and Asp467, the two negatively charged groups highlighted in the preliminary analysis, as well as an additional H-bond to Tyr426. Thus, the two ionized groups of gabapentin engage the protein by up to 8 electrostatic interactions with as many as 7 different amino acids inside the deep pocket, while its cyclohexyl moiety is accommodated in a markedly hydrophobic pocket, lined by aromatic residues Trp181, Trp199, Trp219, Tyr193 and aliphatic residues Val183, Leu430 (the latter located at the tip of the capping loop).

In the case of the MD average structure for compound **3**, the same electrostatic interactions to Arg217, Trp219 and Tyr212 are found for its carboxylic acid moiety. Its amine is also anchored in a double salt bridge to Asp428 and Asp467. However, because compound **3** features a bulkier hydrophobic group, the binding site ceiling has to undergo a slight induced fit to accommodate it. This is primarily achieved by virtue of the movement of residues Trp199 and His143, and by a slight arrangement of the capping loop (see Figure 5).

The MD average structure between α2δ-1 and gabapentin was used as a template for classical docking (rigid receptor) calculations with other aminoacidic derivatives of interest, among them compounds **4** and **5** (Figure 1). In an exhaustive structure-activity study on alpha amino acid ligands, Abbott scientists found compounds **4** and **5,** displaying sharp affinity variations depending on compound chirality.[36] The S enantiomer of compound **4** had a α2δ-1 affinity $K_i$ of 180 nM, whereas its R enantiomer had a Ki of 30 μM, ca. 160-fold lower. Similarly, the S enantiomer of compound **5** was found to have an affinity Ki of 90 nM, whereas its R enantiomer had a Ki >100 μM. Inspection of the binding mode of **4** allows rationalizing the difference (Figure 5). The S enantiomer can engage Arg217 and at the same time Asp428 and Asp467. However, the R enantiomer can only engage Arg217 and Asp467, its amine pointing to the interior of the cavity, away from Asp428 located in the capping loop. The same results were found for compound **5** (data not shown). This suggests that α2δ-1 modulators with an ionized secondary amine, as it is the case for many amino acidic compounds, probably need to properly engage both Asp side chains for good potency. Thus, the optimized structures of these complexes captured a level of detail that could be qualitatively used in a medicinal chemistry program, even though the resolution of the original structure (3.6 Å) was not optimal. It is expected that novel developments in the cryo-EM field, such as new computational protocols to construct atomic models from the density maps to glimpse fine details on the side chains of amino acids[37] (location, protonation states, etc.) will warrant further cases of cryo-EM and *in silico* synergies.

**Conclusions**

The work presented here is one of the first proofs that a latest-generation near-atomic resolution cryo-EM structure of a huge macromolecular complex can be exploited in

computational structure-based drug discovery, in spite of having a modest resolution (3.6 Å). Our exhaustive protein-ligand sampling algorithm has been applied to the α2δ-1 subunit of the calcium channel (together with classical MD simulations), to propose a plausible entry path and final binding mode of several validated modulators. Our simulations provide for the first time a plausible entry path for several aminoacidic structures into the α2δ-1 subunit, suggesting that all these chemical entities access the binding site in a similar manner and bind in the same spot, in the interior of the subunit. Notably, PELE's sampling capabilities properly capture the different induced-fit effects driven by the different chemical entities, successfully reproducing the transitions taking place in the binding events. The simulations are in line with what is known from mutagenesis data, pointing to Arg217 as a crucial interaction point. In addition, they highlight other residues as main anchoring points, in particular Asp428 and Asp467, which H-bond to modulators featuring positively charged groups.

Overall, this study is an example that solved cryo-EM structures can be coupled to computational techniques capable of properly describing protein flexibility and eventually be exploited in the computational study of receptor-ligand interactions at atomic detail.

**Experimental Section**

**Homology modeling**. The homology model of human α2δ-1 (http://www.uniprot.org/uniprot/P54289) was built out of a pairwise sequence alignment with the rabbit counterpart (http://www.uniprot.org/uniprot/P13806) with the SWISS-MODEL.[38] The structural template was downloaded from the Protein Data Bank (PDB ID: 5gjv).[16] It must be borne in mind that not only the resolution of this structure is modest but its overall quality is rather low. Thus, a quality check run on the Molprobity server[39] with PDB entry 5gjv reveals the

number of clashes, Ramachandran and sidechain outliers places this structure in a rather low percentile relative to all structures in its archive (see Supporting Information Table S3), which made the modeling even more challenging (for an in-depth analysis of the structure, see https://files.rcsb.org/pub/pdb/validation_reports/gj/5gjv/5gjv_full_validation.pdf). Our simulations partly corrected the deficiencies, as can be seen in Supporting Information Table S4, for our α2δ-1-Gabapentin complex. In spite of this, the generated human α2δ-1 model was considered highly reliable with respect to its template, as the overall % identity between rabbit and human was found to be 97, that is, their sequences are nearly identical. Further, the % identity is 100% for all residues found within a 10 Å radius of the Arg217 residue, considered crucial for binding in view of mutagenesis experiments.

It is noteworthy that there are differences in the literature regarding amino acid numberings of α2δ-1 subunit due to species differences and also to the renumbering after the removal of the signal peptide in the generation of the mutated mouse. A table and a sequence alignment, where the amino acids cited in the present paper and in many of the references discussed herein are matched to the same residue as found in the Uniprot entry for human α2δ-1, P54289, has been included as Supporting Information Table S1 and S2.

**Protein preparation**. All MD calculations for the α2δ-1 subunit were started from the built homology model of human α2δ-1. The PELE simulations were run on the original α2δ-1 subunit of the cryo-EM pdb entry 5GJV with all missing loops built with Prime[40], which is an algorithm for de novo loop generation. It first builds many solutions by using a dihedral angle-based buildup procedure. The generated loops are clustered, their side-chains are geometry-optimized and subsequently energy-minimized. The algorithm was thoroughly tested with a total of 833 loops and it was found to be quite successful with loops of up to 11 residues. The only exception

for the α2δ-1 model was the long Pro910-Leu969 loop, which was not modeled via Prime and capped at the ends with NME and ACE groups. The structures were refined using the Protein Preparation Wizard[41] of Maestro[42], adding hydrogen atoms, checking the protonation states of ionizable amino acids (at physiological pH, ca. 7.4) and optimizing His, Asn and Gln side chains. The resulting structures were checked by visual inspection.

**Small molecule preparation**. All small molecules modeled in this work were first processed with Ligprep[43], which generates 3D energy-minimized molecular structures and expands tautomeric and ionization states. As expected, all carboxylic and amine groups in molecules **1-5** were predicted to be ionized at physiological pH (7.4), that is, **1-5** were modeled as zwitterions. Gabapentin, pregabalin and molecule **3** were modeled in PELE with OPLS 2005[44] parameters with RESP charges calculated at HF/6-31G**. The latter were also used for gabapentin and compound **3** in the MD simulations and their parameters (including dihedrals) were taken from GAFF[45].

**PELE simulations**. Compounds **1**-**3**, were studied in detail with exhaustive PELE simulations. For this purpose, all compounds were first randomly placed on the surface of the subunit, with a maximum distance of approximately 23-25 Å away from the point taken as reference (see below). PELE is a Monte Carlo (MC) algorithm that uses protein structure prediction techniques.[29] Its algorithm works as follows: each MC move consists of three main steps: i) ligand and protein perturbation; ii) side chain rotamer sampling (all side chains for residues with at least one atom within 5 Å of the ligand's heavy atoms were sampled here); and iii) system energy minimization. The ligand is perturbed by way of rotation and translation. The protein is perturbed based on an energy minimization with constrained displacements along the Cα-atoms following a set of given modes, usually derived from an anisotropic network model (ANM).[46,47]

The resulting structure obtained in the final energy minimization is accepted or rejected by applying a Metropolis criterion. The all-atom force field used is OPLS-2005[48] with addition of the continuum solvation model by Onufriev-Bashford-Case[49]. Since simulating the formation of a protein-ligand complex from its initial to its final state is a high-dimensional problem, exhaustive sampling becomes key. The exhaustiveness is tackled within PELE by running a huge number of independent trajectories in parallel (each running on a single core) with the algorithm explained previously. The calculations described here were mostly run on 256 or even 512 cores, typically lasting 10-15 hours each on a computer cluster using Intel Xeon Gold 6140 CPU processors.

The reference point in all PELE simulations was taken to be the midpoint between atom CD2 of residues Trp199 and Trp219, which was taken to be approximately the center of the cavity next to where Arg217 is located in the cryo-EM of the α2δ-1 subunit. Thus, in order to keep track of the location of all molecules during the simulated trajectories, the distance between the molecule's center of mass and this reference point was monitored (Figure 3).

**Molecular dynamics simulations**. MD simulations of the apo homology model of α2δ-1, and α2δ-1 in complex with gabapentin and compound **3** were carried out in explicit solvent. For the α2δ-1-ligand complexes (gabapentin and compound **3**), their starting structures were generated as follows: both compounds were manually placed inside the cavity of the human homology model according to the PELE generated poses and carefully optimized with a series of energy minimizations and low temperature short simulations which were gradually increased up to 300K. The simulations were carried out with NAMD 2.12[50] with the AMBER ff99sb force field.[51] Explicitly solvated (TIP3P waters[52]) systems were simulated in the NPT ensemble in a truncated octahedron box. The systems were brought to neutrality with the addition of sodium

and chloride ions, which were added to simulate physiological ionic strength. SHAKE[53] was used to constrain chemical bonds, which allowed the use an integration step of 2fs. Total production runs of 600 ns were collected, with snapshots saved every 20 ps. All the analyses from the simulations were carried out with CPPTRAJ.[54] The clustering and extraction of snapshots for the simulations was performed by a hierarchical agglomerative algorithm and average-linkage method based on C pair-wise distances. After clustering, a representative snapshot from the first 10 most populated clusters were taken for each of the two simulations and used for building an ensemble. The MD averaged structures of the α2δ-1-gabapentin and α2δ-1-compound **3** complexes were obtained by a two-step procedure: i) soft restrained minimization (2 kcal/mol constant) of the Cartesian average of snapshots collected over an RMSd equilibrated portion of the trajectories (last 200ns.) and ii) unrestrained minimization of all hydrogen atoms.

**Docking calculations**. All docking calculations on rigid α2δ-1 structures (the original cryo-EM structure and also the MD average structure of the gabapentin-α2δ-1 subunit complex) were carried out with Glide[55] using its Standard Precision scoring function (SP). The two enantiomers of each of compounds **4, 5** were docked on the geometry-optimized MD average structure of the α2δ-1 subunit when bound to gabapentin.

**Supporting Information**

A table and a sequence alignment, where the amino acids cited in the present paper and in many of the references discussed herein are matched to the same residue as found in the Uniprot entry for human α2δ-1, P54289, has been included as Supporting Information Table S1 and S2. Information on the quality of the original cryo-EM structure studied and one of the MD optimized structures is also provided as Supporting Information Tables S3 and S4 respectively.

**Corresponding Author Information**

* R.S: email: [rsoliva@nostrumbiodiscovery.com](mailto:rsoliva@nostrumbiodiscovery.com); telephone. 00346856257

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Abbreviations used**

MD, molecular dynamics. MC, Monte Carlo

**References**

(1) Cheng, Y.; Glaeser, R. M.; Nogales E. How Cryo-EM Became So Hot. *Cell* **2017**, *171*, 1229-1231.

(2) Yu, X.; Jin, L.; Zhou, Z. H. 3.88 Å Structure of Cytoplasmic Polyhedrosis Virus by Cryo-Electron Microscopy. *Nature* **2008**, *453*, 415-419.

(3) Rawat, U. B.; Zavialov, A. V.; Sengupta, J.; Valle, M.; Grassucci, R. A.; Linde, J.; Vestergaard, B.; Ehrenberg, M.; Frank, J. A Cryo-Electron Microscopic Study of Ribosome-Bound Termination Factor RF2. *Nature* **2003**, *421*, 87-90.

(4) Bartesaghi, A.; Aguerrebere, C.; Falconieri, V.; Banerjee, S.; Earl, L. A.; Zhu, X.; Grigorieff, N.; Milne, J. L. S.; Sapiro, G.; Wu, X.; Subramaniam, S. Atomic Resolution Cryo-EM Structure of β-Galactosidase. *Structure* **2018**, *26*, 1-9.

(5) Khoshouei, M.; Radjainia, M.; Baumeister, W.; Danev, R. Cryo-EM Structure of Haemoglobin at 3.2 Å Determined with the Volta Phase Plate. *Nat. Commun.* **2017**, *8*, 16099.

(6) Nguyen, T. H. D.; Tam, J.; Wu, R. A.; Greber, B. J.; Toso, D.; Nogales, E.; Collins, K. Cryo-EM Structure of Substrate-Bound Human Telomerase Holoenzyme. *Nature* **2018**, *557*, 1-6.

(7) Banerjee, S.; Bartesaghi, A.; Merk, A.; Rao, P.; Bulfer, S. L.; Yan, Y.; Green, N.; Mroczkowski, B.; Neitz, R. J.; Wipf, P.; Falconieri, V.; Deshaies, R. J.; Milne, J. L.; Huryn, D.; Arkin, M.; Subramaniam, S. 2.3 Å Resolution Cryo-EM Structure of Human p97 and Mechanism of Allosteric Inhibition. *Science* **2016**, *351*, 871-875.

(8) Bai, X. C.; Rajendra, E.; Yang, G.; Shi, Y.; Scheres, S. H. Sampling the Conformational Space of the Catalytic Subunit of Human γ-Secretase. *eLife* **2015**, *4*, e11182.

(9) Gao, Y.; Cao, E.; Julius, D; Cheng, Y. TRPV1 Structures in Nanodiscs Reveal Mechanisms of Ligand and Lipid Action. *Nature* **2016**, *534*, 347-351.

(10) Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug discovery. *Cell* **2016**, *165*, 1698-1707.

(11) Li, H.; O'Donoghue, A. J.; Van der Linden, W. A.; Xie, S. C.; Yoo, E.; Foe, I. T.; Tilley, L.; Craik, C. S.; Da Fonseca, P. C.; Bogyo, M. Structure- and Function-Based Design of Plasmodium-Selective Proteasome Inhibitors. *Nature* **2016**, *530*, 233-236.

(12) Rawson, S.; Bisson, C.; Hurdiss, D. L.; Fazal, A.; McPhillie, M. J.; Sedelnikova, S. E.; Baker, P. J.; Rice, D. W.; Muench, S. P. Elucidating the Structural Basis for Differing Enzyme Inhibitor Potency by Cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 1795-1800.

(13) Boland, A.; Chang, L.; Barford, D. The Potential of Cryo-Electron Microscopy for Structure-Based Drug Design. *Essays Biochem*. **2017**, *61*, 543-560.

(14) Merino, F.; Rauser, S.; Electron Cryo-microscopy as a Tool for Structure-Based Drug Development. *Angew.Chem. Int. Ed Engl*. **2017**, *56*, 2846-2860.

(15) Subramaniam, S.; Earl L. A.; Falconieri, V.; Milne, J. L.; Egelman, E. H. Resolution Advances in Cryo-EM Enable Application to Drug Discovery. *Curr. Opin. Struct. Biol.* **2016**, *41*, 194-202.

(16) Wu, J.; Yan, Z.; Li, Z.; Qian, X.; Lu, S.; Dong, M.; Zhou, Q.; Yan, N. Structure of the Voltage-Gated Calcium Channel Ca(v)1.1 at 3.6 Å Resolution. *Nature* **2016**, *537*, 191-196.

(17) Li, Z.; Taylor, C.; Weber, M.; Piechan, J.; Prior, F.; Bian, F.; Cui, C.; Hoffman, D.; Donevan S. Pregabalin is a Potent and Selective Ligand for α2δ-1 and α2δ-2 Calcium Channel Subunits. *Eur. J. Pharmacol.* **2011**, *667*, 80-90.

(18) Taylor, C. P.; Angelottib, T.; Faumanc, E. Pharmacology and Mechanism of Action of Pregabalin: The calcium Channel Alpha2-Delta Subunit as a Target for Antiepileptic Drug Discovery. *Epilepsy Res*. **2007**, *73*, 137-150.

(19) Offord, J.; Isom, L. L. Drugging the Undruggable: Gabapentin, Pregabalin and the Calcium Channel **α2δ** Subunit. *Crit. Rev. Biochem. Mol. Biol.* **2015**, *51*, 246-256.

(20) Wang, M.; Offord, J.; Oxender, D. L.; Su, T.-Z. Structural Requirement of the Calcium-Channel Subunit α2δ for Gabapentin Binding. *Biochem. J.* **1999**, *342*, 313-320.

(21) Field, M. J.; Cox, P. J.; Stott, E.; Melrose, H.; Offord, J.; Su, T.-Z.; Bramwell, S.; Corradini, L.; England, S.; Winks, J.; Kinloch, R. A.; Hendrich, J.; Dolphin, A. C.; Webb, T.; Williams, D. Identification of the α2δ-1 Subunit of Voltage-Dependent Calcium Channels as a Molecular Target for Pain Mediating the Analgesic Actions of Pregabalin. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17537-17542.

(22) Lotarski, S. M.; Donevan, S.; El-Kattan, A.; Osgood, S.; Poe, J.; Taylor, C. P.; Offord, J. Anxiolytic-Like Activity of Pregabalin in the Vogel Conflict Test in α2δ-1 (R217A) and α2δ-2 (R279A) Mouse Mutants. *J. Pharmacol. Exp. Ther*. **2011**, *338*, 615-621.

(23) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: the Prediction Accuracy of Sampling Power and Scoring Power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964-12975.

(24) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model*. **2009**, *49*, 1455-1474.

(25) Davis, I. W.; Baker, D. RosettaLigand Docking with Full Ligand and Receptor Flexibility. *J. Mol. Biol.* **2009**, *385*, 381-392.

(26) Ding, F.; Dokholyan, N. V. Incorporating Backbone Flexibility in MedusaDock Improves Ligand-Binding Pose Prediction in the CSAR2011 Docking Benchmark. *J. Chem. Inf. Model.* **2013**, *53*, 1871-1879.

(27) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes,

M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *I$^{st}$ Communications of the ACM* **2008**, *51*, 91-97.

(28) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror R. O.; Seeliger M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc*. **2011**, *133*, 9181-9183.

(29) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **2005**, *1*, 1304-1311.

(30) Grebner, C.; Lecina, D.; Gil, V.; Ulander, J.; Hansson, P.; Dellsen, A.; Tyrchan, C; Edman, K.; Hogner, A.; Guallar, V. Exploring Binding Mechanisms in Nuclear Hormone Receptors by Monte Carlo and X-Ray-Derived Motions. *Biophys J.* **2017**, *112*, 1147-1156.

(31) Hosseini, A.; Alibés, A.; Noguera-Julian, M.; Gil, V.; Paredes, R.; Soliva, R.; Orozco, M.; Guallar, V. Computational Prediction of HIV-1 Resistance to Protease Inhibitors. *J. Chem. Inf. Model.* **2016**, *56*, 915-923.

(32) Kotev, M; Soliva, R; Orozco, M. Challenges of Docking in Large, Flexible and Promiscuous Binding Sites. *Bioorg. Med. Chem.* **2016**, *24*, 4961-4969.

(33) Carlson, H. A.; Smith, R. D.; Damm-Ganamet, K. L.; Stuckey, J. A.; Ahmed, A.; Convery, M. A.; Somers, D. O.; Kranz, M.; Elkins, P. A.; Cui, G.; Peishoff, C. E.; Lambert, M. H.; Dunbar, J. B. Jr. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J. Chem. Inf. Model.* **2016**, *56*, 1063-1077.

(34) Grebner, C.; Iegre, J.; Ulander, J.; Edman, K.; Hogner, A.; Tyrchan, C. Binding Mode and Induced Fit Predictions for Prospective Computational Drug Design. *J. Chem. Inf. Model*. **2016**, *56*, 774-787.

(35) Rawson, D. J.; Stuart, P. G. Proline Derivatives Having Affinity for the Calcium Channel Alpha2-Delta Subunit. PCT Patent WO2004039367, May 13, 2004.

(36) Mortell, K. H.; Anderson, D. J.; Lynch, J. J.; Nelson, S. L.; Sarris, K.; McDonald, H.; Sabet, R.; Baker, S.; Honore, P.; Lee, C. H.; Jarvis, M. F.; Gopalakrishnan, M. Structure–Activity Relationships of α-amino Acid Ligands for the α2δ Subunit of Voltage-Gated Calcium Channels. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1138-1141.

(37) Hryc, C.F.; Chen, D.H.; Afonine, P.V.; Jakana, J.; Wang, Z.; Haase-Pettingell, C.; Jiang, W.; Adams, P.D.; King, J.A.; Schmid, M.F.; Chiu, W. Accurate model Annotation of a Near-Atomic Resolution Cryo-EM Map. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3103-3108.

(38) Biasini, M.; Bienert, S.; Waterhouse A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucl. Acids Res*. **2014**, *42*, W252-W258.

(39) Davis, I.W.; Leaver-Fay, A.; Chen, V.B.; Block, J.N.; Kapral, G.J.; Wang, X.; Murray, L.W.; Arendall, W.B. 3rd; Snoeyink, J.; Richardson, J.S.; Richardson, D.C. MolProbity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res.* **2007**, *35*, W375-83

(40) Jacobson, M.P.; Pincus, D.L.; Rapp, C.S.; Day, T.J.; Honig, B.; Shaw, D.E.; Friesner, R.A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351-367.

(41) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221-234.

(42) Small-Molecule Drug Discovery Suite 2017-1, Schrödinger, LLC, New York, NY, 2017.

(43) Schrödinger Release 2018-2: LigPrep, Schrödinger, LLC, New York, NY, 2018

(44) Banks, J.L.; Beard, H.S.; Cao, Y.; Cho, A.E.; Damm, W.; Farid, R.; Felts, A.K.; Halgren, T.A.; Mainz, D.T.; Maple, J.R.; Murphy, R.; Philipp, D.M.; Repasky, M.P.; Zhang, L.Y.; Berne, B.J.; Friesner, R.A.; Gallicchio, E.; Levy. R.M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comp. Chem*. **2005**, *26*, 1752-1780.

(45) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(46) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J*. **2001**, *80*, 505-515.

(47) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Fold. Des*. **1997**, *2*,173-181.

(48) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc*. **1996**, *118*, 11225-11236.

(49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.

(50)  Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics With NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(51) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712-725.

(52) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(53) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

(54) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

(55) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
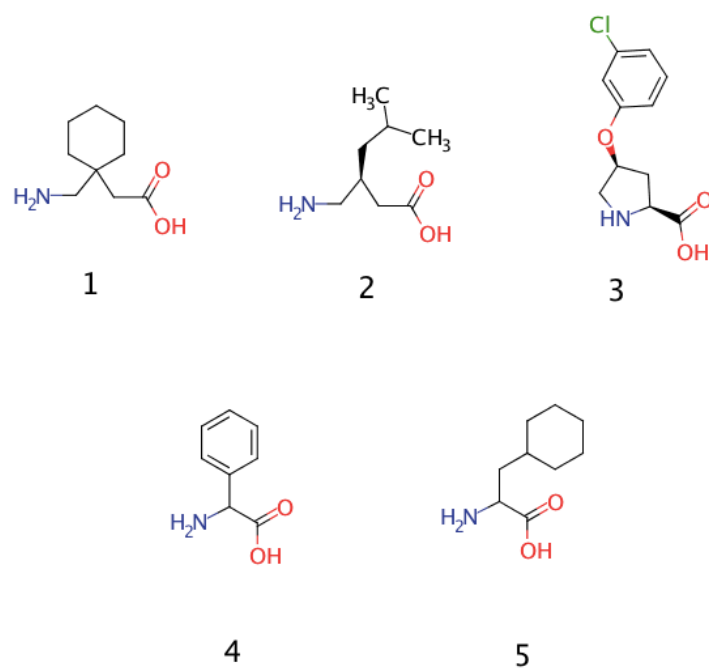
**Figure 1**.  Structures of representative amino acidic α2δ-1 ligands. Approved drugs currently in clinical use are gabapentin (**1**) and pregabalin (**2**). Clinical candidate **3** and compounds **4**, **5** from Mortell et al. [36] were also studied.
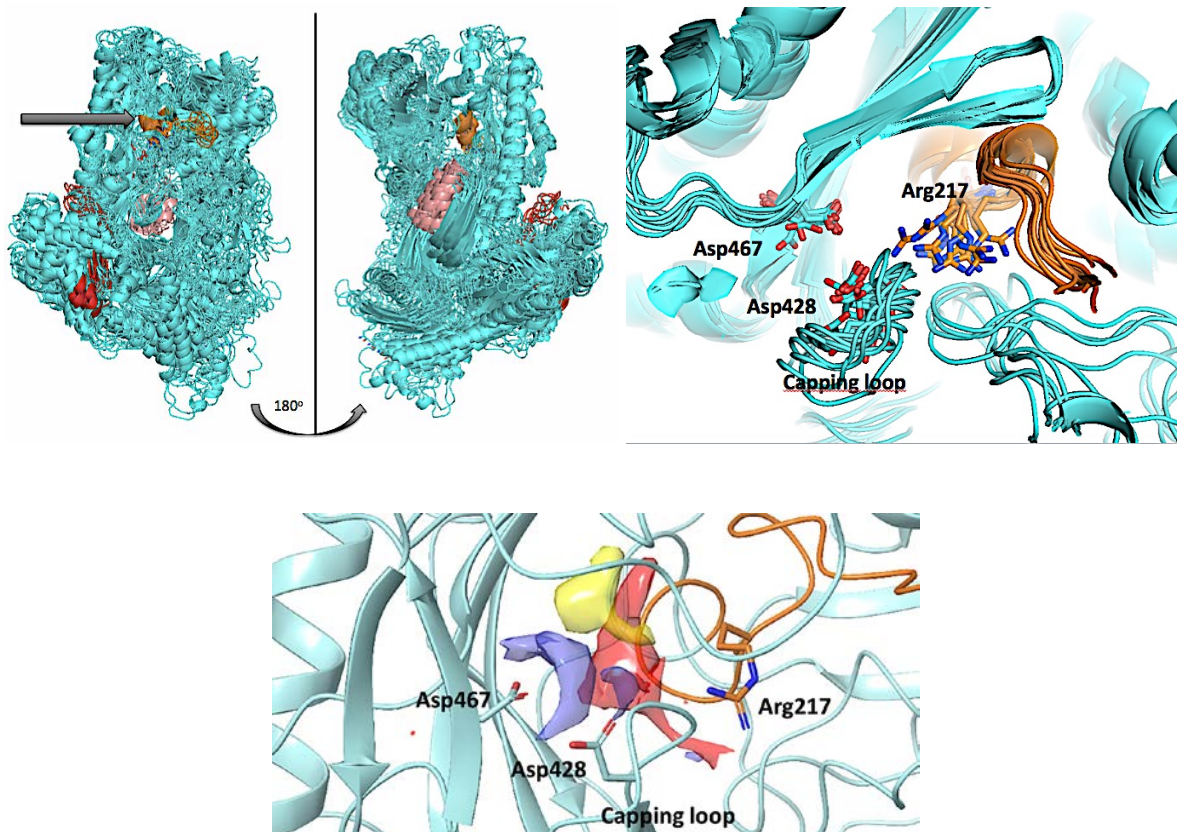
**Figure 2**. Structure superposition of 10 different cluster representatives extracted from an explicitly solvated MD simulation of the whole human α2δ-1 subunit. Segment I is colored in copper, segment II in light pink and segment III in red (only backbone is shown as a ribbon). Asp428, Asp467 and Arg217 residue side chains represented as sticks. Top left) Zoomed-out view of the whole α2δ-1 subunit, with an arrow pointing to Arg217; Top right); Zoomed-in view highlighting Arg217 (segment I), Asp467 and Asp428, the latter located in the capping loop (all other amino acids are removed for clarity). Bottom: SiteMap interaction profiles highlighting the cavity next to Arg217 in the α2δ-1 subunit from the pdb entry 5GJV (hydrophobic, H-bond donor and acceptor probes in yellow, blue and red respectively).
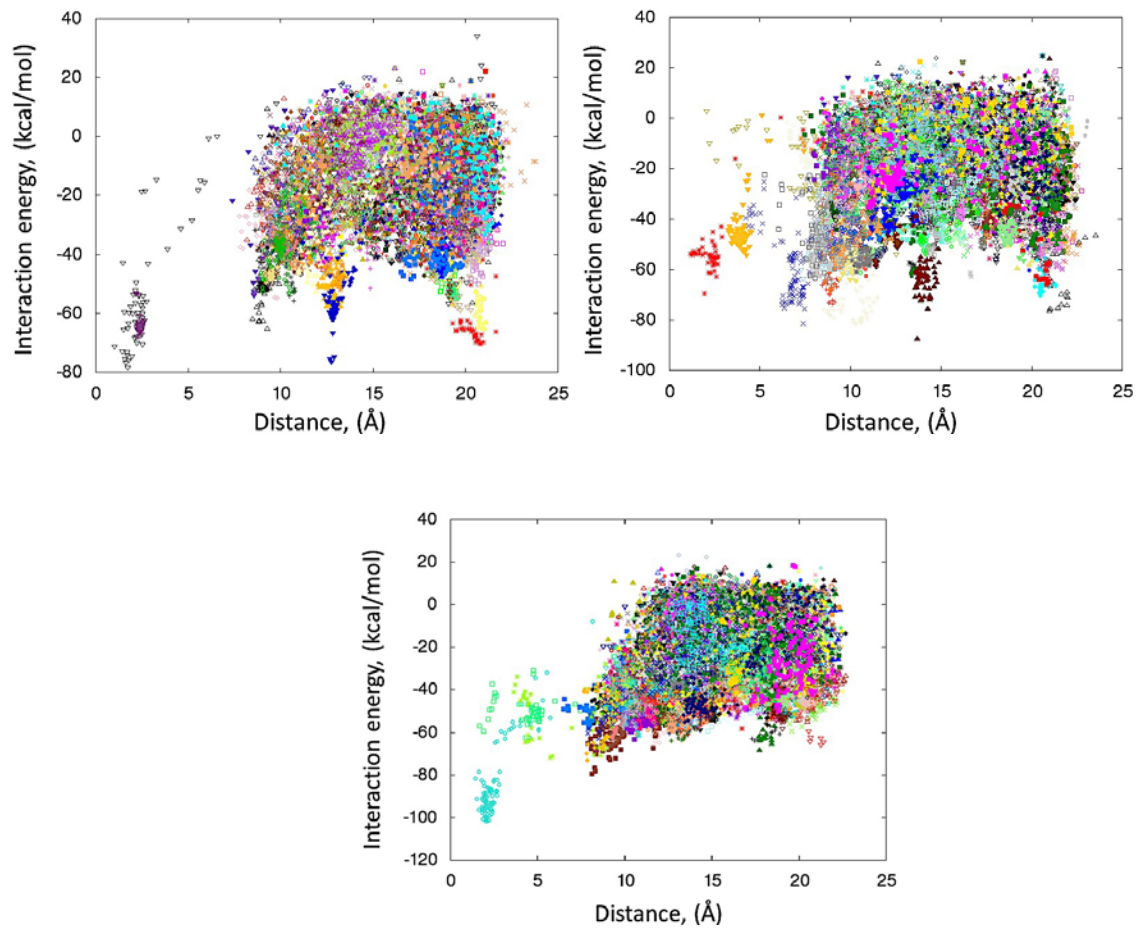
**Figure 3**. Results for PELE simulations for gabapentin (top left), pregabalin (top right) and the proline derivative **3** (bottom). The plots show the correlation of the distance between the ligand center of mass and the reference point in the interior of the binding site, next to Arg217 (x axis, in Å) and the interaction energy (y-axis, in kcal/mol). Each color corresponds to an independent trajectory.
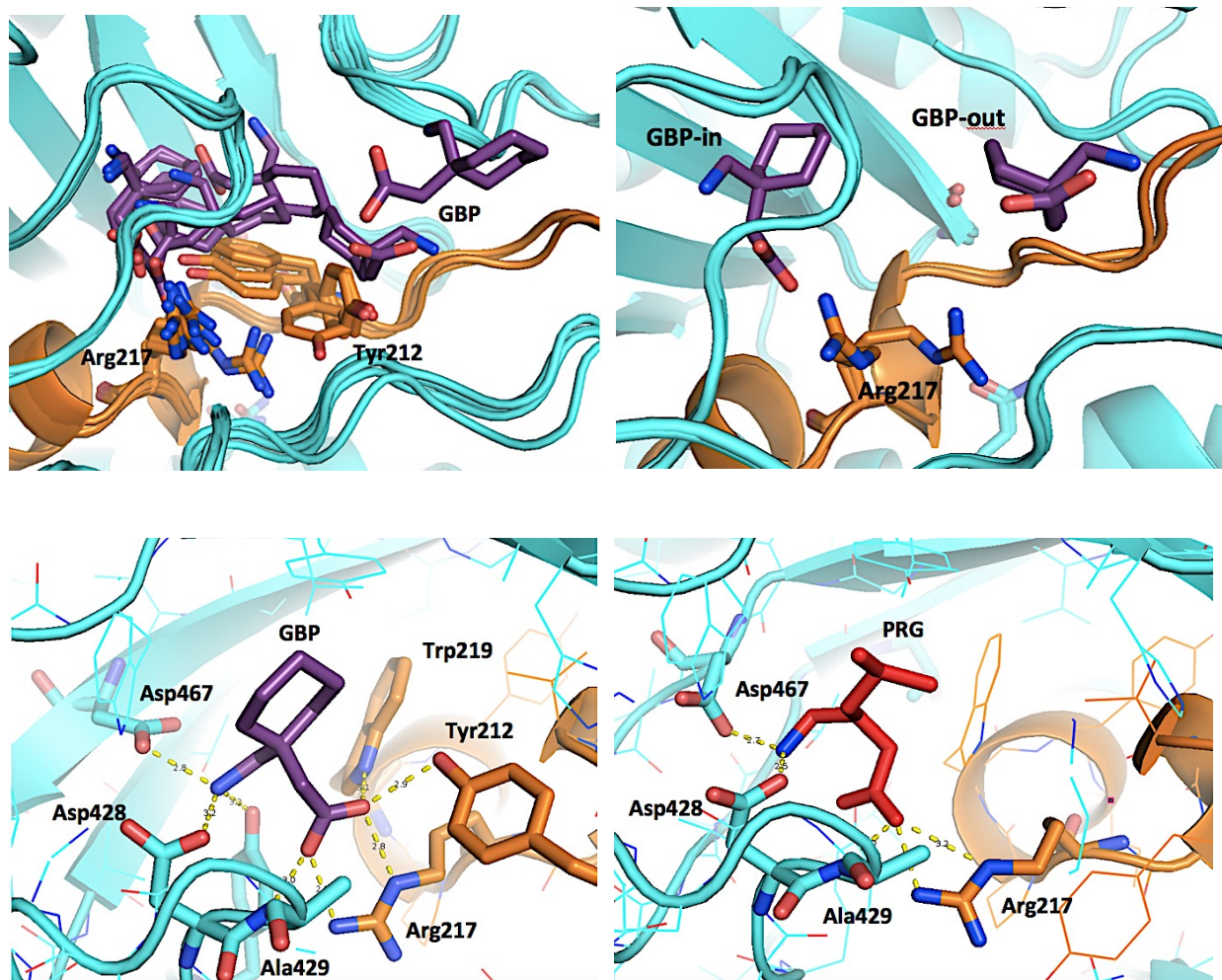
**Figure 4.** Top left: A series of snapshots taken from the full PELE binding trajectory for gabapentin (purple). The molecule's initial structure is next to the exterior of segment I (copper colored). It travels an entry path where it is in close contact with the latter. In order for gabapentin to fully access its final pose, Tyr212 and Arg217 act as a double gating mechanism. Top right: The gating movement by Arg217, which engages gabapentin's carboxylic acid in a closing movement. Bottom left: MD refined gabapentin binding mode, showing 8 electrostatic interaction points with α2δ-1 in dashed lines. Bottom right: One of the lowest energy pregabalin (red) binding modes from the PELE simulation, where the drug is engaging Arg217. Relevant residues are depicted as sticks and the H-bonds as dashed yellow lines (see text for details).
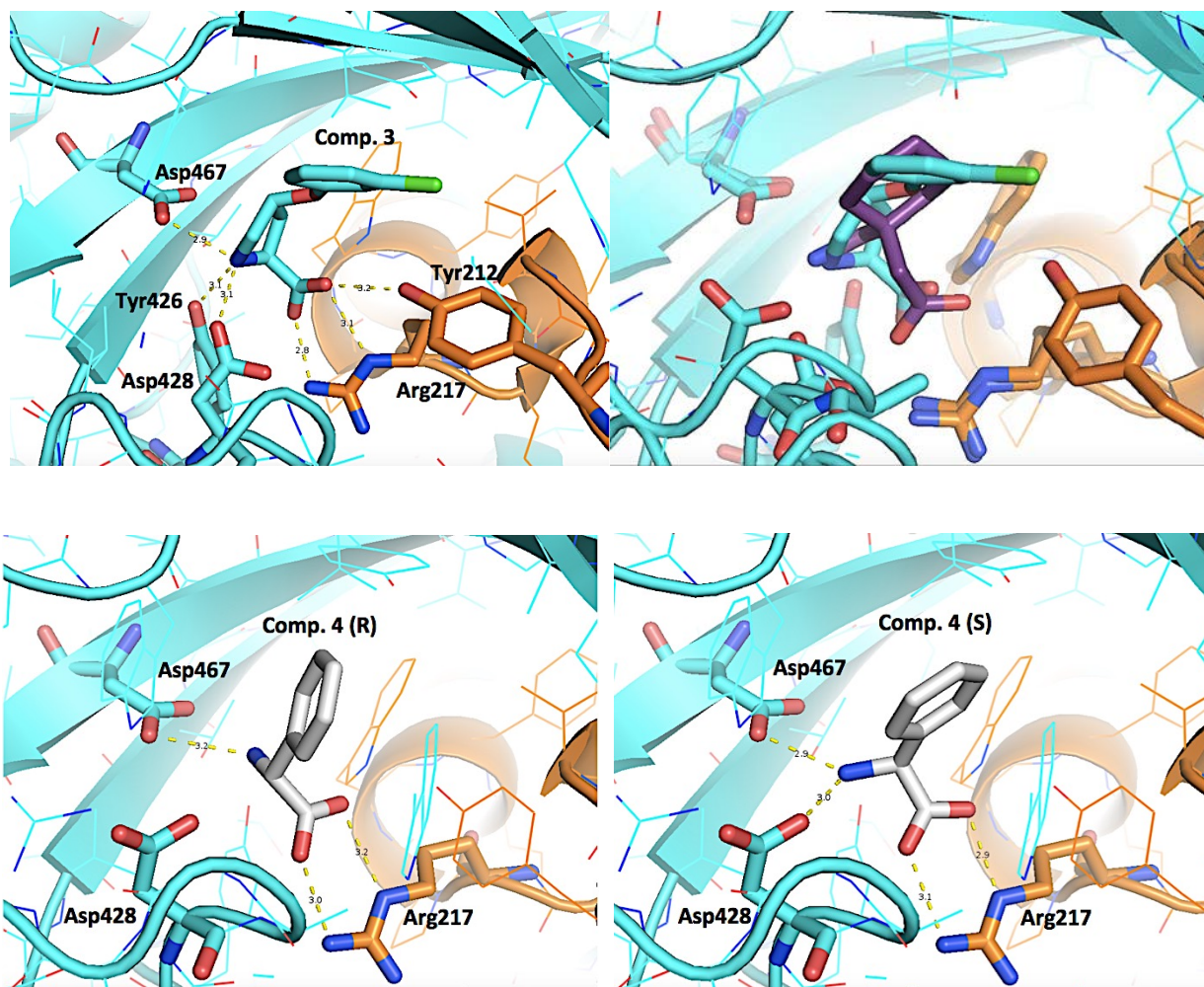
**Figure 5**. Binding mode for compounds **3** and **4**. Top left: MD refined binding mode for compound **3**. Top right: comparison between MD average binding modes for gabapentin (purple) and compound **3** (cyan). Bottom left: inactive R enantiomer of compound **4** docked on the MD average structure from the gabapentin simulation, depicting its only H-bond to the Asp428-Asp467 pair. Bottom right: active S enantiomer of compound **4** docked on the same structure, depicting its two H-bonds to the Asp428-Asp467 pair. Relevant residues are depicted as sticks and the H-bonds as dashed yellow lines (see text for details).

**Table of Contents Graphic**