Computational approaches to assess the binding properties of ligands : the case of the NMDA receptor

Zeynep Pinar Haslak

UNIVERSITÉ DE LORRAINE

École doctorale C2MP Lorraine

Approches numériques pour évaluer les propriétés de liaison de ligands: le cas du récepteur NMDA

Computational approaches to assess the binding properties of ligands: the case of the NMDA receptor

THÈSE

présentée et soutenue publiquement le 13 décembre 2019

pour l'obtention du

Doctorat de l'Université de Lorraine

(Mention Chimie)

par

Zeynep Pinar Haslak

Composition du jury

Rapporteurs :	Prof. Manuel Dauchez, Université de Reims Champagne-Ardenne, Reims, France Prof. F. Aylin Sungur, Istanbul Technical University, Istanbul, Turquie
Examinateurs :	Prof. Safiye Erdem, Marmara University, Istanbul, Turquie Dr. Antonio Monari, Université de Lorraine, Nancy, France Prof. Duygu Avci, Boğaziçi University, Istanbul, Turquie
Invité :	Prof. Viktorya Aviyente, Boğaziçi University, Istanbul, Turquie
Encadrants :	Prof. Gérald Monard, Université de Lorraine, Nancy, France Prof. Ilknur Dogan, Boğaziçi University, Istanbul, Turquie

Laboratoire de Physique et Chimie Théoriques – UMR 7019, Université de Lorraine - CNRS
Boulevard des Aiguillettes B.P. 70239 – F-54506 Vandoeuvre-les-Nancy, FRANCE

Mis en page avec la classe thesul.

Contents

Acknowledgments

I wish to express my deepest appreciation and gratitude to Prof. Viktorya Aviyente for her continuous support, motivation, enthusiasm, immense knowledge and invaluable advices. Since when I was an undergraduate student and from then, I am inspired by her wisdom and I always felt lucky to have the opportunity to be her student.

I am extremely grateful to Prof. Gérald Monard for giving me the opportunity to work in a highly motivated international environment and providing invaluable guidance throughout my PhD studies in France, this will always be a great privilege and honor for me. I would also like to thank him for his help and support both on academic and personal levels. I would like to thank Prof. Ilknur Dogan for her contributions to my work with her critical comments and wise ideas. This dissertation would not have been possible without her help.

I wish to thank deeply the members of the examining committee, Prof. Duygu Avcı, Prof. Safiye Erdem and Prof. Fethiye Aylin Sungur for their valuable time and criticism.

I would also like to acknowledge the invaluable presence of Gamze Tanrıver, who has been a great friend and a companion in this journey. I will never forget the memories we had, especially the ones in the cold Nancy days. I am grateful to Dr. Antonio Monari for his joyful friendship and for helping me to adapt to the French culture. I would like to thank a lot Mrs Séverine Bonenberger for helping me to solve every problem that I had to face in the French bureaucracy. I want to appreciate Mr Florent Sécond, who has always been an admirable companion, for supporting, encouraging and listening to me whenever I needed.

I would like to thank the former and current members of Computational Chemistry and Biochemistry Group in Boğaziçi University, especially to Dr. Berçem Dutağacı, Dr. Sesil Agopcan Çınar and Gamze Bahadır for their contributions to the main chapters of this dissertation. I am also indebted to all members of the Laboratoire de Physique et Chimie Théoriques in Université de Lorraine for making me feel at home. I would like to express my appreciation to the Quantum Chemistry Group of Vrije Universiteit Brussel, especially to Prof. Frank De Proft and Dr. Freija De Vleeshouwer for their contributions to my scientific background and to this dissertation.

I wish to extend my gratitude to Piri Reis University members for their continuous help and friendship, we have always been like a family. I am particularly grateful to our rector Prof. Oral Erdoğan who has supported me and gave the time that I needed to finish my studies.

I would like to express my most sincere gratitude to my family; my dear mother, my dear father and my dear sister, for their endless love, support and patience, for sharing my dreams and being the source of my strength. I am very proud to have such a wonderful family. And last but not least, I would like to thank a lot to my dear friend, my second sister, Melis Demirci for bringing the joy and happiness to my life.

I am thankful to all who have a touch in this journey...

This thesis has been supported by Bogazici University Research Fund BAP (project number 16241 and project code 19B05D7), TUBITAK International Doctoral Research Fellowship (BIDEB 2214/A) and Joint PhD scholarship from the French Embassy in Ankara. This work was granted access to the HPC resources of the National Center for High Performance Computing at Istanbul Technical University (Grant number 5004092016) and Mésocentre Explor.

Résumé

Contexte

Le glutamate, l'un des principaux neurotransmetteurs du cerveau et de la moelle épinière, exerce ses effets post-synaptiques par le biais d'une série de récepteurs membranaires. Les récepteurs du glutamate ioniques (iGluR) sont situés au niveau des membranes des cellules neuronales et jouent un rôle clé dans la plasticité synaptique. Il existe quatre sous-types d'iGluR en fonction de leur pharmacologie et de leur similarité de séquence: le récepteur N-méthyl-D-aspartate (NMDAR), le récepteur de l'acide kaïnique (KAR), l'acide alpha-amino-3-hydroxy-5-méthyl-4-isoxazolepropionique (AMPAR) et le récepteur Delta [1] (Figure 1).

Les récepteurs AMPA et kaïnate interviennent dans la transmission synaptique rapide dans le système nerveux central [2]. Dans le récepteur NMDA, les courants postsynaptiques augmentent et diminuent plus lentement que les récepteurs AMPA et kainate. Le récepteur NMDA est également plus sensible au blocage dépendant de la tension, les canaux ioniques du récepteur permettent le passage des ions calcium en plus des ions monovalents de sodium et de potassium, et il nécessite la liaison simultanée de glycine et de glutamate pour l'activation du canal ionique. [3]. D'autre part, les récepteurs Delta ne répondent absolument pas au glutamate et leur signification fonctionnelle n'a pas encore été comprise [4].

Les monomères d'iGluR sont formés d'un domaine extracellulaire N terminal (NTD), d'un domaine intracellulaire C terminal (CTD), de trois domaines transmembranaires (M1, M2 et M3), d'une boucle de membrane réentrante (P-loop) et d'un ligand domaine de liaison [5] (Figure 2). Le récepteur comprend deux sous-unités GluN1 et deux sous-unités GluN2 ou GluN3. Alors que GluN1 est responsable de l'activation du canal ionique, GluN2 et GluN3 sont responsables des propriétés fonctionnelles telles que l'affinité du glutamate, la désensibilisation des récepteurs et la sensibilisation pharmacologique [6, 7]. Il est connu que le récepteur NMDA joue un rôle clé dans la transmission synaptique excitatrice telle que l'apprentissage, la mémoire et la plasticité synaptique. Le dysfonctionnement du récepteur est donc impliqué dans de nombreux troubles neurodégénératifs comme l'épilepsie, la maladie de Parkinson, la maladie d'Alzheimer et la schizophrénie. Dans des conditions normales de transmission synaptique, le canal ionique du récepteur est bloqué par les ions magnésium. La liaison de la glycine agoniste native à la sous-unité GluN1 et du glutamate co-agoniste à la sous-unité GluN2 active le canal ionique et permet l'écoulement des ions calcium, potassium et sodium dans les neurones. Ce processus est très important pour la formation de la mémoire et la neurotoxicité [8, 9, 10, 11]. Le récepteur peut être inactivé pendant sa période de repos ou exposé à une «mort cellulaire exotoxique» en raison de la stimulation prolongée du récepteur. Ainsi, la sous-unité GluN1 est le site de modulation allostérique qui la rend unique dans l'activation du récepteur [12]. La

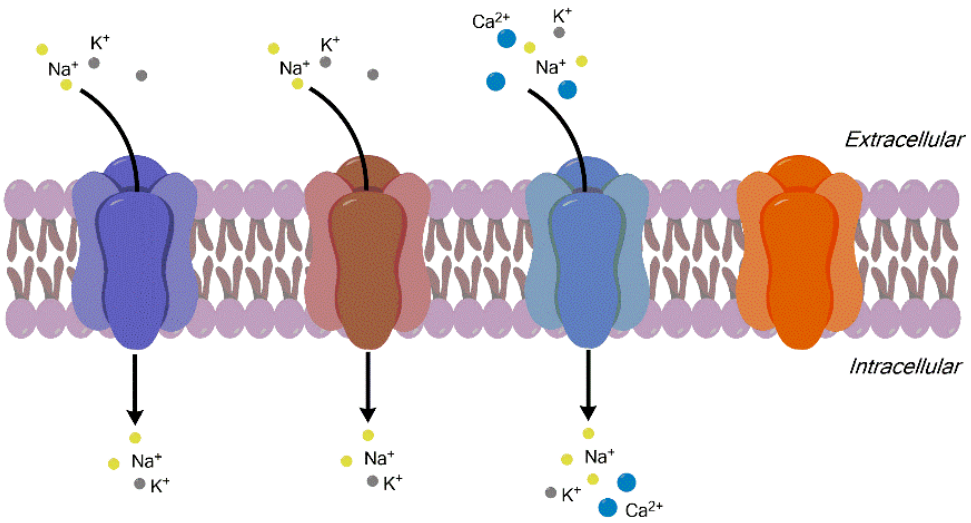


Figure 1: Sous-types de récepteurs ionotropes du glutamate: récepteurs AMPAR, KAR, NMDAR et Delta (de gauche à droite)

régulation de l'activité des récepteurs par l'introduction de ligands agonistes-antagonistes aux sites glycine et glutamate est devenue une stratégie prometteuse pour le traitement des troubles neuropsychiatriques, liés au dysfonctionnement de la NMDA. Par conséquent, les sites de liaison de ces sous-unités sont des cibles importantes de la recherche médicale extensive.

Selon les essais de liaison du ligand et les investigations fonctionnelles, la fonction du récepteur est modifiée par la compétition des ligands en interaction avec la glycine au site de liaison de la glycine [13]. Les agonistes gardent les domaines de liaison du ligand fermés, ce qui entraîne l'ouverture du canal ionique NMDA, tandis que les antagonistes agissent comme un coin entre les domaines de liaison du ligand, empêchant ainsi l'ouverture du canal ionique et la stimulation excessive du récepteur (Figure 3) [14] . Cependant, il est bien connu que l'activation ou le blocage complet du récepteur entraîne de nombreux effets secondaires graves, tels qu'une altération de la mémoire, une neurotoxicité ou des psychotomimétiques [15]. D'autre part, un agoniste partiel peut agir soit comme agoniste complet, soit comme antagoniste, c'est-à-dire qu'en présence d'antagoniste en excès, un agoniste partiel agit comme un agoniste pour réduire les effets d'inhibition totale de l'antagoniste, en fonction de la concentration en glycine à la concentration initiale récepteur. Pour cette raison, l'emploi d'agonistes partiels devient de plus en plus applicable puisque ces composés sont capables de réguler l'activité à un certain niveau [16].

L'affinité de liaison (K_i) et les données d'efficacité d'une molécule de médicament déterminent ensemble la puissance de ce médicament. L'interaction d'un ligand avec son récepteur cible via des forces intermoléculaires telles que des liaisons hydrogène, des interactions de Coulomb et des forces de Van der Waals est caractérisée par son affinité de liaison. Une fois que le ligand se lie à son récepteur, le degré d'activation du récepteur est déterminé par son efficacité [17]. Les antagonistes ont une efficacité presque nulle, inactivant le récepteur, tandis que les agonistes complets ont une efficacité de 100% qui est interprétée comme une activation complète, les agonistes partiels ayant une efficacité allant d'un agoniste à un antagoniste.

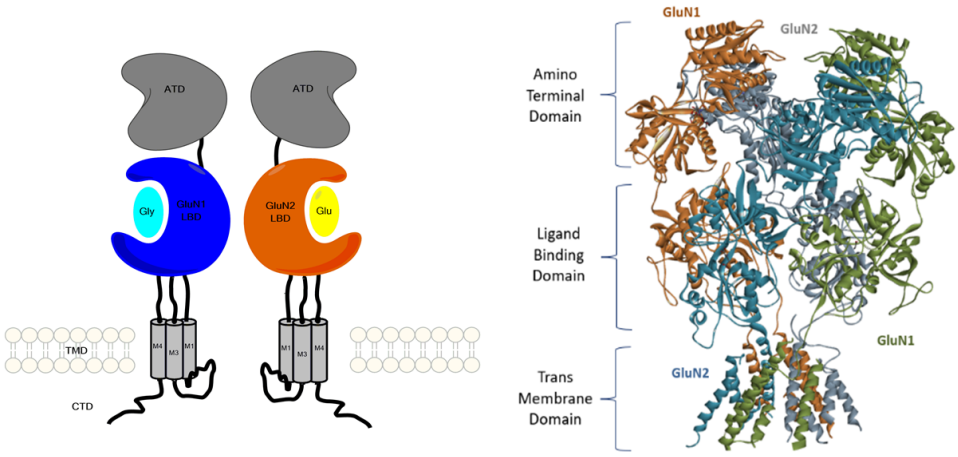


Figure 2: Représentation schématique de la topologie des sous-unités iGluR (à gauche) et du récepteur hétérotétramérique NMDA GluN1-GluN2 (à droite)

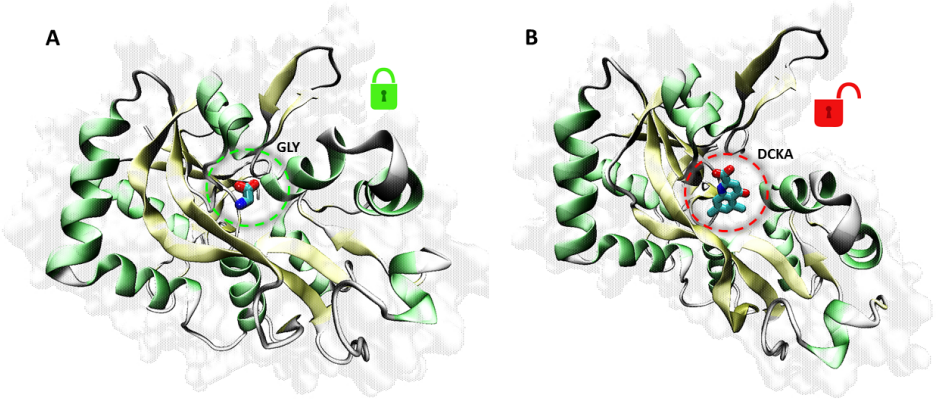


Figure 3: Représentation sous forme de bande dessinée de la sous-unité GluN1 du récepteur NMDA. La structure liée à la glycine (GLY) (code PDB: 1PB7) fait référence à la conformation à fente fermée (A), tandis que la structure liée à l'acide 5,7-dichlorokynurénique (DCKA) (code PDB: 1PBQ) correspond à la conformation à fente ouverte (B)

Un grand nombre de systèmes chimiques et biologiques contiennent des groupes acides et basiques qui régissent les interactions avec l'environnement et donc le fonctionnement du système. À un pH particulier, le pK_a de l'espèce détermine la mesure dans laquelle une espèce ionisable peut être protonée ou déprotonée par les transferts d'hydrogène de / vers l'environnement. La plupart des molécules du médicament sont des acides faibles ou des bases faibles et, lorsqu'elles sont en solution, elles se trouvent à la fois dans leurs états ionisé et non ionisé. Le pK_a d'une molécule de médicament détermine sa solubilité dans les lipides, et donc sa diffusion, car seuls les ligands non chargés peuvent pénétrer dans la membrane cellulaire [18]. En outre, les interactions entre les groupes fonctionnels ionisables d'une molécule de médicament avec les résidus de la protéine, qui affectent l'affinité, l'activité et l'efficacité du ligand, dépendent fortement des pK_a des chaînes latérales du site actif et de la molécule de médicament. D'autre part, les modifications des états de protonation des acides aminés dans les résidus ont un impact direct sur l'établissement de la conformation et de la stabilité de la protéine [19], de la solubilité et du repliement [20], de l'activité catalytique des enzymes [21] et de la capacité de liaison. Par conséquent, obtenir des informations sur les pK_a relatifs des ligands et sur les sites de protonation du récepteur peut expliquer les principales caractéristiques de la liaison du ligand.

Outre les facteurs régissant les interactions d'un ligand avec son environnement, il est important de comprendre les propriétés intrinsèques du ligand. Il est bien connu que les stéréoisomères d'une molécule chirale peuvent présenter différentes activités biologiques. Un ligand peut être stéréosélectif ou régiosélectif en fonction des différences d'environnement électronique et stérique. Pour cette raison, la démonstration de la stabilité de configuration des candidats-médicaments et des mécanismes par lesquels ils racémisent sont des questions importantes. Récemment, Sarigul et Dogan [22] ont synthétisé des dérivés de 5-méthyl et 5-isopropyl-3-(o-aryl)-2-thiohydantoïne sous forme de paires diastéréomériques non racémiques, une classe spécifique de composés considérés comme faisant partie des privilégiés. Les échafaudages dans la découverte de médicaments avec leurs activités agonistes-antagonistes, comme déterminé par les analyses HPLC avec une phase stationnaire chirale et par ^{1}H-NMR. La synthèse de la 5-méthyl-3-phényl-2-thiohydantoïne et de la 5-isopropyl-3-phényl-2-thiohydantoïne, dépourvue d'axes chiraux, a donné des mélanges racémiques. Les chiraux axiaux peuvent exister sous les formes isomères SP, RM, SM et RP. SP / RM est un transoïde, alors que SM / RP est un cisoïde par rapport au substituant en C_5 et au substituant o-aryle. L'utilisation d'outils informatiques pour modéliser les mécanismes de réaction est importante pour comprendre les détails des étapes de racémisation afin de synthétiser des produits énantiomériquement purs.

Objectifs

L'un des problèmes importants dans la conception des médicaments est l'identification de l'activité biologique des ligands des récepteurs. Les mesures de développement, de synthèse et d'activité des ligands ont une importance majeure dans la conception des médicaments. Cependant, il existe certaines limites dans les études expérimentales; la synthèse d'un grand nombre de composés couvrant toutes les molécules potentiellement actives est irréaliste. Les études informatiques pourraient donc apporter une aide précieuse aux études expérimentales sur la conception de ligands pour les récepteurs du glutamate. En combinant les points forts des ap-

proches de la dynamique moléculaire et de la chimie quantique, il est possible d'établir une inspection, une caractérisation et une rationalisation plus ciblées des études de conception de médicaments. Dans le présent travail, nous souhaitons vérifier quelles propriétés intrinsèques de ligands structurellement semblables provoquent une différenciation entre les comportements antagonistes et agonistiques en utilisant divers outils informatiques. De plus, les interactions entre le ligand et le domaine de liaison du ligand seront examinées afin de mieux comprendre les mécanismes d'activation / inhibition du récepteur.

Au chapitre 1, une brève introduction sur le récepteur NMDA, l'activité agoniste-antagoniste des ligands, les facteurs régissant les interactions d'un ligand avec son environnement et l'importance de la sélectivité d'un ligand sont fournis. Une discussion plus détaillée de chacun d'entre eux est donnée dans les chapitres suivants. L'objectif et la portée de la thèse sont décrits au chapitre 2. Au chapitre 3, les principes fondamentaux des calculs effectués, y compris la théorie de la densité fonctionnelle, la dynamique moléculaire et les méthodes d'ancrage, sont introduits. Dans les chapitres pertinents, des explications complètes sur les méthodologies sont présentées. Après un bref aperçu des méthodes théoriques, les résultats de cette étude seront présentés en 4 chapitres. Dans les deux chapitres suivants, nous cherchons à différencier les agonistes, les antagonistes et les agonistes partiels basés sur les descripteurs de Quantum Chemical (chapitre 4) et les énergies libres de Gibbs de liaison (chapitre 5). Plusieurs propriétés moléculaires qui pourraient jouer un rôle dans la liaison du ligand à la sous-unité glycine GluN1 de NMDA et les énergies libres de liaison calculées de Gibbs sont en outre utilisées pour établir un lien entre les efficacités et les affinités de liaison des ligands. La prédiction des constantes de dissociation acide d'acides aminés dans les protéines et les ligands nous permet d'avoir des informations sur l'affinité de liaison et l'efficacité du ligand à sa protéine cible. Au chapitre 6, nous verrons comment les charges atomiques des acides carboxyliques peuvent être liées à la prédiction du pK_a des ligands. Un cas particulier sera présenté au chapitre 7 afin de comprendre la stéréosélectivité des ligands de la thiohydantoïne, qui sont de puissants antagonistes des récepteurs aux androgènes, en explorant diverses voies de synthèse mécanistiques. Les conclusions cruciales tirées de chaque chapitre et les conclusions sont mises en évidence au chapitre 8.

Résultats

Discrimination de l'activité agoniste et antagoniste de ligands se liant au récepteur NMDA

Les relations quantitatives structure-activité (QSAR) dérivent des modèles qui corrèlent les activités biologiques du ligand avec sa structure moléculaire soit par des paramètres physicochimiques (analyse de Hansch), par des variables indicatrices codant différentes caractéristiques structurelles (analyse de Free Wilson), ou par des profils de propriétés moléculaires tridimensionnels. des composés (analyse comparative du champ moléculaire, CoMFA). 1962 peut être considérée comme l'année de naissance de la méthodologie QSAR moderne avec la première publication de Corwin Hansch [23]. Les descripteurs chimiques quantiques codent des informations chimiques sur les propriétés intrinsèques des ligands, qui peuvent ensuite être associées à leurs activités biologiques. Différentes méthodologies ont été appliquées au cours des dernières années afin d'associer les propriétés structurelles des ligands à leurs activités biologiques. Un

modèle CoMFA a été suggéré en 2003 par Tikhonova *et al.* pour la sélectivité des ligands entre les récepteurs AMPA et NMDA [10]. Cela a été une approche importante dans le problème de la sélectivité protéine-ligand pour la synthèse de nouveaux médicaments à ciblage biologique. Afin de prédire les valeurs de pK_i des agonistes partiels se liant à la sous-unité NMDA GluN1, Cheng *et al.* a développé un modèle QSAR dérivé de la fonction de distribution radiale dans lequel le volume et la masse de Van der Waals ont été utilisés comme descripteurs [24]. Les activités expérimentales ont été reproduites à l'aide du modèle proposé. Ainsi, il sera possible d'avoir des informations sur l'activité de molécules qui n'ont pas encore été synthétisées. Sur la base de descripteurs semi-empiriques quantiques de la chimie, de l'état électro-topologique, de la propriété moléculaire et de l'indice d'ombre, Zhu *et al.* construit un modèle informatique fiable pour la classification des agonistes et antagonistes du récepteur 5-HT1A par apprentissage automatique et par algorithme génétique [25]. L'électronégativité, la polarisabilité atomique, le volume et la masse de van der Waals, la charge atomique et le coefficient de partage (logP) en tant que descripteurs semblent jouer un rôle important dans les études QSAR les plus récentes visant à identifier les agonistes et les antagonistes du récepteur A3 de l'adénosine [26]. Djeradi *et al.* a utilisé les indices de Fukui comme descripteurs afin de prédire les valeurs IC50 des flavonoïdes [27].

En ce qui concerne le récepteur NMDA, Yosa *et al.* ont montré en 2009 que la liaison des agonistes et des agonistes partiels à la sous-unité GluN1 du récepteur NMDA pouvait être différenciée des antagonistes par leurs énergies LUMO [28]. Dans leur document, les auteurs indiquent toutefois que les deux classes de composés étudiés, à savoir les composés agonistes (partiels) versus antagonistes, appartiennent à des classes de composés chimiques distinctes, de sorte qu'il n'existe aucun chevauchement structurel entre elles afin de permettre la création de LUMO similaires. trouvé parmi les deux populations. Ils suggèrent de tester leurs résultats pour les antagonistes qui ressemblent chimiquement au groupe d'agonistes existants, ou vice versa. Dans ce chapitre, les propriétés intrinsèques de ligands sélectifs de sous-unités de GluN1 ressemblant structurellement pour le récepteur NMDA causent une différenciation entre le comportement antagoniste et le comportement agonistique ont été vérifiées par des calculs de chimie quantique. Plusieurs propriétés moléculaires globales et locales pouvant jouer un rôle dans la liaison de ligand à la sous-unité GluN1 de NMDA ont été examinées et ces descripteurs ont ensuite été utilisés pour discriminer la nature d'agonisme-antagonisme partiel des ligands. De plus, il a été vérifié si ces descripteurs chimiques pouvaient fournir un lien entre les caractéristiques chimiques des molécules et leur efficacité, c'est-à-dire le degré d'effet agonistique.

30 ligands à activité antagoniste [28, 29] et 30 ligands à activité agoniste partielle [30] ont été sélectionnés dans la littérature. Toutes les structures ont été optimisées avec le programme Gaussian09 en utilisant le niveau de théorie B3LYP / 6-311G **, le meilleur conformateur pour chaque molécule est ensuite utilisé pour le calcul des descripteurs. Afin de tester la dépendance fonctionnelle des descripteurs, la même procédure a été appliquée à un ensemble de ligands choisis de manière aléatoire en utilisant M06-2X / 6-311G **. Les descripteurs globaux et locaux de la chimie quantique ont été utilisés fréquemment car ils constituent un outil utile pour rationaliser les aspects locaux de la réactivité et de l'activité [31, 32]. Les énergies HOMO et LUMO, la dureté, la fréquence d'étirage du carbonyle, les charges atomiques et les constantes de protection isotropes ont été calculées, puis associées à leurs affinités pour le NMDA site de liaison de la glycine.

Les énergies HOMO et LUMO en tant que descripteurs moléculaires ont déjà été utilisées

pour distinguer les agonistes des antagonistes de différents récepteurs. Les énergies HOMO des agonistes partiels ont été calculées pour être en moyenne inférieures à celles des antagonistes, alors que l'inverse a été observé pour les énergies LUMO, conformément aux résultats de Yosa *et al.* [28]. Cependant, aucune distribution bimodale claire n'a été trouvée. Lorsque nous nous sommes concentrés sur des structures très similaires dans les ensembles antagonistes (Anta1, Anta2, Anta3) et agonistes partiels, certaines tendances ont pu être observées. Pour les ligands du groupe Anta1, les énergies HOMO et LUMO se situent dans une plage assez large. Il existe une grande variation dans les énergies HOMO de l'ensemble Anta2, alors que les énergies LUMO sont très similaires. Cela peut être retracé aux groupes amino substitués sur le cycle phényle. Aucune différence dans les énergies HOMO et LUMO n'est observée pour les ligands du groupe Anta3. La substitution sur le cycle saturé à 7 chaînons n'affecte pas les orbitales les plus occupées et les plus inoccupées, car ces orbitales proviennent du système π du ligand. Une variation légèrement plus grande dans les énergies HOMO que dans LUMO est observée pour les agonistes partiels. Le HOMO et le LUMO sont tous deux délocalisés sur les cycles aromatiques et, dans le cas du LUMO, le groupe carbonyle est également inclus. Lorsque ces orbitales sont comparées aux orbitales respectives d'un agoniste complet tel que la D-sérine, une image différente apparaît. Pour la D-sérine, HOMO et LUMO sont situés principalement sur le groupe carboxyle. Il en résulte une énergie HOMO plus profonde et une énergie LUMO plus élevée par rapport à tous les agonistes partiels examinés et en particulier aux antagonistes. Cela signifie qu'il y a une différence distincte entre les agonistes partiels considérés dans cette étude et les agonistes complets. Les agonistes partiels et les antagonistes ont tous deux HOMO et LUMO présents sur les cycles aromatiques et leurs substituants. Puisque les antagonistes et les agonistes partiels ne peuvent pas être différenciés complètement en utilisant l'énergie et la localisation de leurs HOMO et LUMO, nous nous tournons vers un descripteur moléculaire qui combine les deux propriétés, à savoir la dureté chimique. À partir de l'approximation des noyaux gelés pour la dureté chimique, il est connu que les molécules dures ont un large intervalle d'énergie HOMO-LUMO et que les molécules douces ont un faible intervalle d'énergie HOMO-LUMO. Dans ce travail, la dureté chimique a été calculée par la méthode des différences finies. Les agonistes partiels semblent être des espèces plus dures que les antagonistes. Basé sur le principe des acides et bases durs et mous [33, 34, 35], selon lequel les acides durs préfèrent se lier aux bases dures et les acides mous aux bases molles, cela pourrait indiquer que les deux types de ligands interagissent plus ou moins fortement avec les résidus sur les sites de liaison du récepteur, en fonction de leurs valeurs de dureté.

Les forces intermoléculaires telles que l'attraction coulombienne, la liaison hydrogène, l'empilement π – π, les contacts de van der Waals, les interactions hydrophobes, etc., forment la base de la liaison ligand-récepteur. Ces interactions sont principalement activées par les groupes fonctionnels phényle, amide, amine, céto ou carboxylate présents en tant que sites de liaison de ligand sur le récepteur NMDA [36, 37, 38, 39]. Des études d'amarrage de la sous-unité NMDA GluN1 dans un complexe avec divers ligands montrent que l'oxygène de carbonyle est le noyau de liaison primaire puisque la plupart des régions riches en électrons sont localisées autour de celle-ci. Afin de comprendre les différences entre les centres de réactivité dans les interactions ligand – récepteur d'antagonistes et d'agonistes partiels, des descripteurs locaux ont été appliqués aux atomes d'azote, d'hydrogène et d'oxygène du groupe amide. Le comportement chimique et la réactivité d'une molécule peuvent être interprétés en étudiant la distribution de charge dans le système. La reconnaissance ligand-récepteur est principalement

dictée par les interactions entre atomes partiellement chargés, telles que la liaison hydrogène, la liaison halogène et d'autres interactions électrostatiques. Deux analyses de population différentes ont été utilisées, à savoir Mulliken et Natural Population Analysis (NPA). L'oxygène amide et l'azote des agonistes partiels sont plus chargés négativement que ceux des antagonistes. Ceci suggère une capacité plus forte d'accepteur et de donneur de liaisons hydrogène du groupe amide des agonistes partiels.

Le dernier descripteur local sondé est la fréquence d'étirement du carbonyle. Les agonistes partiels, qu'ils soient à l'état neutre ou zwitterionique, ont des fréquences voisines de 1650 cm^{-1}, alors que les molécules antagonistes ont des fréquences égales ou supérieures à 1700 cm^{-1}. Le nombre d'onde inférieur observé dans les agonistes partiels peut être attribué à une plus grande capacité de liaison hydrogène de l'oxygène carbonyle. L'efficacité des ligands du récepteur NMDA est calculée par rapport à l'effet maximal de la glycine (= 100%). Cela signifie que les agonistes complets ont une efficacité de 100%, les antagonistes, une efficacité de 0% et les agonistes partiels, une efficacité de 0 à 100%. Dans la partie suivante, il est confirmé quels descripteurs moléculaires précédemment rationalisés distinguent véritablement les ligands agonistes partiels des antagonistes en introduisant un modèle de régression linéaire multiple (MLR) liant les descripteurs moléculaires respectifs à l'efficacité des ligands. Le modèle peut être décrit par l'équation suivante:

$$\widehat{y_i} = 91.88\eta + 2314\varepsilon_{HOMO} - 1653\varepsilon_{LUMO} + 69.40 \quad (1)$$

Pour la plupart des ligands, la concordance entre modèle et expérience est relativement bonne, souvent dans les limites des écarts expérimentaux. Il ressort clairement de cette discussion que l'efficacité telle que prédite par le modèle introduit peut être qualifiée d'efficacité intrinsèque du ligand, en considérant uniquement les caractéristiques de ce ligand et non la topologie des sites de liaison du récepteur. Cependant, lorsque des groupes latéraux volumineux supplémentaires sur les systèmes cycliques aromatiques du ligand augmentent encore plus le volume du ligand, il peut être nécessaire de corriger l'efficacité intrinsèque. Cette correction peut améliorer ou aggraver l'efficacité intrinsèque, en fonction de l'adaptation de ces groupes latéraux dans les domaines de liaison au ligand et de l'importance des interactions de liaison à l'intérieur.

Les résultats de cette expérience DFT peuvent être résumés comme suit: (1) La combinaison des énergies HOMO et LUMO dans la dureté chimique permet de séparer les deux types de ligands. (2) les distributions de charge NPA des ligands ont montré que l'oxygène d'amide et l'azote des agonistes partiels étaient plus chargés négativement que ceux des antagonistes. Par conséquent, il faut s'attendre à ce que le groupe amide des agonistes partiels soit plus capable d'accepter les liaisons hydrogène et de donner des donneurs, ce qui a été confirmé par les fréquences plus basses d'étirement du carbonyle. (3) La MLR a montré que, pour tous les ligands, la dureté chimique était le facteur le plus important et le plus important en ce qui concerne l'efficacité. Cependant, l'énergie LUMO semble avoir plus d'influence sur l'ensemble des agonistes partiels, alors que pour les antagonistes, ce rôle est joué par l'énergie HOMO. (4) Nous ne pouvons pas tester le modèle de MLR suggéré car nous ne disposons pas de suffisamment de données expérimentales, mais les prévisions pourraient constituer une ligne directrice pour les expérimentateurs. Néanmoins, l'ensemble de formation utilisé pour construire le modèle est une bonne représentation de l'ensemble de données complet des agonistes partiels et des antagonistes, même si aucun antagoniste n'a été inclus dans l'ensemble de formation. Dans

l'ensemble, cette étude a montré que les calculs de la chimie quantique pouvaient constituer une aide précieuse pour les études expérimentales sur la conception de ligands pour les récepteurs du glutamate.

Utilisation d'outils de calcul hybrides pour rationaliser la liaison d'agonistes et d'antagonistes dans le recepteur NMDA

La protéine NMDA effectue des réarrangements conformationnels de l'état ouvert à l'état fermé lors de la liaison d'un agoniste au domaine de liaison du ligand en forme de coques flexible de la protéine [40]. L'efficacité du ligand est directement liée à la quantité de fermeture de la fente, de sorte que les agonistes complets établissent de fortes interactions avec les résidus du domaine de liaison du ligand en fermant étroitement la coque. Des fluctuations se produisent entre des états partiellement ouverts et fermés lors de la liaison d'agonistes partiels, car ces types de ligands exercent des forces plus faibles [41]. Les propriétés fonctionnelles qui déterminent le rôle d'un récepteur dans la transmission du signal sont déterminées par les caractéristiques structurelles du canal ionique et par le mécanisme du flux ionique. Les structures cristallographiques des récepteurs du glutamate contenant de nombreuses informations fonctionnelles montrent que la région de liaison au ligand communique par de courtes liaisons avec les hélices formant le canal ionique [41, 42]. La région de liaison au ligand est constituée de régions S1 rigides et de S2 flexibles et les modifications de la région S2 affectent directement les spirales connectées au canal ionique. Le modèle viable actuel du mécanisme d'activation du canal ionique suggère l'ouverture du canal ionique par des transitions conformationnelles après la fermeture du domaine de liaison du ligand lors de la liaison d'un agoniste [43, 44]. Pour comprendre les mécanismes d'activation ou d'inhibition du récepteur, il est important d'élucider les comportements conformationnels du domaine de liaison au ligand lors de la liaison de différents ligands.

Le domaine isolé de liaison aux agonistes de la sous-unité GluN1 a été cristallisé en tant que complexe avec des agonistes complets glycine (code PDB: 1PB7) et D-sérine (code PDB: 1PB8) [45], agonistes partiels D-cyclosérine (code PDB: 1PB9) [45], 1- acide aminocyclopropane-1-carboxylique (ACPC) (code PDB: 1Y20) et 1-aminocyclobutane-1- acide carboxylique (ACBC) (code PDB: 1Y1Z) [40]; antagonistes de la cycloleucine (code PDB: 1Y1M) [40], acide 5,7-dichlorokynurénique (DCKA) (code PDB: 1PBQ) [45] et 1-thioxo-1,2-dihydro- [1,2,4] triazolo [4,3-a] quinoxaline-4 (5H) -one (TK40) (code PDB: 4KFQ) [46]. Ces études suggèrent que les angles entre les régions S1-S2 dans trois types de ligands différents sont différents. Les antagonistes des sites glycine, DCKA et cycloleucine, stabilisent la conformation ouverte de cette région. La cycloleucine, un ligand de plus grande taille comparé aux agonistes partiels ACPC et ACBC, provoque l'ouverture du domaine de liaison du ligand, permettant ainsi aux molécules d'eau d'entrer et d'éliminer les effets hydrophobes entre les chaînes latérales de Phe484 et le groupe indole de Trp731. Enfin, la conformation en fente ouverte est stabilisée pour permettre des changements de conformation dans le canal ionique et pour réguler le flux du récepteur. En particulier, les effets des agonistes partiels sur la fonction du récepteur n'ont pas encore été pleinement compris. Dans le cas d'agonistes partiels du récepteur AMPA, il est connu que la quantité de fermeture de la fente est directement proportionnelle à l'activité du ligand [41, 47, 48]. D'autre part, il a été rapporté que certains agonistes partiels du récepteur

NMDA exercent le même degré de fermeture de la fente que celui observé chez les agonistes [40, 45, 49, 50, 51, 52].

L'amarrage par ligand est l'une des techniques les plus courantes pour estimer l'affinité de liaison d'un ligand à son récepteur cible [53]. Un processus d'amarrage de base vise à prédire la configuration (conformation, position et orientation) la plus favorable d'un ligand à l'intérieur de la cavité de liaison d'un récepteur [54]. Des liaisons hydrogène formées entre le ligand et la protéine ainsi que des interactions de Van der Waals et Coloumbic expliquent le score d'ancrage [55]. Plus de 60 logiciels d'amarrage sont utilisés. Malgré la facilité d'utilisation de ces logiciels d'accueil, des fonctions de notation médiocres sont obtenues en raison du nombre limité de degrés de liberté conformationnels, du manque de flexibilité des protéines et de l'utilisation de modèles de solvants implicites. Il est important de développer des méthodes permettant d'estimer avec précision l'énergie libre de la liaison, ce qui permet de prédire la force de liaison de tout médicament candidat avant sa synthèse. Par conséquent, il est plus intéressant d'appliquer des simulations de dynamique moléculaire à différentes poses issues du schéma d'arrimage et de calculer les énergies libres de liaison avec des méthodes telles que la perturbation d'énergie libre (FEP) [56], l'intégration thermodynamique (TI) [57], la réponse linéaire (LR) [58], méthodes de la surface moléculaire généralisée (MM-GB/SA) et de la mécanique moléculaire-Poisson-Boltzmann (MM-PB/SA) [59, 60]. Parmi ces méthodes, les plus précises et les plus rigoureuses sont FEP et TI [61]. Malgré leur précision, ils ont été peu utilisés dans la conception de médicaments en raison de leur convergence uniquement pour des ligands assez similaires et un coût de calcul [62, 63]. Les méthodes MM-GB/SA et MM-PB/SA combinant l'énergie mécanique et les modèles de solvatation implicites sont simples et plus rapides que le FEP et ont été largement utilisées pour les calculs d'énergie libre en chimie médicinale computationnelle [64, 65]. De nombreuses études de simulation en dynamique moléculaire ont été menées afin de comprendre les comportements conformationnels des récepteurs du glutamate avec une série de ligands et leurs interactions [5, 16, 51, 66, 67, 68, 69]. Il a été démontré que les distances de liaison hydrogène entre domaines (IHB) pouvaient être utilisées comme indicateurs de l'activité biologique (agonisme-antagonisme) des ligands du récepteur du kainate [67]. Suite à cette étude, une relation a été établie entre les distances de IHB et l'efficacité expérimentale de divers agonistes de NMDA GluN1 (partiel): plus la distance de IHB est petite, plus le ligand [16] est efficace. Récemment, Cifci *et al.* a pu montrer la corrélation entre les énergies de liaison et les valeurs de concentration inhibitrice demi-maximale (IC50) du système Phosphodiestérase IV-B [70]. Ils pourraient reproduire les valeurs expérimentales de la IC50 avec le protocole suggéré.

Dans ce chapitre, l'énergie libre de liaison de Gibbs pour les complexes récepteur naturel-ligand (NMDAR-ligand) a été calculée et corrélée à leur efficacité expérimentale ou à leur affinité de liaison pK_i. Le but ultime est de prédire les énergies de liaison des ligands proposés et enfin de différencier les agonistes, les antagonistes et les agonistes partiels en fonction de leur énergie de liaison de Gibbs estimée. Toutes les structures cristallines de rayons X disponibles du récepteur NMDA avec la sous-unité GluN1 dans la base de données PDB ont été collectées; toutes ces sources de protéines ont une homologie de séquence supérieure à 95%. Les domaines de liaison au ligand GluN1 ont été isolés et les ligands ont été retirés. Moyenne carrée des écarts par rapport à la structure cristalline de référence (code PDB: 1PB7) ont été calculés pour tous les PDB afin de les regrouper par rapport à leurs différences de conformation. La plupart des cas distincts ont été choisis en tant que représentants de différentes conformations de récepteurs

pour l'accostage; 1PB7 et 1PB8 représentant la conformation fermée du domaine de liaison au ligand; 1PB9, 1Y1Z et 1Y20 représentant la conformation partiellement fermée; 1Y1M, 1PBQ, 4KFQ et 4KCC représentant la conformation ouverte. Huit ligands avec des efficacités variables allant de 0 à 100 (%) ont été choisis pour les calculs d'énergie libre de liaison en tant qu'entraînement et deux agonistes partiels de l'étude d'Urwyler *et al.* [30] ont été sélectionnés de manière aléatoire pour l'ensemble de tests.

Les interactions du domaine de liaison du ligand de la sous-unité GluN1 du récepteur NMDA avec divers ligands ont été étudiées et les réponses conformationnelles du domaine de liaison du ligand à l'activité du ligand (agonisme contre antagonisme) ont été étudiées. Huit ligands de l'ensemble d'entraînement ont été attachés à neuf PDB et l'énergie de liaison la plus basse pour chaque procédure d'accostage a été sélectionnée comme meilleure pose. Un ajustement de régression linéaire entre l'énergie libre Autodock la plus basse des valeurs de liaison et les constantes d'affinité expérimentales des ligands a été obtenu pour l'ensemble d'apprentissage, mais aucune relation n'a été trouvée. La même étude a été réalisée afin de trouver un lien entre l'efficacité des ligands et la plus faible énergie libre de liaison Autodock et des correspondances relativement bonnes ont été observées avec des coefficients de corrélation de 0,748 et 0,658 dans deux modèles suggérés, ce qui signifie qu'AutoDock est capable de prédire la configuration la plus favorable des ligands dans le domaine de liaison du ligand, mais ne parvient pas à estimer l'énergie libre expérimentale des valeurs de liaison.

40 simulations indépendantes de la dynamique moléculaire ont été réalisées pour chaque meilleure pose à partir du processus d'accostage. Des simulations gratuites de MD ont été appliquées, mais elles ont entraîné un dépliement partiel des protéines. Par conséquent, une force harmonique de 10 kcal mol^{-1} $\mathrm{Å}^{-2}$ a été appliquée sur les atomes de l'épine dorsale de la protéine. Cinq modèles de solvatation différents ont été utilisés pour chaque calcul ($\mathrm{GB}^{\mathrm{HCT}}$, $\mathrm{GB}^{\mathrm{OBC}}$, $\mathrm{GB}^{\mathrm{OBC2}}$, $\mathrm{GB}^{\mathrm{neck}}$, $\mathrm{GB}^{\mathrm{neck2}}$). Les modèles de solvatation $\mathrm{GB}^{\mathrm{HCT}}$ et $\mathrm{GB}^{\mathrm{OBC}}$ ont réussi à reproduire les structures co-cristallines originales en donnant les énergies de liaison les plus basses dans cinq calculs (glycine: 1PB7, ACPC: 1Y20, D-cyclosérine: 1PB9, TK40: 4KFQ, DCKA: 1PBQ). Les corrélations entre l'efficacité expérimentale et les valeurs d'affinité des ligands avec les énergies libres de liaison calculées ont été vérifiées, mais des linéarités médiocres ont été obtenues avec tous les modèles de solvatation. Cela peut être dû au fait que la fonctionnalité du récepteur est étroitement liée aux sous-unités NMDA, y compris le canal ionique. Ceci doit être pris en compte lors de la réalisation des simulations. D'autre part, il existe une bonne corrélation entre l'activité antagonisme-agonisme du ligand et l'état de conformation du récepteur. Par ces corrélations, on peut distinguer l'activité d'agonisme d'un ligand d'une activité antagoniste, mais pas celle d'agonistes d'agonistes partiels.

Utilisation de charges atomiques pour décrire les pK_{a} des acides carboxyliques

La solubilité d'un ligand médicamenteux est régie par les pK_{a} des sites acides et basiques de la molécule. Lorsque le ligand se trouve dans l'environnement protéique cible, son affinité, son activité et son efficacité dépendent des pK_{a} des résidus en interaction du site actif.

L'acide carboxylique est le principal groupe fonctionnel acide dans les systèmes biologiques. Le glutamate et l'aspartate ont des groupes acide carboxylique dans leurs chaînes latérales et

ces groupes aident à maintenir le peptide ensemble par des liaisons hydrogène. Plus de 30% des résidus ionisables (32% des résidus Arg, 19% des résidus Asp, 13% des résidus Glu et 6% des Lys) [71] sont enfouis dans les cavités hydrophobes, ce qui limite le contact avec solvant [72]. La matrice protéique étant hétérogène, les fluctuations de l'environnement électrostatique modifient les interactions entre les charges enfouies, ce qui entraîne à son tour des modifications des affinités des sites de protonation pour l'ionisation; et ainsi leurs valeurs de pK_a sont réajustées [73]. Finalement, dans les parties polaires de la protéine, le pK_a des groupes acides des résidus passe aux valeurs les plus élevées et le pK_a des groupes de base aux valeurs les plus basses de celles des acides aminés isolés [74]. Les liaisons hydrogène entre les groupes fonctionnels de l'acide aminé et les atomes de la chaîne latérale ou du squelette ont également tendance à entraîner des déviations du pK_a; surtout lorsque le nombre de liaisons hydrogène augmente et si elles sont rigides, l'effet est plus important, de sorte que les pK_a des chaînes latérales acides sont perturbés au-dessus de leurs valeurs intrinsèques de pK_a et que l'inverse est observé pour les groupes de base [20, 75]. La formation de ponts salins entre deux résidus, qui contribuent à la stabilité des protéines, aurait également pour résultat des valeurs de pK_a plus basses ou plus élevées avec les mêmes tendances en termes d'effets de polarisation et de liaison hydrogène [76].

Les constantes de dissociation des composés organiques peuvent être déterminées expérimentalement par des techniques potentiométriques, spectrophotométriques, chromatographiques, électrophorétiques, calorimétriques, conductimétriques et NMR, mais parmi elles, seuls les titrages par NMR estiment les valeurs de pK_a avec précision lorsqu'ils sont appliqués à des acides aminés. Dans le cas de protéines membranaires non solubles, où la NMR est inapplicable, il est nécessaire de disposer de structures cristallines aux rayons X mais le nombre de structures présentes est limité dans la littérature [77]. D'autre part, une résolution médiocre des cristaux et un manque de flexibilité conformationnelle limitent les applications de cristallographie aux rayons X. Une méthode plus fiable pour la production de structures protéiques cristallographiques à haute résolution est la méthode de diffraction neutronique. Malgré sa grande précision dans la détermination des emplacements de l'hydrogène sur les acides aminés, cette technique est malheureusement beaucoup plus limitée en NMR et en cristallographie aux rayons X, car un petit nombre d'instruments sont disponibles pour les mesures et nécessite de longues durées de collecte de données. et des volumes plus importants de cristaux [78].

En raison des difficultés rencontrées lors de la mesure de pK_a de molécules dans des milieux complexes de grande taille et des intermédiaires de courte durée de vie avec des moyens expérimentaux, comme indiqué ci-dessus, le besoin d'estimations précises du pK_a au moyen d'applications d'applications théoriques est inévitable. Les caractéristiques qui déterminent les acidités de différentes classes de composés chimiques peuvent être expliquées par la structure moléculaire. La méthode traditionnelle de calcul des pK_a est basée sur les variations d'énergie libre dans le cycle thermodynamique. En règle générale, les inétarctions électrostatiques sont obtenues en résolvant numériquement l'équation de Poisson de Boltzmann linéarisée (LPBE). Malgré le grand nombre de prédictions de pK_a réussies en utilisant les énergies de déprotonation et les énergies libres de solvatation [79, 80, 81, 82], ces calculs échouent généralement en raison de l'instabilité de l'acide ou de sa base conjuguée en phase gazeuse. et les différences de conformation entre les calculs en phase solvant et en phase gazeuse [83]. Outre les méthodes empiriques telles que PROPKA et les méthodes basées sur l'équation de Poisson-Boltzmann, l'équation de Borné généralisée, QM/MM ou la dynamique moléculaire ou une combinaison

d'un ou de plusieurs; QSPR est une technique largement utilisée dans laquelle plusieurs descripteurs moléculaires sont liés avec succès à des pK_a de molécules organiques telles que l'état topologique [84, 85], le type d'atome [86, 87], le groupe philicity [88], la longueur et la fréquence de la liaison [89, 90], potentiel de surface maximal [91], énergies HOMO et LUMO [92, 93], charge atomique [94, 95]. Parmi ceux-ci, le concept de charge atomique partielle est étroitement lié à l'acidité et à la basicité relatives d'une molécule.

Un modèle MLR a été développé par Dixon et Jurs avec une précision de 0,5 unité pour le calcul des pK_a des oxyacides en utilisant les charges atomiques empiriques des atomes d'une molécule [95]. Le modèle est basé sur les modifications des charges sigma et pi lors du passage de l'état neutre à l'état ionique, concernant la résonance et les effets inductifs des atomes proches. Citra a construit quatre modèles de régression linéaire en utilisant les charges atomiques partielles sur les atomes d'oxygène et d'hydrogène impliqués dans la déprotonation et l'ordre des liaisons O-H pour l'ensemble des phénols, des alcools et des acides carboxyliques aromatiques et non aromatiques [96]. Différentes combinaisons de différents niveaux de théories, d'ensembles de base et de modèles de charge ont été testées par Vařeková *et al.* Afin de créer un modèle pour les phénols [97]. Récemment, Ugur *et al.* ont utilisé une approche similaire avec une étude approfondie pour prédire les acides aminés pK_a dans les protéines et ont mis au point un protocole précis en calculant la charge atomique sur la forme anionique des alcools et des thiols [98]. Parmi les fonctionnels DFT testés, les bases, les méthodes semi-empiriques, les modèles de solvatation et de charge, ils ont observé que la meilleure combinaison est le calcul de la charge NPA dans le modèle CPCM au niveau B3LYP / 3-21G (R^2 = 0.995) pour les alcools et M06-2X/6-311G (R^2 = 0.986) niveau théorique pour les thiols afin de reproduire les pK_a expérimentaux. De plus, ils ont testé la stabilité des pK_a calculés dans les acides aminés par des calculs MM-MD et DFT-MD.

En ce qui concerne les applications réussies des charges de gestion de la qualité en tant que descripteurs, notre étude visait à proposer un protocole précis pour la prévision rapide de pK_a d'acides carboxyliques. D'après la littérature, un total de 59 composés d'acide carboxylique avec des pK_a allant de 0,65 à 5,12 ont été sélectionnés [99, 100]. L'un de nos critères de sélection a été d'identifier des molécules capables de représenter le plus grand nombre possible de pK_a expérimentaux. La plupart de ces molécules sont aussi des molécules petites et plutôt rigides. Pour éviter les molécules flexibles, il faut éviter le minimum global lors de l'optimisation de la géométrie, ce qui provoquerait des erreurs systématiques dans les prédictions du pK_a. Un ensemble de formation de 30 petites molécules et un ensemble de test de 29 petites molécules ont été extraits de l'ensemble.

Huit différents fonctionnels de densité (BLYP [101, 102], B3LYP [101, 103], OLYP [101, 104], PBE [105], M06-2X [106, 107]) et quinze ensembles de base différents ont été utilisés. Pour interpréter l'environnement solvant aqueux, le modèle de solvatation basé sur la densité (SMD [108]), le modèle de continuum polarisable (PCM [109]) et le modèle de solvant à conducteur polarisable (CPCM [110]) ont été utilisés avec la constante diélectrique (ε) de 78,5. Trois types différents de modèles de charge atomique ont été testés: analyse de population de Mulliken [111], analyse de population de Löwdin [112], analyse de population naturelle (NPA [113]). Sauf indication contraire, tous les calculs de charge ont été effectués sur les géométries optimisées des formes anioniques des molécules (après prise en compte ou non de l'effet solvant) qui ne contiennent aucune fréquence imaginaire. Afin de tester la stabilité de la prédiction, des simulations de dynamique moléculaire ont été appliquées à un ensemble de pep-

tides (ACE-ASP-NME, ACE-GLU-NME, ACE-ALA) définis dans le champ de force Amber.

La relation linéaire entre les charges atomiques et les pK_a expérimentaux dépend de nombreux facteurs: le choix de la méthode DFT, le choix de l'ensemble de base, l'utilisation (ou non) d'un modèle de solvant implicite, le type de modèle de charge atomique, et lequel les charges atomiques sont considérées. De l'étude globale actuelle, il a été trouvé que la meilleure combinaison de tous ces facteurs est de considérer la charge atomique en oxygène la plus élevée de chaque fragment carboxylate calculée avec NPA au niveau M06L/6-311G(d,p) en utilisant le SMD implicite modèle de solvant. Dans ce qui suit, nous présentons la relation linéaire entre les charges atomiques expérimentales de pK_a et de fragment carboxylate calculées à l'aide du cadre théorique présenté ci-dessus. Ensuite, en utilisant ces résultats comme référence, nous discutons du choix du descripteur de charge, du modèle de charge, du modèle du solvant, de la fonction DFT et du jeu de base en modifiant l'un de ces paramètres, les autres restant fixés à leur meilleure combinaison.

Une équation linéaire est obtenue par un ajustement des moindres carrés

$$pKa = a \cdot \max q(O_1), q(O_2) + b \tag{2}$$

où a et b sont les paramètres ajustés et $\max q(O_1), q(O_2)$ est la charge calculée sur les atomes d'oxygène qui a la valeur maximale. Les pK_a prédits sont calculés en important $q(O)$ d'une molécule donnée dans l'équation paramétrée. On a trouvé que R^2 était 0,955 et qu'on n'observait pas de valeurs aberrantes fortes pour l'ensemble de molécules d'entraînement. La différence maximale entre le pK_a prédit et le pK_a expérimental parmi toutes les molécules était de 0,60 unité. Ces résultats indiquent une forte corrélation avec les pK_a expérimentaux et les charges sur O^-.

Afin de construire un modèle prédictif de pK_a pour les acides carboxyliques, la sélection du descripteur de charge approprié a été effectuée en vérifiant les charges atomiques sur les atomes des groupes carboxylate. Par conséquent, charge sur l'atome de carbone $Q = q(C)$, somme moyenne des charges sur deux atomes d'oxygène $Q = 0.5 * q(O_1) + q(O_2)$, charge minimale sur deux atomes d'oxygène $Q = \min q(O_1), q(O_2)$ et la somme des charges sur les atomes de carbone et d'oxygène $Q = q(C) + q(O_1) + q(O_2)$ ont également été utilisés en tant que descripteurs. Comme il n'est pas possible de faire la distinction entre O_1 et O_2, leur somme ou leur sélection de taxe minimale a été évaluée. Fait intéressant, la prédictivité pK_a de $Q = \min q(O_1), q(O_2)$ n'est pas aussi précise que $Q = \max q(O_1), q(O_2)$ et $Q = 0,5 * q(O_1) + q(O_2)$. D'autre part, il n'y a pas de corrélation trouvée entre la charge sur l'atome de carbone et le pK_a expérimental. De plus, l'addition de la charge atomique du carbone à la somme des charges O_1 et O_2 diminue la linéarité avec R^2 = 0,536. Des calculs de charges alternatifs ont également été effectués sur les mêmes géométries optimisées que celles utilisées pour les calculs de charges NPA. Tous les modèles de charge testés semblent avoir une corrélation beaucoup plus faible avec le pK_a expérimental; 0,733 pour l'analyse de population de Mulliken et 0,788 pour l'analyse de population de Löwdin. Pour ces modèles, presque toutes les molécules se sont avérées être écartées de la ligne d'appareillage. Les calculs PCM et CPCM sont aussi précis que les calculs SMD avec $R^2 \geq 0,930$. Le modèle de prédictivité en phase gazeuse est plus pauvre (R^2 = 0,826) par rapport aux autres modèles où les méthodes de solvatation PCM, CPCM et SMD ont été appliquées. Le modèle SMD diffère des modèles PCM et CPCM en prenant en compte les énergies de dispersion-répulsion en plus de l'énergie électronique. Ces termes supplémentaires semblent contribuer à trouver le minimum global d'optimisation de la géométrie et à

l'attribution des charges atomiques. Les plus grands MADs et MAX-ΔpK_a's ont été trouvés pour les combinaisons du jeu de base 3-21G avec toutes les fonctions sauf M06L. Pour le reste des combinaisons fonctionnelles et des ensembles de base, les valeurs MADs et MAX-ΔpK_a's sont acceptables. Dans tous les sous-ensembles, les plus grands MAD sont obtenus avec les fonctions M06-2X ou OLYP. La précision est diminuée lorsque des fonctions diffuses ont été ajoutées à l'ensemble de base pour n'importe laquelle des fonctions DFT. Les plus petites MAD ont été trouvées pour les combinaisons de tous les jeux de base avec la fonction M06L (sauf 6-31G) et parmi toutes les méthodes testées, M06L/6-311G(d,p) a donné le résultat le plus précis.

La stabilité de la prédiction a été testée en utilisant des formes dipeptidiques d'aspartate, de glutamate et d'acétylalanine. Des calculs de charge en un point ont été effectués sur des géométries obtenues à partir de simulations de dynamique moléculaire courte. Les valeurs moyennes de pK_a prédites obtenues sont en très bon accord avec les valeurs expérimentales de trois des peptides et près de 95% des prédictions se situent à l'intérieur de ± 1 unité de pK_a.

Élucidation de l'atroposélectivité dans la synthèse de dérivés de thiohydantoïne axialement chiraux

Les hydantoïnes, une classe d'imides cycliques, ont un large éventail d'activités biologiques et agricoles; ils sont également d'importants précurseurs dans la synthèse de plusieurs acides aminés et dérivés de l'acide pyruvique en raison de la diversité des méthodes de synthèse utilisées pour modifier la double liaison exocyclique[114]. Les analogues du soufre des hydantoïnes, les thiohydantoïnes, sont les échafaudages privilégiés de la chimie organique de synthèse moderne en raison de leurs propriétés anticancérigènes, [115] antiviraux, [116] antimicrobiens, [117] antithyroïdiens, [118] hypolipidémiques, [119, 120] applications antimutagènes [121, 122] en plus de leur utilisation comme pesticides, [123] réactifs pour colorants, [124] agents anti-ulcéreux et anti-inflammatoires, [125] agents C-terminaux le séquençage des protéines [126] et la complexation métal-cation [127]. L'importance des thiohydantoïnes provient de la présence d'atomes de soufre et d'oxygène qui jouent le rôle d'accepteurs de protons à liaison hydrogène et d'un atome d'hydrogène lié à l'azote qui joue le rôle de donneur de liaison hydrogène. [128] Il existe un certain nombre de méthodes de synthèse des thiohydantoïnes dans la littérature. [129] Une étude récente portant sur le transfert de la chiralité dans le cadre de la synthèse asymétrique a été réalisée via le traitement d'alpha-aminoesters énantiomériquement purs avec des électrophiles appropriés. [130] Sarigul et Dogan ont synthétisé un certain nombre de 2-thiohydantoïnes atropoisomères axialement chirales enrichies en diastéréoisomère via la stratégie de pool chirale par la réaction de sels d'esters d'acides L-amino et d'isothiocyanates d'o-aryle en présence de triéthylamine à reflux dans le dichlorométhane [22]. Les composés synthétisés possèdent un centre chiral en position C_5 et un axe chiral le long de N_{sp2} -C_{aryle} avec un substituant ortho attaché au cycle phényle. Les 2-thiohydantoïnes non axialement chirales sont des produits racémiques quels que soient les substituants R_1 et R_2. Il est bien connu que, lorsque la taille du substituant ortho augmente, la rotation autour de la liaison N_{sp2} -C_{aryle} est plus limitée et cette rotation empêchée donne lieu à des composés atropisomères. [131, 132, 133, 134, 135, 136, 137, 138, 139, 140] Les données expérimentales de Sarigul et Dogan montrent que le raccrochage de C_5 dans la 5-méthyl-3-o-bromophényl-2-thiohydantoïne

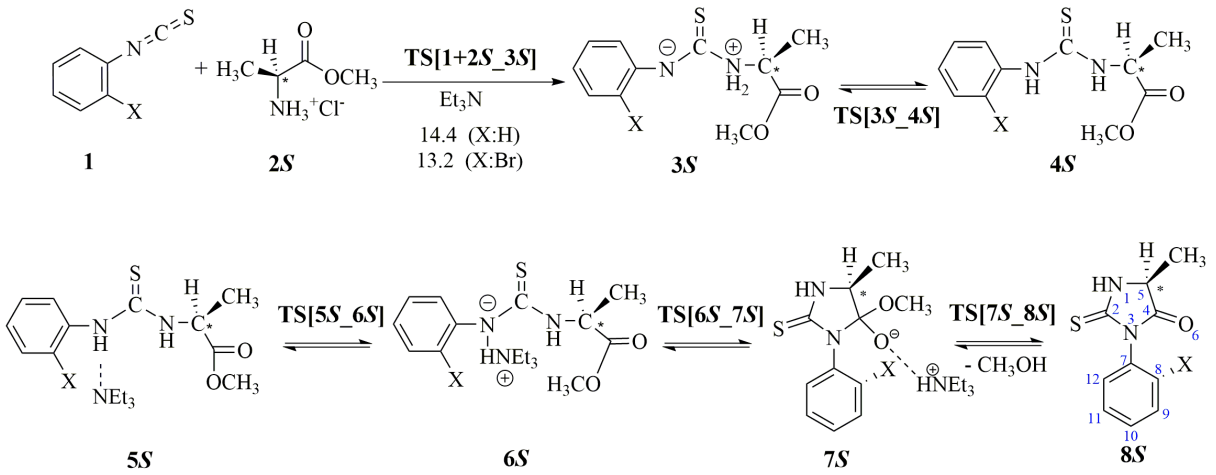


Figure 4: Mécanisme suggéré pour la synthèse du produit cyclisé **8S**.

est évité par le volumineux substituant o-aryle et que la chiralité du sel de l'ester méthylique de L-alanine est préservée. dans le produit final (la configuration SP est dominante dans laquelle l'atome de brome est transoïde par rapport au méthyle en C_5).

L'utilisation d'outils informatiques pour modéliser les mécanismes de réaction est importante pour comprendre les détails des étapes de racémisation afin de synthétiser des produits énantiomériquement purs. Le but de cette étude est de rationaliser l'atroposélectivité observée dans la synthèse des dérivés de 2-thiohydantoïne axialement chiraux en combinant les données actuelles et précédentes. Plusieurs questions sont ici abordées: la racémisation a-t-elle lieu avant ou après la cyclisation? La racémisation se produit-elle à l'aide de la triéthylamine déjà présente dans le milieu réactionnel ou du méthanol qui est un produit secondaire, ou par tautomérisation céto-énol non assistée? Des mécanismes plausibles ont été proposés pour la racémisation et la cyclisation du produit d'addition intermédiaire (3S) formé à partir de l'attaque initiale de l'azote de l'acide aminé sur le carbone isothiocyanate de thiocarbonyle. Cette étude est composée de deux parties: Dans la partie I, les mécanismes de synthèse ont été étudiés par les méthodes DFT afin de répondre aux questions posées en modélisant les différentes possibilités de formation de 2-thiohydantoïnes chirales. Dans la partie II, la synthèse ainsi que le mécanisme de conversion stéréoscopique de nouveaux dérivés atroposélective de la 2-thiohydantoïne - 5-benzyl-3-aryl-2-thiohydantoïne et de la 5-isobutyl-3-aryl-2-thiohydantoïne - ont été traités à l'aide d'outils informatiques.

La toute première étape du mécanisme de synthèse proposé commence par l'addition nucléophile de l'acide aminé (2S) à l'isothiocyanate de o-aryle (1) donnant l'intermédiaire 3S (Figure 4). À partir de 3S, nous proposons un itinéraire suivant la transformation de 3S à 6S, puis la formation du produit cyclisé 8S. En raison de la présence d'une séparation de charge dans 3S, le transfert de protons de l'azote thioamide chargé positivement (N_1) à l'azote arylamine chargé négativement N_3) se produit comme dans la structure donnant le rendement attendu 4S (Figure 4). Nos calculs montrent que cette étape est une réaction très exothermique. 6S qui est formé à partir de 4S via une complexation avec triéthylamine (5S) et une extraction d'hydrogène sur N_3 est légèrement exergonique par rapport au matériau de départ 3S. Les réactions de cyclisation peuvent se produire spontanément par chauffage ou nécessiter un catalyseur tel qu'un

acide de Bronsted ou un acide de Lewis ou une espèce de base.[141] Dans le cas de la cyclisation de 6S, la triéthylamine protonée dans 6S colle sur le méthoxy oxygène, une liaison N_3-C_4 se produit dans la première étape (TS [6S_7S]) et cette barrière de fermeture de cycle est très faible. Dans la deuxième étape, la libération de méthanol par transfert d'hydrogène à partir de triéthylamine protonée se produit très rapidement pour TS [7S_8S]. Les résultats expérimentaux de Sarigul et Dogan sont en harmonie avec les données de cyclisation calculées, car ils rapportent une étape de cyclisation très rapide.

Afin de préciser si l'énolisation est un processus viable dans la racémisation des thiohydantoïnes ou non, le transfert intramoléculaire de protons du stéréocentre C_5 à un atome d'oxygène lié doublement à un cycle hétérocyclique est modélisé avec et sans catalyseur basique. La tautomérisation céto-énolique sans catalyseur n'est pas cinétiquement favorable avec une barrière d'activation très élevée. Comme une mole de méthanol est libérée au cours de l'étape de cyclisation (TS [7S_8S]), elle peut extraire l'atome d'hydrogène au niveau du stéréocentre (C_5) hydrogène en transférant son hydrogène OH à un atome d'oxygène lié à un noyau hétérocyclique, ce qui donne l'énolate 8-Enol. Le méthanol agit comme un catalyseur basique qui contribue explicitement à la déprotonation et à la reprotonation de l'abaissement de la barrière d'énolisation en C_5, mais reste trop élevé pour qu'une réaction se produise dans les conditions de la réaction.

Etant donné que la tautomérisation des céto-énols dans le cycle thiohydatoïne avec ou sans l'aide de méthanol n'est pas responsable de la racémisation en raison des barrières d'activation élevées qui ne peuvent pas se produire dans les conditions de la réaction, il est suggéré de procéder à une déprotonation-reprotonation médiée par la triéthylamine en C_5. La barrière d'activation pour la déprotonation en C_5 par la triéthylamine est relativement basse et peut être conservée dans des conditions de réaction. H-10 (R_1 = CH3, X = H) n'est que de 0,9 kcal / mol d'énergie inférieure à celle du H-TS [9S_10] et du H-TS [10_9R], ce qui rend les réactions directes et inverses également favorables du point de vue thermodynamique et spontané. Br-10 (R_1 = CH3, X = Br) peut continuer via Br-TS [10_9R] avec un processus sans barrière et spontané; mais la réaction inverse via Br-TS [9S_10] peut également avoir lieu dans des conditions de réaction. Même si les deux conversions de Br-9S à Br-9R ou de Br-9R à Br-9S sont réalisables dans des conditions de réaction, la voie en arrière est légèrement plus favorisée que la voie en avant, ce qui donne un rapport S/R de 88/12 selon Fonction de distribution de Boltzmann, qui est en très bon accord avec le rapport expérimental rapporté (S/R: 86/14). En outre, plusieurs mécanismes de racémisation possibles des 3S, 4S et 6S avec et sans triéthlyamine ont été évalués afin de comprendre si la racémisation se produit avant ou après la cyclisation. La formation de 8S et 8R par la cyclisation de 3S, 3R, 4S et 4R a également été discutée. Bien que les obstacles à surmonter soient accessibles à température ambiante pour les énantiomérisations de Br-3S à Br-3R, Br-4S à Br-4R, Br-6S à Br-6R, la transformation de Br-8SP à Br-8RP est très favorable en raison de la barrière d'activation inférieure non compétitive. En revanche, si la racémisation se produisait avant la cyclisation, les produits les plus abondants en Br-8R auraient l'atropo-isomère M, car les données de calcul ont montré que les conformères les plus stables de Br-3R, Br-4R et Br-6R avaient des atomes de brome dans le synperiplanar. conformations par rapport au groupe méthyle dans le stéréocentre. Ainsi, la conformation synperiplanaire de l'atome de Br est maintenue à partir de Br-3S jusqu'à la thiohydantoïne finale formée, ce qui donne un atropoisomère Br-8SP dans lequel l'atome de brome est transoïde par rapport au méthyle en C_5. La rotation du cycle bromoaryle, Br-8SP à Br-8SM, a une barrière déter-

minée expérimentalement, 27,8 kcal/mol [22], et cette barrière relativement élevée explique le fait que le Br-8SP a une abondance de 83% alors que le Br-8SM n'est que de 3%. Une fois que la racémisation aura lieu en C_5 de Br-8SP, la Br-8RP sera formée et la rotation du noyau bromoaryle (RP en RM) nécessite 27,5 kcal / mol.

La tautomérisation des céto-enols serait l'un des principaux mécanismes responsables de la racémisation des produits cyclisés [142, 143, 144]. Par conséquent, afin de clarifier si l'énolisation est un processus viable de racémisation ou non, nous avons modélisé le transfert intramoléculaire de protons de C_5 à O_6 avec et sans catalyseur basique. Nous avons sélectionné la molécule Benzyle-8SP comme composé modèle en raison de la présence de brome en position ortho du noyau aryle et du groupe benzyle en C_5. La tautomérisation céto-énolique sans catalyseur n'est pas entraînée cinétiquement en raison de la barrière d'activation très élevée. Nous proposons donc un procédé de racémisation à relais de protons, dans lequel des molécules d'éthanol explicites assistent le transfert intramoléculaire de protons de C_5 à O_6, formant l'intermédiaire de l'énol. Nous avons modélisé toute la réaction avec une et deux molécules d'éthanol explicites. Les états de transition localisés pour la déprotonation / reprotonation en C_5 du noyau hétérocyclique via une assistance à l'éthanol comportent une structure cyclique à six chaînons entre le Benzyle-8SP et la molécule d'éthanol. À la suite d'une réorganisation concertée des liaisons dans laquelle un double transfert de protons se produit entre C_5-O éthanol-O6, un complexe énol-éthanol est formé. Par ailleurs, la déprotonation par l'assistance de deux molécules d'éthanol a une barrière d'activation beaucoup plus faible, dans laquelle les structures à l'état de transition comprennent une structure cyclique à huit chaînons impliquant un triple transfert de protons. Des liaisons hydrogène plus courtes dans deux mécanismes assistés par l'éthanol indiquent des états de transition plus stabilisés pour l'énolisation et la cétonisation. La barrière d'activation expérimentale pour la racémisation de Benzyle-8 (SP à RP) est de 24,4 kcal/mol; a suggéré deux mécanismes assistés à l'éthanol avec une énergie d'activation de 25,5 kcal/mol (l'étape déterminante du taux est l'étape de reprotonation) correspond très bien à la valeur déterminée expérimentalement.

Enfin, nous avons calculé les proportions relatives de Benzyle-8SP en termes de rapport isomère S/R selon le mécanisme de racémisation suggéré dans la première partie de cette étude, le rapport obtenu de 56/44 est en très bon accord avec les données rapportées de manière expérimentale. Les méthodes de calcul utilisées dans cette étude peuvent être utilisées avec confiance pour prédire la chiralité dans la synthèse atroposélective de dérivés de thiohydantoïne axialement chiraux.

Chapter 1

Introduction

Glutamate, one of the major neurotransmitters in the brain and spinal cord, exerts its postsynaptic effects through a series of membrane receptors. Ionotropic glutamate receptors (iGluRs) are located at the membranes of neuronal cells and play key roles in synaptic plasticity. There are four different subtypes of iGluRs based on their pharmacology and sequence similarity: N-methyl-D-aspartate receptor (NMDAR), kainic acid receptor (KAR), α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPAR) and Delta receptor [1] (Figure 1.1).

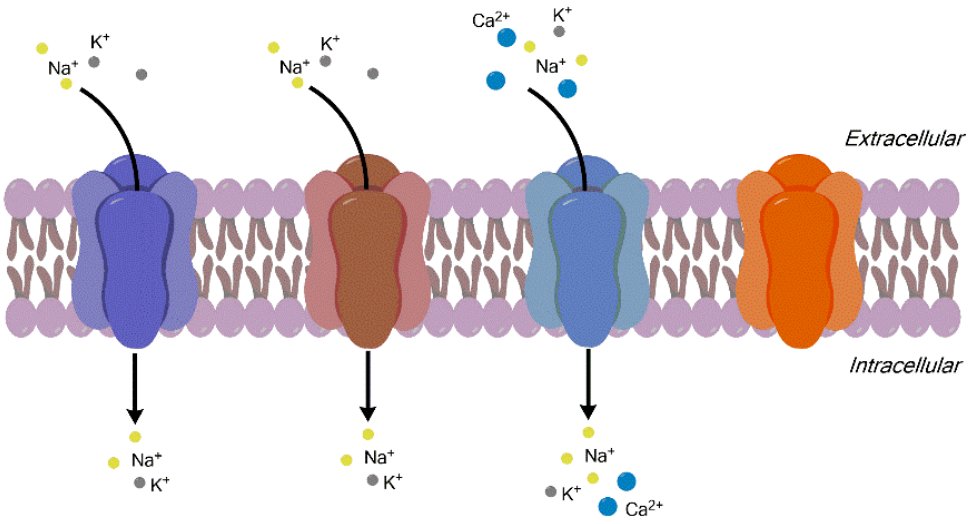


Figure 1.1: Subtypes of ionotropic glutamate receptors: AMPAR, KAR, NMDAR and Delta receptor (from left to right)

AMPA and kainate receptors mediate fast synaptic transmission in the central nervous system [2]. In the NMDA receptor, postsynaptic currents increase and decrease more slowly compared to AMPA and kainate receptors. The NMDA receptor is also more sensitive to voltage-dependent blocking, the receptor's ion channels permit the passage of the calcium ions in addition to monovalent sodium and potassium ions, and it needs the simultaneous binding of glycine and glutamate for the activation of the ion channel [3]. On the other hand, Delta receptors are

completely unresponsive to glutamate and their functional significance has not yet been understood [4].

Monomers of iGluRs are formed of an extracellular N terminal domain (NTD), an intracellular C terminal domain (CTD), three transmembrane domains (M1, M2 and M3), a re-entrant membrane loop (P-loop), and a ligand binding domain (LBD) [5] (Figure 1.2). The receptor consists of two GluN1 and two GluN2 or GluN3 subunits. While GluN1 is responsible from the activation of the ion channel, GluN2 and GluN3 are responsible from the functional properties such as glutamate affinity, receptor desensitization and pharmacological sensitization [6, 7]. It is known that NMDA receptor plays a key role in excitatory synaptic transmission such as learning, memory and synaptic plasticity, thus the malfunctioning of the receptor is involved in many neurodegenerative disorders like epilepsy, parkinson's disease, alzheimer and schizophrenia. Under normal conditions of synaptic transmission the ion channel of the receptor is blocked by the magnesium ions. The binding of native agonist glycine to the GluN1 subunit and co-agonist glutamate to the GluN2 subunit activates the ion channel and permits the flux of calcium, potassium and sodium ions into neurons and this process is very important in memory formation and neurotoxicity [8, 9, 10, 11]. The receptor can be inactivated during its resting period or can be exposed to 'excitotoxic cell death' due to the prolonged stimulation of the receptor. Thus, the GluN1 subunit is the allosteric modulatory site which makes it unique in receptor activation [12]. Regulation of receptor activity by introducing agonist-antagonist ligands to the glycine and glutamate sites became a promising strategy for the treatment of neuropsychiatric disorders, which are related to dysfunction of NMDA. Hence, the binding sites of these subunits are important targets of extensive medical research.

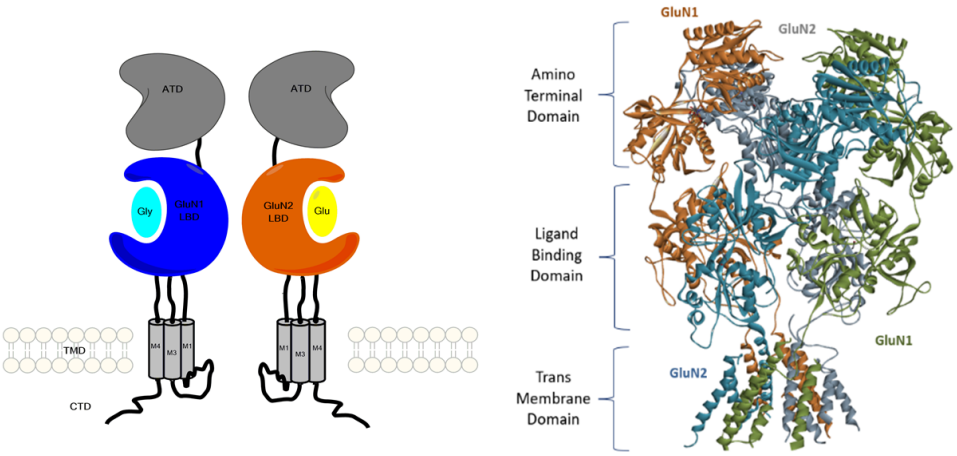


Figure 1.2: Schematic representation of iGluR subunit topology (left) and heterotetrameric GluN1-GluN2 NMDA receptor (right)

According to the ligand binding essays and functional investigations, the receptor function is modified by the competition of the interacting ligands with glycine at the glycine binding site [13]. Agonists keep the LBDs closed, resulting in the opening of the NMDA ion channel, while antagonists act as a wedge between the LBDs, as such preventing the opening of the ion channel

and overstimulation of the receptor (Figure 1.3)[14]. However, it is well known that complete activation or blockage of the receptor has many serious side effects like memory impairment, neurotoxicity or psychotomimetics [15]. On the other hand, a partial agonist can act either as a full agonist or antagonist, i.e. in the presence of excess antagonist, a partial agonist acts like an agonist to reduce the full inhibition effects of the antagonist, depending on the glycine concentration at the receptor. For this reason, employment of partial agonists is becoming more applicable since these compounds are able to regulate the activity at a certain level [16].

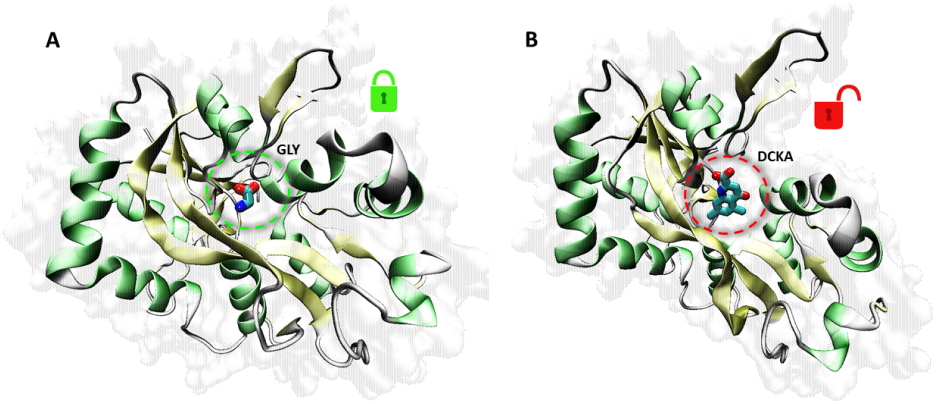


Figure 1.3: Cartoon representation of the GluN1 subunit of the NMDA receptor. Glycine (GLY) bound structure (PDB code: 1PB7) refers to the closed-cleft conformation (A), whereas 5,7-dichlorokynurenic acid (DCKA) bound structure (PDB code: 1PBQ) refers to open-cleft conformation (B) [45]

Binding affinity (K_i) and efficacy data of a drug molecule together determine the potency of that drug. Interaction of a ligand with its target receptor through intermolecular forces such as hydrogen bonds, Coulombic interactions and Van der Waals forces is characterized with its binding affinity. Once the ligand binds to its receptor, the extent to which the receptor is activated is determined by its efficacy [17]. Antagonists have almost zero efficacy, inactivating the receptor, while full agonists have 100% efficacy which is interpreted as full activation, with partial agonists having efficacy scaling up to an agonist and down to an antagonist.

A large number of chemical and biological systems contain acidic and basic groups which govern the interactions with the surroundings and therefore the function of the system. At a particular pH, the extent to which an ionizable species can be protonated or deprotonated by the hydrogen transfers from/to the environment is determined by the pK_a of the species. Most of the drug molecules are weak acids or weak bases and when they are in solution they are in their both ionized and nonionized states. The pK_a of a drug molecule effects its solubility in lipid, and thus its diffusion, since only the uncharged ligands can penetrate into the cell membrane [18]. Besides, the interactions between the ionizable functional groups of a drug molecule with

the residues of the protein, which effects the affinity, activity and efficacy of the ligand, is higly dependent on the pK_a's of the side chains in the active site and the drug molecule. On the other hand, the changes in the protonation states of the aminoacids in the residues have a direct impact on establishing the protein conformation and stability [19], solubility and folding [20], catalytic activity of enzymes [21] and binding ability. Therefore, gaining insights of relative pK_a's of ligands and the protonation sites of the receptor may explain the key features of ligand binding.

Besides the factors governing the interactions of a ligand with its environment, it is important to understand the intrinsic properties of the ligand. It is well known that, stereoisomers of a chiral molecule may show different biological activities. A ligand can be stereoselective or regioselective depending on the differences in electronic and steric environment. For this reason, the demonstration of configurational stabilities of drug candidates and mechanisms by which they racemize are important issues. Recently, Sarigul and Dogan [22] synthesized 2-thiohydantoin derivatives as racemic and nonracemic diastereomeric pairs of atropisomers, a specific class of compounds which are considered to be among the privileged scaffolds in drug discovery with their agonistic-antagonistic activities. The synthesis of 5-methyl-3-phenyl-2-thiohydantoin, which lacks chiral axes, gave racemic mixtures. The axially chiral 5-methyl-3-*o*-bromophenyl-2-thiohydantoin may exist in SP, RM, SM and RP isomeric forms. SP/RM is transoid, whereas is SM/RP is cisoid with respect to the substituent at C_5 and the *o*-aryl substituent. Using computational tools to model the reaction mechanisms is important for understanding the details of the racemization steps in order to synthesize enantiomerically pure products.

One of the important issues in drug design is the identification of the biological activity of receptor ligands. Development, synthesis and activity measurements of ligands have a major importance in drug design. However, there are certain limits in experimental studies; synthesis of a large number of compounds to cover all the potentially active molecules is unrealistic. Computational studies could therefore provide a valuable aid to experimental studies on ligand design for glutamate receptors. By combining the strengths of Molecular Dynamics and Quantum Chemical approaches, a more focused inspection, characterisation and rationalization of the drug design studies is allowed to be established.

Chapter 2

Objective and scope

The proper function of N-methyl-D-Aspartate (NMDA) receptor is critical in memory formation and synaptic plasticity; complete activation or blockage of the receptor has many serious side effects like memory impairment, neurotoxicity or psychotomimetics. Therefore, regulation of receptor activity by introducing agonist and antagonist ligands to the glycine site (GluN1 subunit) became a promising strategy for the treatment of neuropsychiatric disorders. The identification of the biological activity (agonism versus antagonism) of receptor ligands is one of the crucial issues in drug design. In the present work, we wish to verify which intrinsic properties of structurally resembling ligands cause differentiation between antagonistic and agonistic behaviour by using various computational tools. The interactions between the ligand and the ligand binding domain will be inspected in order to gain insights into the activation/inhibition mechanisms of the receptor. Moreover, the stereoselectivity of a class of ligands will be elucidated.

In the previous section of the dissertation, a brief introduction about the NMDA receptor, agonism-antagonism activity of ligands, factors governing the interactions of a ligand with its environment and importance of the selectivity of a ligand are provided. A more detailed discussion of each one will be given in the following chapters.

In Chapter 3, the fundamental principles of the performed computations, including Density Functional Theory, Molecular Dynamics and docking methods, are introduced. Within the relevant chapters comprehensive explanations on the methodologies are presented.

Following the brief overview of the theoretical methods, the results of this study will be introduced in 4 chapters. In the following two chapters, we aim to differentiate between agonists, antagonists and partial agonists based on Quantum Chemical descriptors (Chapter 4) and binding Gibbs free energies (Chapter 5). Several molecular properties that could play a role in ligand binding to the glycine GluN1 subunit of NMDA and calculated binding Gibbs free energies are further used to provide a link between the efficacies and binding affinities of the ligands. Prediction of the acid dissociation constants of amino acids in proteins and ligands allows us to have information about the binding affinity and efficacy of the ligand to its target protein. In Chapter 6, we will explore how atomic charges of carboxylic acids can be related to the prediction of pK_a of the ligands. A special case will be presented in Chapter 7 in order to understand the stereoselectivity of thiohydantoin ligands, which are potent androgen receptor antagonists, by investigating various mechanistic synthesis pathways. The crucial conclusions drawn from each chapter and concluding remarks are highlighted in Chapter 8.

Chapter 3

Theoretical background

This chapter provides basic principles of most commonly used theoretical approaches in this dissertation including Quantum Mechanics, Molecular Dynamics and Molecular Docking methods.

3.1 Quantum Mechanics

3.1.1 The Schrödinger Equation

Quantum mechanics provides mathematical tools in order to describe properties of microscopic systems. The behaviour of electrons in molecules under the influence of the electromagnetic field exerted by nuclear charges is described by Schrödinger equation

$$\hat{H}\psi = E\psi \tag{3.1}$$

where the total energy of the system E is expressed as an eigenvalue of the Hamiltonian H and ψ is the wave function which is the mathematical function that describes the spatial distribution of electrons and nuclei in the system. The Hamiltonian operator is the sum of kinetic energy operator and potential energy operator:

$$\hat{H} = \hat{T} + \hat{V} \tag{3.2}$$

For a particle with mass m, the kinetic energy operator ($\hat{T}$) can be expressed in three dimensions as:

$$\hat{T} = -\frac{\hbar^2}{2m}\nabla^2 \tag{3.3}$$

where

$$\nabla^2 = \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right] \tag{3.4}$$

and

$$\hbar = \frac{h}{2\pi} \tag{3.5}$$

For an n particle system, the kinetic energy operator is represented as:

$$\hat{T} = -\frac{\hbar^2}{2m_i}\sum_{i=1}^{n}\nabla^2 \tag{3.6}$$

The potential energy operator for a system of charged particles i and j is

$$\hat{V}(\mathbf{r}) = \sum_{i>j}\frac{Z_i Z_j e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{3.7}$$

where the atomic number of particle is represented as Z, charge on the electron is e and the distance between particles is r.

For a molecular system composed of electrons and nuclei, the potential energy operator in Equation 3.7 becomes:

$$\hat{V}(\mathbf{r}) = -\sum_{A,i}\frac{Z_A e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_A - \mathbf{r}_i|} + \sum_{A>B}\frac{Z_A Z_B e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_A - \mathbf{r}_B|} + \sum_{i>j}\frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{3.8}$$

In Equation 3.8, the electron-nuclei attraction is defined with the first term, nuclei-nuclei repulsion is accounted in the second term and electron-electron repulsion is represented in the last term.

Therefore Equation 3.2 can be written as:

$$\hat{H} = T_n + T_e + V_{e-n} + V_{n-n} + V_{e-e} \tag{3.9}$$

The Schrödinger equation can be solved accurately for systems of two particles. For many particle systems, the motions of all of the particles are considered as correlated and thus some approximations are needed to solve the Schrödinger equation. The nuclei are much heavier than electrons therefore their motions are much slower than electron motions and the electronic wave function depends upon the nuclear positions but not upon their velocities. Born-Oppenheimer Approximation assumes that the positions of the nuclei can be considered to be fixed. This allows the separation of the Hamiltonian into nuclear $\hat{\mathscr{H}}_n$ and electronic $\hat{\mathscr{H}}_{el}$ parts. Equation 3.9 can be written as:

$$\hat{H} = T_n + V_{n-n} + \hat{\mathscr{H}}_{el} \tag{3.10}$$

The nuclear kinetic energy term in Equation 3.9 is therefore neglected, and nuclear-nuclear potential energy term is considered as a constant. Thus, the electronic structure of a molecule can be computed by the electronic Schrödinger equation:

$$\hat{\mathscr{H}}_{el}\psi_{el} = E_{el}\psi_{el} \tag{3.11}$$

And the total energy of the system is given by:

$$E_{total} = E_{el} + V_{n-n} \tag{3.12}$$

One of the main approximation methods used in quantum mechanics in order to solve Equation 3.11 is the variational method which allows us to estimate the energy of the ground state

of a many particle system. Variational principle states that the expectation value of the Hamiltonian which is computed with any trial wave function (Φ) is always higher than or equal to the energy of the ground state (ε_0). Mathematically,

$$\frac{\int \Phi^* \hat{H} \Phi}{\int \Phi^* \Phi} \geq \varepsilon_0, \tag{3.13}$$

An approximation to the ground state can be found by varying Φ inside a given set of functions and looking for the function that minimizes the expectation value of the Hamiltonian. Hartree-Fock Theory and Density Functional Theory are based on variational principle.

3.1.2 Hartree-Fock Theory

In the Hartree-Fock method, the electrons are considered as occupying single-particle orbitals making up the many-electron wave function. By an effective potential each electron experiences the presence of electrons in other orbitals. The overall electronic wave function of a system composed of N number of electrons is defined as a Slater determinant which is constructed by antisymmetrized product of one-electron wave functions, satisfying the Pauli Exclusion Principle.

$$\Psi^{SD}(\vec{x_1},\vec{x_2},...,\vec{x_N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x_1}) & \chi_2(\vec{x_1}) & \dots & \chi_N(\vec{x_1}) \\ \chi_1(\vec{x_2}) & \chi_2(\vec{x_2}) & \dots & \chi_N(\vec{x_2}) \\ \vdots & \vdots & & \vdots \\ \chi_1(\vec{x_N}) & \chi_2(\vec{x_N}) & \dots & \chi_N(\vec{x_N}) \end{vmatrix} \tag{3.14}$$

A spin orbital (χ) is simply the product of a spatial orbital (φ) and the spin function (g) in a given coordinate r. The value of $g(m_s)$ can be either α or β depending on the value of the quantum number m_s.

$$\chi_i(r) = \varphi(r) g(m_s) \tag{3.15}$$

Molecular orbital coefficients are varied according to the variational principle and the overall wave function is optimized in an iterative manner until no further changes occur. This procedure is called the Self-Consistent Field (SCF). Hartree-Fock equation is used for the minimum energy calculations of the corresponding orbital of energy ε_i:

$$\hat{F} = \varepsilon_i \chi_i \tag{3.16}$$

The Fock operator for each electron i ($\hat{F}_i$) is expressed by:

$$\hat{F}_i = -\frac{1}{2}\nabla_i^2 \sum_k^{nuclei} \frac{Z_k}{r_{ik}} + V_{HF(i)} \tag{3.17}$$

$V_{HF(i)}$ is the Hartree-Fock potential which accounts for the average repulsive potential experienced by each electron due to the other electrons. The SCF strategy uses some guess wave functions to construct the Fock operator, and then solve the Schrödinger equation. The procedure is then iterated by using the output functions as new input functions or with more sophisticated methods until the input and output functions are the same. If the convergence fails, the

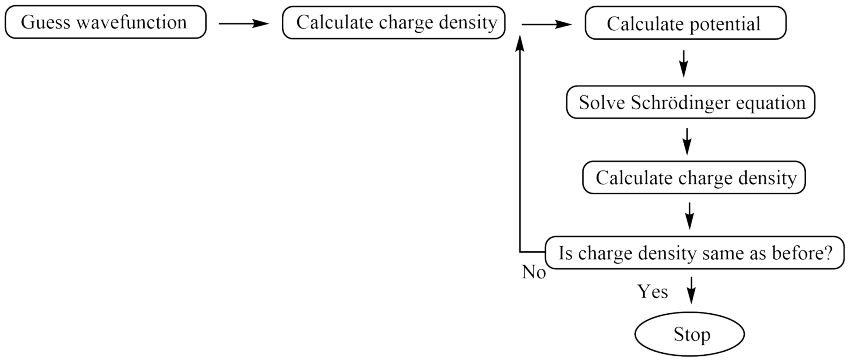


Figure 3.1: General procedure for Self Consistent Field strategy

trial functions are varied and the process is iterated upon till self-consistency is attained to yield numerical solutions to the Hartree-Fock potential.

Electrons as charged particles expose Coulomb repulsions and the motion of one electron has an impact on the motion of the others. Hartree-Fock theory stems from the description of the dynamic electron correlation as induced by their instantaneous mutual repulsion. Methods based on the calculation of the wave function including the electron correlation effects like Configuration Interaction, Moller-Plesset Perturbation Theory and Coupled Cluster are computationally very expensive. Density functional methods are potentially capable of very accurate results at low cost.

3.1.3 Semi-Empirical Methods

Semi-empirical methods are established upon the HF equations and they involve experimental data, approximations or assumptions. Since the core orbitals do not contribute to the chemical properties as much as valence orbitals, they are not treated by semi-empirical methods. In these methods, many of the integrals are either manipulated or parametrized in order to overcome the computational cost. Thus, parametrization quality determines the accuracy of these methods. Several semi-empirical methods have been developed on parametrization with different approximations.

Most of the modern semi-empirical methods are based on the Zero Differential Overlap approximation (ZDO) which neglets the overlap between different basis functions centered on different atoms:

$$\chi_{\mu}^{A}(i)\chi_{\mu}^{B}(i) = 0 \quad if \quad A \neq B \tag{3.18}$$

The various ZDO models can be grouped according to their approximations for the one and two electron integrals:

CNDO: complete neglect of differential overlap [145]

INDO: intermediate neglect of differential overlap model [146]

NDDO: neglect of diatomic differential overlap model [147]

Many methods have been developed based on NDDO approximation. In the MNDO method (modified neglect of diatomic overlap) [148], the repulsions between atoms which are separated by their van der Waals distances are overestimated. AM1 (Austin Model 1) [149] treats the hydrogen bondings better than MNDO but the misrepresentation of the hydrogen bonds is still a problem. PM3 (parametrized model number 3) [150] method uses the same equations and formalism as the AM1 method but the number of parameters for each element is different.

3.1.4 Density Functional Theory

For the past 30 years, Density Functional Theory (DFT) has been the method of choice for the quantum mechanical simulations of many-body systems. DFT, in contrast to Hartree-Fock Theory which deals directly with the wave function, is based on the electron density. The foundations of the current application of DFT is the Kohn and Sham formalism. In the Kohn-Sham approach, a fictitious non-interacting system of electrons is constructed with single particle orbital wave functions in which the same external potential energy function acts on all non-interacting electrons. Then the energy functional can be divided into:

$$E[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \tag{3.19}$$

where $E_{KE}[\rho(r)]$ is the kinetic energy of the non-interacting electrons (Equation 3.20), $E_H[\rho(r)]$ represents the electron-electron Coulombic energy which does not take into account the correlation between motions of electrons (Equation 3.21) and $E_{XC}[\rho(r)]$ corresponds to the exchange-correlation energy which accounts for the correction to the kinetic energy arising from the interacting nature of electrons of opposite spin (correlation term) and non-classical corrections to the repulsion energy between electrons of the same spin (exchange term) (Equation 3.22).

$$E_{KE}[\rho(r)] = \sum_i^N \int \psi_i(r) \; -\frac{\nabla^2}{2} \psi_i(r) dr \tag{3.20}$$

where $\psi_i(r)$ represents the Kohn-Sham molecular orbitals.

$$E_H[\rho(r)] = \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 \tag{3.21}$$

$$E_{XC}[\rho(r)] = \int \rho(r)\varepsilon_X[\rho(r)]dr + \int \rho(r)\varepsilon_C[\rho(r)]dr \tag{3.22}$$

The wave function of this system with N number of electrons can be expressed as a Slater determinant of one-electron functions χ_i and the electron density can be defined as:

$$\rho(r) = \sum_i^N < \chi_i | \chi_i > \tag{3.23}$$

From the Equations 3.19, 3.20, 3.21 and 3.22, the hamiltonian h_i^{KS} of the Kohn Sham equations

$$h_i^{KS} \chi_i = \varepsilon_i \chi_i \tag{3.24}$$

can be expressed as:

$$h_i^{KS} = -\frac{\nabla^2}{2} - \sum_k^{nucleus} \frac{Z_k^2}{|r_i - r_k|} + \int \frac{\rho(r)}{|r_i - r_j|} dr + V_{XC} \tag{3.25}$$

The exchange correlation potential V_{XC} is obtained by differentiating Equation 3.22.

$$V_{XC}[\rho(r)] = \varepsilon_X[\rho(r)] + \varepsilon_C[\rho(r)] + \frac{\partial \varepsilon_X(r)}{\partial \rho} + \frac{\partial \varepsilon_C(r)}{\partial \rho} \tag{3.26}$$

Since the evaluation of the exchange-correlation term is not easy, one can use approximations. This term includes the contributions of exchange and correlation together with the difference in kinetic energy between the real system and the fictitious non-interacting system. Construction of $E_{XC}[\rho(r)]$ by Local Density Approximation (LDA) which assumes that the electron density is uniform everywhere gives:

$$E_{XC}^{LDA}[\rho(r)] = E_X^{LDA}[\rho(r)] + E_C^{LDA}[\rho(r)] = \int \rho(r)\varepsilon_X^{LDA}\rho(r)dr + \int \rho(r)\varepsilon_C^{LDA}\rho(r)dr \tag{3.27}$$

where

$$E_X^{LDA}[\rho(r)] = C_X \int \rho(r)^{4/3} dr \tag{3.28}$$

and ε_X^{LDA} representing the exchange energy per electron:

$$\varepsilon_X^{LDA} = C_X \rho^{1/3} \tag{3.29}$$

with C_X being a constant equal to 0.7386. The correlation energy $E_C[\rho(r)]$ is based on the results of Monte Carlo calculations for a homogeneous electron gas of different densities [151, 152, 153, 154].

Local Spin Density Approximation (LSDA) allows different orbitals for electrons with different spins for the spin polarized systems. $E_{XC}^{LSDA}[\rho(r)]$ is expressed as:

$$E_{XC}^{LSDA}[\rho(r)] = E_{XC}^{LSDA}[\rho_\alpha(r), \rho_\beta(r)] = E_X^{LSDA}[\rho_\alpha(r), \rho_\beta(r)] + E_C^{LSDA}[\rho_\alpha(r), \rho_\beta(r)] \tag{3.30}$$

where $E_X^{LSDA}[\rho_\alpha(r),\rho_\beta(r)]$ is

$$E_X^{LSDA}[\rho_\alpha(r), \rho_\beta(r)] = -2^{1/3} C_X \int [\rho_\alpha^{4/3}(r) + \rho_\beta^{4/3}(r)] dr \tag{3.31}$$

and exchange energy per electron ε_X^{LSDA} is described as:

$$\varepsilon_X^{LSDA} = -2^{1/3} C_X [\rho_\alpha^{1/3} + \rho_\beta^{1/3}] \tag{3.32}$$

In contrast to LDA methods, generalized gradient approximation (GGA) assumes that the electron density is inhomogeneous and includes gradients for the variation of ρ with position.

$$E_{XC}^{GGA}[\rho_\alpha(r), \rho_\beta(r)] = \int f(\rho_\alpha(r), \rho_\beta(r), \Delta\rho_\alpha, \Delta\rho_\beta) \tag{3.33}$$

where f is a function of ρ_α and ρ_β, and their gradients. E_{XC}^{GGA} is divided into exchange and correlation parts and each part is modelled independently.

The most popular E_X functionals are PW86, B88, PW91, PBE and the E_C functionals are LYP, PW91, PBE, P86 [155] and any exchange functional can be used with any correlation functional; BLYP functional [101, 102] is an example for combination of B88 exchange functional and Lee-Yang-Parr correlation functional [101, 156]. Hybrid DFT functionals incorporate a portion of non-local Hartree-Fock exchange (HFX) with local/semi-local DFT/GGA exchange in the E_{XC} term obtained from Kohn-Sham orbitals.

$$E_{XC}[\rho(r)] = a_0 E_X^{HFX}[\{\psi_i\}] + (1-a_0)E_X^{DFT}[\rho(r)] + E_C^{DFT}[\rho(r)] \tag{3.34}$$

where a_0 coefficient denotes the fraction of HFX [157]. The hybrid functionals used in this dissertation are B3LYP [101, 103], OLYP [101, 104], PBE [105], PBE0 [158], M06 [106, 107], M06L [107, 159] and M06-2X [106, 107].

3.1.5 Basis Sets

A basis set is a set of functions used to build the molecular orbitals, which are expanded as a linear combination with coefficients to be determined. Slater type orbitals (STO) are the natural basis functions due to their similarity with the eigenfunctions of the hydrogen atom. They are simply defined as:

$$S_v(r) = e^{-\varsigma v r} \tag{3.35}$$

Nevertheless, difficulties in mathematical integration has restricted the use of STOs. Therefore Gaussian Type Orbitals (GTO) were suggested as an alternative where the exponential $e^{-\varsigma v r^2}$ is evaluated rather than $e^{-\varsigma v r}$ of the STOs in Equation 3.35. Contracted GTOs which are constructed from primitive Gaussian functions are recommended to be used since GTOs do not represent the electron density of the real situation as well as the STOs. The simplest form of contracted Gaussians are the STO-nG minimal basis sets which attempt to approximate STOs by n-Primitive Gaussians. Although they are much cheaper, they provide rough results as they include only one contracted Gausssian per atomic orbital. Split valence basis sets were introduced by Pople which exhibit more flexibility in the valence basis functions than in the core since valence electrons are more active in chemical bonding [160]. For example, the split valence double-ς basis set uses only one basis function for each core atomic orbital, and uses two basis functions for the valence atomic orbitals. 3-21G, 6-21G, 4-31G, 6-31G are the split valence double-ς basis sets and 6-311G is the split valence triple-ς basis set. The first number represents the number of primitive gaussians used in the core functions and the numbers after the hypen are the number of functions used for the valence orbitals.

Including polarization functions in the basis set improves the flexibility of the basis set and allows the atomic electron densities to be polarized in order to represent the electron density better in bonding regions. Polarization functions add higher angular momentum orbital to heavy atoms and they are denoted in Pople's sets by an asterisk (*) or (d). Polarization can also be added on light atoms like hydrogen and helium by a second asteriks or (d,p). In circumstances where electrons may be far from the nucleus, diffuse basis functions are added to the basis set to broaden the electron distributions. Diffuse functions are important when considering anions, excited states and other large, soft molecular systems to accurately represent the tail portion

of the atomic orbitals, which are distant from the atomic nuclei. One plus (+) adds a set of diffuse *s* and *p* orbitals to the heavy atoms and the second plus adds a set of diffuse *s* functions to hydrogen.

The level of DFT functional and the basis sets should be selected according to the properties and the behavior of the system modeled. Diffuse functions can also be added along with polarisation functions. This leads, for example, to the 6-31+G*, 6-31++G*, 6-31+G** and 6-31++G** basis sets.

3.1.6 Atomic Charges

Although many different methods have been developed for assigning quantitative values to the amount of electron density for each atom in a molecule, the partial atomic charge on an atom is not a quantum mechanical observable. The partial atomic charge on a positively charged atomic center (Z_A) which is shielded by an electron cloud can be determined by [161]:

$$q_A = Z_A - \int \rho_A(r) dr \tag{3.36}$$

The total number of electrons (N) is the integration over all electron density part in Equation 3.36 and it is expressed as:

$$N = \sum_{\mu}^{AO} (PS)_{\mu\mu} \tag{3.37}$$

where $\boldsymbol{P}$ is the electron density and $\boldsymbol{S}$ is the overlap population summed over all atomic orbitals. In Mulliken population analysis [111], all electronic charge contributions from atomic orbitals are summed up and electronic overlap clouds between two atoms are divided between atomic orbitals which contribute to the corresponding overlap. The charge q_A is then expressed as the difference between the number of electrons on the free atom which is equal to the atomic number Z_A and the total number of electrons:

$$q_A = Z_A - \sum_{\mu \in A}^{AO} (PS)_{\mu\mu} \tag{3.38}$$

Mulliken population analysis is applicable if basis functions centered on the nuclei are used. Since the partial charges calculated from Mulliken analysis are very sensitive to basis set size, comparisons at different level of theories have no meaning. Löwdin charge [112] is an effort for improvement of Mulliken charge by eliminating the problem of overlap partition. Thus Löwdin charges are much less sensitive to basis functions compared to Mulliken charges.

$$q_A = Z_A - \sum_{\mu \in A}^{AO} (S^{1/2} P S^{1/2})_{\mu\mu} \tag{3.39}$$

Another orbital based atomic charge is the Natural Population Analysis (NPA) [113] in which the orbitals are orthogonalized and localized to form one or two center natural bond orbitals. These orbitals are partitioned to core, valence and Rydberg orbitals in order to provide the most accurate Lewis structure of the molecule.

Hirshfeld population analysis [162] partitions the molecular electron density $\rho_{mol}(r)$ into a sum of atomic densities $\rho_A(r)$ according to the Equation 3.40. Then the atomic charges are computed by using the Equation 3.36

$$\rho_A(r) = \frac{\rho_A^0(r)}{\sum_B \rho_B^0(r)} \rho_{mol}(r) \tag{3.40}$$

The last population analysis method used in this dissertation is the CM5 scheme [163] which builds upon the Hirshfeld charges by including parametrized-charge dependent observables such as dipole moment.

3.1.7 Solvation Models

In quantum chemical calculations, the effect of the solvent should be taken into account for reactions that occur in solution phase. For this purpose, solvation models are useful in order to mimic the solvent environment of molecular systems. Description of the solvent effects can be done by explicitly or implicitly. In explicit solvation models all solvent molecules are explicitly represented, whereas in implicit solvation models the solvent is represented as a continuous medium.

In implicit solvation models, the solute molecules are placed in a cavity in the uniform and polarizable continuum medium with fixed dielectric constant and the solvation free energy of a system (ΔG_{sol}) is computed by:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{cav} + \Delta G_{disp} \tag{3.41}$$

in which electrostatic solute-solvent interaction energy is represented by ΔG_{elec} term, free energy required to form the solute cavity is the ΔG_{cav} term and the dispersion energy due to the solvent-solvent dispersion forces corresponds to ΔG_{disp} term.

Polarizable Continuum Model (PCM) [164] has been implemented in various quantum mechanical calculation packages. The solute is embedded in a cavity based on union of spheres centered on each atom with a radii defined by van der Waals radius and the polarization charges are placed into the small domains called tesserae. In Conductor like PCM (CPCM) model [165], the surrounding medium is modeled as a conductor instead of a dielectric. Integral Equation Formalism PCM (IEFPCM) model [166, 167] overcomes the problem of assumption of entire encapsulation of solute charge density in the cavity by reformulation of dielectric PCM. The solvation model based on density (SMD)[108] applies the IEFPCM protocol, solves the non-homogeneous Poisson equation using atomic Coulomb radii. The non-electrostatic contributions are calculated on the basis of a parameterised function which includes terms for atomic and molecular surface tensions as well as the solvent accessible surface area.

3.2 Molecular Mechanics

Molecular Mechanics (MM) treats atoms as perfect spheres and the bonds between atoms as springs. The energy of molecular systems which contain significant number of atoms is calculated by classical mechanics due to the very high computational costs arising from quantum

mechanical calculations. The potential energy of the system is defined by force fields which are parametrized mathematical expressions and consists of the addition of the bond stretching energy (V_{str}), the energy due to the bond-angle bending (V_{bend}), the energy due to the internal rotation about bonds (torsion) (V_{tors}), the energy due to the van der Waals (V_{vdW}) and electrostatic (V_{elec}) interactions. First three terms in Equation 3.42 represent the energies due to the bonded interactions and last two terms represents the energies due to the non-bonded interactions.

$$V = \sum_{bonds} V_{str} + \sum_{angles} V_{bend} + \sum_{dihedrals} V_{tors} + \sum_{pairs} V_{vdW} + \sum_{pairs} V_{elec} \tag{3.42}$$

One of the force fields that was developed for proteins, nucleic acids, and many related organic molecules in condensed phases is the Assisted Model Building with Energy Refinement (AMBER) [168] and the potential energy is given by:

$$\begin{aligned} V = & \sum_{bonds} K_{str}(r-r_0)^2 + \sum_{angles} K_{bend}(\theta-\theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1+cos(n\phi-\gamma)] \\ & + \sum_{pairs} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \sum_{pairs} \frac{q_i q_j}{\varepsilon R_{ij}} \end{aligned} \tag{3.43}$$

where K_{str} is the bond force constant and K_{bend} is the angle bending force constant, r is the bond length, θ is the bond angle, V_n is the amplitude of the barrier in the potential energy surface of rotations about dihedral angle ϕ, over periodicity of n, with the minimum phase angle of γ, R_{ij} is the distance between the centers of the non-bonded atoms i and j, q is the point charge on atoms i and j and ε is the dielectric constant of the medium. MM provides static ideal geometry of a system by finding the minimum of a potential energy surface. Since the interactions of atoms are dynamic and the molecules are in continuous motion in reality, one can obtain motion dependent conformational changes through Molecular Dynamics simulations.

3.2.1 Molecular Dynamics Simulations

Molecular Dynamics (MD) calculates the time dependent behavior of a molecular system. Detailed information on the fluctuations and conformational changes of proteins and nucleic acids is achieved by integrating Newton's laws of motion. The force acting on an accelerating particle with mass m_i is given by:

$$\mathbf{F_i} = m_i \mathbf{a_i} \tag{3.44}$$

Force is derived from the potential energy function with respect to the internal coordinates:

$$F_i = -\frac{\partial U(r)}{\partial r_i} \tag{3.45}$$

Combining Equation 3.44 and Equation 3.45 gives the Newton's second law:

$$-\frac{\partial U(r)}{\partial r_i} = m_i \frac{\partial^2 \mathbf{r}}{\partial t^2} \tag{3.46}$$

Since the motions of all the particles are coupled together, analytical solution to the Equation 3.46 is not possible. There are many methods in order to perform step-by-step numerical

integration of the coupled ordinary differential equations such as Verlet Algorithm [169], Leap-Frog Algorithm [170] and Velocity-Verlet Algorithm [171].

In MD simulations, infinitely large systems are represented by a set of boundary conditions using a small part called a unit cell. This unit cell is replicated throughout the space in all three cartesian directions to form an infinite lattice. All the periodic copies of the particles, which are called images, move solidary in the same direction during the simulation so that the number of particles within the simulation box is always conserved. This procedure is called periodic boundary conditions (PBC) and the Ewald sum technique is commonly used to efficiently threat long range interactions [172].

Thermodynamic properties of the systems are obtained by using statistical ensembles. Ensemble is the probability distribution of the states of a system and it is defined by a small set of parameters such as the pressure (P), the temperature (T), energy (E), volume (V), chemical potential (μ) and the number of particles (N). The most common ensembles in molecular dynamics are microcanonical ensemble, canonical ensemble, isobaric-isothermal ensemble and grand canonical ensemble. In the microcanonical ensemble (NVE) the system is isolated with constant number of particles, volume and energy. In the canonical ensemble (NVT) the number of particles in the system, the volume and the temperature are fixed. The isobaric-isothermal ensemble (NPT) is characterized by constant number of atoms, pressure and temperature. The thermodynamic state of grand canonical ensemble (μVT) is described by a fixed chemical potential, volume and temperature.

For the simulations of biological systems, the canonical ensemble is preferred due to its computational efficiency. Different thermostat algorithms are used in order to adjust constant temperature during the simulations. These algorithms use the relationship between the kinetic energy (E_K), and the instantaneous temperature (T):

$$E_K = \frac{N_{df}}{2} k_\beta T = \frac{1}{2} \sum_i^{N_{df}} \frac{p_i^2}{m_i} \tag{3.47}$$

where N_{df} represents the number of degrees of freedom, k_β is the Boltzmann's constant, p_i and m_i express the linear momentum and the mass of the particle i respectively.

In the Andersen thermostat [173] the system is thermally coupled with a fictitious heat bath that imposes the desired temperature. The method is stochastic and randomly alters the velocity of thermostated particles by choosing from a Maxwell-Boltzmann distribution. The Berendsen thermostat [174] uses velocity rescaling periodically to maintain the temperature. In this approach the system is assumed to be weakly coupled to a heat bath and the velocities of particles are rescaled at each time step based on targeted average kinetic energy. In the Langevin thermostat [160] at each time step all particles are exposed to a random force and by a constant friction their velocities are lowered. The momentum of the particles follow a Maxwell-Boltzmann distribution at the desired temperature.

3.2.2 Binding Free Energy Calculation Methods

The free energy behaviour of chemical and biochemical processes is among the most critical thermodynamic quantities. Protein-ligand binding constants and membrane-water partition coefficients can be calculated accurately by the knowledge of the associated free energy changes.

Most of the free energy calculation approaches are generally based on estimating the relative free energy differences between two states since ΔG values reflect the difference in the thermodynamic properties.

Free Energy Perturbation (FEP) [56] and Thermodynamic Integration (TI) [57] are the most accurate and rigorous methods among the ones that directly calculate the free energy difference and these calculations are based on the changes in the molecular structure during the molecular dynamics or Monte-Carlo simulations. However, these methods are computationally very expensive. Molecular Mechanics-Poisson Boltzmann/Surface Area (MM-PB/SA) and Molecular Mechanics-Generalized Born/Surface Area (MM-GB/SA) methods [59, 60] are computationally more efficient in terms of cost and simplicity. These methods require the direct simulation of end-points of only the bound and unbound states by using implicit solvation. The procedure can briefly be described in three steps. In the first step, coordinate sampling is prepared for the MD simulations of the protein-ligand complex in order to sample configurations for energy analysis. In the second step, gas phase potential energies and solvation free energies are calculated from an ensemble average. In the final step, estimated change in entropy is calculated. Then the final binding free energy is calculated by:

$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuum} + \Delta G^0_{solv,complex} - \Delta G^0_{solv,ligand} - \Delta G^0_{solv,receptor} \tag{3.48}$$

and $\Delta G^0_{bind,vacuum}$ is obtained by:

$$\Delta G^0_{bind,vacuum} = \Delta E^0_{MolecularMechanics} - T\Delta S^0 \tag{3.49}$$

The solvation energy term G_{solv} is typically obtained by solving the Poisson-Boltzmann (PB) equation or by using the generalized Born (GB) model.

3.3 Molecular Docking

Molecular docking is one of the most used computational tools for understanding and predicting the molecular recognition of a ligand by its target receptor both structurally and energetically. The docking process involves the prediction of the ligand's most favourable configuration (pose) and assessment of its binding affinity inside the binding cavity of a receptor. Search algorithms implemented in docking software sample the conformations of the ligand in the active site of the protein and scoring functions rank these conformations.

Computationally generating all possible conformations is very expensive, thus search algorithms are used to create an optimum number of configurations that include the experimentally determined binding modes. Search methods can either be systematic or stochastic. Systematic search methods change the conformation of the ligands gradually by varying the structural parameters sligthly [175]. The energy landscape of conformational space is explored and then convergence of the minimum energy for the most likely binding mode is achieved. Unfortunately in some cases the algorithm may fail to converge to a local minima instead of global minima. Systematic search methods are more suitable for rigid protein-protein dockings and are used in softwares such as DOT [176], GRAMM [177] and ZDOCK [178]. In contrast to systematic search algorithms, stochastic methods change the conformation of the ligands by random modifications [175]. A wide range of conformational ensembles are generated and thus

probability of finding a global minimum is increased. Stochastic search methods are commonly used in flexible ligand-protein dockings and Monte Carlo [179, 180] and Genetic Algorithm [181, 182, 183] belong to the class of stochastic methods.

The distinction between the binding poses from the non-binding poses or active compounds from inactive compounds is enabled by scoring functions. Binding free energies are estimated by these functions adopting various assumptions and simplifications and can be force field based, empirical or knowledge based [184]. Force field based scoring functions calculate the binding energy by summing the bonded (bond stretching, angle bending, dihedral variation) and non-bonded (electrostatic and van der Waals interactions) terms using the equations of classical mechanics [185]. These methods fail in estimating of the entropic contribution. DOCK [186] and GoldScore [182] are examples of force field based scoring functions. In empirical scoring functions, each energy component involved in the formation of the ligand-receptor complex (hydrogen bonding, ionic and apolar interactions, hydrophobic effect and binding entropy) is multiplied by a coefficient and then summed up to give the binding energy [185]. Coefficients are obtained from regression analysis fitted to a series of ligand-protein complexes with known binding affinities, thus the accuracy of the obtained data relies on the accuracy of the data used to develop the model. Surflex [187] and FlexX [188] are examples derived from empirical scoring functions. Knowledge based scoring functions utilize the interatomic contact frequencies and distances between the ligand and protein by the analysis of complex crystal structures to obtain a general potential function. PMF [189] and DrugScore [190] are examples of knowledge-based functions.

Chapter 4

Discrimination of the agonist and antagonist activity of ligands binding to the NMDA receptor

4.1 Introduction

Quantitative structure-activity relationships (QSAR) derive models which correlate the ligand's biological activities with its molecular structure either by physicochemical parameters (Hansch analysis), by indicator variables encoding different structural features (Free Wilson analysis), or by three-dimensional molecular property profiles of the compounds (comparative molecular field analysis, CoMFA). 1962 may be considered as the year of birth of modern QSAR methodology with the first QSAR publication of Corwin Hansch [23].

Quantum chemical descriptors encode chemical information about intrinsic properties of ligands, which then can be associated with their biological activities. Different methodologies have been applied over the recent years in order to associate structural properties of ligands with their biological activities. A CoMFA model was suggested in 2003 by Tikhonova *et al.* for selectivity of ligands between AMPA and NMDA receptors [10]. This has been an important approach in protein-ligand selectivity problem for the synthesis of novel biotargeting drugs. In order to predict pK_i values of partial agonists binding to the NMDA GluN1 subunit, Cheng *et al.* developed a QSAR model derived from Radial Distribution Function in which Van der Waals volume and mass were used as descriptors [24]. The experimental activities were reproduced by the model they suggested. Thus, it would be possible to have information about activity of molecules which have not yet been synthesized. Based on semi-empirical quantum-chemical, electro-topological state, molecular property and shadow index descriptors, Zhu *et al.* constructed a reliable computational model for the classification of agonists and antagonists of the 5-HT1A receptor by machine learning and genetic algorithm techniques [25]. Electronegativity, atomic polarizability, van der Waals volume and mass, atomic charge and partition coefficient (logP) as descriptors appear to be important in the most recent QSAR studies performed to identify both agonist and antagonists for the A3 Adenosine receptor [26]. Djeradi *et al.* used Fukui indices as descriptors in order to predict IC_{50} values of flavonoids [27].

Concerning the NMDA receptor, Yosa *et al.* showed in 2009 that GluN1 subunit selective

agonists and partial agonists could be differentiated from antagonists by their LUMO energies [28]. In their paper, the authors state, however, that the two classes of compounds they investigated, i.e. (partial) agonist versus antagonist compounds, belong to separate chemical compound classes, so that no structural overlap exists between them to allow for similar LUMOs to be found amongst the two populations. They suggest to test their results for antagonists that chemically resemble the existing agonists set, or vice versa.

In this study, intrinsic properties of structurally resembling GluN1 subunit selective ligands for NMDAR that cause differentiation between antagonistic and agonistic behavior were verified by quantum chemical calculations. Several global and local molecular properties that could play a role in ligand binding to the GluN1 subunit of NMDA were examined and these descriptors were then used to discriminate the partial agonism-antagonism nature of the ligands. In addition, it was verified whether these chemical descriptors could provide a link between the chemical traits of molecules and their efficacies, i.e. the degree of agonistic effect.

4.2 Methodology

30 ligands with antagonist activity [28, 29] and 30 ligands with partial agonist activity[30] have been selected from the literature (2D structures are given in Table 4.1 and Table 4.2) All possible conformations of the ligands were located with the semi-empirical PM3 method [150] by using the SPARTAN software [191]. Free rotations around single bonds were taken into account and all the geometries corresponding to stationary points have been considered for further study. It is known that DFT methods [192, 193] give better accuracies than semi-empirical methods in geometry optimizations and prediction of structural, electrostatic and spectroscopic properties of organic molecules [194]. All the stationary points emerging from free rotations around single bonds have been evaluated with PM3 and were re-optimized with the Gaussian09 program package [195], using B3LYP/6-311G**, the best conformer for each molecule was used further for the calculation of the descriptors. The B3LYP [101, 102] functional is known to be good in reflecting the geometries of organic compounds to eventually detect the conformation with the lowest energy. This is especially important to account for intramolecular hydrogen bonds, which play a role in the electronic charge distribution of the various conformers for the same species. The same procedure was applied to a randomly selected set of ligands (189, 192, 210, 213, 217, 29, 41, 24, 38, 44; Z9, Z15, Z19, Z22, Z26, Z29, Z35, Z39, Z46, Z51) by using M06-2X/6-311G** to test the functional dependency of the descriptors used. ‘Z’ prefix is used within the text for partial agonists which are in their zwitterionic states. Note that all the results reported in this study were obtained with the B3LYP/6-311G** unless otherwise mentioned. The biological environment is water; therefore, the optimizations were performed with the Polarizable Continuum Model (PCM) [166, 167, 196].

Table 4.1. Two-dimensional representations of selected antagonist molecules (*[28]; **[29]).

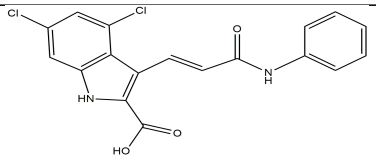
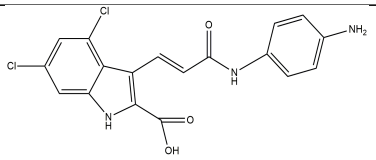
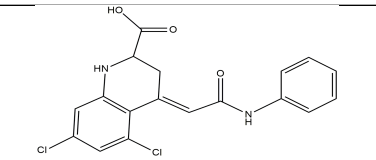
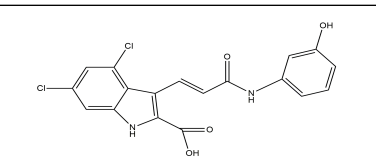
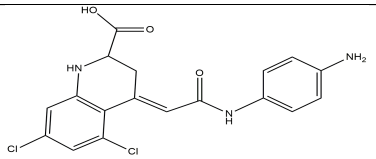
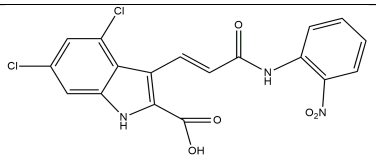
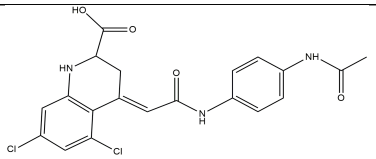
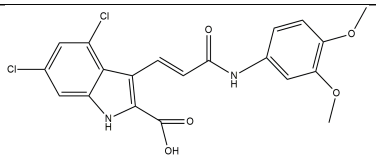
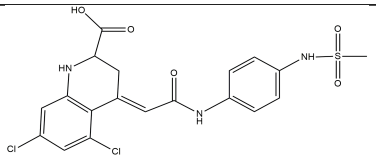
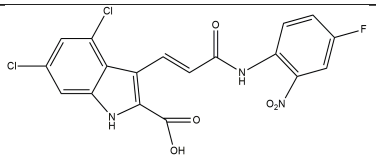
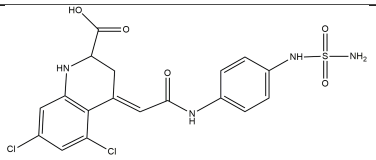
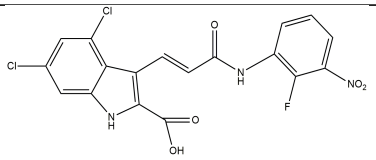
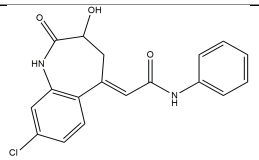
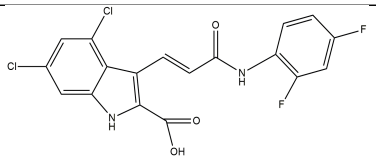
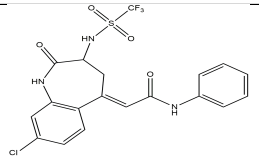
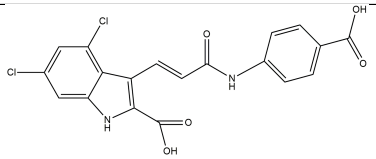
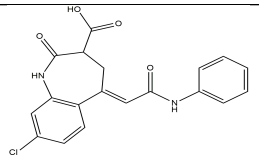
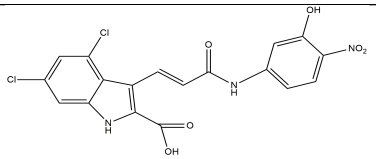
ID	2D Structure	pK_i	ID	2D Structure	pK_i
189*		8.5	24**		8.9
190*		8.2	26**		8.4
191*		7.8	28**		7.6
192*		8.5	29**		8.1
194*		8.2	31**		7.5
195*		8.2	32**		8.3
210*		5.9	33**		8.3
211*		5.6	34**		7.2
212*		7.5	35**		8.0

Table4.1 – continued from previous page

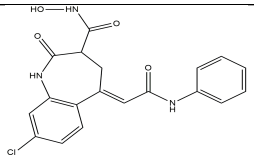
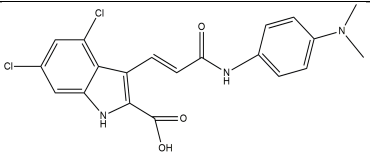
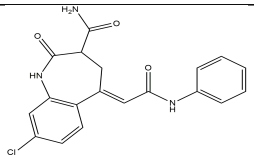
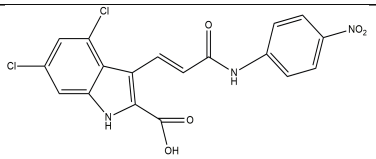
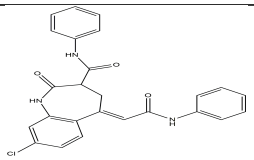
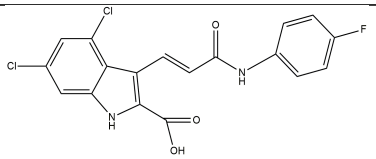
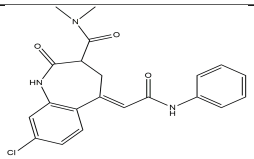
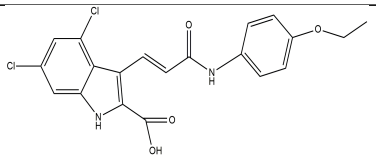
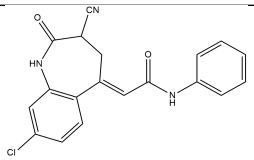
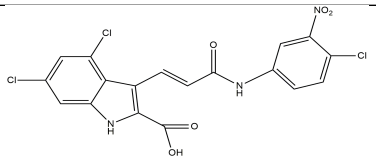
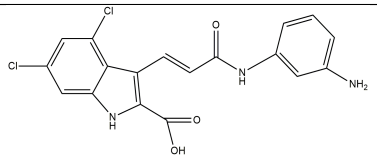
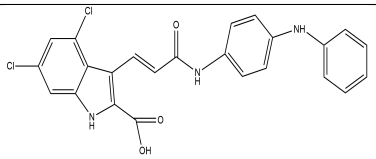
ID	2D Structure	pK_i	ID	2D Structure	pK_i
213*		7.4	36**		7.9
214*		7.4	37**		7.0
215*		7.1	38**		8.2
216*		4.6	39**		8.3
217*		6.5	41**		6.9
22**		8.3	44**		6.7

Table 4.2. Two-dimensional representations of selected partial agonist molecules [30]

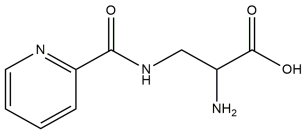
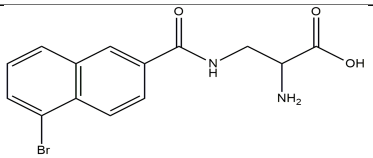
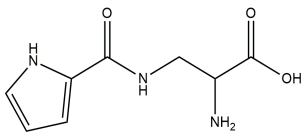
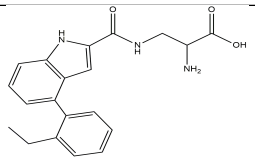
ID	2D Structure	pK_i	ID	2D Structure	pK_i
8		4.4	25		6.6
9		4.9	26		7.6

Table4.2 – continued from previous page

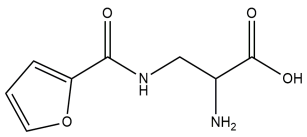
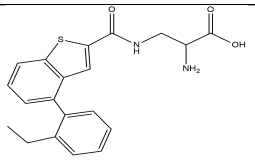
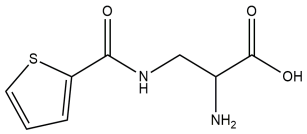
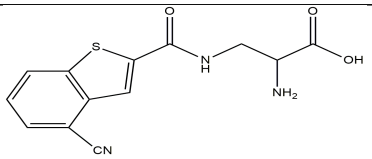
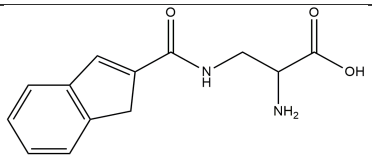
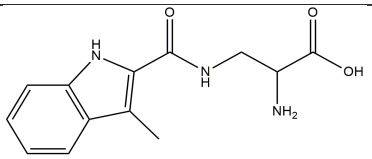
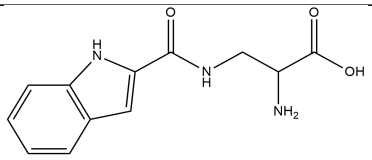
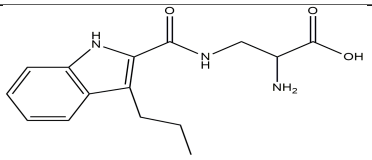
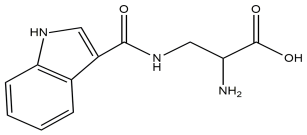
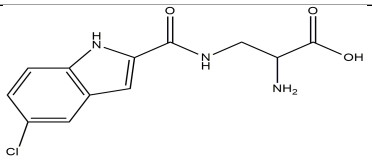
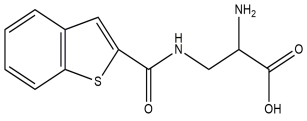
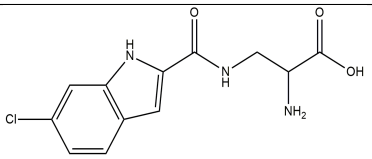
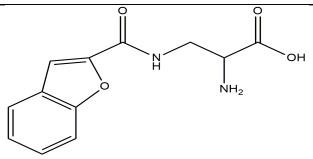
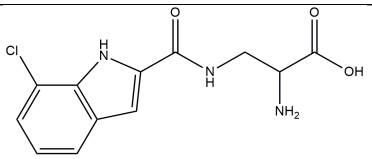
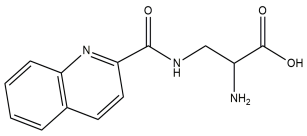
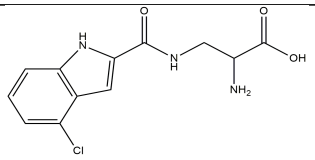
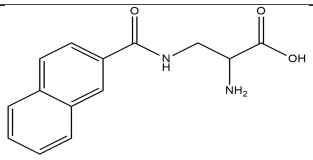
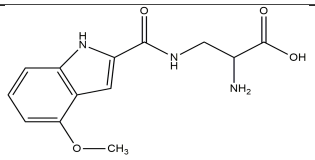
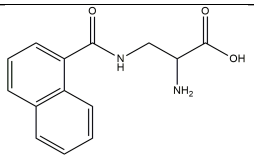
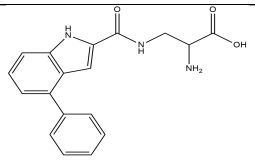
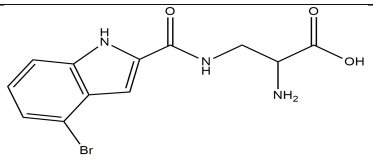
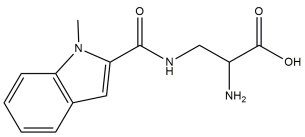
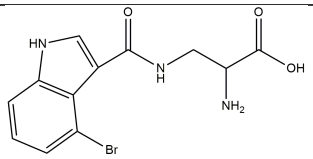
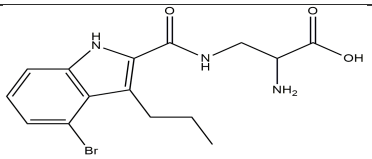
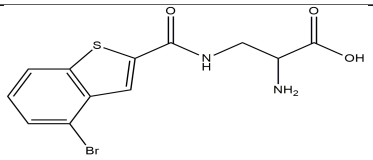
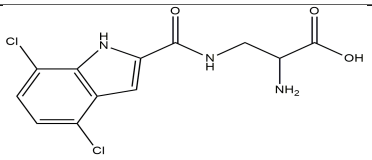
ID	2D Structure	pK_i	ID	2D Structure	pK_i
10		4.9	27		6.2
11		5.1	29		5.9
13		5.5	32		5.5
14		5.3	33		6.3
15		5.4	34		4.7
17		5.9	35		4.9
18		5.2	36		5.5
19		5.6	37		6.0
20		5.9	38		5.9

Table4.2 – continued from previous page

ID	2D Structure	pK_i	ID	2D Structure	pK_i
21		5.9	39		5.6
22		7.0	46		4.9
23		4.3	47		6.5
24		6.4	51		6.8

Global and local quantum chemical descriptors have been used frequently since they are useful tools in rationalizing the local aspects of reactivity and activity [31, 32]. HOMO and LUMO energies, hardness, carbonyl stretching frequency, atomic charges (Mulliken [111]/NPA [113]/Hirshfeld [162]/CM5 [163]) and isotropic shielding constants were calculated, and then associated with their affinities for the NMDA glycine binding site.

The chemical hardness, η, is defined as the second order derivative of the energy with respect to the number of electrons at constant external potential [197].

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(r)} \tag{4.1}$$

The chemical hardness expresses the resistance of the chemical potential of a system to a change in the number of electrons, so to how easily the system can accommodate additional charge. This descriptor can be computed using the finite difference approach, allowing for orbital relaxation upon charge intake:

$$\eta \cong I - A \tag{4.2}$$

with I the vertical ionisation potential and A the vertical electron affinity, or using the frozen core approximation:

$$\eta \cong \varepsilon_{LUMO} - \varepsilon_{HOMO} \tag{4.3}$$

Magnetic properties can be useful in creating a clearer picture about the electronic structure of a molecule [198]. Nuclear magnetic resonance (NMR) chemical shifts reflect to what extent the nucleus j is shielded by the electrons around the nuclei from an applied external magnetic field $\mathbf{B_o}$. The actual magnetic field $\mathbf{B}_j$ at nucleus j is therefore always less than the external field

$\mathbf{B_0}$:

$$B_j = (1 - \sigma_j)B_o \tag{4.4}$$

In Equation 4.4, σ_j refers to the shielding tensor of nucleus j. The isotropic magnetic chemical shielding constant is calculated with the following formula:

$$\sigma_{j,iso} = \frac{1}{3}(\sigma_{j,xx} + \sigma_{j,yy} + \sigma_{j,zz}) \tag{4.5}$$

Increasing the electron density around a nucleus produces a stronger induced magnetic field that opposes the external field, shielding the nucleus more and resulting in lower chemical shifts in the NMR spectrum. Conversely, as the electron density decreases, the nucleus is deshielded, and the peak appears more to the left of the spectrum. In the present study, isotropy calculations were performed with the gauge independent atomic orbital method (GIAO) [199, 200].

4.3 Results and Discussion

Chemical descriptors are associated with chemical reactivity and selectivity, and are therefore often used to estimate and quantify molecular behaviour. In this study, molecular descriptors have been applied to help discriminating between partial agonist and antagonist nature. The investigated molecules have a similar structural pattern, since all the ligands contain at least one amide group, and 1-3 aromatic rings. The amide group is important in receptor recognition with a carbonyl group that can form a hydrogen bond with a positively charged hydrogen donor site and an amino group that can form a hydrogen bond with a negatively charged hydrogen acceptor site of the receptor. All the investigated antagonists contain a substituted aryl group on the amine side of their backbone. Regarding the substituents on the carbonyl side, three types of antagonists can be identified, indicated as Anta1, Anta2 and Anta3 (Figure 4.1).

Concerning the set of partial agonists, several ligands with partial agonistic behaviour derived from the full agonist D-serine were introduced in 2009 by Urwyler *et al.* [30]. The hydroxyl group in D-serine was replaced by an amide group to retain the hydrogen bond donor function of the hydroxyl group, while allowing for further substitution (see R in Figure 4.2). It has been reported that ligands containing an amino acid group are found to be in their zwitterionic state at physiological pH [201, 202, 203, 204]. Therefore, chemical descriptor calculations were performed for both the neutral and zwitterionic states, and the resulting descriptor values are compared separately with those for the antagonists. The novel class of partial agonists at the glycine site on the N-Methyl-D-aspartate (NMDA) receptor complex are displayed in Figure 4.2.

On the average, the size of the partial agonists considered in this work is somewhat smaller than that of the studied antagonists as can be observed in Figure 4.3, but still significantly larger than full agonists such as D-alanine, D-serine and aminocyclopropane carboxylic acid (ACC). Both smaller and larger partial agonists display intrinsic activities well over 50% relative to the maximal effect of glycine (100%) [30]. This suggests that agonist behaviour is not governed solely by structural topology. Several global and local molecular descriptors were thus examined for the differentiation of partial agonists from antagonists.

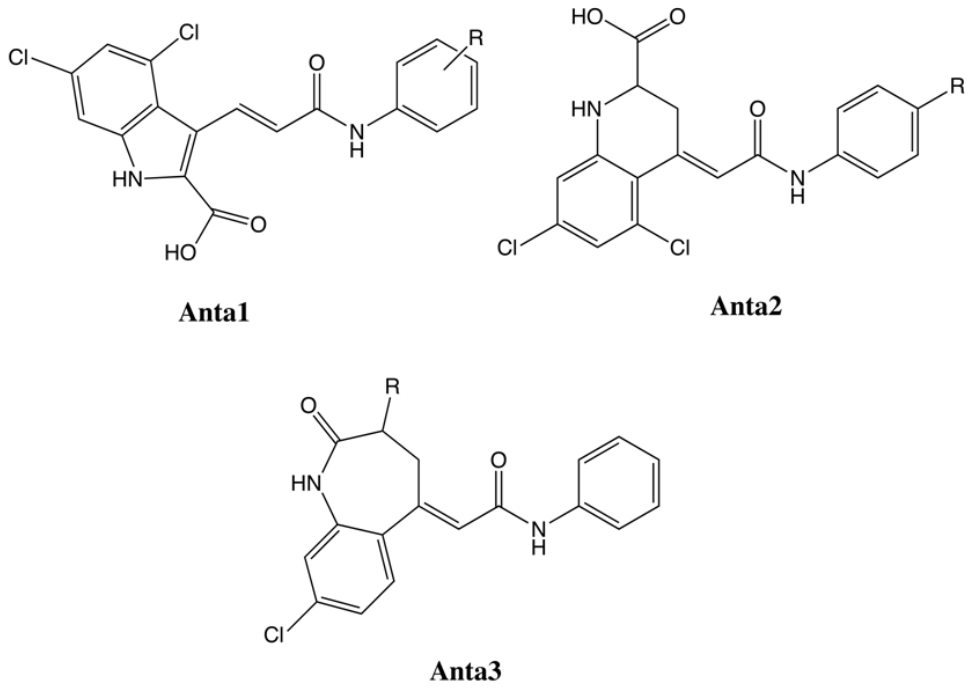


Figure 4.1: The three types of antagonists under investigation.

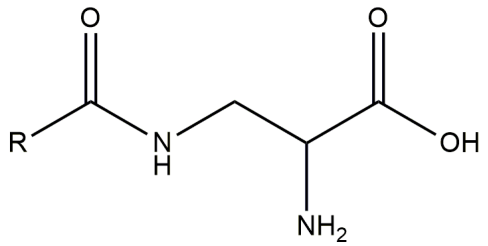


Figure 4.2: 3-acylamino-2-aminopropionic acid derivatives as partial agonists under investigation.

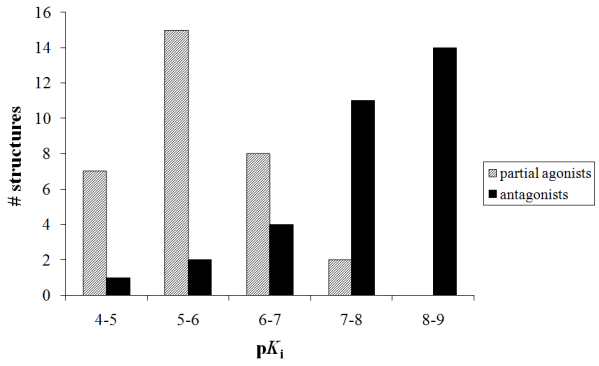


Figure 4.3: Partition of the number of partial agonists and antagonists according to the number of heavy atoms present in their structures.

4.3.1 Global Molecular Descriptors

Firstly, the HOMO and LUMO energies of the listed structures have been considered. These molecular descriptors have been used before, to distinguish between agonists and antagonists of different receptors. It was found that the HOMO energy was one of four quantum mechanical properties necessary to reliably classify 5-HT1A receptor agonists and antagonists, next to other descriptor classes such as electro-topological state, molecular property, and shadow index descriptors [25]. In 2009, Yosa *et al.* reported that it is a necessary, but not sufficient, condition for agonists to possess a significantly higher LUMO energy than the average LUMO energy of the antagonists of the NMDA receptor [28]. In their set of ligands, antagonists' LUMO orbitals are generally present in the ring system, so the LUMO are the delocalized π orbitals of these rings, whereas the agonists' LUMO orbitals are located on the carbonyl and carboxyl groups. The authors, however, mentioned that this distinction in LUMO energies might exist because there is no structural overlap between the two sets of ligands. Figure 4.4 depicts the HOMO and LUMO energies, respectively, for the list of partial agonists and antagonists, with the partial agonists in their zwitterionic and neutral states.

The results for the neutral state partial agonists are very similar with the zwitterionic state. We observe that the HOMO energies of the partial agonists are on average lower than those of the antagonists, whereas the reverse is seen for the LUMO energies, in accordance with the results of Yosa *et al.* [28]. However, no clear bimodal distribution is found. When we focus on highly similar structures within the antagonist (Anta1, Anta2, Anta3) and partial agonist sets, some trends can be observed (Figure 4.5).

Anta1 set contains the antagonist ligands 189, 22, 24, 26, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, and 44. For these ligands, the HOMO and LUMO energies are within quite a wide range; see blue colored marks in Figure 4.5. When the phenyl ring in Figure 4.1 is substituted by a strongly electron-donating group such as NH_2 or NMe_2 (ligand 36 in Figure 4.6 and Figure 4.7), the HOMO is significantly higher in energy than for the reference ligand 189. The effect is the largest when the amino group is in para position and the smallest in meta. When the phenyl ring contains a strongly electron-withdrawing group such as NO_2 (ligand 32

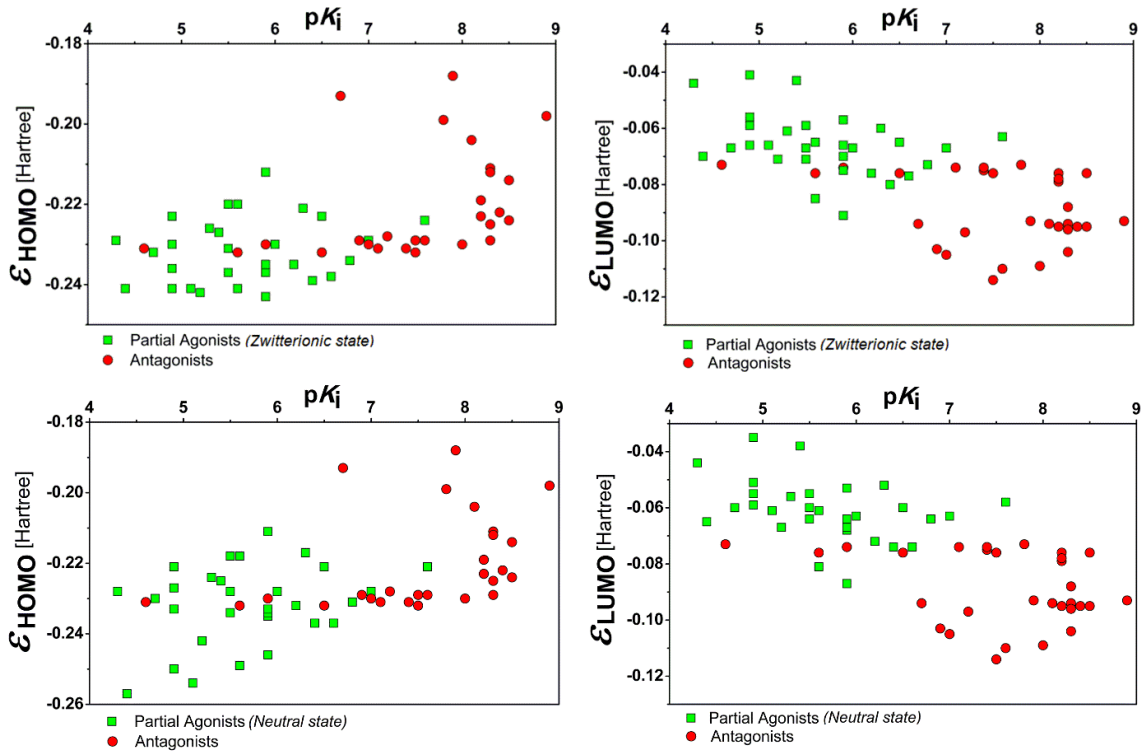


Figure 4.4: HOMO (left) and LUMO (right) energies for the partial agonists in their zwitterionic (up) and neutral (down) states and the antagonists in a.u. (B3LYP/6-311G**).

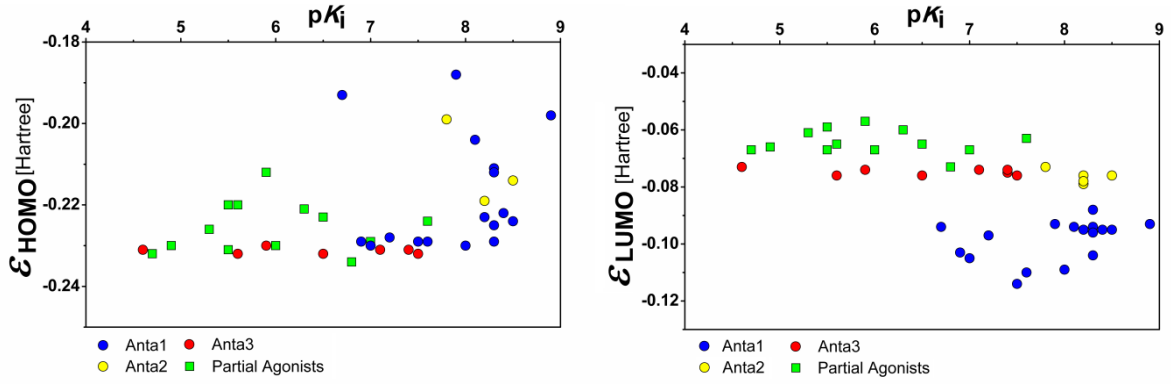


Figure 4.5: HOMO (left) and LUMO (right) energies for a selection of partial agonists (in their zwitterionic state) and the antagonist sets Anta1, Anta2, Anta3 (B3LYP/6-311G**).

in Figure 4.6 and Figure 4.7), the LUMO is clearly stabilized compared to ligand 189. The effect is the largest when the nitro group is in the meta position and the smallest in the ortho position. Figure 4.6 and Figure 4.7 show that for the ligands 32 and 36 the HOMO and LUMO are localized on different ends of the structure and include the π orbitals of the nearby lying ring system(s).

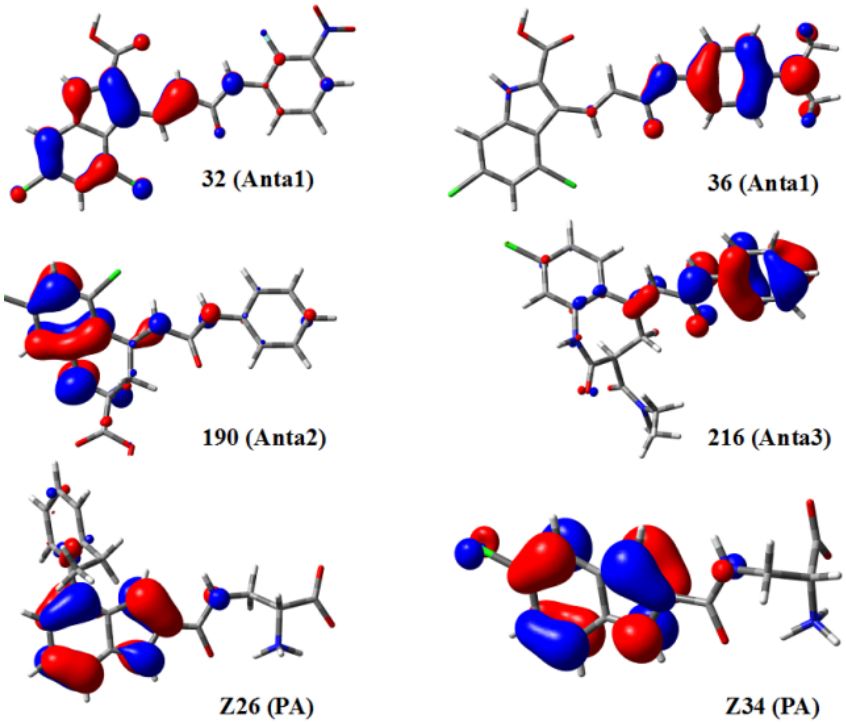


Figure 4.6: Highest Occupied Molecular Orbital (HOMO) (MO isovalue 0.04) for antagonists 32 (pK_i = 8.3, ε_{HOMO} = -0.2291 a.u.), 36 (pK_i = 7.9, ε_{HOMO} = -0.1884 a.u.), 190 (pK_i = 8.2, ε_{HOMO} = -0.2310 a.u.), 216 (pK_i = 4.6, ε_{HOMO} = -0.2309 a.u.), and partial agonists Z26 (pK_i = 7.6, ε_{HOMO} = -0.2238 a.u.), Z34 (pK_i = 4.7, ε_{HOMO} = -0.2324 a.u.) (B3LYP/6-311G**).

Anta2 set contains the antagonists 190, 191, 192, 194, and 195, indicated by the yellow marks in Figure 4.5. There is a large variation in HOMO energies, whereas the LUMO energies are very similar. This can be traced back to the substituted amino groups on the phenyl ring (*R* in Figure 4.1). The HOMO and LUMO of the reference Anta2 ligand are depicted in Figure 4.6 and Figure 4.7, respectively. For the amino substituted versions, the HOMO looks like a combination of ligand 190 and ligand 36 from the Anta1 set.

Anta3 set contains the antagonists 210, 211, 212, 213, 214, 215, 216, and 217. No dissimilarity in HOMO and LUMO energies is observed; see red marks in Figure 4.5. Substitution on the saturated 7-membered ring (Figure 4.1) does not affect the highest occupied and lowest unoccupied orbitals, since those orbitals originate from the π system of the ligand. Antagonist 216 is given as an example in Figure 4.6 and Figure 4.7.

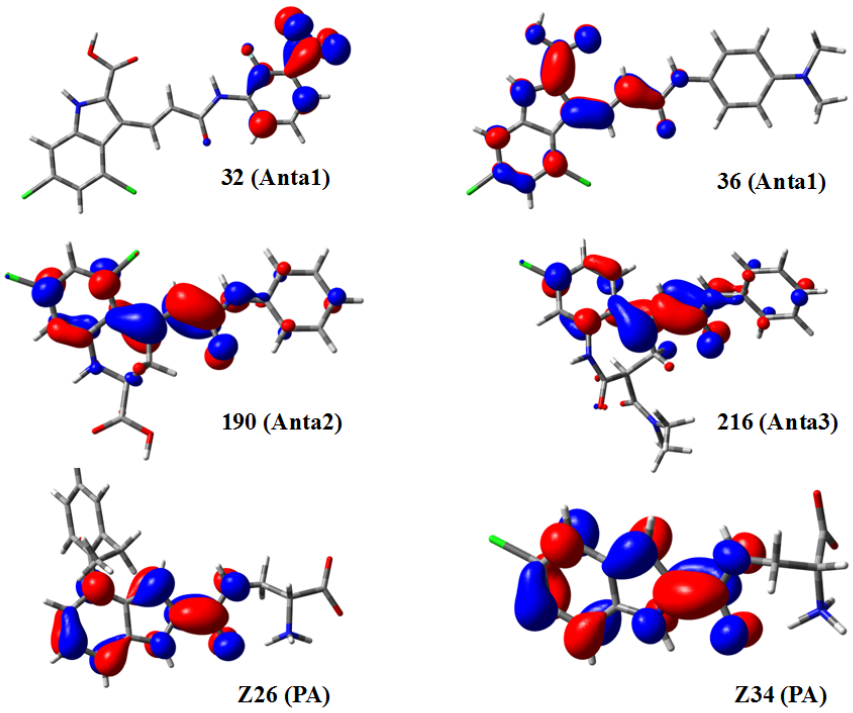


Figure 4.7: Lowest Unoccupied Molecular Orbital (LUMO) (MO isovalue 0.04) for antagonists 32 (pK_i = 8.3, ε_{LUMO} = -0.1038 a.u.), 36 (pK_i = 7.9, ε_{LUMO} = -0.0925 a.u.), 190 (pK_i = 8.2, ε_{LUMO} = -0.0723 a.u.), 216 (pK_i = 4.6, ε_{LUMO} = -0.0727 a.u.), and partial agonists Z26 (pK_i = 7.6, ε_{LUMO} = -0.0626 a.u.), Z34 (pK_i = 4.7, ε_{LUMO} = -0.0666 a.u.) (B3LYP/6-311G**).

Finally, we also consider a selection of partial agonists for a more detailed comparison, containing the ligands 14, 22, 26, 32, 33, 34, 35, 36, 37, 38, 39, 47, and 51. We observe a slightly larger variation in HOMO than in LUMO energies (Figure 4.5). Both the HOMO and the LUMO are delocalized over the aromatic rings and in the case of the LUMO also the carbonyl group is included, as shown for partial agonists Z26 and Z34 in Figure 4.6 and Figure 4.7. Notably, the additional aromatic ring in Z26 does not participate in forming the HOMO or LUMO. When we compare these orbitals to the respective orbitals for a full agonist like D-serine, as depicted in Figure 4.8 for both the neutral and the zwitterionic states, a different picture appears. For D-serine, the HOMO and LUMO are located mainly on the carboxyl group, as indicated before by Yosa *et al.* [28]. This results in a deeper-lying HOMO energy and higher-lying LUMO energy compared to all our examined partial agonists and especially antagonists. This means that there is a distinct difference between the partial agonists we consider and full agonists. Both our partial agonists and antagonists have HOMO and LUMO present on the aromatic rings and their substituents. This observation is different from Yosa's *et al.*, since the authors did not consider partial agonists with aromatic rings.

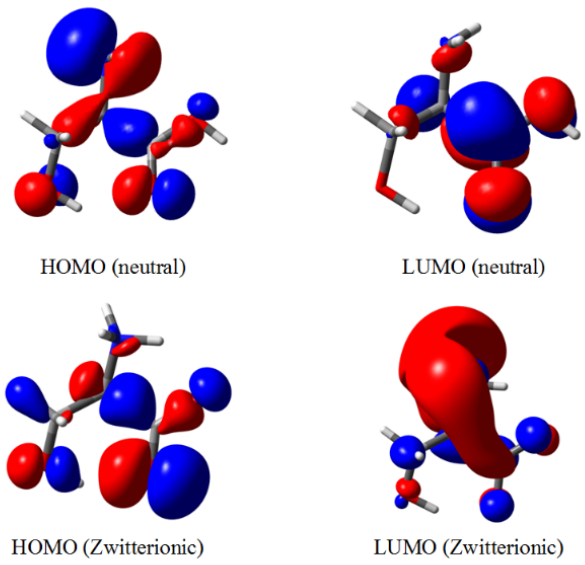


Figure 4.8: Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) (MO isovalue 0.04) for the full agonist D-serine in neutral state (pK_i = 6.5, ε_{HOMO} = -0.2525 a.u., ε_{LUMO} = -0.0140 a.u.) and in zwitterionic state (pK_i = 6.5, ε_{HOMO} = -0.2540 a.u., ε_{LUMO} = -0.0130 a.u.) (B3LYP/6-311G**).

In addition, note that, although they might play a role, the HOMO and LUMO energies do not univocally determine the pK_i values, as can be witnessed in Figure 4.4 and Figure 4.5, and from the comparison in orbital energies of Anta1 36 versus PA Z26, and Anta1 32 versus Anta2 190.

Since the antagonists and partial agonists cannot be differentiated completely using the energy and location of their HOMO and LUMO, we turn to a molecular descriptor that combines

both properties, namely the chemical hardness. From the frozen core approximation for the chemical hardness (see Equation 4.3), it is known that hard molecules have a large HOMO-LUMO energy gap, and soft molecules have a small HOMO-LUMO energy gap. In this work, we have computed the chemical hardness via the finite difference approach in Equation 4.2 using B3LYP for all set of ligands and M06-2X for selected set of ligands (Figure 4.9).

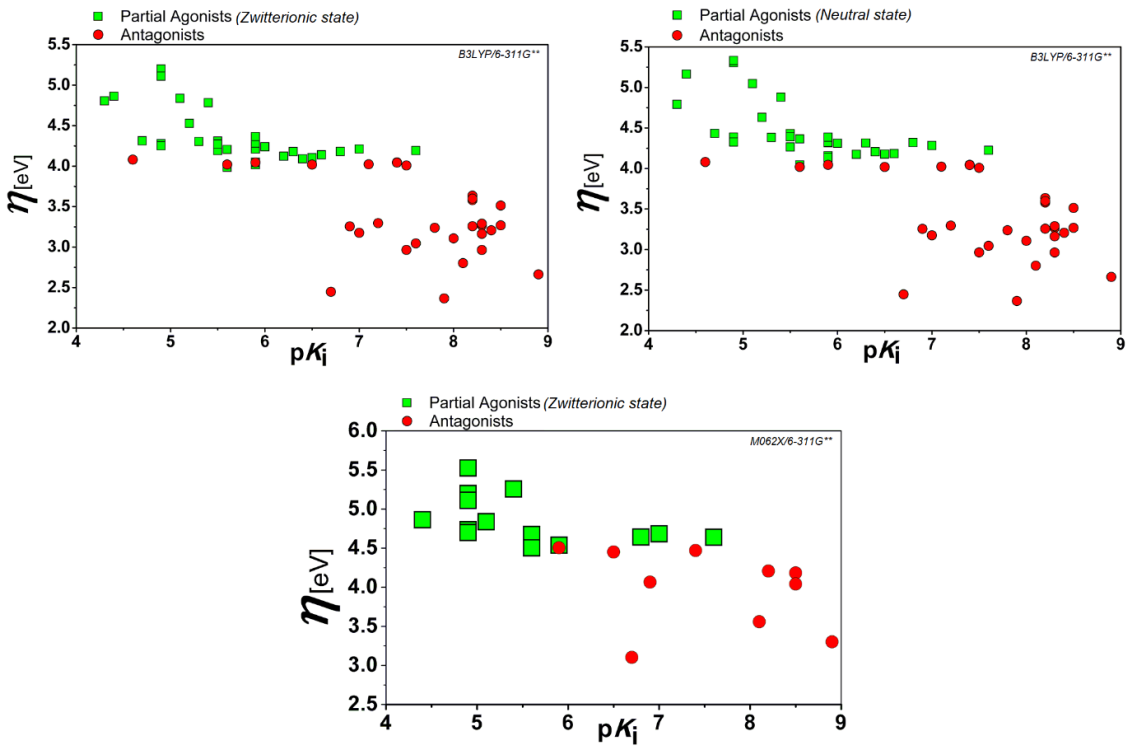


Figure 4.9: Chemical hardness computed via the finite difference approach for the antagonists and partial agonists in eV.

Although the correlation in Figure 4.9 does not show a clear bimodal distribution, the partial agonists are more or less separated from the antagonists. Partial agonists seem to be harder species than antagonists. Based on the hard and soft acids and bases (HSAB) principle [33, 34, 35], which states that hard acids prefer to bind to hard bases and soft acids to soft bases, this could indicate that the two types of ligands interact stronger or weaker with the residues on the binding sites of the receptor, depending on their hardness values. The M06-2X/6-311G** calculations also show the same trend with a slight separation but with no clear bimodal distribution. In harmony with B3LYP results, antagonists are softer species than partial agonists according to M06-2X calculations. When we go back to the HOMO-LUMO energy gap description of the chemical hardness, a very good correlation is found between the two approaches in Equation 4.2 and Equation 4.3 (Figure 4.10). Accordingly, partial agonists display a larger energy gap than antagonists.

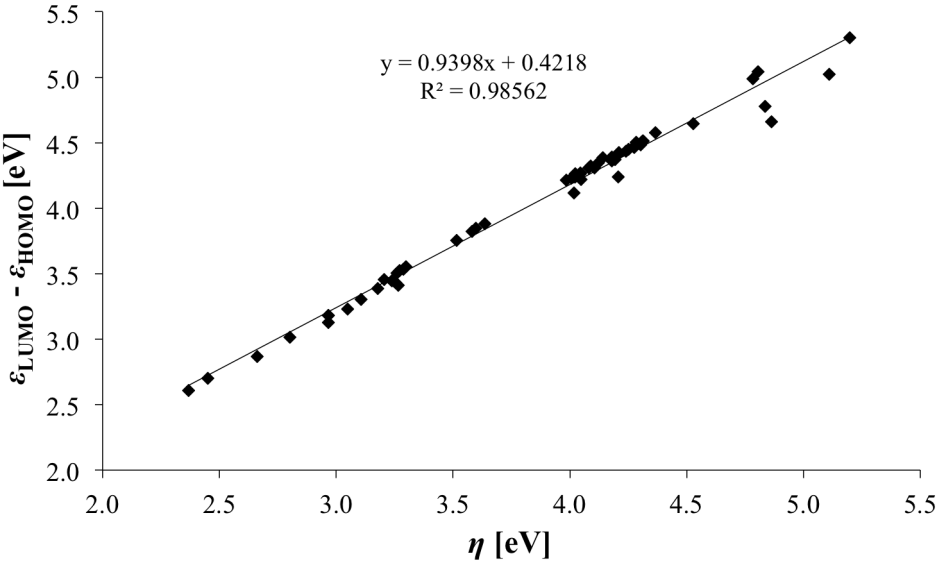


Figure 4.10: Correlation plot for the chemical hardness computed via the finite difference approach and the frozen core approximation for the partial agonists in their zwitterionic form (B3LYP/6-311G**).

4.3.2 Local Molecular Descriptors

Intermolecular forces such as Coulombic attraction, hydrogen bonding, π-π stacking, van der Waals contacts, hydrophobic interactions, etc., form the basis of ligand-receptor binding. These interactions are mainly activated by the phenyl, amide, amine, keto, or carboxylate functional groups present as ligand binding sites on the NMDA receptor [36, 37, 38, 39]. Docking studies of the NMDA GluN1 subunit in complex with various ligands (1PB9 [45], 3QEM [39], 4KFQ [46]; Protein Data Bank) show that the carbonyl oxygen is the primary binding core since most electron-rich regions are localized around it. In order to understand the differences between reactivity centers in ligand-receptor interactions of antagonists and partial agonists, local descriptors were applied to the nitrogen, hydrogen and oxygen atoms of the amide group (Figure 4.1 and Figure 4.2). Note, however, that some antagonist molecules contain more than one amide group. For a reliable comparison with the partial agonists, the amide group on the chain is selected rather than the one in the ring.

Chemical behavior and reactivity of a molecule can be interpreted by investigating the charge distribution around the system. Ligand-receptor recognition is mostly driven by interactions between partially charged atoms such as hydrogen bonding, halogen bonding, and other electrostatic interactions. Figure 4.11 depicts the amide's oxygen partial charge separation between the examined partial agonists (in their zwitterionic and neutral states) and antagonists. Two different population analyses were used, namely Mulliken and Natural Population Analysis (NPA). Although the values are rather different, both schemes show approximately the same trend for the carbonyl oxygen atom, antagonist ligands having less negative oxygen partial charges than partial agonists. This differentiation is for an important part due to the charge

separation present in the amino acid group of the partial agonists. This can be verified by looking into the charge plots comparing the neutral partial agonists and the antagonists. Note that for both types of ligands the charges are spread over several hundredths of atomic units. This indicates that there is a significant effect on the oxygen charge from the adjacent substituted aromatic rings (*R* in Figure 4.1 and Figure 4.2).

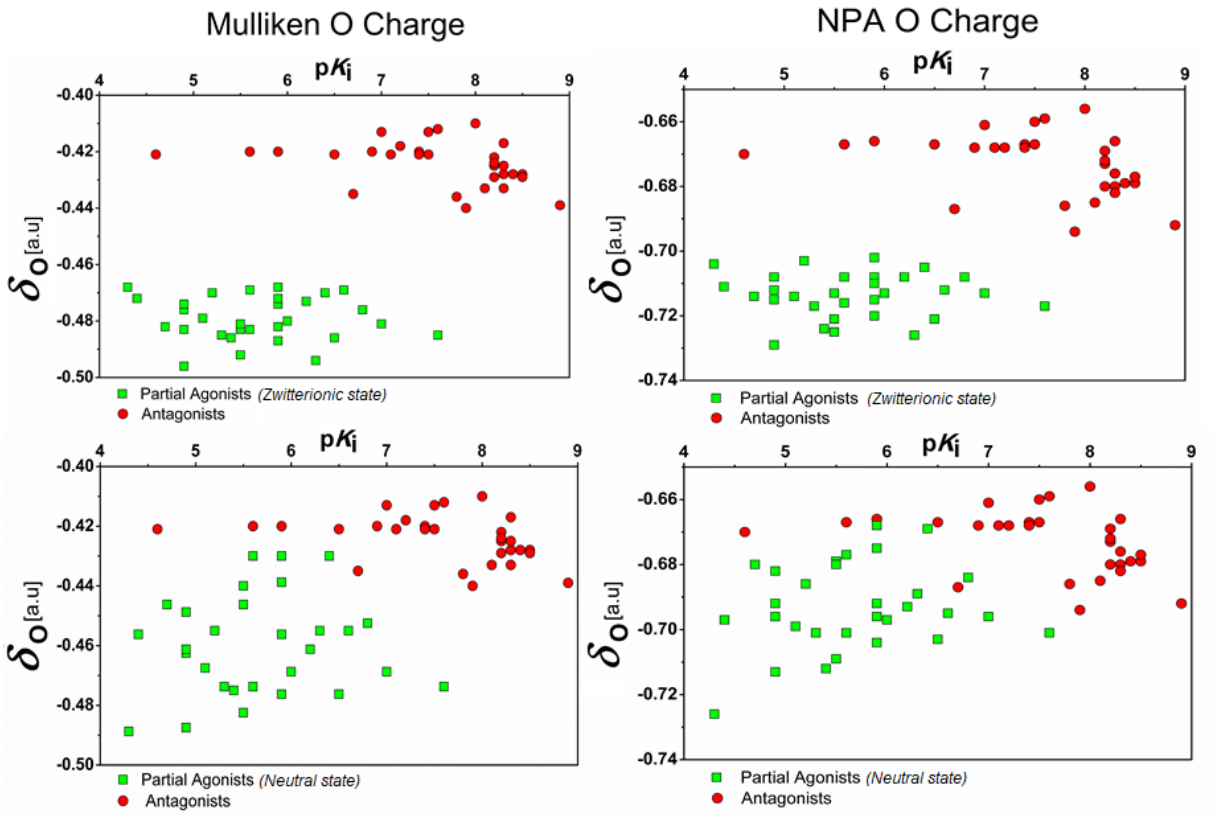


Figure 4.11: Amide oxygen partial charges using the Mulliken and Natural Population Analysis (NPA) schemes for the partial agonists and the antagonists (B3LYP/6-311G**).

Surprisingly, the trend for the charge on the nitrogen atom in the amide group is opposite for Mulliken and NPA schemes, though both population analyses clearly show differentiation (Figure 4.12). Based on the Mulliken charges, the amide nitrogen of the antagonists is richer in electron population than that of the partial agonists. The reverse is true for the NPA charges.

The same Mulliken charge distribution trend is also obtained by the M06-2X/6-311G** calculations (Figure 4.13).

To resolve this discrepancy for the nitrogen atom, we have computed the partial charge on the amide nitrogen using two additional population schemes, namely the Hirshfeld scheme and Charge Model 5 (CM5), which is a recently introduced extension of the Hirshfeld population analysis (Figure 4.14). Again, the absolute values differ from the other population scheme values. The trends, however, are very similar to the results obtained with the NPA scheme.

The discrepancy for the nitrogen atom between Mulliken and NPA has been further inves-

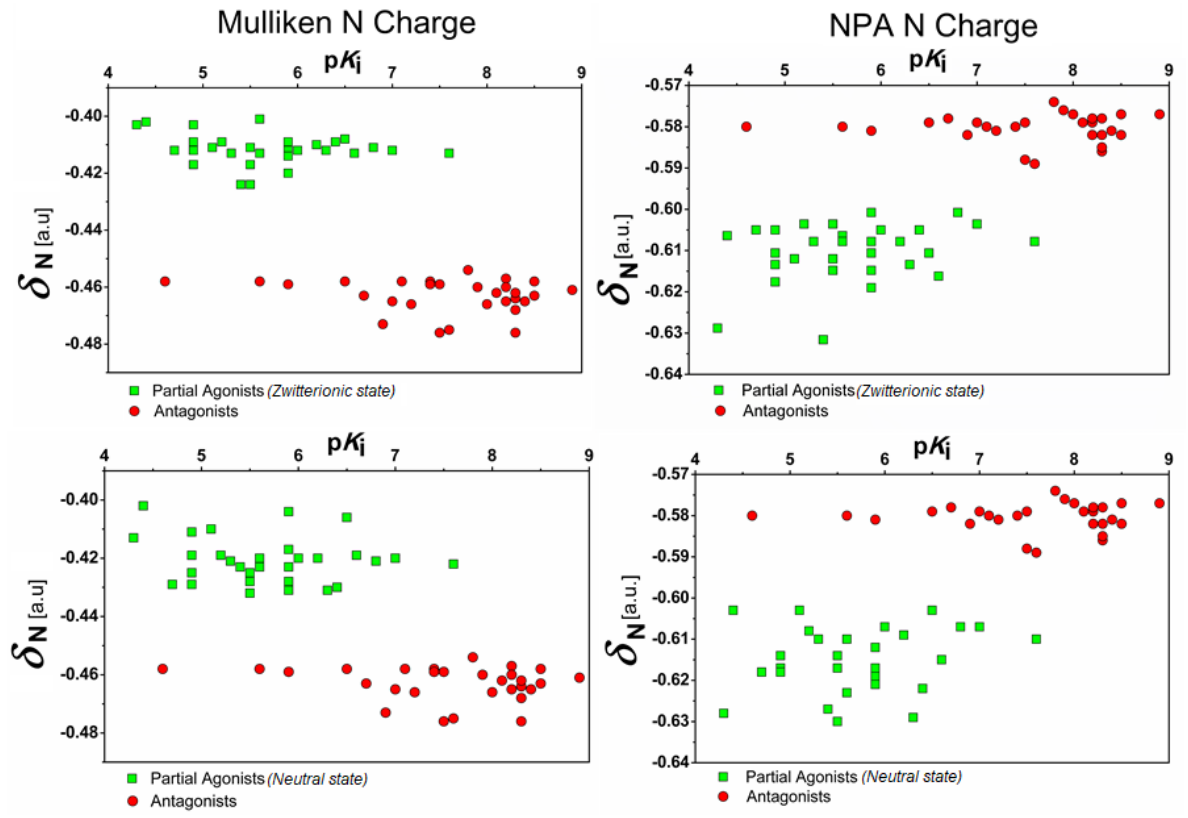


Figure 4.12: Amide nitrogen partial charges using the Mulliken and Natural Population Analysis (NPA) schemes for the partial agonists and the antagonists (B3LYP/6-311G**).

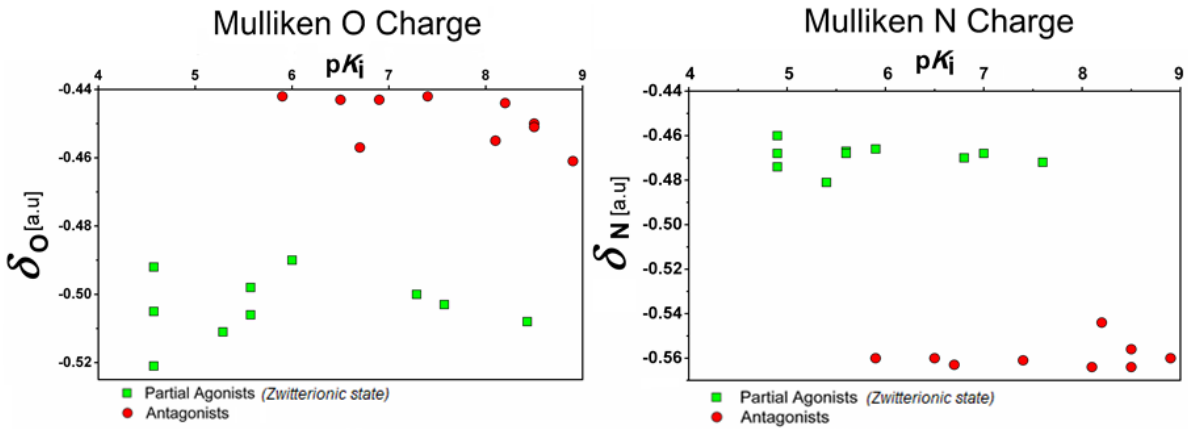


Figure 4.13: Oxygen and amide nitrogen partial charges using the Mulliken scheme for the selected set of partial agonists and the antagonists (M06-2X/6-311G**).

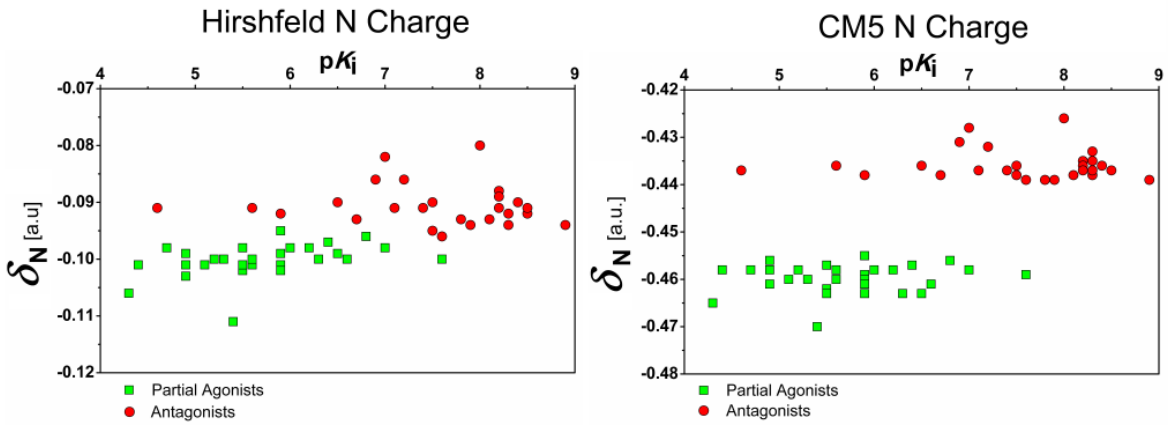


Figure 4.14: Amide nitrogen partial charges using the Hirshfeld and Charge Model 5 (CM5) population schemes for the partial agonists (in their zwitterionic state) and the antagonists (B3LYP/6-311G**).

tigated by magnetic descriptors. Since the calculation of atomic isotropies ($\sigma_{j,iso}$) using DFT is independent of performing population analyses, comparison of ^{17}O and ^{15}N isotropies in both antagonists and partial agonists would give insights into their partial charge trends (Figure 4.15). The more positive isotropic constants $\sigma|$O$|$ and $\sigma|$N$|$ in the partial agonist ligands show that their amide oxygen and amide nitrogen are more shielded compared to the respective nuclei in the antagonists. According to these results, there is less electron density around the antagonist's amide oxygen and nitrogen nuclei than in case of the partial agonists. This is in accordance with the NPA partial charge results.

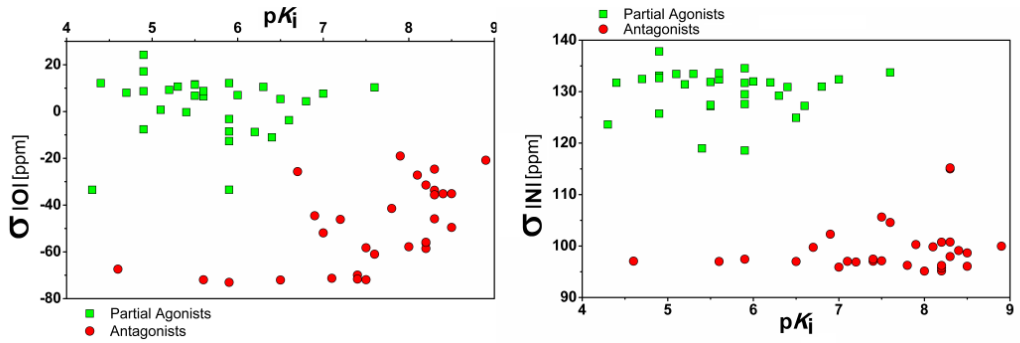


Figure 4.15: Isotropic shielding constant for the amide oxygen (left plot) and nitrogen (right plot) versus pK_i for the partial agonists (in their zwitterionic state) and the antagonists (B3LYP/6-311G**).

The amide oxygen and nitrogen of the partial agonists are therefore more negatively charged than those of the antagonists. This suggests stronger hydrogen bonding acceptor and donor

ability of the partial agonists' amide group. Remarkably, a separation based on the hydrogen atom bonded to the amide nitrogen is not observed.

The last local descriptor probed is the carbonyl stretching frequency (Figure 4.16). The partial agonists, whether they are in the neutral or zwitterionic state, have frequencies around 1650 cm^{-1}, whereas the antagonist molecules have frequencies around or above 1700 cm^{-1}. The lower wavenumber observed in partial agonists can be attributed to a stronger hydrogen bonding ability of the carbonyl oxygen.

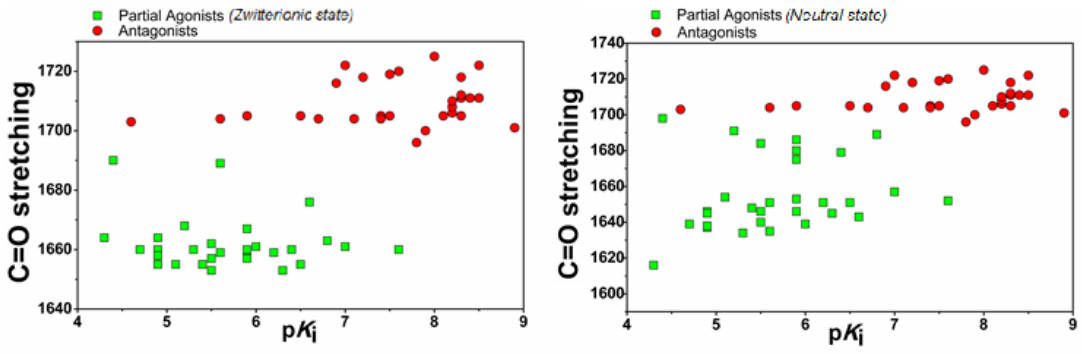


Figure 4.16: Carbonyl stretching frequency (in cm^{-1}) for the partial agonists and the antagonists (B3LYP/6-311G**).

The question now arises as to which of the discriminating molecular descriptors can be used to predict the efficacy of a ligand. The efficacy of NMDA receptor ligands is calculated relative to the maximal effect of glycine (=100%). This means that full agonists have an efficacy of 100%, antagonists an efficacy of 0%, and partial agonists an efficacy between 0 and 100%. In the next part, we will confirm which of the previously rationalized molecular descriptors are truly distinguishing partial agonist ligands from antagonist ones by introducing a multiple linear regression model linking the respective molecular descriptors to the efficacy of the ligands.

4.3.3 Linking Molecular Descriptors to Ligand's Efficacy

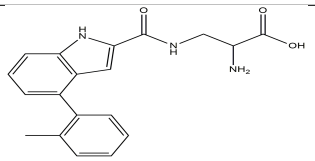
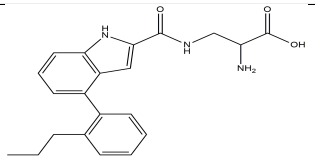
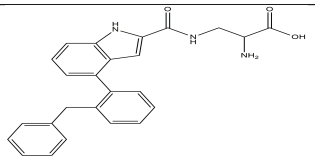
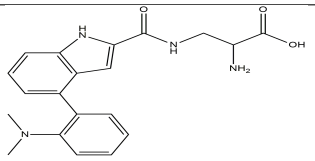
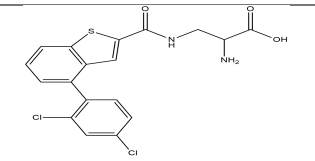
In their paper, Urwyler *et al.* listed the efficacy for 15 possible partial agonists. Ten of the ligands already belong to the examination set (Z9, Z10, Z11, Z14, Z22, Z26, Z29, Z33, Z47, Z51). For the remaining five, all relevant molecular descriptor values have been additionally computed (Z40, Z41, Z43, Z44, Z31). The additional molecular structures are given in Table 4.3.

The data to be modeled are denoted as the **y**-vector, in this case the ligand efficacies. The molecular descriptors are denoted as the **X**-matrix. A quantitative structure-activity relationship, where efficacy (**y**) is a function of one or more descriptors (**X**), can be obtained using multiple linear regression (MLR) analysis. The model can be described by a simple mathematical equation of the form:

$$y = \hat{y} + e = X \cdot c + e \tag{4.6}$$

where **y** is the experimentally measured efficacy vector, $\hat{y}$ is the efficacy vector predicted by the model, **c** is the regression coefficients vector, and **e** is the so-called error vector. Then,

Table 4.3: Two-dimensional representations of additional partial agonist molecules [30]

ID	2D Structure	pK_i
40		6.5
41		7.3
43		6.0
44		5.8
31		4.7

through a minimization of the residual sum of squares (RSSQ), which corresponds to the squared difference between the efficacies predicted by the model ($\widehat{y_i}$) and the experimental efficacies (y_i), the regression coefficients can be estimated. The molecular properties considered for this regression are the chemical hardness, the HOMO and LUMO energies, and the amide's oxygen charge using the NPA scheme, indicating that at least four regression coefficients will be estimated. In addition, a fifth regression coefficient is estimated, the constant in the model. Notably, this constant is not the error vector in Equation 4.6 but a value to shift the sum of the regression terms as close as possible to the actual efficacies. The MLR analysis is performed under constraints for the chemical hardness and the oxygen charge regression coefficients. Both descriptors have to follow the observations made in their respective partial agonist/antagonist differentiation plots. For the chemical hardness, this implies that its regression coefficient needs to be positive, since the η values of the partial agonists are larger than those of the antagonists. For the oxygen charge, the reverse is applicable: its regression coefficient needs to be negative, since the q_O values of the antagonists are less negative than those of the partial agonists. We have used the ligands Z9, Z10, Z11, Z14, Z22, Z26, Z29, Z33, Z47 and Z51 to estimate the model; their experimental pK_i and efficacy values and the calculated properties are presented in Table 4.4.

Table 4.4: pK_i values and efficacies for a set of partial agonists and one antagonist (Z31), also their respective chemical hardness values (η), HOMO and LUMO energies (ε_{HOMO} and ε_{LUMO}), and amide's oxygen NPA charges (q_O) are given (B3LYP/6-311G**).

ligand	**pK_i**	**efficacy** (%)	η **[eV]**	ε_{HOMO} **[a.u.]**	ε_{LUMO} **[a.u.]**	q_O **[a.u.]**
Z9	4.9	63±7	5.198	-0.236	-0.041	-0.729
Z10	4.9	81±5	5.110	-0.241	-0.056	-0.708
Z11	5.1	63±3	4.837	-0.241	-0.066	-0.714
Z14	5.3	42±1	4.304	-0.226	-0.061	-0.717
Z22	7.0	45±4	4.210	-0.229	-0.067	-0.713
Z26	7.6	55±2	4.193	-0.224	-0.062	-0.717
Z29	5.9	22±2	4.018	-0.243	-0.091	-0.702
Z33	6.3	28±1	4.180	-0.220	-0.060	-0.726
Z47	6.5	38±2.5	4.104	-0.223	-0.065	-0.721
Z51	6.8	32±2	4.181	-0.234	-0.073	-0.708
Z40	6.5	62±4	4.190	-0.224	-0.063	-0.716
Z41	7.3	64±1	4.207	-0.224	-0.062	-0.717
Z43	6.0	55±10	4.086	-0.222	-0.063	-0.718
Z44	5.8	59±2	3.730	-0.206	-0.062	-0.717
Z31	4.7	0	4.132	-0.238	-0.078	-0.707

The equation of the MLR model becomes:

$$\widehat{y}_i = 91.88\eta + 2314\varepsilon_{HOMO} - 1653\varepsilon_{LUMO} + 69.40 \tag{4.7}$$

First of all, the regression coefficient for the oxygen charge is zero. This denotes that the oxygen charge does not determine or influence the efficacy of the ligands. The equation can be further interpreted as follows: a preliminary efficacy of 70% is assumed and the molecular property contributions provide corrections to this value. The harder is the ligand, the higher is its efficacy, as observed in Figure 4.9. The effect of the chemical hardness is being compensated partly by the orbital energy terms in Equation 4.7. This suggests that the HOMO cannot be too low, i.e. the ionization potential should be small enough, whereas the LUMO cannot be too high, i.e. the electron affinity should be sufficiently positive. Table 4.5 compares the efficacies obtained by the model with the experimentally measured ones and lists the contribution terms of the various descriptors in Equation 4.7 for each ligand.

Table 4.5: Measured efficacies (with their experimental deviation), efficacies obtained by the model, and the difference between the measured and model efficacies (Δeff) (in %) for a set of partial agonists and one antagonist (Z31). Also the contributions from their chemical hardness values (η), HOMO and LUMO energies (ε_{HOMO}, ε_{LUMO}), and amide's oxygen NPA charges (q_O) are given.

Ligand	**eff**$_{exp}$	**eff**$_{model}$	**Δeff**	$\mathbf{c}_1\ \eta$	$\mathbf{c}_2\ \varepsilon_{HOMO}$	$\mathbf{c}_3\ \varepsilon_{LUMO}$	$\mathbf{c}_4\ q_O$
Z9	63±7	69	-6	478	-545	67	0
Z10	81±5	75	6	470	-557	93	0
Z11	63±3	64	-1	444	-558	109	0
Z14	42±1	43	-1	395	-523	101	0
Z22	45±4	37	8	387	-531	111	0
Z26	55±2	40	15	385	-518	103	0
Z29	22±2	28	-6	369	-561	151	0
Z33	28±1	44	-16	385	-509	98	0
Z47	38±2.5	37	1	377	-516	107	0
Z51	32±2	32	0	384	-542	120	0
Z40	62±4	40	22	385	-518	104	0
Z41	64±1	41	23	387	-518	103	0
Z43	55±10	37	18	375	-513	105	0
Z44	59±2	38	21	343	-477	103	0
Z31	0	27	-27	380	-552	129	0

For most ligands, the agreement between model and experiment is relatively good, often within experimental deviations. Exceptions are Z26 and Z33. Based on the diversity in the

various molecular descriptor contributions listed in Table 4.5, the chemical hardness seems indeed the most essential descriptor. Moreover, correlating the chemical hardness with the experimentally measured efficacy for the partial agonists used to perform the MLR analysis, but excluding the ones that do not follow that trend (Z26 and Z33), already results in a correlation coefficient R^2 of 0.85. Including also the HOMO and LUMO energies further improves the correlation with an R^2 value of 0.93 (Figure 4.17).

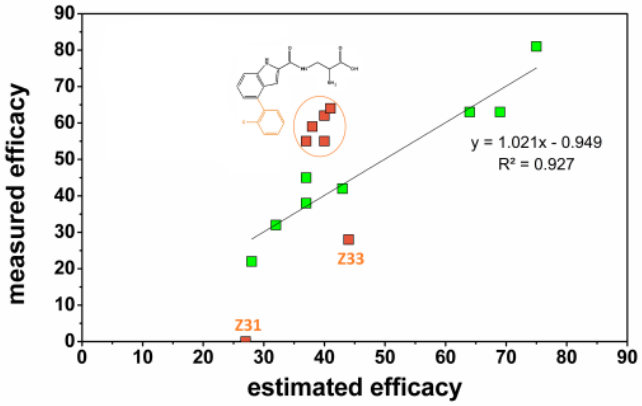


Figure 4.17: Correlation between the estimated and measured efficacies for the partial agonists listed in Table 4.5, with indication of the ligands that deviate because of topological reasons.

When the efficacy of the ligands have been estimated in Table 4.5 which were not included in the MLR equation derivation, namely Z40, Z41, Z43, Z44, and Z31, relatively big differences have been observed with their experimental values. Z40 to Z43, however, are very similar in structure to Z26. Based on their electronic properties, a lower efficacy is suggested, and this is despite their bulky aryl substituent on one of the aromatic rings. Urwyler *et al.* attributed the increase in ligand activity of these partial agonists to the existence of a separate attachment site within a specifically shaped lipophilic area of the glycine site that can be reached by the alkyl group of the aryl substituent. It may well be that this hydrophobic pocket is essential for the ligand to bind properly, making the ligand-domain complex more compact. On the other hand, the efficacy of ligands Z31 and Z33 is very much overestimated. Again, this can probably be linked to the structural topology of the ligand with respect to the attachment sites in the receptor. A relatively long propyl side chain on the five-membered part of the indole ring is noticed in ligand Z33. Although Urwyler *et al.* observed an increase in glycine site affinity for ligand Z33 w.r.t. the reference ligand Z14, there is no mention of a possible hydrophobic pocket that can be reached [30]. This could mean that, although this particular bulky partial agonist links the two domain lobes, it evidently prevents the full closure of the receptor cleft. The same propyl side chain on the indole ring is present in ligand Z47. However, the pK_i values indicate that the bromine located on the indole nucleus is more favorable for glycine site affinity (compare ligands Z14, Z22 and Z47) than the n-propyl side chain. In other words, the long propyl substituent in Z47 is not required to bind effectively to the NMDA receptor and therefore this substituent can adapt a more favorable conformation. Internal rotations with and

within the propyl substituent indeed confirm that its position hardly affects the energy of the ligand. Z31, on the other hand, contains a 2,4-dichlorophenyl group on the benzothiophene part of the structure. The ligand has a very low glycine site affinity (pK_i value of 4.7), which was attributed to the increasing steric bulk at the sulfur atom of the bicyclic ring system in combination with the large 2,4-dichlorophenyl substituent located on the other side of the six-membered ring [30]. It is not surprising that this enormous steric bulk acts as a wedge between the ligand binding domains.

It is clear from this discussion that the efficacy as predicted by the model introduced in Equation 4.7 can be labeled as the intrinsic efficacy of the ligand, only considering the characteristics of that ligand and not the topology of the binding sites on the receptor. However, when additional bulky side groups on the ligand's aromatic ring systems increase the ligand volume even more, the intrinsic efficacy might need to be corrected. This correction can improve or worsen the intrinsic efficacy, depending on the fitting of these side groups into the ligand binding domains and the importance of the binding interactions within.

Using the proposed MLR model, the intrinsic efficacy for the remaining partial agonists, as well as for the listed antagonists have been predicted. The results for the former are presented in Table 4.6 and for the latter in Table 4.7.

Table 4.6. Intrinsic efficacies (in %) obtained by the model for the set of partial agonists and the contributions from their chemical hardness values (η), and HOMO and LUMO energies (ε_{HOMO}, ε_{LUMO})

Ligand	$\mathbf{c}_1\ \eta$	$\mathbf{c}_2\ \varepsilon_{HOMO}$	$\mathbf{c}_3\ \varepsilon_{LUMO}$	$\mathbf{eff}_{model}$
Z8	447	-557	115	74
Z9	478	-545	67	69
Z10	470	-557	93	75
Z11	444	-558	109	64
Z13	396	-549	118	35
Z14	395	-523	101	43
Z15	439	-525	72	56
Z17	387	-549	123	31
Z18	416	-559	117	44
Z19	386	-557	141	39
Z20	394	-544	115	34
Z21	401	-543	110	37
Z22	387	-531	111	37
Z23	441	-529	72	53
Z24	376	-554	133	24
Z25	380	-550	127	26
Z26	385	-518	103	40

Table4.6 – continued from previous page

Ligand	$\mathbf{c}_1\ \eta$	$\mathbf{c}_2\ \varepsilon_{HOMO}$	$\mathbf{c}_3\ \varepsilon_{LUMO}$	$\mathbf{eff}_{model}$
Z27	379	-545	125	29
Z29	369	-561	151	28
Z32	385	-509	98	43
Z33	386	-510	98	44
Z34	396	-538	110	38
Z35	393	-532	109	39
Z36	393	-535	110	38
Z37	389	-532	111	38
Z38	372	-491	95	44
Z39	366	-509	107	34
Z46	391	-515	98	43
Z47	377	-516	107	37
Z51	384	-542	120	32
max-min	112	70	83	50

For the partial agonists, the intrinsic efficacies are ranging from 24 to 75%. Ligands with a high intrinsic degree of partial agonism are Z8-Z11, and to a lesser degree Z15 and Z23, whereas Z24, Z25, Z27, and Z29 are nearly antagonistic in nature. Because of the bulky aryl substituent on the one side (the same as for Z26) and the larger sulfur atom (w.r.t. an NH-group) on the other side, the actual efficacy of Z27 might even be lower than predicted by the model, suggesting that we might be dealing with an antagonist in this case. Concerning the antagonists, the intrinsic efficacies of the Anta1 set are all very small, with values ranging from 1 to 11% (Table 4.7). For the Anta2 and Anta3 sets, somewhat larger values were obtained, though still no more than 30%. Also in these cases, the structural topology of the ligands could play a role. Agarwal and Taylor found for the 5-HT1A receptor that agonists tend to be "flatter", i.e. more coplanar, than antagonists, for which greater bulk is protruding out of the plane of the aromatic ring system [205]. Except for the amino acid group at the end, the structures of our partial agonists are completely planar. This is also the case for the Anta1 antagonists. The structures of the Anta2 and Anta3 ligands, however, are out-of-plane because of the presence of nonaromatic six- and seven-membered rings, respectively, which could explain the observed zero efficacy.

Table 4.7. Intrinsic efficacies (in %) obtained by the model for the set of antagonists and the contributions from their chemical hardness values (η), and HOMO and LUMO energies (ε_{HOMO}, ε_{LUMO})

Ligand	$\mathbf{c}_1\ \eta$	$\mathbf{c}_2\ \varepsilon_{HOMO}$	$\mathbf{c}_3\ \varepsilon_{LUMO}$	$\mathbf{eff}_{model}$
189	300	-518	156	8
190	334	-507	126	23

Table4.7 – continued from previous page

Ligand	$\mathbf{c}_1\ \eta$	$\mathbf{c}_2\ \varepsilon_{HOMO}$	$\mathbf{c}_3\ \varepsilon_{LUMO}$	$\mathbf{eff}_{model}$
191	298	-461	120	26
192	323	-495	126	23
194	329	-508	130	21
195	331	-507	128	22
210	372	-533	121	30
211	369	-536	125	28
212	368	-536	126	28
213	372	-535	123	29
214	371	-534	122	29
215	370	-534	123	28
216	375	-534	120	30
217	369	-537	126	28
22	245	-458	153	9
24	280	-530	182	2
26	300	-530	172	11
28	303	-527	161	6
29	225	-446	155	3
31	272	-488	155	9
32	272	-530	189	1
33	302	-522	158	8
34	257	-473	155	9
35	299	-529	170	9
36	295	-513	157	8
37	286	-532	180	2
38	292	-531	174	4
39	217	-436	153	4
41	299	-517	156	8
44	273	-487	154	10

Finally, an intrinsic efficacy value of 124% is computed for the full agonist D-serine, well above the 100% limit for full agonists (calculated chemical hardness value of 6.698 eV, HOMO and LUMO energies of -0.252 and -0.014 a.u., respectively). Overall, two data sets have been constructed, one for partial agonists and one for antagonists, that structurally look very similar in size, presence of at least one amide group, inclusion of one or more aromatic rings. Unfortunately, only a very limited set of experimental data is available to test the suggested MLR model.

4.4 Conclusions

In this study, differentiation of partial agonists and antagonists for glutamate receptors has been attempted using both global and local quantum chemical descriptors. The design of partial agonists for NMDA receptor is an attractive approach as they have less side effects. The outcome of this DFT experiment can be summarized as: (1) Combining HOMO and LUMO energies in the chemical hardness results in a separation between the two types of ligands. (2) the NPA charge distributions of the ligands have shown the amide oxygen and nitrogen of the partial agonists to be more negatively charged than those of antagonists. Therefore, stronger hydrogen bonding acceptor and donor ability of the partial agonists' amide group is to be expected and this was confirmed by the lower carbonyl stretching frequencies. (3) The MLR has shown that for all the ligands the chemical hardness is the most differentiating and important factor regarding the efficacy. However, the LUMO energy seems to be more influential for the set of partial agonists, while for the antagonists this role is played by the HOMO energy. (4) The suggested MLR model cannot be tested by us since not enough experimental data are available, but the predictions might be a guideline for experimentalists. Nonetheless, the training set used for constructing the model is a good representation of the full data set of both partial agonists and antagonists, even though no antagonists were included in the training set. Overall, this study has shown that quantum chemical calculations can provide a valuable aid to experimental studies on ligand design for glutamate receptors.

Chapter 5

Usage of hybrid computational tools to rationalize the binding of agonists and antagonists in the NMDA receptor

5.1 Introduction

The NMDA protein makes conformational rearrangements from open to closed state upon binding of an agonist to the protein's flexible clamshell-shaped ligand binding domain (LBD) [40]. The ligand's efficacy is directly related to the amount of cleft closure such that full agonists make strong interactions with the residues in the LBD tightly closing the clamshell. Fluctuations occur between partially open and closed states upon binding of partial agonists since these type of ligands exert weaker forces [41]. The functional properties that determine the role of a receptor in signal transmission are determined by the structural characteristics of the ion channel and the mechanism of the ion flux. The crystallographic structures of glutamate receptors that contain many functional information show that the ligand binding region communicates through short links with the helices forming the ion channel [41, 42]. The ligand binding region consists of rigid S1 and flexible S2 regions and the changes in the S2 region directly affect the spirals connected to the ion channel. The current viable model for the activation mechanism of the ion channel suggests the opening of the ion channel by conformational transitions following the closure of the LBD upon binding of an agonist [43, 44]. To understand the activation or inhibition mechanisms of the receptor, it is important to elucidate the conformational behaviours of the LBD upon binding of different ligands.

The isolated agonist-binding domain of the GluN1 subunit was crystallized as a complex with full agonists glycine (PDB code:1PB7) and D-serine (PDB code:1PB8)[45], partial agonists D-cycloserine (PDB code:1PB9)[45], 1- aminocyclopropane-1-carboxylic acid (ACPC) (PDB code:1Y20) and 1-aminocyclobutane-1- carboxylic acid (ACBC) (PDB code:1Y1Z)[40]; antagonists cycloleucine (PDB code:1Y1M)[40], 5,7-dichlorokynurenic acid (DCKA) (PDB code:1PBQ)[45] and 1-thioxo-1,2-dihydro-[1,2,4]triazolo[4,3-a]quinoxalin-4(5H)-one (TK40) (PDB code:4KFQ)[46]. These studies suggest that the angles between the S1-S2 regions in three different ligand types are different. Glycine-site antagonists DCKA and cycloleucine stabilize the open conformation of this region. Cycloleucine, a larger ligand in size compared to partial

agonists ACPC and ACBC, causes the opening of the LBD, thus allows the water molecules to enter and eliminates the hydrophobic effects between the side chains of Phe484 and the indole group of Trp731. Finally the open-cleft conformation is stabilized to provide conformational changes in the ion channel and to regulate the receptor flow. Especially the effects of partial agonists on the receptor function have not yet been fully understood. In case of AMPA receptor partial agonists, it is known that the amount of cleft closure is directly proportional to the activity of the ligand [41, 47, 48]. On the other hand, it is reported that some NMDA receptor partial agonists exert the same degree of cleft closure as observed in agonists [40, 45, 49, 50, 51, 52].

Ligand-docking is one of the most popular techniques in estimating the binding affinity of a ligand to its target receptor [53]. A basic docking process aims to predict the most favourable configuration (conformation, position and orientation) of a ligand inside the binding cavity of a receptor [54]. Hydrogen bonds formed between the ligand and the protein as well as Van der Waals and Coloumbic interactions account for the docking score [55]. There are more than 60 docking softwares in use such as FlexX [188], GOLD [182], AutoDock [206], Surflex [187], ICM [207], LigandFit [208], FRED [209], Glide [210], AutoDock Vina [211], MOE-Dock [212], LeDock [213], rDock [214], UCSF Dock [215]. Despite the ease of use of these docking softwares, poor scoring functions are obtained due to the restricted number of conformational degrees of freedom, lack of protein flexibility and utilization of implicit solvent models. It is important to develop methods that can accurately estimate the free energy of binding, $\Delta G_{binding}$, which allows the prediction of the binding strength of any drug candidate prior to its synthesis. Therefore it is more interesting to apply molecular dynamics simulations to different poses coming from docking scheme and to calculate the binding free energies with methods such as Free Energy Perturbation (FEP) [56], Thermodynamic Integration (TI) [57], Linear Response (LR) [58], Molecular Mechanics-Generalized Born Surface Area (MM-GB/SA) and Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PB/SA) methods [59, 60]. Among these methods, the most accurate and rigorous ones are FEP and TI [61]. Despite their accuracy, they have found little use in drug design due to their convergence only for rather similar ligands and computational cost [62, 63]. The MM-GB/SA and MM-PB/SA methods that combine molecular mechanics energy and implicit solvation models, are simple and faster than FEP and have been widely used for free energy calculations in computational medicinal chemistry [64, 65].

There are large number of molecular dynamics simulation studies that have been carrying out in order to understand the conformational behaviours of glutamate receptors with a series of ligands and their interactions [5, 16, 51, 66, 67, 68, 69]. It was shown that interdomain hydrogen bonding (IHB) distances could be used as indicators for the biological activity (agonism-antagonism) of the kainate receptor ligands [67]. Following this study, a relationship was derived between IHB distances and experimental efficacies of various NMDA GluN1 (partial) agonists: the smaller the IHB distance, the more effective the ligand [16]. Recently, Cifci *et al*. was able to show the correlation between the binding energies and half maximal inhibitory concentration (IC_{50}) values of Phosphodiesterase IV-B system [70]. They could reproduce the experimental IC_{50} values with the suggested protocol.

In this study, the binding Gibbs free energy for the natural receptor-ligand complexes (NMDAR-ligand) are calculated and correlated with their experimental efficacy or binding affinity pK_i. The ultimate goal is to predict the binding energies of the ligands proposed and finally to differentiate between agonists, antagonists and partial agonists based on their estimated binding

Gibbs free energies.

5.2 Methodology

5.2.1 Protein and Dataset Preparation

All available X-ray crystal structures of NMDA receptor with GluN1 subunit in the PDB database have been collected (1PB7, 1PB8, 1PB9, 1PBQ, 1Y1M, 1Y1Z, 1Y20, 2A5T, 4KCC, 4KFQ, 4NF4, 4NF5, 4NF6, 4NF8, 5DEX, 5I57, 5I58, 5I59, 5JTY, 5U8C, 5VIH, 5VII, 5VIJ). All these protein sources have sequence homology larger than 95%. GluN1 ligand binding domains have been isolated and the ligands have been removed. Root Mean Square Deviations from the reference crystal structure (PDB code: 1PB7) have been calculated for all PDBs in order to group them with respect to their conformational differences (Table 5.1). Most distinct ones were selected as representatives of different receptor conformations for docking; 1PB7 and 1PB8 representing the closed conformation of LBD; 1PB9, 1Y1Z and 1Y20 representing the partially closed conformations; 1Y1M, 1PBQ, 4KFQ and 4KCC representing the open conformations (Table 5.2).

Table 5.1: RMSD values (in Å) of the NMDA X-ray crystal structures with respect to their conformational differences in their GluN1 subunit.

PDB ID	RMSD	PDB ID	RMSD	PDB ID	RMSD	PDB ID	RMSD
1PB7	0.000	2A5T	0.080	5U8C	0.118	4KCC	1.691
5VII	0.065	4NF8	0.091	1Y1Z	0.119	4KFQ	1.726
5VIJ	0.065	5I57	0.097	5DEX	0.162	5JTY	1.927
4NF5	0.067	5I59	0.103	5I58	0.168	4NF4	1.957
1PB8	0.077	1Y20	0.104	1PB9	0.188	1PBQ	2.135
4NF6	0.079	5VIH	0.111	1Y1M	0.810		

Table 5.2: Families of representative NMDA crystal structures with respect to their conformational differences in GluN1 subunit.

1PB7	1PB8	1Y20	1Y1Z	1PB9	1Y1M	4KCC	4KFQ	1PBQ
5VII	4NF6	5I57	5U8C	5DEX				5JTY
5VIJ	2A5T	5I59		5I58				4NF4
4NF5	4NF8	5VIH						

The extent of the receptor activation is determined by the bound ligand's efficacy [17]. Antagonists inactivate the receptor which is interpreted having 0% efficacy, agonists fully activate the receptor with 100% efficacy and partial agonists have efficacies scaling up to an agonist and down to an antagonist. In fact, there is a direct relation between the conformations of the ligand binding domain of a receptor and activation/inhibition of the ion channel.

Experimental binding affinity (K_i) is a measure of the interaction of a ligand with its target receptor. The relationship between the binding energy and the binding affinity can be represented as:

$$\Delta G_{binding} = -RTlog_{10}K_i \tag{5.1}$$

where R is the ideal gas constant and T is the temperature.

$$pK_i = -log_{10}K_i \tag{5.2}$$

Therefore Equation 5.1 becomes:

$$\Delta G_{binding} = -RT(pK_i) \tag{5.3}$$

From Equation 5.3 we observe a linear dependency between ΔG and pK_i. Eight ligands with varying efficacies from 0 to 100 (%) were chosen for binding free energy calculations as the training set and their 2D chemical structures are given in Table 5.3.

Two partial agonists from the study of Urwyler [30] were randomly selected for the test set (Table 5.4).

5.2.2 Docking Procedure

In order to obtain the most favorable orientation of eight guest ligands given in Table 5.3 in the nine host protein structures which have been chosen with respect to the Table 5.2 (1PB7 [45], 1PB8 [45], 1Y20 [40], 1PB9 [45], 1Y1Z [40], 1Y1M [40], 4KCC [51], 4KFQ [46], 1PBQ [45], docking process has been applied by using the AutoDock v4.2 software [216] and different poses have been obtained. For each ligand 20 independent runs were performed. Grid maps were calculated with AutoGrid program before docking of ligands with a grid box of 50 Å x 50 Å x 50 Å which contains the ligand binding domain with interacting residues. Gasteiger charges and solvation parameters were assigned using AutoDock Tools. In order to find optimal conformation with the lowest binding energy Lamarckian Genetic Algorithm is chosen as the docking search parameter. The population size is set to 150; maximum number of energy evaluations is set to 10 000 000; the number of generations is set to 27 000. The remaining parameters are set as the default values. At the end of each run the docked poses of each ligand were clustered based on their RMSD values and the best pose with the lowest binding energy (ΔG_{bind}) was selected.

5.2.3 Molecular Dynamics Simulations

Molecular dynamics simulations have been performed for each lowest energy host:guest complex conformations obtained from the docking process. Hydrogen atoms were added to the systems with the tleap module of AMBER17 [217]. The systems were solvated with TIP3P

Table 5.3: 2D chemical sketches and experimental efficacies of the selected ligands for the training set.

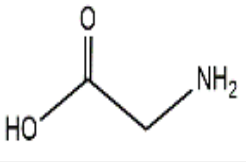
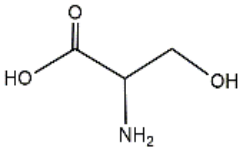
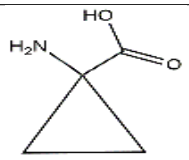
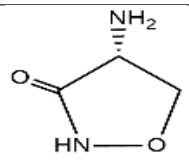
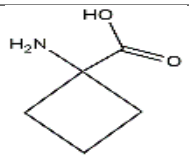
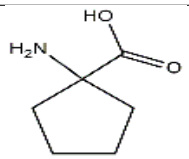
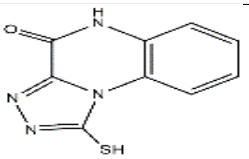
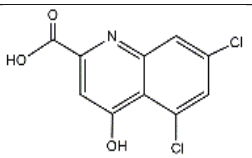
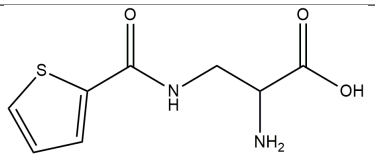
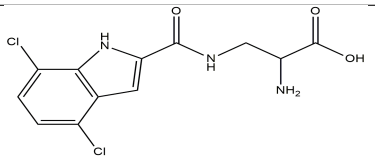
Ligand ID	2D Structure	Activity	Efficacy (%)	pK_i
Glycine		Agonist	100	6.9
D-serine		Agonist	98	6.5
ACPC		Partial Agonist	80	6.8
D-cycloserine		Partial Agonist	75	3.9
ACBC		A Partial gonist	42	4.8
Cycloleucine		Antagonist	0	2.7
TK40		Antagonist	0	7.1
DCKA		Antagonist	0	7.4

Table 5.4: 2D chemical sketches and experimental efficacies of the selected ligands for the test set.

Ligand ID	2D Structure	Activity	Efficacy (%)	pK_i
11		Partial Agonist	63	5.1
51		Partial Agonist	32	6.8

[218] water molecules in a cubic box of 8 Å. Chlorine and sodium ions were added to neutralize the systems. The Amber ff14SB force field [219] was used to model the NMDA protein with general Amber force field GAFF [220] parameters to model the ligands. In order to constrain the bond lengths the SHAKE algorithm [221] was chosen. For the minimization steps, the Andersen temperature coupling algorithm was applied to ensure a constant temperature ensemble (NVT). The time step was set to 2 fs.

A short gas phase minimization was performed to optimize the hydrogen atom positions for 100 ps with a force constant of 200 kcal mol^{-1} $Å^{-2}$ applied on all heavy atoms before solvation. Minimization was carried out in 5 steps. In the first step, only the hydrogen atoms were allowed to move for 100 ps at 10 K with a force constant of 50 kcal mol^{-1} $Å^{-2}$ applied on all heavy atoms on the protein. In the second, third and fourth steps, water molecules were allowed to move for 100 ps at 10 K with a force constant of 10 kcal mol^{-1} $Å^{-2}$ applied on all heavy atoms on the protein. In the final step, the thermostat temperature was increased from 10 K to 300 K for the next 2 ns with the same force applied on heavy atoms. After the equilibration, 40 independent NPT (isothermal-isobaric ensemble) MD simulations were performed for each ligand:protein complex during 1 ns at 300 K with different initial velocities.

5.2.4 Molecular Mechanics-Generalized Born Surface Area (MM-GB/SA) Post Processing

Converged MM-GB/SA results were obtained by averaging multiple independent trajectories. In MM-GB/SA method, the free energy of binding between a ligand and a receptor to form a complex is calculated as

$$\Delta G_{binding} = G_{complex} - G_{receptor} - G_{ligand} \tag{5.4}$$

Each free energy is calculated from the formula

$$G = E_{MM} + G_{solvation} - TS_{MM} \tag{5.5}$$

Where E_{MM} is the molecular mechanics gas phase energy of the reactant, $G_{solvation}$ is the solvation energy coming from a continuum representation of the solvent for the polar part with a relation to the solvent accessible surface area for the non-polar part, TS_{MM} is the product of entropy and absolute temperature. In this study, the entropy term was not included in our calculations due to the high computational cost. The aim of this study is to relate the computed free energy values to experimental activity of ligands, not to estimate the experimental binding free energies of the ligands.

E_{MM} and $G_{solvation}$ terms were calculated by MMGBSA.py module of Amber17. Five different solvation models have been evaluated for the calculation of polar solvation free energy: GB^{HCT} [222, 223, 224], GB^{OBC} [225], GB^{OBC-2} [225], GB^{neck} [226], and GB^{neck2} [227]. The Linear Combinations of Pairwise Overlaps (LCPO) method [228] was used to approximate the hydrophobic contribution.

5.3 Results and Discussion

5.3.1 Docking of Ligands

We have selected nine NMDA receptor X-ray crystal structures from the PDB database; two in complex with an agonist ligand (glycine and D-serine in 1PB7 and 1PB8, respectively), three in complex with a partial agonist ligand (D-cycloserine, ACBC and ACPC in 1PB9, 1Y1Z and 1Y20, respectively), three in complex with an antagonist ligand (cycloleucine, DCKA and TK40 in 1Y1M, 1PBQ and 4KFQ, respectively) and one apo structure (4KCC). We removed the ligands from the crystal structures to form the host proteins and docked eight guest ligands to these nine crystal structures (from close to open conformations of the ligand binding domain) with AutoDock v4.2. At the end of each docking process, 20 docked poses of each ligand were clustered based on their RMSD values. The best pose (top ranked) with the lowest binding energy (ΔG_{bind}) was selected (Table 5.5).

This multiple docking process was performed in order to find a correlation between the binding energies and the activity of ligands (agonism vs antagonism). We could reproduce the expected ligand-receptor interactions but in some cases we failed to obtain the lowest binding energies for the original protein structures (see Table 5.5: in red, the lowest binding energy for a given ligand; in orange, the binding energy of a given ligand towards its respective host structure). For glycine, ACBC, TK40 and DCKA the best host proteins are 1PB7, 1Y1Z, 4KFQ and 1PBQ respectively, which match their experimental complex. For D-serine, ACPC, D-cycloserine and cycloleucine, although the lowest binding energies from AutoDock results do not correspond to their experimental protein match (numbers in orange color in Table 5.5), we observed very close binding energies. The interacting residues with the ligands are given in Figure 5.1.

From Figure 5.1 we observe that all ligands fit well in ligand binding domain of the NMDA receptor by the docking procedure applied. Two randomly selected partial agonists from the paper of Urwyler [30] are also docked to all nine crystal structures (Table 5.6).

A linear regression fit between the lowest Autodock $\Delta G_{binding}$ values and the experimental affinity constants of the ligands, namely pK_i's, was obtained for the eight ligands (Figure 5.2). No linear correspondance exists with a correlation coefficient of 0.038 which means that there

Table 5.5: Binding free energies of each docking study (*kcal/mol*).

	GLYCINE	**D-SERINE**	**ACPC**	**D-CYCLO SERINE**	**ACBC**	**CYCLO LEUCINE**	**TK40**	**DCKA**
Efficacy (%) pK_i **PDB code**	*100* 6.9	*98* 6.5	*80* 6.8	*75* 3.9	*42* 4.8	*0* 2.7	*0* 7.4	*0* 7.1
1PB7	-6.10	-6.95	-5.42	-6.30	-4.82	-3.90	-5.63	-5.79
1PB8	-6.10	-7.12	-6.99	-6.85	-7.39	-3.91	-5.65	-5.82
1Y20	-6.02	-6.92	-7.02	-6.37	-7.47	-5.68	-5.86	-6.47
1PB9	-5.91	-7.19	-6.79	-6.64	-6.56	-3.76	-5.58	-5.77
1Y1Z	-5.84	-6.60	-7.12	-6.50	-8.03	-8.45	-6.16	-6.21
1Y1M	-5.46	-5.82	-6.43	-5.79	-7.17	-7.72	-6.22	-7.31
4KCC	-4.87	-5.26	-5.48	-5.08	-5.91	-6.67	-6.97	-6.67
4KFQ	-4.09	-4.27	-4.75	-4.32	-5.18	-5.57	-7.71	-7.57
1PBQ	-3.81	-3.97	-4.46	-3.71	-4.81	-5.08	-6.83	-8.76

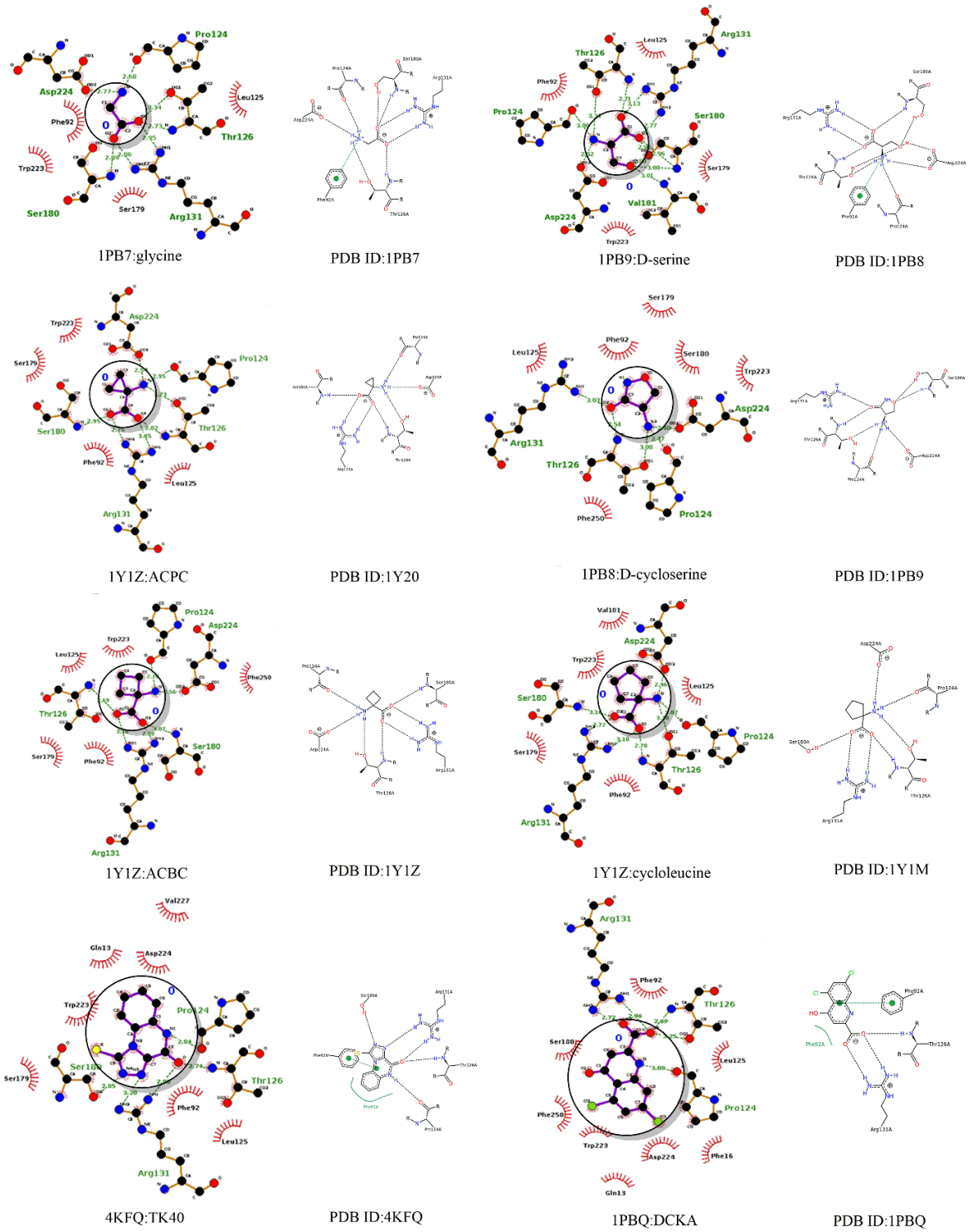


Figure 5.1: Ligand-receptor interactions of the docking structures and the interactions taken from corresponding original PDB crystal structures.

Table 5.6: Binding free energies of ligands 11 and 51 (in *kcal/mol*).(NA indicates when a ligand could not enter into the LBD due to its large size therefore the AutoDock binding energy has no meaning.)

	11	**51**
Efficacy (%) pK_i **PDB code**	*63* 5.1	*32* 6.8
1PB7	NA	NA
1PB8	NA	NA
1Y20	NA	NA
1PB9	-6.31	NA
1Y1Z	NA	NA
1Y1M	-8.53	-10.16
4KCC	-6.62	-8.93
4KFQ	-6.07	-7.18
1PBQ	-5.64	-6.76

is no relationship between the AutoDock ΔG scoring function and the pK_i values in case of NMDA. Another linear regression fit between the lowest Autodock $\Delta G_{binding}$ values and the experimental efficacies of the ligands was obtained for all the ligands in the training set and for five ligands (glycine, D-serine, ACPC, D-cycloserine and ACBC) discarding the antagonists since they have zero efficacies (Figure 5.3). Relatively good correspondances are observed with correlation coefficients of 0.658 and 0.748 which means that AutoDock is capable of predicting the most favorable configuration of the ligands inside the LBD, but fails in estimating experimental $\Delta G_{binding}$ values.

The two derived models have been applied to the ligands in the test set (Table 5.7). The efficacies of the ligands are very much underestimated especially with Model1 which has a higher correlation coefficient R^2 (0.748 vs 0.658 in Model2). While the ligands 11 and 51 have partial agonism activity, according to these estimations they are supposed to be antagonists. Prediction with Model2 gives a more realistic efficacy at least for ligand 11, but 51 is still predicted to be an antagonist.

5.3.2 MM-GB/SA Binding Free Energies

MM-GB/SA approach can give more reliable binding free energies when compared with the docking methods. Following the protocol developed by Cifci *et al.* [70] we performed multiple independent molecular dynamics of host-guest systems starting from the AutoDock poses. The average binding energies of the receptor:ligand complexes were obtained by post processing 40

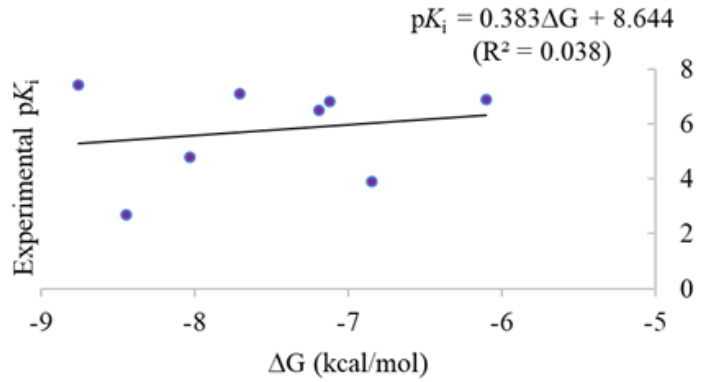


Figure 5.2: Correlations between experimental pK_i values and the lowest ΔG scores obtained from a multiple docking scheme by AutoDock. (The black line represent the linear fit between the lowest AutoDock $\Delta G_{binding}$ values and the experimental pK_i's.)

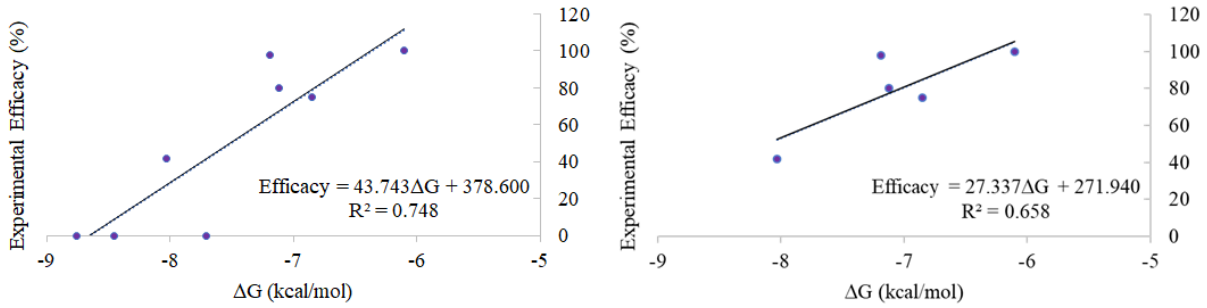


Figure 5.3: Correlations between experimental efficacies and the lowest ΔG scores obtained from a multiple docking scheme by AutoDock for all ligands in the training set (a) and for five ligands with agonist and partial agonist activity (b). (The black line represent the linear fit between the lowest AutoDock $\Delta G_{binding}$ values and the experimental efficacies.)

Table 5.7: Intrinsic efficacies (in %) obtained by the Model1 and Model2 for the set of ligands in the test set

	11	**51**
Eff_{exp}	*63*	*32*
Eff_{Model1} = 43.743ΔG + 378.600	-5	-66
Eff_{Model2} = 27.337ΔG + 271.940	39	-6

independent 1 ns MD runs using MM- GB/SA approach.

When independent constraint-free molecular dynamics simulations were performed, the protein structure unfolded. This is due to the fact that the X-ray LBD is only a fraction of the total protein assembly and it is not stable when it is modeled isolated. To prevent the unfolding of the protein we have applied constraint by restraining the backbone atoms (C, C_α N, O) into their initial positions with an harmonic force of 10 kcal mol^{-1} $Å^{-2}$. Following the equilibration and sampling steps with restraints, binding free energies have been calculated with the MM-GB/SA methodology. Five different solvation models have been utilized (GB^{HCT}, GB^{OBC}, GB^{OBC2}, GB^{neck} and GB^{neck2}) and the results are given in Table 5.8 for each ligand in each target protein. Protein's and ligand's maximum deviations from the starting structure have also been provided. PDB codes written in bold denote the original host of the corresponding ligand and the red colored numbers denote the lowest binding free energies for the set of host structures in each solvation model.

We sometimes observed high RMSD values for the ligand; for example 12.52 Å for glycine :4KCC complex. That means that the ligand left the LBD during the simulation. It can be explained by the fact that the LBD in 4KCC is open and the ligand is relatively small. The reverse case is valid for cycloleucine:1PB7 complex with 24.83 Å RMSD value; the large ligand is prevented from entering the closed cleft conformation of the LBD. Glycine gave the lowest binding energy with 1PB7 (GB^{HCT} model), ACPC with 1Y20 (GB^{OBC2}, GB^{OBC} and GB^{neck} models), D-cycloserine with 1PB9 (GB^{HCT} and GB^{OBC} models), TK40 with 4KFQ (GB^{HCT}, GB^{OBC}, GB^{OBC2}, GB^{neck} and GB^{neck2} models) and DCKA with 1PBQ (GB^{HCT}, GB^{OBC}, GB^{OBC2}, GB^{neck} and GB^{neck2} models) which match the experimental crystal structures of the ligands. Among the solvation models used in free energy calculations, GB^{HCT} and GB^{OBC} models seem to be suitable to be utilized in this system. We failed to reproduce the lowest binding energies for D-serine, ACBC and cycloleucine with their original crystal structures 1PB8, 1Y1Z and 1Y1M respectively.

Table 5.8. Binding free energies calculated by MM-GB/SA method with different solvation models and corresponding maximum root mean square deviations from the beginning conformations of the protein and the ligand (10 kcal mol^{-1} $Å^{-2}$ explicit constraint force introduced).

Ligand	**Target Protein**	**GB^{HCT}**	**GB^{OBC}**	**GB^{OBC2}**	**GB^{neck}**	**GB^{neck2}**	**RMSD *max* Protein**	**RMSD *max* Ligand**
GLYCINE	**1PB7**	-51.80	-47.80	-54.40	-46.61	-38.52	0.19	0.72
	1PB8	-49.93	-46.02	-52.26	-44.73	-37.11	0.19	0.86
	1Y20	-51.66	-49.31	-56.31	-49.57	-38.55	0.19	1.03
	1PB9	-50.41	-47.21	-53.33	-45.99	-39.79	0.19	0.91
	1Y1Z	-51.37	-47.42	-53.69	-44.24	-36.76	0.19	0.76
	1Y1M	-39.61	-36.20	-40.28	-31.82	-30.34	0.19	1.19
	4KCC	-11.93	-9.27	-10.35	-8.43	-8.84	0.19	12.52
	4KFQ	-27.19	-22.80	-25.33	-16.93	-18.20	0.19	1.84
	1PBQ	-27.61	-22.24	-24.36	-15.87	-18.50	0.19	1.74

Table5.8 – continued from previous page

Ligand	Target Protein	GB^{HCT}	GB^{OBC}	GB^{OBC2}	GB^{neck}	GB^{neck2}	RMSD *max* Protein	RMSD *max* Ligand
D-SERINE	1PB7	-54.51	-54.87	-62.02	-54.74	-44.53	0.19	1.16
	1PB8	-54.20	-56.12	-63.35	-56.76	-46.12	0.19	1.10
	1Y20	-56.42	-57.57	-64.48	-57.35	-45.89	0.19	0.95
	1PB9	-50.57	-53.41	-60.64	-54.01	-42.39	0.19	1.30
	1Y1Z	-58.60	-60.96	-68.85	-63.71	-49.47	0.19	0.73
	1Y1M	-40.11	-45.30	-51.05	-44.68	-35.39	0.19	1.40
	4KCC	-30.03	-30.03	-33.61	-27.40	-24.96	0.19	1.73
	4KFQ	-29.82	-25.53	-28.36	-18.07	-19.77	0.19	1.47
	1PBQ	-29.34	-25.44	-28.07	-17.66	-20.15	0.19	2.31
ACPC	1PB7	-57.67	-53.65	-59.65	-48.14	-38.15	0.19	1.30
	1PB8	-55.48	-53.01	-59.30	-48.01	-40.31	0.19	0.76
	1Y20	-56.42	-55.17	-61.85	-51.78	-40.73	0.19	0.92
	1PB9	-56.23	-54.52	-61.19	-50.32	-42.73	0.19	0.94
	1Y1Z	-56.56	-53.87	-60.14	-46.76	-39.32	0.19	0.99
	1Y1M	-46.74	-44.53	-49.13	-36.90	-32.99	0.19	1.37
	4KCC	-28.56	-27.97	-32.01	-26.63	-23.19	0.18	2.11
	4KFQ	-30.53	-27.05	-30.36	-20.07	-19.24	0.19	2.18
	1PBQ	-29.12	-24.75	-27.55	-17.77	-17.99	0.19	1.78
D-CYCLOSERINE	1PB7	-47.53	-34.71	-33.25	-15.03	-27.06	0.19	1.71
	1PB8	-49.31	-33.55	-30.19	-11.63	-28.70	0.19	1.49
	1Y20	-48.52	-36.20	-34.35	-18.32	-29.11	0.19	1.30
	1PB9	-51.23	-36.19	-33.33	-14.20	-31.48	0.19	1.12
	1Y1Z	-46.49	-33.72	-31.65	-15.86	-26.69	0.19	1.34
	1Y1M	-39.61	-27.92	-16.71	-10.59	-21.26	0.19	1.80
	4KCC	-17.76	-10.01	-8.88	-3.83	-8.05	0.19	5.03
	4KFQ	-21.78	-16.00	-16.07	-11.70	-13.18	0.19	13.99
	1PBQ	-19.70	-12.57	-11.73	-4.64	-10.23	0.19	4.16
ACBC	1PB7	-55.15	-51.19	-57.46	-48.32	-39.24	0.19	0.74
	1PB8	-57.61	-53.35	-60.08	-49.92	-41.87	0.19	0.61
	1Y20	-58.19	-54.80	-61.82	-52.30	-41.02	0.19	0.94
	1PB9	-57.21	-53.04	-59.98	-49.55	-41.62	0.19	0.75
	1Y1Z	-57.76	-54.06	-60.99	-50.68	-39.86	0.19	0.78
	1Y1M	-46.42	-43.09	-48.03	-39.52	-36.29	0.19	1.14
	4KCC	-31.61	-30.56	-34.80	-29.82	-25.00	0.19	2.57
	4KFQ	-36.04	-32.41	-35.67	-29.07	-25.97	0.19	1.26
	1PBQ	-25.62	-22.02	-24.26	-17.48	-17.97	0.19	1.51

Table5.8 – continued from previous page

Ligand	Target Protein	GB^{HCT}	GB^{OBC}	GB^{OBC2}	GB^{neck}	GB^{neck2}	RMSD *max* Protein	RMSD *max* Ligand
CYCLOLEUCINE	1PB7	-4.75	-2.45	-1.75	1.00	-3.02	0.19	24.83
	1PB8	-24.42	-18.31	-19.47	-14.09	-15.07	0.19	1.46
	1Y20	-60.76	-57.14	-63.92	-55.45	-42.89	0.19	1.02
	1PB9	-16.94	-12.75	-12.75	-6.60	-9.94	0.19	9.76
	1Y1Z	-59.34	-55.11	-61.56	-52.16	-41.26	0.19	0.93
	1Y1M	-49.39	-45.80	-50.63	-42.59	-39.09	0.19	0.94
	4KCC	-35.59	-33.79	-37.70	-32.62	-28.07	0.19	1.79
	4KFQ	-37.25	-33.26	-36.47	-27.63	-25.62	0.19	2.14
	1PBQ	-34.35	-29.44	-32.05	-22.20	-23.54	0.19	2.02
TK40	1PB7	-21.48	-18.35	-18.87	-17.51	-13.36	0.19	3.72
	1PB8	-22.25	-18.92	-19.48	-17.60	-15.21	0.19	2.96
	1Y20	-22.13	-18.60	-19.05	-17.38	-14.44	0.19	3.61
	1PB9	-20.56	-17.58	-18.16	-16.87	-12.35	0.19	4.31
	1Y1Z	-24.05	-21.07	-22.07	-19.74	-17.31	0.19	3.34
	1Y1M	-24.28	-21.81	-23.90	-20.82	-14.69	0.19	3.61
	4KCC	-27.25	-23.62	-25.43	-22.98	-19.49	0.19	3.17
	4KFQ	-34.97	-31.21	-33.19	-31.11	-27.05	0.19	1.54
	1PBQ	-29.05	-24.79	-26.29	-24.62	-21.49	0.19	1.86
DCKA	1PB7	-21.45	-19.13	-21.98	-19.39	-12.87	0.19	2.50
	1PB8	-22.78	-19.21	-21.70	-18.96	-15.72	0.19	2.30
	1Y20	-22.91	-19.68	-22.33	-19.45	-14.61	0.19	2.35
	1PB9	-22.71	-19.57	-22.36	-19.93	-14.43	0.19	2.74
	1Y1Z	-22.44	-19.33	-22.09	-20.07	-14.61	0.19	1.79
	1Y1M	-22.96	-21.01	-23.45	-19.91	-15.53	0.19	3.89
	4KCC	-27.42	-24.01	-25.66	-22.62	-18.81	0.19	1.49
	4KFQ	-29.80	-30.22	-33.65	-29.42	-20.38	0.19	2.24
	1PBQ	-42.36	-40.76	-46.15	-41.84	-31.20	0.19	1.63

In order to find a correlation between the calculated MM-GB/SA lowest $\Delta G_{binding}$ values and the experimental binding affinity (Figure 5.4) or the efficacy values (Figure 5.5) of the ligands, we applied linear regression to the training set.

From the correlation graphs in Figure 5.4, a very slight improvement in establishing a linearity between ΔG and pK_i is observed only in GB^{HCT} model, which is still poor for constructing a method for pK_i prediction.

The correlation graphs in Figure 5.5 give poorer relationships between the efficacies of the ligands and the binding free energies calculated by MM-GB/SA approach compared to AutoDock results. GB^{HCT} and GB^{neck2} models give the highest R^2 values but the suggested MM-GB/SA protocol fails in estimating the efficacies of the NMDA GluN1 sub-site specific ligands. On the other hand, the calculated binding free energies between each ligand in the training set and each X-ray crystal structure give correlations with the activity of ligands (agonism-

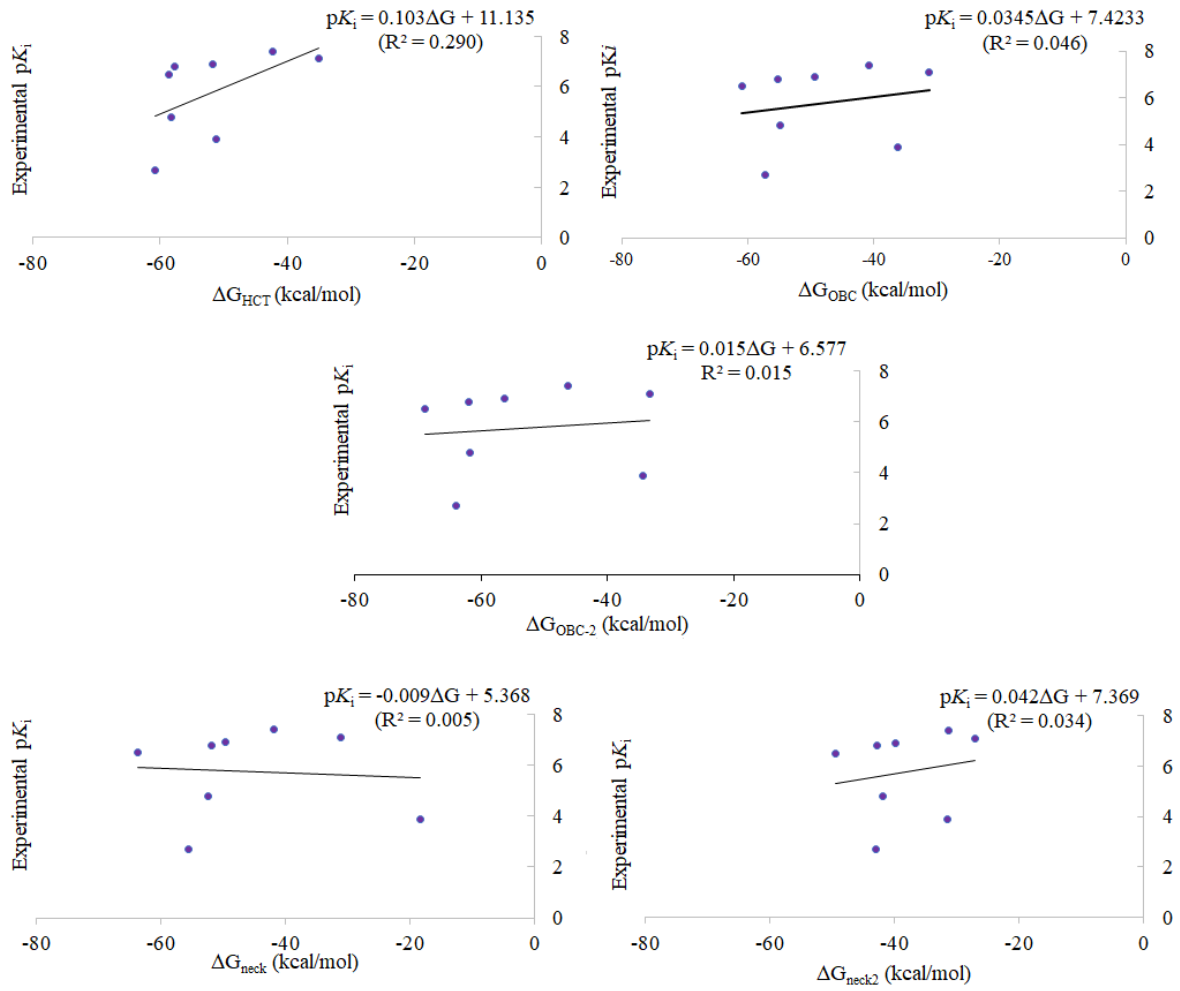


Figure 5.4: Correlations between experimental binding affinities (pK_i) and lowest MM-GB/SA averaged $\Delta G_{binding}$ free energies of eight ligands in the training set. (The black lines represent the linear fit between MM-GB/SA $\Delta G_{binding}$ values and the experimental affinities.)

.

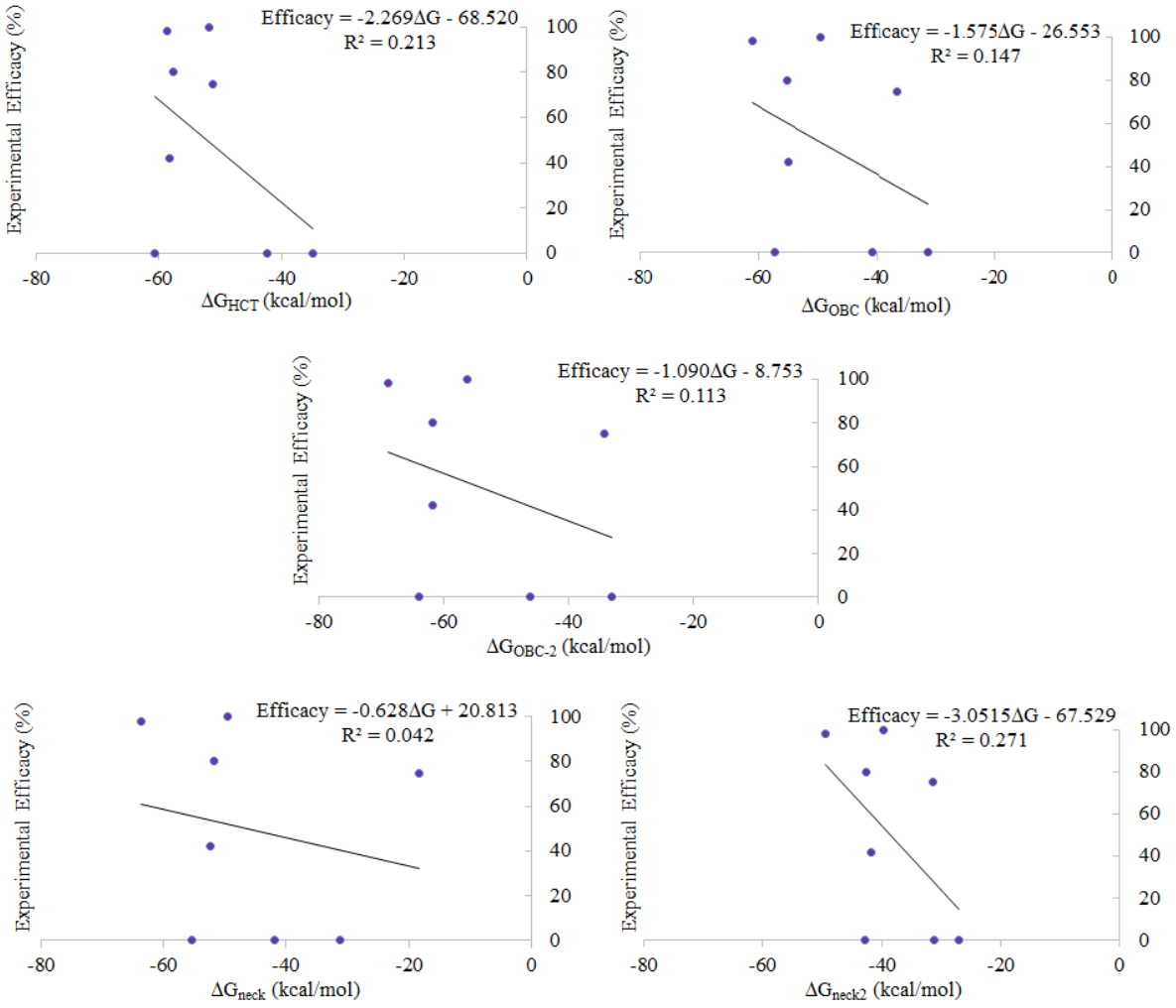


Figure 5.5: Correlations between experimental efficacies (%) and lowest MM-GB/SA averaged $\Delta G_{binding}$ free energies of eight ligands in the training set. (The black lines represent the linear fit between MM-GB/SA $\Delta G_{binding}$ values and the experimental efficacies.)

.

antagonism) (Figure 5.6). By these correlations one can discriminate a ligand's agonism activity from antagonism activity, but not agonists from partial agonists. There is a clear trend in agonism nature with a positive slope (glycine, D-serine, ACPC, D-cycloserine and ACBC), while in antagonism nature there is a negative slope (cycloleucine, TK40 and DCKA).

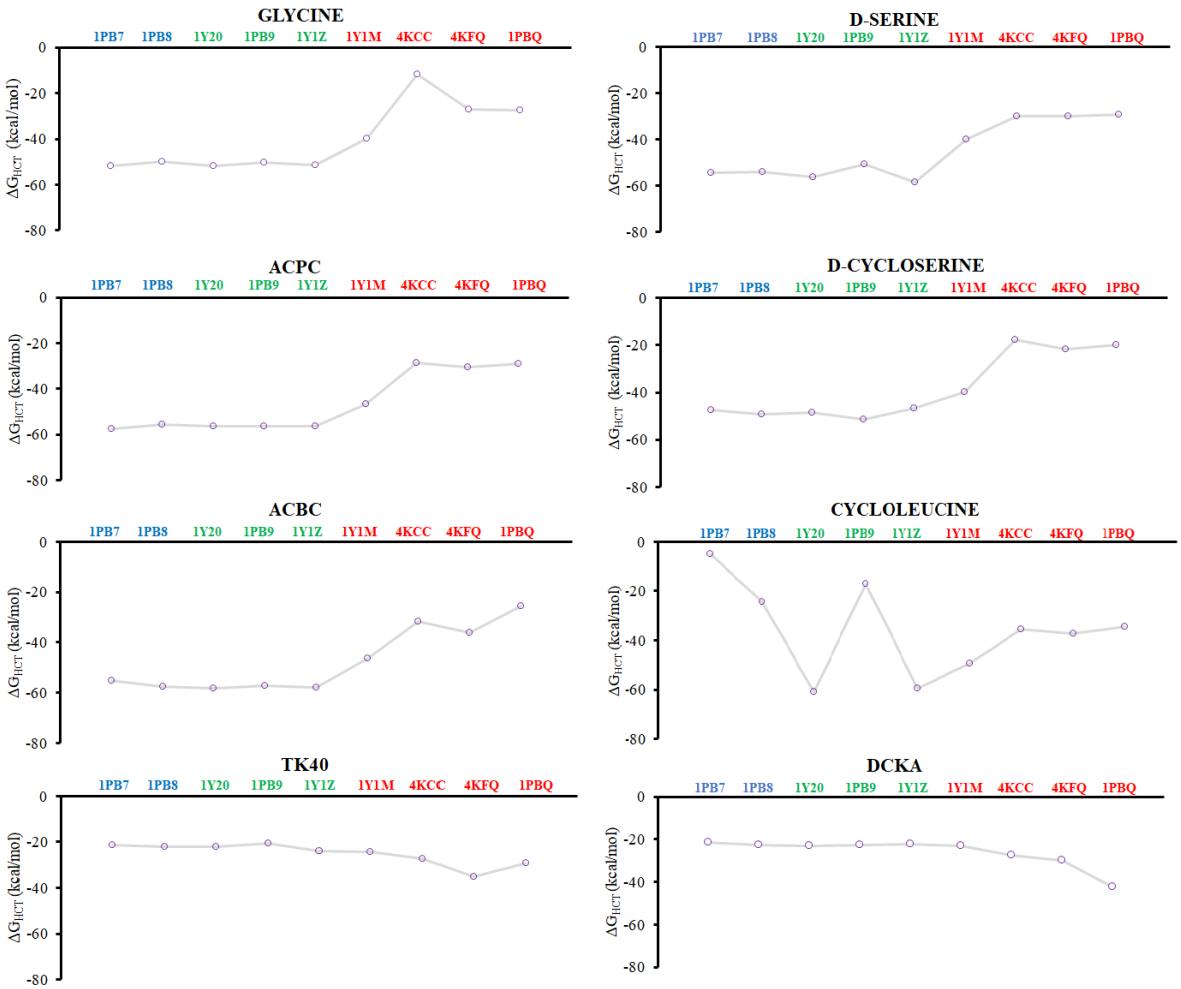


Figure 5.6: Correlation of binding free energies (solvation model GB^{HCT}) to the corresponding X-ray crystal structures. The PDBs are sequenced from closest conformation of LBD (1PB7) to the open conformation (1PBQ).

.

5.4 Conclusions

In this study, the interactions of the NMDA receptor GluN1 subunit ligand binding domain with various ligands have been studied and the conformational responses of the ligand binding domain to the ligand's activity (agonism vs antagonism) have been investigated. All available X-

ray crystal structures from PDB database have been collected. GluN1 LBDs have been isolated and the ligands have been removed. RMS fit to a reference crystal structure (1PB7) has been performed and grouped with respect to their differences in LBDs as open, partially open and closed conformations.

Eight ligands in the training set have been docked to nine PDBs and lowest binding energy for each docking procedure was selected as the best pose. A linear regression fit between the lowest Autodock $\Delta G_{binding}$ values and the experimental affinity constants of the ligands were obtained for the training set but no relationship was found. The same study was performed in order to find a link between efficacy of ligands and the lowest Autodock $\Delta G_{binding}$ values and relatively good correspondances were observed with correlation coefficients of 0.748 and 0.658 in two suggested models which means that AutoDock is capable of predicting the most favorable configuration of the ligands inside the LBD, but fails in estimating experimental $\Delta G_{binding}$ values.

40 independent molecular dynamics simulations have been performed for each best pose starting from the docking process. We first applied free MD but it resulted in partial unfolding of the proteins. Therefore an harmonic force of 10 kcal mol^{-1} $Å^{-2}$ has been applied to the protein's backbone atoms, constrained MD was succesfully performed. Five different solvation models have been used for each calculation. GB^{HCT} and GB^{OBC} solvation models were succesful in reproducing the original co-crystal structures by giving the lowest binding energies in five calculations (glycine:1PB7, ACPC:1Y20, D-cycloserine:1PB9, TK40:4KFQ, DCKA:1PBQ). Correlations between experimental efficacy and affinity values of ligands with the calculated binding free energies have been checked, but poor linearities have been obtained with all solvation models. This might be because the receptor's functionality is tightly linked to the NMDA subunits including the ion channel. This should be taken into account when performing the simulations. On the other hand, there is a good correlation between the ligand's agonism-antagonism activity and the receptor's conformational state. By these correlations one can discriminate between a ligand's agonism activity from antagonism activity, but not agonists from partial agonists.

Unfortunately modeling the full NMDA assembly will yield a biomolecular system of several hundreds thousands of atoms but also a membrane and water as a solvent. Limited computational resources do not allow the simulation of the whole receptor at the moment.

Chapter 6

Using atomic charges to describe carboxylic acid pKa's

6.1 Introduction

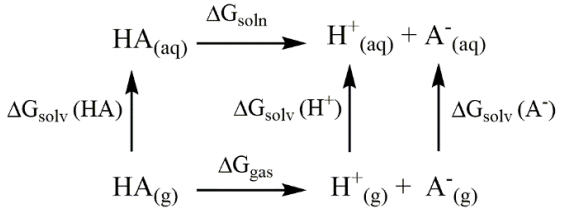
The solubility, lipophilicity and permeability of a drug ligand in cell membrane is governed by the pK_a's of the acidic and basic sites within the molecule . When the ligand is in the target protein environment, its affinity, activity and efficacy is dependent on the pK_a's of the interacting residues of the active site.

Carboxylic acids are the main acidic functional groups in biological systems. Glutamate and aspartate have carboxylic acid groups in their side chains and these groups help in holding the peptide together by hydrogen bonds. More than 30% of the ionizable residues (32% of the Arg residues, 19% of the Asp residues, 13% of the Glu residues, and 6% of the Lys) [71] are buried inside the hydrophobic cavities which limits the contact with solvent [72]. Since the protein matrix is heterogeneous, the fluctuations in the electrostatic environment alter the interactions between buried charges which in turn leads to modifications in the affinities of the protonation sites for ionization; and thus their pK_a values are re-adjusted [73]. Eventually, in polar parts of the protein the pK_a of the acidic groups in the residues shifts to the higher values and the pK_a of the basic groups shifts to the lower values from that of the isolated amino acids [74]. Hydrogen bondings between the amino acid's functional groups and the side chain or the backbone atoms also tend to result in pK_a deviations; especially when the number of H-bonds increases and if they are rigid the effect is larger such that the pK_a for acidic side chains are perturbed above their intrinsic pK_a values and for the basic groups the reverse is observed [20, 75]. Salt-bridge formation between two residues, which contributes to protein stability, is also reported to result in lower or higher pK_a values with the same trends in polarization and hydrogen bonding effects [76].

Dissociation constants of organic compounds can be determined experimentally by potentiometric, spectrophotometric, chromatographic, electrophoresis, calorimetric, conductometric and NMR techniques but among them only NMR titrations estimate the pK_a values accurately when applied to amino acids. In case of non-soluble membrane proteins, where NMR is inapplicable, one needs X-ray crystal structures but there is only a limited number of structures present in the literature [77]. On the other hand, poor resolution of the crystals and lack of conforma-

tional flexibility restrict the X-ray crystallography applications. A more reliable technique for producing high resolution crystallographic protein structures is neutron diffraction method. Despite its strong accuracy in determining hydrogen locations on the amino acids, this technique is much more limited in use with respect to NMR and X-ray crystallography, since there is a little number of instruments available for measurements and it needs long data collection times and larger volumes of crystals [78].

Due to the difficulties in pK_a measurements of molecules in large complex mediums and short-lived intermediates with experimental means as discussed above, the need for accurate pK_a estimations by the applications of theoretical approaches is necessary. The features that determine the acidities of different classes of chemical compounds can be explained by the molecular structure. The traditional method for the calculation of pK_a's is based on the free energy changes in the thermodynamic cycle (Figure 6.1) in which desolvation of HA to gas phase and then deprotonation of HA is considered. Typically electrostatic interactions are obtained by numerically solving the linearized Poisson-Boltzmann equation (LPBE).



$$pK_a = \frac{\Delta G_{soln}}{RT \ln(10)}$$

$$\Delta G_{soln} = \Delta G_{gas} + \Delta G_{solv}(H^+) + \Delta G_{solv}(A^-) - \Delta G_{solv}(HA)$$

Figure 6.1: Thermodynamic cycle.

Despite the enormous number of successful pK_a predictions by using the deprotonation energies and solvation free energies [79, 80, 81, 82], these calculations usually fail in its purpose due to the instability of the ion in gas phase and the conformational differences between the solvent and gas phase calculations [83]. Besides empirical methods such as PROPKA and the methods based on Poisson-Boltzmann equation, Generalized Born equation, QM/MM or Molecular Dynamics or a combination of one or more; quantitative structure property relationship (QSPR) is a widely used technique in which several molecular descriptors are successfully linked to pK_a's of organic molecules such as topological state [84, 85], atom type [86, 87], group philicity [88], bond length and frequency [89, 90], maximum surface potential [91], HOMO and LUMO energies [92, 93], atomic charge [94, 95]. Among them, the concept of partial atomic charges is closely related to the relative acidity and basicity of a molecule.

A Multiple Linear Regression model was developed by Dixon and Jurs with an accuracy of 0.5 units for the calculation of pK_a's of oxyacids by using the empirical atomic charges of atoms in a molecule [95]. The model is based on the changes in the σ and π charges upon going from the neutral to ionic state, concerning the resonance and inductive effects of nearby atoms.

Citra constructed four linear regression models by using the partial atomic charges on oxygen and hydrogen atoms which are involved in deprotonation and O-H bond order for the set of phenols, alcohols and aromatic and non-aromatic carboxylic acids [96]. Various combinations of different level of theories, basis sets and charge models were tested by Vareková *et al.* in order to create a model for phenols [97]. Recently, Ugur *et al.* made use of a similar approach with an extended study for the prediction of amino acid pK_a's in proteins and developed an accurate protocol by computing the atomic charge on the anionic form of alcohols and thiols [98]. Among the tested DFT functionals, basis sets, semiempirical methods, solvation and charge models, they observed the best combination is NPA charge calculation in CPCM model at the B3LYP/3-21G (R^2=0.995) level of theory for alcohols and M06-2X/6-311G (R^2=0.986) level of theory for thiols in order to reproduce the experimental pK_a's. Moreover, they tested the stability of the calculated pK_a's in amino acids by MM-MD and DFT-MD calculations.

Regarding the successful applications of QM charges as descriptors, in this study we aim to suggest an accurate protocol for the fast prediction of pK_a's of carboxylic acids.

6.2 Methodology

6.2.1 Experimental Database

From literature [99, 100], we have selected a total of 59 carboxylic acid compounds with pK_a's ranging from 0.65 to 5.12. One of our criterion of selection has been to identify molecules that are able to represent the widest range of experimental pK_a's as possible. Most of these molecules are also small and rather rigid molecules. The reason for avoiding flexible molecules is to overcome the risk of failing to obtain the global minimum during geometry optimization, which would raise systematical errors in pK_a predictions [100]. A training set of 30 small molecules (see Table 6.1 and Figure 6.2) and a test set of 29 small molecules (see Table 6.2 and Figure 6.3) have been extracted from the ensemble.

Table 6.1. Carboxylic Acid Training Set: CAS Number, Molecule Name, Experimental pK_a, Predicted pK_a, and Differences between Experimental and Predicted pK_a values

CAS Number	Molecule Name	pK_a (exp.)	pK_a (pred.) [a]	Δ pK_a
129-66-8	2,4,6-Trinitrobenzoic acid	0.65[99]	0.63	-0.02
610-30-0	2,4-Dinitrobenzoic acid	1.42[100]	1.39	-0.03
471-25-0	Propiolic acid	1.89[100]	1.85	-0.04
552-16-9	2-Nitrobenzoic acid	2.21[100]	2.60	0.39
1460-34-0	α-Keto-β-methylvaleric acid	2.30[100]	2.18	-0.12
590-93-2	2-Butynoic acid	2.62[99]	2.52	-0.10
298-12-4	2-Oxoacetic acid	2.98[100]	2.38	-0.60
69-72-7	2-Hydroxybenzoic acid	2.98[99]	2.89	-0.09
122-59-8	Phenoxyacetic acid	3.17[99]	3.62	0.45
88-14-2	2-Furoic acid	3.27[100]	3.52	0.25

Table6.1 – continued from previous page

CAS Number	Molecule Name	pK_a (exp.)	pK_a (pred.) [a]	Δ pK_a
62-23-7	4-Nitrobenzoic acid	3.43[99]	3.45	0.02
480-63-7	2,4,6-Trimethylbenzoic acid	3.55[100]	3.74	0.19
625-45-6	Methoxyacetic acid	3.57[100]	3.93	0.36
1877-72-1	3-Cyanobenzoic acid	3.60[99]	3.66	0.06
33445-07-7	Isopropoxyacetic acid	3.69[100]	3.97	0.28
64-18-6	Formic acid	3.76[100]	3.86	0.10
627-03-2	Ethoxyacetic acid	3.84[100]	3.96	0.12
488-93-7	3-Furoic acid	3.90[99]	4.05	0.15
99-06-9	3-Hydroxybenzoic acid	4.08[99]	4.07	-0.01
93-09-4	2-Naphtoic acid	4.16[99]	4.04	-0.12
190965-42-5	3-Propoxybenzoic acid	4.20[100]	4.08	-0.12
99-04-7	3-Methylbenzoic acid	4.25[99]	4.18	-0.07
103-82-2	Phenylacetic acid	4.31[99]	4.37	0.06
99-50-3	3,4-Dihydroxybenzoic acid	4.48[99]	4.41	-0.07
79-31-2	Isobutyric acid	4.64[100]	4.74	0.10
1759-53-1	Cyclopropanecarboxylic acid	4.83[99]	4.64	-0.19
142-62-1	Hexanoic acid	4.85[99]	4.94	0.09
6202-94-4	trans-2-Methylcyclopropane-carboxylic acid	5.00[100]	4.75	-0.25
6142-57-0	cis-2-Methylcyclopropane-carboxylic acid	5.02[100]	4.72	-0.30
541-47-9	3-Methyl-2-butenoic acid	5.12[100]	4.64	-0.48

[a] pK_a values are computed for each molecule on the anionic form, optimized with M06L/6-311G(d,p) and SMD, using the highest NPA atomic charge of the two oxygen atoms of the carboxylate fragment

Table 6.2. Monocarboxylic Acid Test Set: CAS Number, Molecule Name, Experimental pK_a, Predicted pK_a, and Differences between Experimental and Predicted pK_a values

CAS Number	Molecule Name	pK_a (exp.)	pK_a (pred.) [a]	Δ pK_a
625-75-2	Nitroacetic acid	1.48[99]	2.19	0.71
372-09-8	Cyanoacetic acid	2.47[99]	3.34	0.87
127-17-3	Pyruvic acid	2.60[100]	2.59	-0.01
5699-58-1	Acetopyruvic acid	2.61[100]	2.11	-0.50
121-92-6	3-Nitrobenzoic acid	3.46[99]	3.57	0.11

Table6.2 – continued from previous page

CAS Number	Molecule Name	pK_a (exp.)	pK_a (pred.) [a]	Δ pK_a
619-65-8	4-Cyanobenzoic acid	3.55[99]	3.62	0.07
2516-93-0	Butoxyacetic acid	3.66[100]	3.96	0.30
54497-00-6	Propoxyacetic acid	3.69[100]	3.94	0.25
50-21-5	2-Hydroxypropanoic acid	3.83[100]	3.95	0.12
79-14-1	Hydroxyacetic acid	3.83[99]	3.90	0.07
118-90-1	2-Methylbenzoic acid	3.98[100]	3.79	-0.19
586-38-9	3-Methoxybenzoic acid	4.09[100]	4.07	-0.02
65-85-0	Benzoic acid	4.19[100]	4.12	-0.07
2529-39-7	2,3,4,5-Tetramethylbenzoic acid	4.22[100]	4.06	-0.16
86-55-5	1-Naphtoic acid	3.69[99]	3.71	0.02
79-10-7	Acrylic acid	4.25[99]	3.95	-0.30
1077-07-2	3-Allylbenzoic acid	4.32[100]	4.12	-0.20
99-94-5	4-Methylbenzoic acid	4.37[99]	4.27	-0.10
5438-19-7	4-Propoxybenzoic acid	4.46[100]	4.43	-0.03
100-09-4	4-Methoxybenzoic acid	4.50[99]	4.42	-0.08
1498-96-0	4-Butoxybenzoic acid	4.53[100]	4.43	-0.10
99-96-7	4-Hydroxybenzoic acid	4.58[100]	4.45	-0.13
64-19-7	Acetic acid	4.76[100]	4.81	0.05
107-92-6	Butyric acid	4.82[100]	4.90	0.08
109-52-4	Pentanoic acid	4.86[100]	4.93	0.07
79-09-4	Propanoic acid	4.87[99]	4.87	-0.00
98-89-5	Cyclohexanecarboxylic acid	4.90[100]	5.05	0.15
3400-45-1	Cyclopentanecarboxylic acid	4.99[99]	4.93	-0.06
75-98-9	Trimethylacetic acid	5.05[100]	4.63	-0.42

[a] pK_a values are computed for each molecule on the anionic form, optimized with M06L/6-311G(d,p) and SMD, using the highest NPA atomic charge of the two oxygen atoms of the carboxylate fragment

6.2.2 Quantum Mechanical Calculations

All of the Quantum Mechanical (QM) calculations were performed using the Gaussian 09 program package [195]. Eight different density functionals (BLYP [101, 102], B3LYP [101, 103], OLYP [101, 104], PBEPBE [105], PBE1PBE [158], M06 [106, 229], M06L [159, 229], M06-2X [106, 229]) and 15 different basis sets were used. To interpret the aqueous solvent environment, the solvation model based on density (SMD [108]), the polarizable continuum model

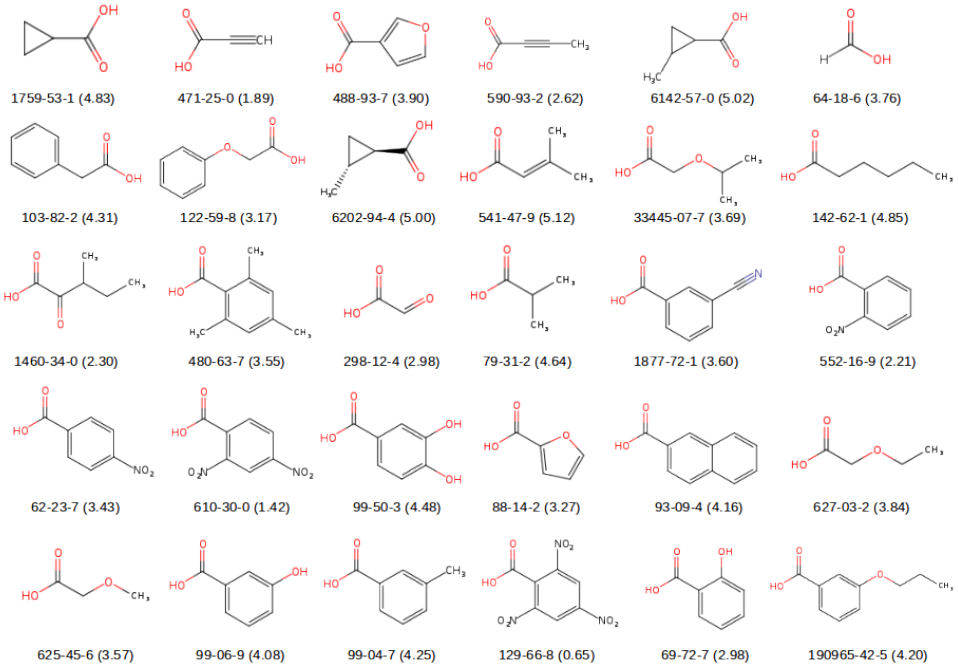


Figure 6.2: 2D structures of the carboxylic acids selected for the training set. Naming of the molecules is based on their CAS registry numbers and their experimental pK_a values are provided in parenthesis.

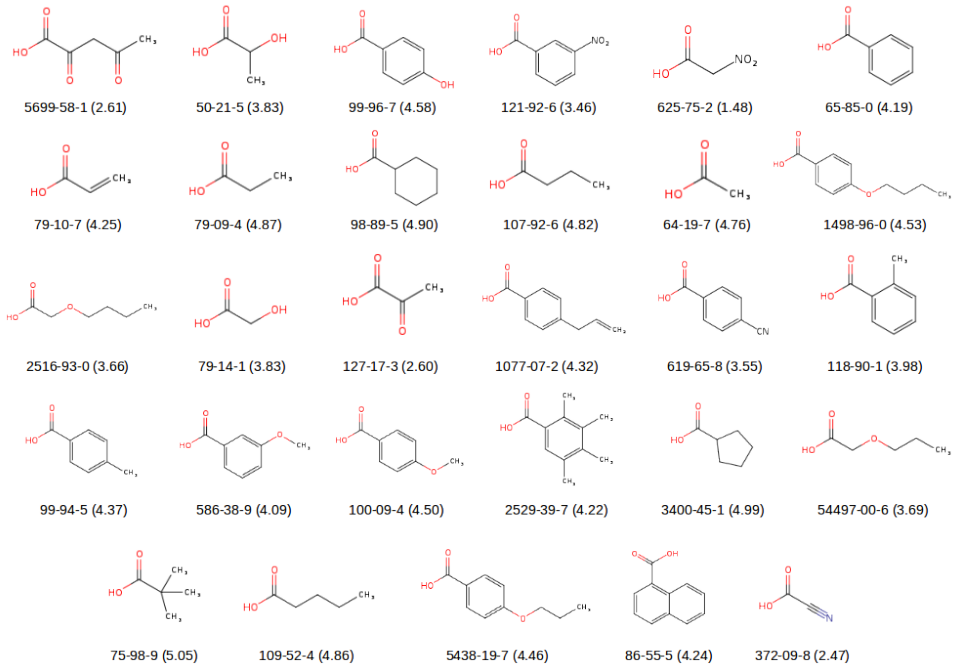


Figure 6.3: 2D structures of the carboxylic acids selected for the test set. Naming of the molecules is based on their CAS registry numbers and their experimental pK_a values are provided in parenthesis.

(PCM[109]), and the polarizable conductor solvent model (CPCM[110]) were used with a dielectric constant (ε) of 78.5. Three different types of atomic charge models were tested: Mulliken population analysis [111], Löwdin population analysis [112], Natural Population Analysis (NPA)[113]. Compared to the study from Ugur *et al.* [98], Electrostatic Potential (ESP) derived atomic charges, like the Merz-Kollman (MK) model [230] and the CHelpG model [231] are not reported here since preliminary studies have shown us that, as in the cases of thiols and alcohols, they do not perform better than NPA atomic charges (data not shown). Unless otherwise stated, all the charge calculations were performed on the optimized geometries (after including or not the solvent effect) that do not contain any imaginary frequency.

6.2.3 Molecular Dynamics Calculations

Molecular dynamics simulations have been performed using the AMBER biomolecular package [217]. All simulated molecules have been modeled with the AMBER ff14SB protein force field [219]. The aqueous polar environment was mimicked by the implicit modified generalized Born model with α, β, γ are 1.0, 0.8, and 4.8555 as implemented in AMBER 18 (igb = 5). Following minimization, the systems were heated up to 300 K using the Langevin thermostat during 50 ps with a collision frequency Υ = 10 ps^{-1}, and a timestep of 1 fs. Then, NVT production runs were performed for another 150 ps using the same thermostat algorithm. From each of these molecular dynamics, 1500 frames were extracted, one every 0.1 ps.

6.3 Results and Discussion

6.3.1 Linearity of the Relationship Between Experimental pK_a's and Atomic Charges

The linear relationship between atomic charges and experimental pK_a's depends on many factors: the choice of the DFT method, the choice of the basis set, the use (or not) of an implicit solvent model, the type of the atomic charge model, and which atomic charges are considered. From the overall present study, we have found that the best combination of all these factors is to consider the highest oxygen atomic charge of each carboxylate fragment computed with NPA at the M06L/6-311G(d,p) level using the SMD implicit solvent model. In what follows, we present the linear relationship between experimental pK_a's and atomic charges computed using the theoretical framework detailed above. Then, using these results as a reference, we discuss the choice of charge descriptor, charge model, solvent model, DFT functional and basis set by changing one of these parameters while the others remain fixed to their best combination.

For each molecule of the training set, a geometry optimization was performed at the M06L/6-311G(d,p) level using the SMD implicit solvent model. We ensure that no imaginary frequency remains for any molecule. Atomic charges were computed using the natural population analysis. For each carboxylate fragment, we extracted the highest of the two oxygen atomic charges and we compared it with the experimental pK_a of the corresponding molecule. Figure 6.4 shows the relationship between experimental pK_a and computed NPA charge for the training set. A linear equation is obtained by a least-square fit:

$$\mathrm{p}K_a = a \cdot Q + b \quad \text{with} \quad Q = \max\{q(\mathrm{O}_1), q(\mathrm{O}_2)\} \tag{6.1}$$

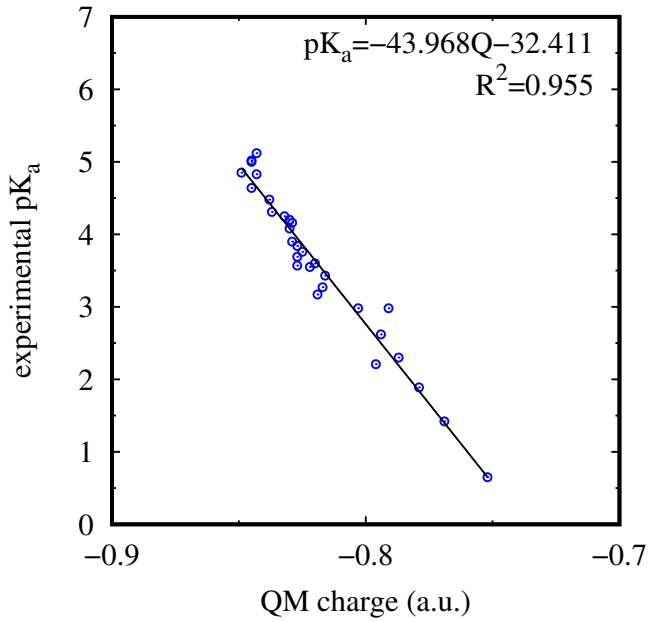


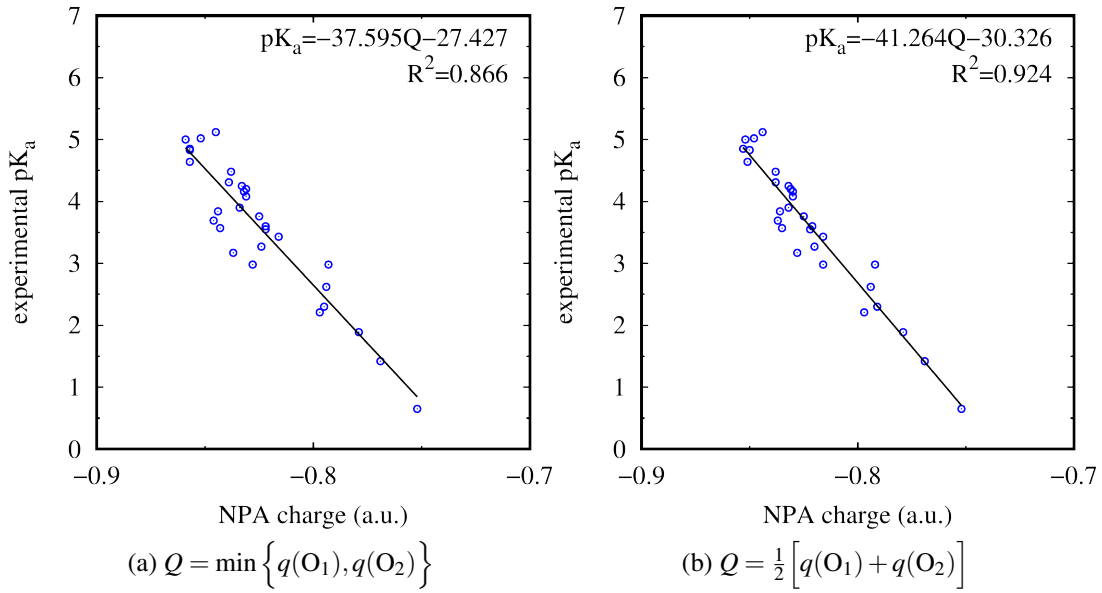
Figure 6.4: Linear regression between calculated NPA atomic charges and experimental pK_a. Calculations were done using M06L/6-311G(d,p)//SMD.

where a and b are the fitted parameters and $Q = \max\{q(O_1), q(O_2)\}$ is the highest atomic charges of the two carboxylate oxygens, respectively. The parameters a and b and the squared Pearson correlation coefficient (R^2) are also illustrated in Figure 6.4. The predicted pK_a's are computed using Equation 6.1 (i.e., by reporting $\max\{q(O_1), q(O_2)\}$ of a given molecule into the parametrized equation).

For carboxylate molecules, the R^2 value has been found to be 0.955. No strong outlier molecule was observed for the training set. The maximum difference between the predicted and experimental pK_a among all the molecules was found as -0.60 units (see Table 6.1). These results indicate a strong correlation between experimental pK_a's and the oxygen charges. In order to analyze the influence of the charge descriptor, charge model and solvent model on the quality of the fit, the same protocol was applied with four other charge descriptors, two other charge models, two other solvent models and gas phase calculations.

6.3.2 Influence of the Charge Descriptor

Compared to alcohols and thiols that were analyzed by Ugur *et al.*[98], the negative charge of the base form in the case of carboxylate can be shared between different atoms: the carbon and the two oxygen atoms of the carboxylate fragment. Thus, there are different ways to extract atomic charges for this fragment and then to compare them with experimental pK_a's. We have analyzed different atomic extraction schemes for the negative charge Q of the carboxylate



(a) $Q = \min\left\{q(\mathrm{O_1}), q(\mathrm{O_2})\right\}$ (b) $Q = \frac{1}{2}\left[q(\mathrm{O_1}) + q(\mathrm{O_2})\right]$

(c) $Q = q(\mathrm{C}) + q(\mathrm{O_1}) + q(\mathrm{O_2})$ (d) $Q = q(\mathrm{C})$

Figure 6.5: Effect of the charge descriptor on the linear regression between calculated atomic charges and experimental pK_a's. Calculations were done with M06L/6-311G(d,p)//SMD: (a) Minimum atomic charge on O_1 and O_2; (b) Average sum of atomic charges on O_1 and O_2; (c) Sum of atomic charges on C, O_1 and O_2; (d) Atomic charge on C.

fragment composed of atoms C, O_1 and O_2:

$$Q = \max\{q(O_1), q(O_2)\} \quad (6.2)$$

$$Q = \min\left\{q(O_1), q(O_2)\right\} \quad (6.3)$$

$$Q = \frac{1}{2}\left[q(O_1) + q(O_2)\right] \quad (6.4)$$

$$Q = q(C) + q(O_1) + q(O_2) \quad (6.5)$$

$$Q = q(C) \quad (6.6)$$

From the two oxygen atomic charges, it is possible to extract the highest value (Equation 6.2), the lowest value (Equation 6.3) or the average (Equation 6.4). The carbon atomic charge can also be taken into account via the sum of all 3 atomic charges (Equation 6.5) or by itself (Equation 6.6).

Figure 6.5 shows the relationship between carboxylate atomic charges expressed by Equations 6.3-6.6 and experimental pK_a's using M06L/6-311G(d,p)//SMD. When the lowest (i.e., the most negative) oxygen atomic charge is considered, the linear relationship is less accurate than with the highest oxygen atomic charge scheme: R^2 = 0.866 for the "min" scheme *vs.* $R^2 = 0.955$ for the "max" scheme, respectively. This is somewhat unexpected, since if one considers a proton, one could expect it to be more attracted by the most negative oxygen atoms. Therefore, one could expect that the $Q = \min\left\{q(O_1), q(O_2)\right\}$ scheme should better reflect the experimental pK_a's. In all our linear regressions with different density functionals, basis sets, etc., we have never found a better regression with the scheme $Q = \min\left\{q(O_1), q(O_2)\right\}$ than with its $Q = \max\{q(O_1), q(O_2)\}$ counterpart. As a consequence, the scheme $Q = \frac{1}{2}\left[q(O_1) + q(O_2)\right]$ that computes the average of the two oxygen atomic charges is placed in between the two previous scheme with $R^2 = 0.924$.

Another possibility to search for a relationship between experimental pK_a and atomic charge is to take into account the atomic charge on the carboxylate carbon. Figure 6.5(d) shows the (lack of) relationship between the carbon atomic charges and experimental pK_a's. With a $R^2 = 0.055$, the carbon charge cannot be regarded as a descriptor of the experimental pK_a. As a consequence,when the three atomic charges on the carboxylate fragment are considered together (Equation 6.5), the correlation coefficient ($R^2 = 0.536$) is worse than when the carbon atom is not included.

6.3.3 Influence of the Charge Model

In a pK_a prediction model, the variations in the pK_a during the dissociation process should be reflected precisely by the electronic changes. Three different charge schemes were tested for their predictivity power to generate charges that associate with the experimental pK_a's: NPA as well as Mulliken and Löwdin population analysis. These methods are based on charge partition schemes and define the atomic orbitals by wave functions. In the Mulliken population analysis, the calculated electron density is equally shared through the adjacent atoms in a molecule without taking into account the electronegativity and polarizability differences in atom types. Löwdin population analysis is very similar to the Mulliken method with only difference in usage

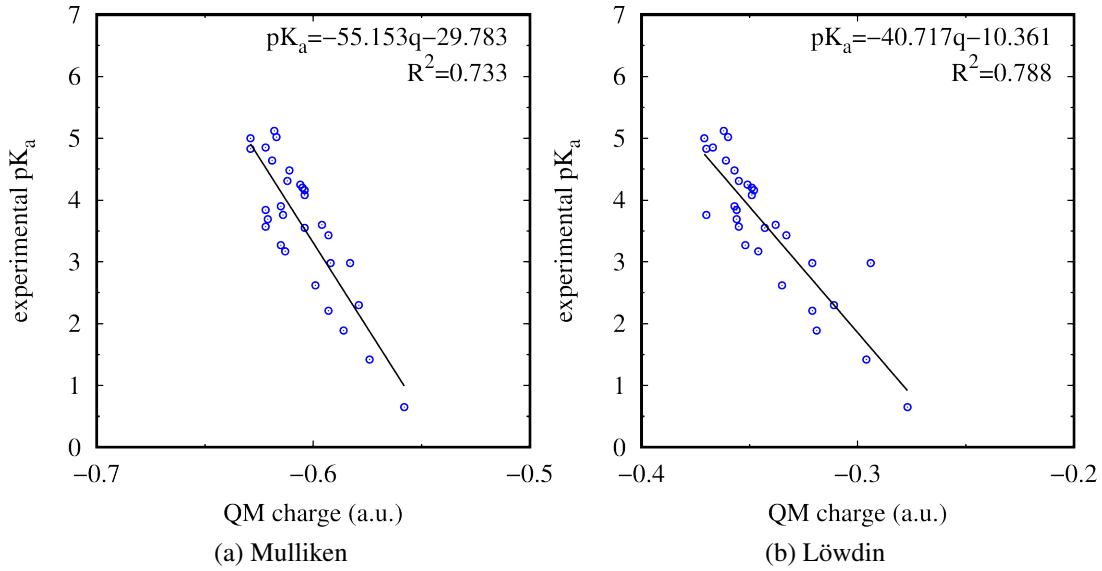


Figure 6.6: Effect of the charge model on the linear regression between calculated atomic charges and experimental pK_a's. Calculations were done with M06L/6-311G(d,p)//SMD: (a) Mulliken atomic charge model; (b) Löwdin atomic charge model

of orthogonal basis functions. Neither Löwdin or Mulliken schemes are able to reproduce the values of the dipole moments and they are both dependent on the basis set that is used. Natural population analysis localizes and classifies the orbitals into core, valence and Rydberg each of which contribute differently to the density. This partititoning of the atomic orbitals makes the NPA method less basis set dependent than its counterparts.

The strength of a carboxylic acid is determined by the strength of its conjugate base and the strength of a base is proportional to the charge density on the carboxylate oxygens. The lesser the charge density on the oxygen atoms means more stability and thus it becomes a weaker base and finally a stronger acid. Figure 6.6 presents the linear regressions between the highest oxygen atomic charge and experimental pK_a for the training set at the M06L/6-311G(d,p)//SMD using the Mulliken population analysis (Figure 6.6(a)) and the Löwdin population analysis (Figure 6.6(b)). The charge analysis show that the oxygen charges become more negative with increasing pK_a, suggesting that an oxygen atom with more associated electron density readily accepts a proton; indication of a stronger conjugate base and thus a weaker acid. On the other hand, Mulliken and Löwdin charges give R^2 coefficients lower than that of NPA with values of 0.733 and 0.788 respectively. This result is similar to those obtained for alcohols and thiols by Ugur *et al.*[98]: atomic charges extracted from natural population analysis are more linearly correlated to pK_a's than using the Mulliken's or Löwdin's schemes. Using Equation 6.1, the calculated pK_a of the strongest outlier is 1.35 unit different from the experimental pK_a when Löwdin charges are used. In case of Mulliken scheme, all predicted pK_a's are within ± 1 unit range, no strong outliers are observed.

6.3.4 Influence of the Solvent Model

The description of the surrounding environment that the charged species is exposed to accounts for the ideal charge derivation scheme. Implicit solvent models offer some advantages for modeling the interactions between the solute and solvent. In this part of the study, we have tested the accuracy of PCM and CPCM implicit solvation models in addition to SMD model calculations. Besides, due to its smaller computational costs, gas phase calculations have been taken into consideration. Figure 6.7 presents the linear regression fits of CPCM, PCM and gas phase calculations using NPA charges and the DFT methods as discussed in the previous sections.

Both PCM and CPCM calculations are as accurate as SMD calculations with $R^2 \geq 0.930$ (Figures 6.4 and 6.7). The predictivity of gas phase model is poorer ($R^2 = 0.826$) compared to other models where PCM, CPCM and SMD solvation methods are applied since in this study we have extracted the water phase acidities rather than gas-phase proton affinities. SMD model is different from PCM and CPCM models in considering the dispersion-repulsion energies in addition to electronic energy. These additional terms seem to contribute in finding the global minimum in geometry optimizations and assigning the atomic charges. Maximum deviations of the predicted pK_a's from the experimental pK_a's are found to be 0.75, 0.80 and 1.13 units for PCM, CPCM and gas phase calculations respectively.

6.3.5 Density Functionals and Basis Set Benchmarks

A deep analysis of DFT functionals and basis sets influence on pK_a prediction capability for carboxylic acids have been performed by applying the same protocol to the training set. Highest NPA charge on the oxygen atoms of carboxylate fragment calculated at various level of theories with SMD model were extracted to obtain R^2, a and b values in Equation 6.1 from the linear fit with experimental pK_a's. In Figure 6.8 for each combination of DFT functional and basis set, the Mean Absolute Deviations (MADs) are presented as box representations. The differences between the experimental and predicted pK_a's (ΔpK_a) have been calculated for each level of theory and the maximum value of this difference (MAX-ΔpK_a) is represented as black colored lines in Figure 6.8.

All of the DFT methods gave strong correlations between calculated NPA atomic charges and experimental pK_a's with R^2 range of $0.702 \leq R^2 \leq 0.955$. The largest MADs and MAX-ΔpK_a's were found for the combinations of 3-21G basis set with all the functionals except M06L. Removing the 3-21G basis set combinations from the benchmark study, we obtained high accuracy range of MAD and ΔpK_a values ($0.17 \leq$ MAD ≤ 0.36 and $0.56 \leq$ MAX-ΔpK_a ≤ 1.13). The power of the predictivity slightly diminishes with the addition of diffuse functions to the basis set for any of the DFT functionals (i.e. 6-31+G* has higher MAD and MAX-ΔpK_a compared to 6-31G*). On the other hand, polarization functions did not cause any significant improvement. Regarding the performance of the functionals, in all subsets the largest MADs were obtained with either M06-2X or OLYP functionals. The smallest MADs were found for the combinations of all basis sets with the M06L functional (except 6-31G) and among all the tested methods M06L/6-311G(d,p) gave the most accurate result with MAD value of 0.174. When we applied the Equation 6.1 to the test set, the MAD value for the predicted pK_a's was found to be 0.199 and the MAX-ΔpK_a was found to be 0.87.

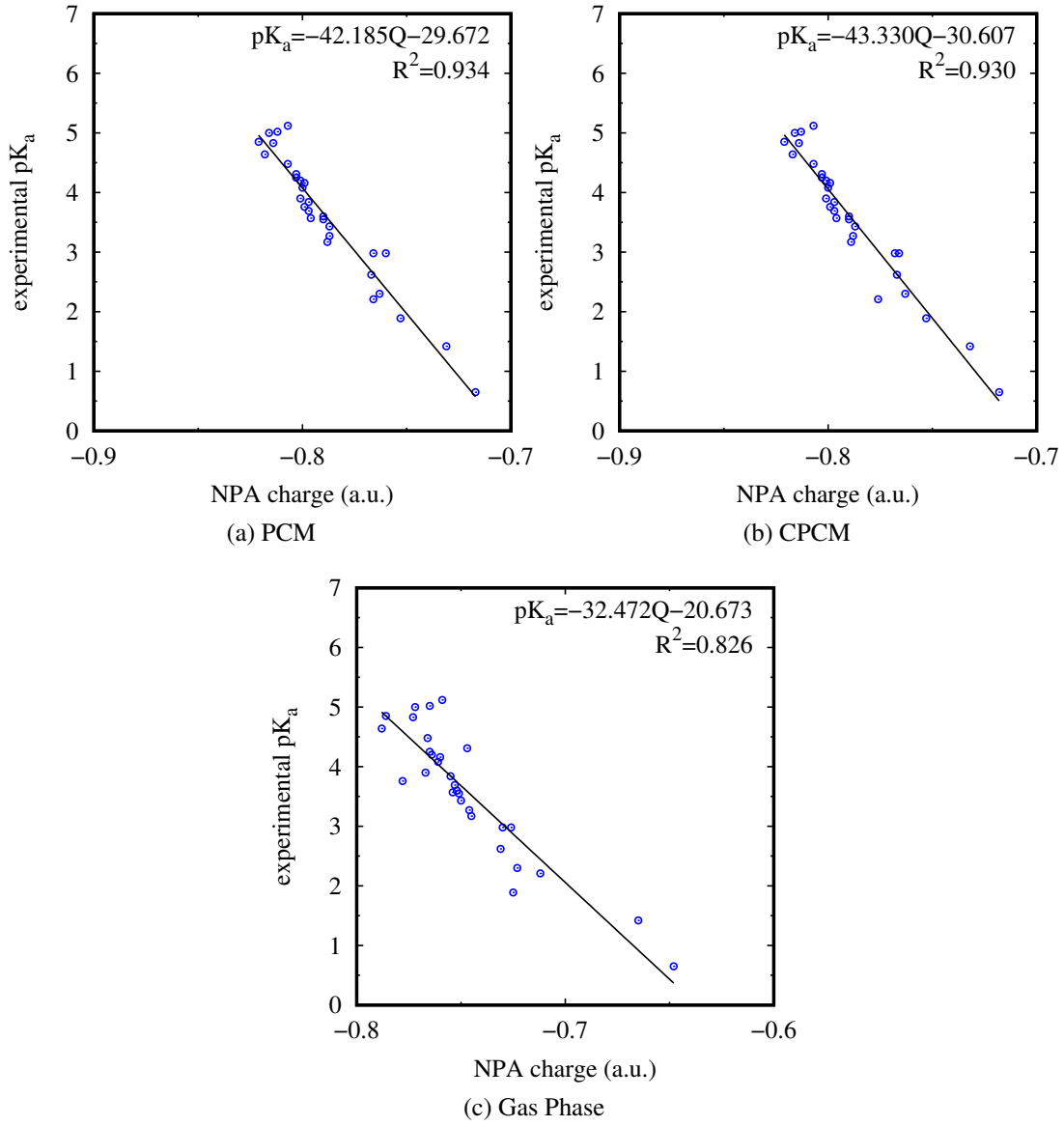


Figure 6.7: Effect of the implicit solvent model on the linear regression between calculated atomic charges and experimental pK_a's. Calculations were done with M06L/6-311G(d,p): (a) PCM model (b) CPCM model (c) gas phase.

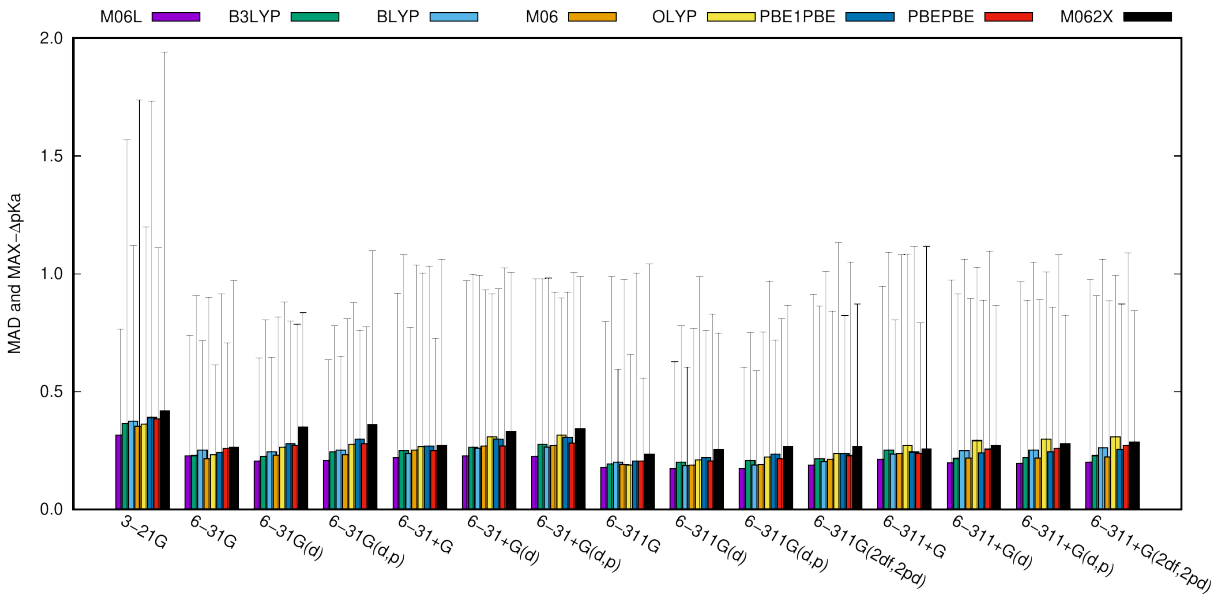


Figure 6.8: Mean Absolute Deviation (MAD) and maximum difference between predicted and experimental pK_a (MAX-ΔpK_a for eight different DFT functionals and fifteen different basis sets considered in this work. Geometry optimizations and NPA charge calculations were done using the SMD model.

The average predicted pK_a over all the methods has been calculated in order to have an overview on the efficiency of the level of theory. The minimum and maximum predicted pK_a's among all the methods (except 3-21G basis set due to its large MAD and MAX-ΔpK_a) were added to the average predicted pK_a of each molecule as error bars. The predicted pK_a is plotted versus experimental values for both training and test sets (Figure 6.9). Minimum, maximum and average values of the predicted pK_a were found to be within the range of ± 1 unit compared to the experimental value.

6.3.6 Stability of the Prediction Along Geometry Changes

The stability of the calculated pK_a's with respect to geometrical changes is crucial for the pK_a predictions of proteins. Short molecular dynamics simulations (150 ps) for N-acetyl alanine and dipeptide forms of aspartate and glutamate were performed in order to provide multiple geometries around the optimum structures and to establish the variability of the pK_a prediction with respect to geometrical changes. A total of 1500 frames were extracted from these MD simulations and single point NPA charge calculations were performed on these geometries by using SMD with M06L/6-311G(d,p) method. The predicted pK_a's were obtained using a and b values derived from the fit. The experimental pK_a's (pK_a [aspartate]=3.94 [232], pK_a [glutamate]=4.25 [233], pK_a [alanine]=3.67 [233]) were taken as a reference and the fluctuations of the calculated pK_a's with respect to geometrical changes were observed. The average value over all the frames were calculated and found to be in very good agreement with the experimental values for three of the peptides (red line in Figure 6.10). Almost 95% of the predictions are within ± 1 pK_a unit. These results point out that the suggested protocol can accurately and

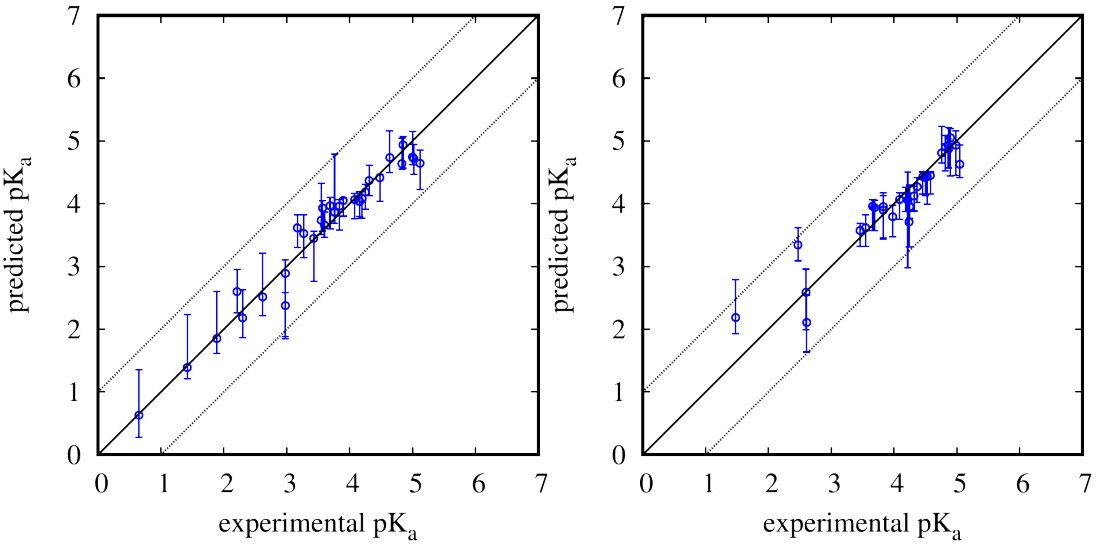
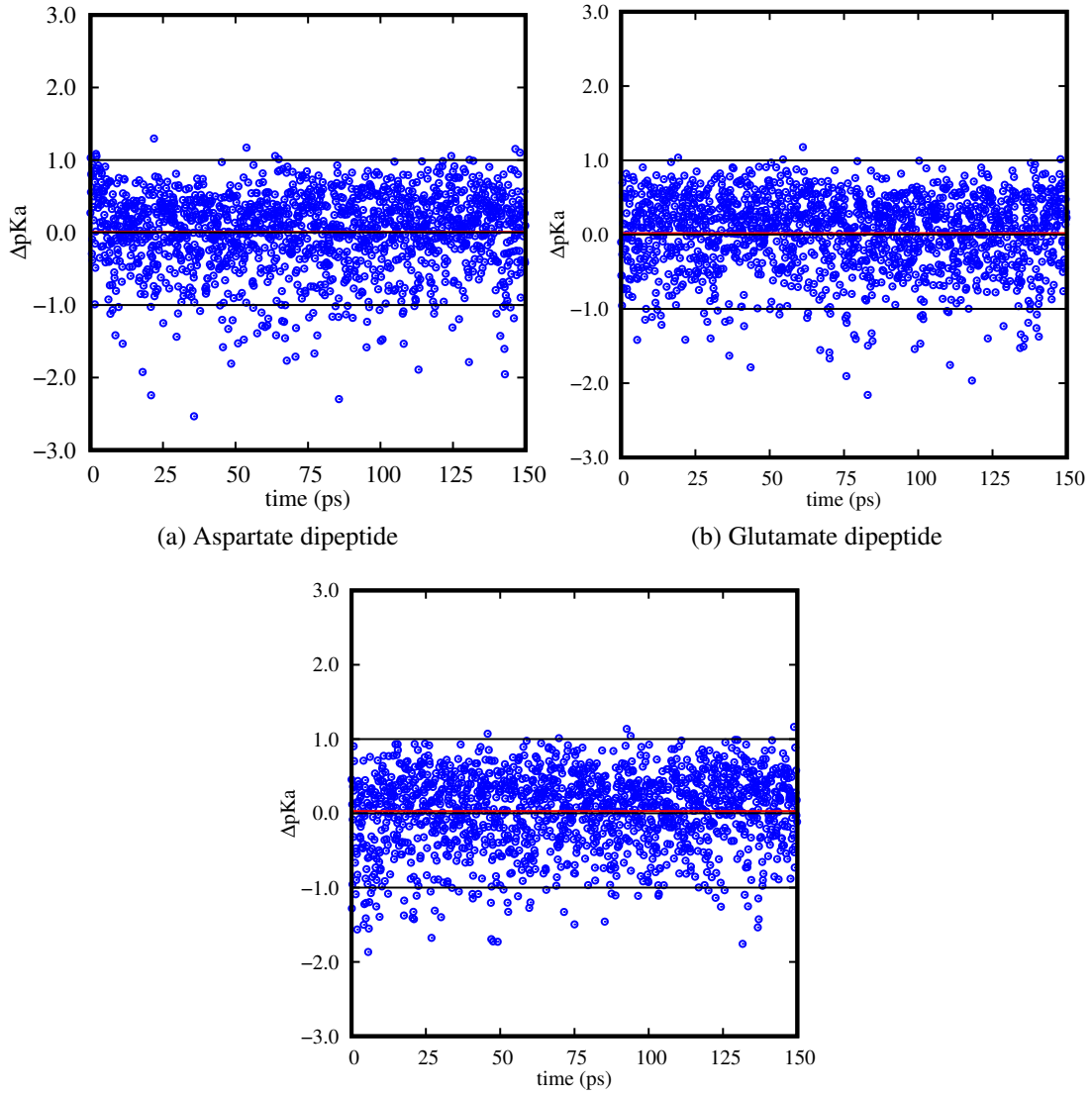


Figure 6.9: Predicted pK_a over all the DFT functionals and basis sets (3-21G discarded) versus experimental pK_a for the Training (a) and Test (b) sets (solvation model=SMD, charge model=NPA). Circles show the average pK_a, and the error bars denote minimum and maximum predicted pK_a.

efficiently predict pK_a's of aspartate, glutamate and alanine in solution.

6.3.7 Application of Protocol to the NMDAR Ligands

Glycine is the natural agonist in the NMDA receptor and its presence is a prerequisite for channel activation. Therefore the drug-design strategy for selective high affinity agonists and antagonists at the GluN1 subunit of NMDA receptor suggests the presence of an acidic moiety to mimic the receptor binding of glycine [234]. We applied our protocol to the set of partial agonists and antagonists listed in Table 4.2 and Table 4.1 which are potential NMDA GluN1 subunit selective ligands in order to predict the pK_a's of their carboxylate fragment. The pK_a's of agonists and partial agonists should be calculated by considering their zwitterionic states in physiological pH; this is confirmed by applying the protocol to the both anionic and zwitterionic states of glycine ligand which has an experimentally determined pK_a value 2.35 [235]. The predicted pK_a for anionic form is 4.73 whereas the predicted pK_a for zwitterionic state is 2.68. Taking this information into account, calculations have been performed on anionic forms of antagonists and zwitterionic state of partial agonists. NPA charges on the oxygen atoms of carboxylate fragments have been calculated at M06L/6-311G(d,p) level of theory with SMD model and the predicted pK_a's are computed using Equation 6.1. The results are given in Table 6.3. Our results indicate that all partial agonists and antagonists bear pK_a's lower than 4.00. This means that all considered ligands for NMDAR GluN1 subunit are in a carboxylate form at physiological pH in aqueous solution.



(a) Aspartate dipeptide

(b) Glutamate dipeptide

(c) N-Acetyl-Alanine

Figure 6.10: Deviations of predicted pK_a with respect to geometrical changes. Geometries were obtained from aqueous phase MD calculations. M06L/6-311G(d,p) method was used for single point NPA calculations using SMD. The red line shows the numerical average of the pK_a deviations.

Table 6.3. Predicted pK_a's of selected set of potential partial agonists and antagonists that are represented in Table 4.2 and Table 4.1

Ligand ID (Partial Agonists)	pK_a (pred.)	Ligand ID (Antagonists)	pK_a (pred.)
8	2.87	189	2.56
9	2.86	190	3.72
10	2.88	191	3.75
11	2.90	192	3.74
13	2.92	194	3.70
14	2.90	195	3.72
15	2.97	212	4.01
17	2.89	22	2.57
18	2.84	24	2.63
19	2.85	26	2.58
20	2.92	28	2.46
21	2.95	29	2.61
22	2.69	31	2.46
23	2.92	32	2.53
24	2.91	33	2.55
25	2.94	35	2.52
26	2.92	36	2.66
27	2.91	37	2.37
29	2.87	38	2.58
32	2.45	39	2.63
33	2.90	41	2.56
34	2.90	44	2.61
35	2.91		
36	2.89		
37	2.87		
38	2.90		
39	2.93		
46	2.90		
47	2.94		
51	2.87		

The protonation states of the receptor, ligand and complex should be taken into account in order to perform successful docking, screening and free energy calculations. The role of

protonation states in the initial steering is understood by the accurate pK_a predictions of the ionizable groups of the unbound receptor and the ligand. In the next step of this study, this protocol will be transferred to the prediction of pK_a's of aspartate, glutamate and alanine within proteins.

6.4 Conclusions

In this study, a protocol has been suggested in order to obtain fast and accurate pK_a predictions for small carboxylic acids. According to the suggested protocol, pK_a's are computed by using the equation derived from the linear regression of the experimental pK_a's with the atomic charges on the carboxylate fragment. Five charge descriptors, three charge models, three solvent models, gas phase calculations and several DFT methods (combination of eight DFT functionals and fifteen basis sets) were tested. Among those, NPA charge calculations performed with the SMD solvation model on optimized geometries gave the most accurate results. The best combination of DFT functionals and basis sets were found to be M06L/6-311G(d,p) ($R^2 = 0.955$). The strongest linearity is found by selecting the maximum atomic charge on carboxylic oxygen atoms and relating it to the experimental pK_a. Molecular dynamics simulations have been performed for a set of aspartate, glutamate and alanine peptides in order to test the stability of the prediction. The protocol was applied to a randomly selected set of frames which were extracted from MD simulations and the calculations showed that the predicted pK_a's were scattered within ± 1 unit from the experimental value. The suggested protocol will be transferred to the pK_a prediction of aspartate, glutamate and alanine within protein environment. By reporting the calculated atomic charge of the carboxylate form in to the linear relationship derived in this work, it should be possible to estimate the pK_a's of aspartate, glutamate and alanine.

Chapter 7

Elucidation of the atroposelectivity in the synthesis of axially chiral thiohydantoin derivatives

7.1 Introduction

Hydantoins, a class of cyclic imides, have a broad range of biological and agricultural activities; they are also important precursors in the synthesis of several amino acids and pyruvic acid derivatives [114]. Sulfur analogs of hydantoins, thiohydantoins, are the privileged scaffolds in modern synthetic organic chemistry due to their anticarcinogenic [115], antiviral [116], antimicrobial [117], antithyroidal [118], hypolipidemic [119, 120], antimutagenic [121, 122] applications besides their usage as pesticides [123], reagents for dyes [124], anti-ulcer and anti-inflammatory agents [125], C-terminal protein sequencing [126] and metal-cation complexation agents [127]. The importance of thiohydantoins comes from the presence of sulfur and oxygen atoms which act as H-bond proton acceptors and hydrogen atom bonded to nitrogen which acts as a H-bond donor [128].

There are a number of methods for the synthesis of thiohydantoins in the literature [129]. Sarigul and Dogan synthesized a number of diastereomerically enriched axially chiral atropoisomeric 2-thiohydantoins via the chiral pool strategy by the reaction of L-amino acid ester salts and *o*-aryl isothiocyanates in the presence of triethyl amine under reflux in dichloromethane [22] (Figure 7.1). The synthesized compounds possess a chiral center at C_5 position and a chiral axis along N_{sp^2}-C_{aryl} with an ortho substituent attached to the phenyl ring (Figure 7.2). The nonaxially chiral 2-thiohydantoins are racemic products irrespective of the R_1 and R_2 substituents. It is a well known phenomenon that as the size of the ortho substituent gets larger, the rotation around the N_{sp^2}-C_{aryl} bond is more restricted, and this hindered rotation gives rise to atropisomeric compounds [131, 132, 133, 134, 135, 136, 137, 138, 139, 140]. The experimental data of Sarigul and Dogan [22] shows that C_5 racemization in 5-methyl-3-*o*-bromophenyl-2-thiohydantoin is avoided by the bulky *o*-aryl substituent and eventually chirality of L-alanine methyl ester salt is preserved in the final product (SP configuration is dominant in which bromine atom is transoid with respect to C_5 methyl).

Using computational tools to model the reaction mechanisms is important for understanding

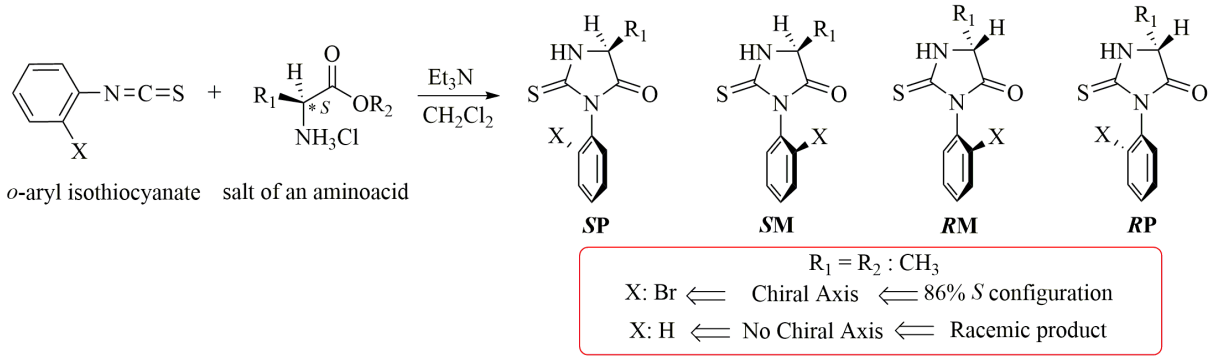


Figure 7.1: Synthesis of atroposelective compounds by Sarigul and Dogan. [22]

the details of the racemization steps in order to synthesize enantiomerically pure products. The aim of this study is to rationalize the atroposelectivity observed in the synthesis of the axially chiral 2-thiohydantoin derivatives and compare the results with the current and the previous experimental data [22]. Herein, several questions are addressed: Does racemization occur before or after cyclization? Does racemization occur via the assistance of triethyl amine which is already present in the reaction medium or methanol which is a side product or through unassisted keto-enol tautomerization? Plausible mechanisms have been proposed for the racemization and cyclization of the intermediate adduct (3*S*) formed from the initial attack of nitrogen of the amino acid to the isothiocyanate thiocarbonyl carbon (Figure 7.2).

This study is composed of two parts: In Part I, the synthesis mechanisms have been investigated by DFT methods in order to answer the addressed questions by modeling the various possibilities for the formation of chiral 2-thiohydantoins. In Part II, the stereo conversion mechanism of atroposelective 2-thiohydantoin derivatives have been addressed with computational tools.

7.2 Methodology

Geometry optimizations and frequency calculations were performed at the M06-2X/6-311+ G(d,p) level of theory by using the Gaussian09 program package. The 6-311++G(3df,3pd) basis set was used for Br. The M06-2X functional [106, 107] is known to be efficient in reflecting the properties of organic reaction mechanisms by including the contribution of dispersive effects [236, 237, 238, 239, 240, 241, 242]. The nature of the corresponding reactants and products was verified by performing intrinsic reaction coordinate (IRC) calculations on the transition state structures [243]. All free energies were calculated at room temperature. Implicit solvation by dichloromethane or ethanol using the integral equation formalism polarizable continuum model (IEFPCM) has been taken into account in the calculations [196, 244]. The charges reported were calculated with Mulliken Population Analysis scheme unless otherwise stated.

7.3 Results And Discussion

7.3.1 Computational Approach to the Elucidation of the Atroposelectivity in the Synthesis of 2-Thiohydantoin Derivatives

Recently, Sarigul and Dogan [22] have reported the synthesis of several 2-thiohydantoin derivatives where it was stated that although the synthesis started with an enantiomerically pure amino acid (single stereoisomer having an *S* configuration) the product thiohydantoin was obtained as a racemic mixture if the thiohydantoin was not axially chiral. If, on the other hand, the product thiohydantoin was axially chiral, the racemization took place at various extents. In this part of the study, DFT calculations were performed for modeling possible mechanisms in order to investigate if the observed racemization occurs before or after the cyclization and the factors that drive the racemization, together with the influence of axial chirality on the stereo outcome of the reaction. For this purpose, we have selected the racemic 5-methyl-3-phenyl-2-thiohydantoin (X:H, R_1:CH_3 in Figure 7.1; hereafter denoted as H-8*S* and H-8*R*) and the nonracemic axially chiral 5-methyl-3-*o*-bromophenyl-2-thiohydantoin (X:Br, R_1:CH_3 in Figure 7.1; hereafter denoted as Br-8*S* and Br-8*R*) as model compounds since they represent very specific distributions of their isomers: H-8*S*:H-8*R*=50:50 and Br-8*S*M:Br-8*R*P:Br-8*S*P:Br-8*R*M=3:14:83:0.

The very first step of the proposed mechanism starts with the nucleophilic addition of the amino acid (2*S*) to the *o*-aryl isothiocyanate (1) yielding intermediate 3*S* (Figure 7.2). We propose a route following 3*S* → 6*S* transformation and then the formation of the cyclized product 8*S*.

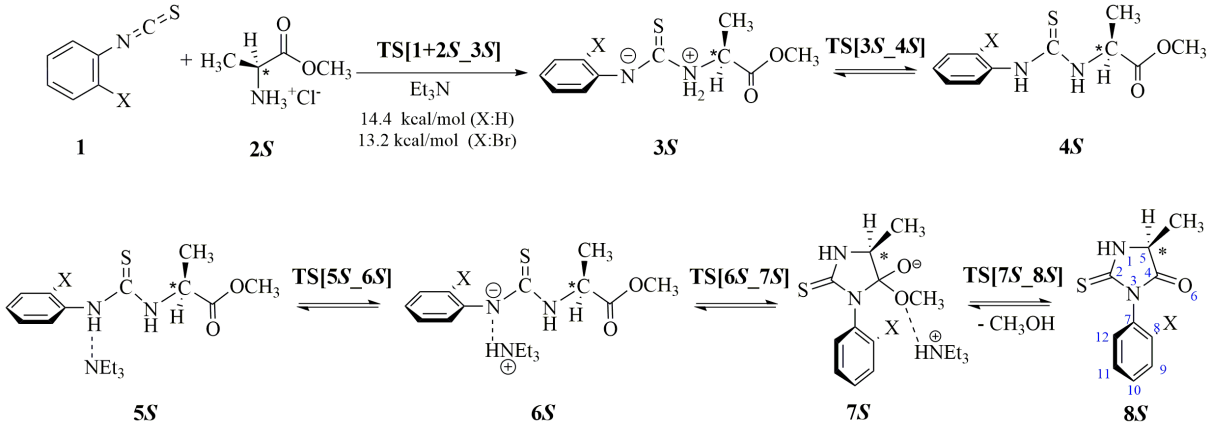


Figure 7.2: Suggested mechanism for the synthesis of the cyclized product 8*S*.

Cyclization The study started with locating the lowest energy conformer of H-3*S* and Br-3*S* (Figure 7.3). Steric effects and electrostatic interactions between the phenyl ring and lone pairs on N_3 atom disfavor the planar conformation and H-3*S* has a torsion angle (C_2-N_3-C_7-C_8) of 56.2°. The corresponding torsion angles are 102.9°in Br-3*S* and -87.4°in Br-3*S*-2. In Br-3*S*, Br has favorable attractive interactions with hydrogen atom bonded to C_5 (3.44 Å) and the

closest hydrogen of the methyl group (3.19 Å) at the stereocenter C_5; these interactions are absent in Br-3*S*-2 which is 1.2 kcal/mol higher in energy. Sarigul and Dogan have reported the relative proportions of all synthesized isomers of Br-8 (*S*M:*R*P:*S*P:*R*M as 3:14:83:0) in which 97% of the products were in P configuration [22]. The formation of Br-8 starting from Br-3*S* gives a transoid conformation of Br with respect to the C_5 methyl (Br-8*S*P), whereas cyclization from Br-3*S*-2 would give a cisoid conformation of Br with respect to the C_5 methyl (Br-8*S*M). Hence, transfer of central chirality at stereogenic carbon to the axial chirality in the final product as suggested by Kawabata *et al.* [245], took place where the central chirality at C_5 of Br-3*S* (the lowest energy conformer) is to be expected to orient the chiral axis in the final product Br-8*S*P.

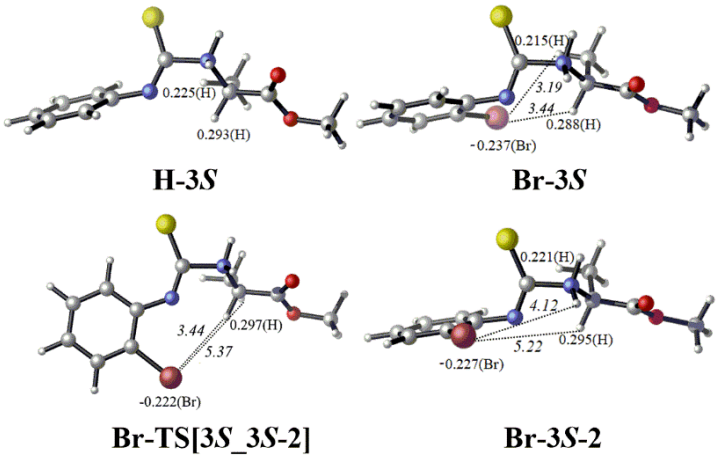


Figure 7.3: Charges (Mulliken Population Analysis) and distances (italics) for H-3*S*, Br-3*S*, Br-3*S*-2 and the rotational transition state Br-TS[3*S*_3*S*-2]

Due to the presence of charge separation in 3*S*, proton transfer from the positively charged thioamide nitrogen (N_1) to the negatively charged arylamine nitrogen (N_3) occurs as expected yielding structure 4*S* (Figure 7.2). Our calculations show that this step is a very exothermic reaction for both X:H and Br (-20.5 kcal/mol for H-4*S* and -19.2 kcal/mol for Br-4*S*) with respect to H-3*S* and Br-3*S* respectively (Figure 7.4). Steric effects and electrostatic interactions between the phenyl ring and hydrogen on N_3 disfavor the planar conformation with a torsion angle (C_2-N_3-C_7-C_8) of 55.6°in H-4*S*; this angle is 86.1°in Br-4*S* and -82.0°in Br-4*S*-2 (Figure 7.5). The relative stability (0.6 kcal/mol) of Br-4*S* as compared to Br-4*S*-2 can be ascribed to the steric encumbrance in the latter. So, the synperiplanar conformation of bromine atom with respect to methyl group attached to stereocenter C_5 in Br-3*S* is retained in Br-4*S*.

Upon the formation of 4*S*, there are three possible sites of deprotonation by TEA: N_1, N_3 or C_5 (Figure 7.6). Locating the transition states for hydrogen abstraction from N_1, N_3 and C_5 then optimizing the final formed anions showed that H-6*S* (13 kcal/mol higher in energy with respect to H-4*S*) is the most stable anion when compared to H-18*S* (16.2 kcal/mol higher in energy with respect to H-4*S*) and H-15 (27.1 kcal/mol higher in energy with respect to H-4*S*); this is due to the resonance stabilization of the resulting anion (H-6*S*) through N_3-aryl ring (the deprotonation of 4*S* at C_5 will be discussed in the part entitled as 'Reactions of 4*S*'). The activation barriers

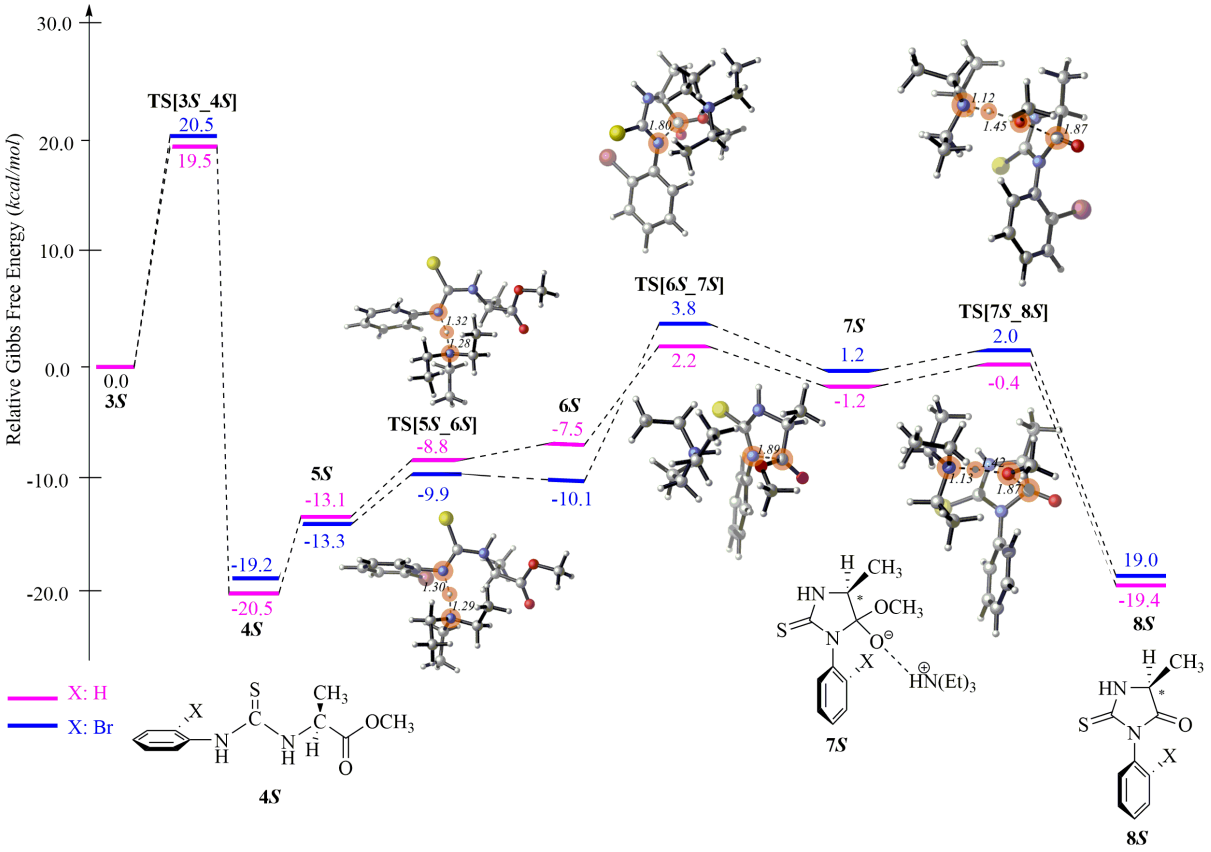


Figure 7.4: Free energy profile for the suggested mechanism.

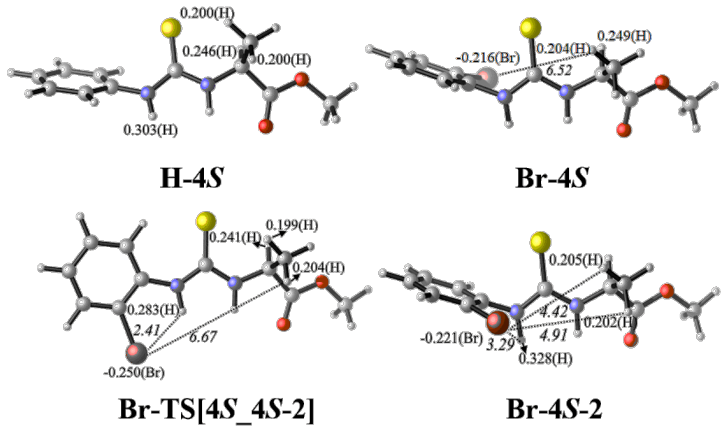


Figure 7.5: Charges (Mulliken Population Analysis) and distances (italics) for H-4*S*, Br-4*S*, Br-4*S*-2 and the rotational transition state Br-TS[4*S*_4*S*-2].

for each proton abstraction step from complex structures of 4*S* (5*S*, 14*S* and 17*S*) are displayed in Figure 7.6.

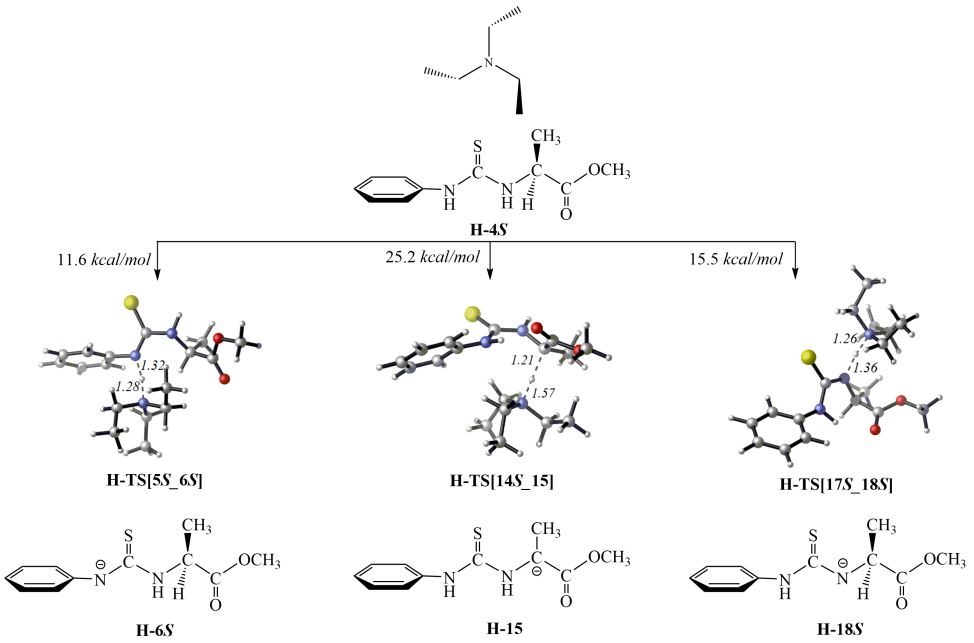


Figure 7.6: Deprotonation pathways of 4*S*

The formation of H-6*S* and Br-6*S* from 4*S* via complexation with TEA (5*S*) and hydrogen abstraction from N_3 has the corresponding activation barriers; 11.6 kcal/mol and 9.3 kcal/mol respectively (Figure 7.4). Formation of H-6*S* and Br-6*S* are slightly exergonic with respect to the starting materials H-3*S* and Br-3*S* respectively.

In structure H-6*S* the planar conformation is not favored (C_2-N_3-C_7-C_8=63.9°) due to the steric effects and electrostatic interactions between the phenyl ring and lone pairs on N_3 as in structure H-3*S* (Figure 7.7). The same torsion angles are 118.6°in uncomplexed Br-6*S* and -88.1°in Br-6*S*-2. In Br-6*S*, Br has favorable attractive interactions with the closest hydrogen of the methyl group (3.05 Å) at the stereocenter C_5; these interactions are absent in Br-6*S*-2 which is 1.7 kcal/mol higher in energy. As a result, the synperiplanar conformation of bromine atom with respect to methyl group attached to C_5 in Br-3*S* which is retained in Br-4*S*, is still in the synperiplanar position in Br-6*S*. Therefore, as Br-6*S* undergoes cyclization, brominated aryl ring will tend to keep its conformation, thus yielding Br-8*S*P in which Br is transoid with respect to the C_5 methyl.

Cyclization reactions may occur spontaneously, by heating or may require a catalyst like a Bronsted or Lewis acid or base [141]. In the case of the cyclization of 6*S*, the protonated TEA in 6*S* sticks on the methoxy oxygen, N_3-C_4 bond formation occurs in the first step (TS[6*S*_7*S*], Figure 7.2) and this ring closure barrier is low (9.7 kcal/mol for H-TS[6*S*_7*S*] and 13.9 kcal/mol for Br-TS[6*S*_7*S*]) (Figure 7.4). In the second step, methanol release by the hydrogen transfer from protonated TEA occurs very fast with less than 1.0 kcal/mol for H-TS[7*S*_8*S*] and Br-TS[7*S*_8*S*P]. The experimental findings by Sarigul and Dogan are in harmony with the compu-

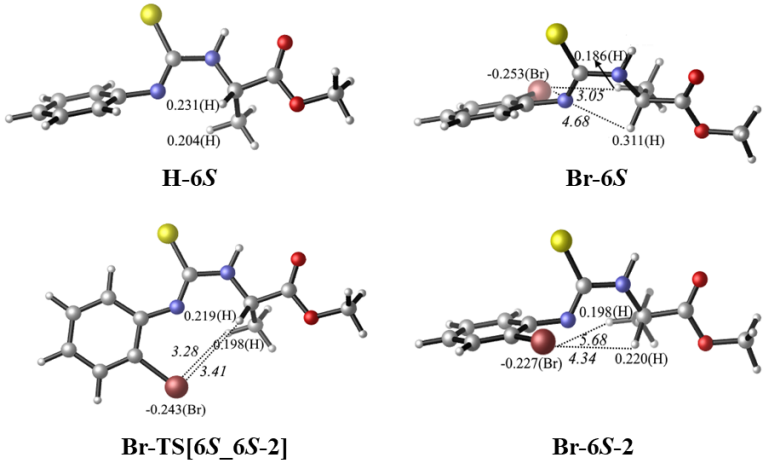


Figure 7.7: Charges (Mulliken Population Analysis) and distances (italics) for H-6*S*, Br-6*S*, Br-6*S*-2 and the rotational transition state Br-TS[6*S*_6*S*-2].

tational cyclization data, since they reported a very fast cyclization step.

Racemization After Cyclization Sarigul and Dogan [22] have reported variable extents of racemization at C_5 of the heterocyclic ring in their set of synthesized axially chiral thiohydantoin derivatives (Figure 7.1). In order to clarify whether the enolization is a viable process in racemization of thiohydantoins or not, the intramolecular proton transfer from C_5 to O_6 is modeled with and without a base catalyst. The upper path (red line in Figure 7.8) shows that keto-enol tautomerization without a catalyst is not kinetically favorable with a very high activation barrier (77.0 kcal/mol). As one mole of methanol is released during the cyclization step (TS[7*S*_8*S*], Figure 7.2), it can abstract the hydrogen atom at stereocenter (C_5) by transferring its OH hydrogen to O_6 yielding the enolate H-8*S*-Enol. Methanol acts like a base catalyst which explicitly assists the deprotonation and reprotonation of C_5 lowering the enolization barrier to 37.4 kcal/mol (green line in Figure 7.8) which is still too high for a reaction to occur under the reaction conditions.

Since keto-enol tautomerization in the thiohydatoin ring with or without the assistance of methanol has been found to be unplausable due to the high activation barriers, triethyl amine mediated deprotonation-reprotonation at C_5 is suggested (Figure 7.9). The activation barrier for the deprotonation at C_5 by TEA is 14.3 kcal/mol for H-TS[9*S*_10] and 13.6 kcal/mol for Br-TS[9*S*P_10]; these are relatively low barriers that can be reached under the reaction conditions. H-10 is only 0.9 kcal/mol lower in energy with respect to the H-TS[9*S*_10] and H-TS[10_9*R*]; both the forward and reverse reactions are equally favored. Even though both interconversions from Br-9*S*P to Br-9*R*P or Br-9*R*P to Br-9*S*P are attainable under reaction conditions, the backward path is slightly (1.2 kcal/mol) more favored than the forward path yielding an *S*/*R* ratio of

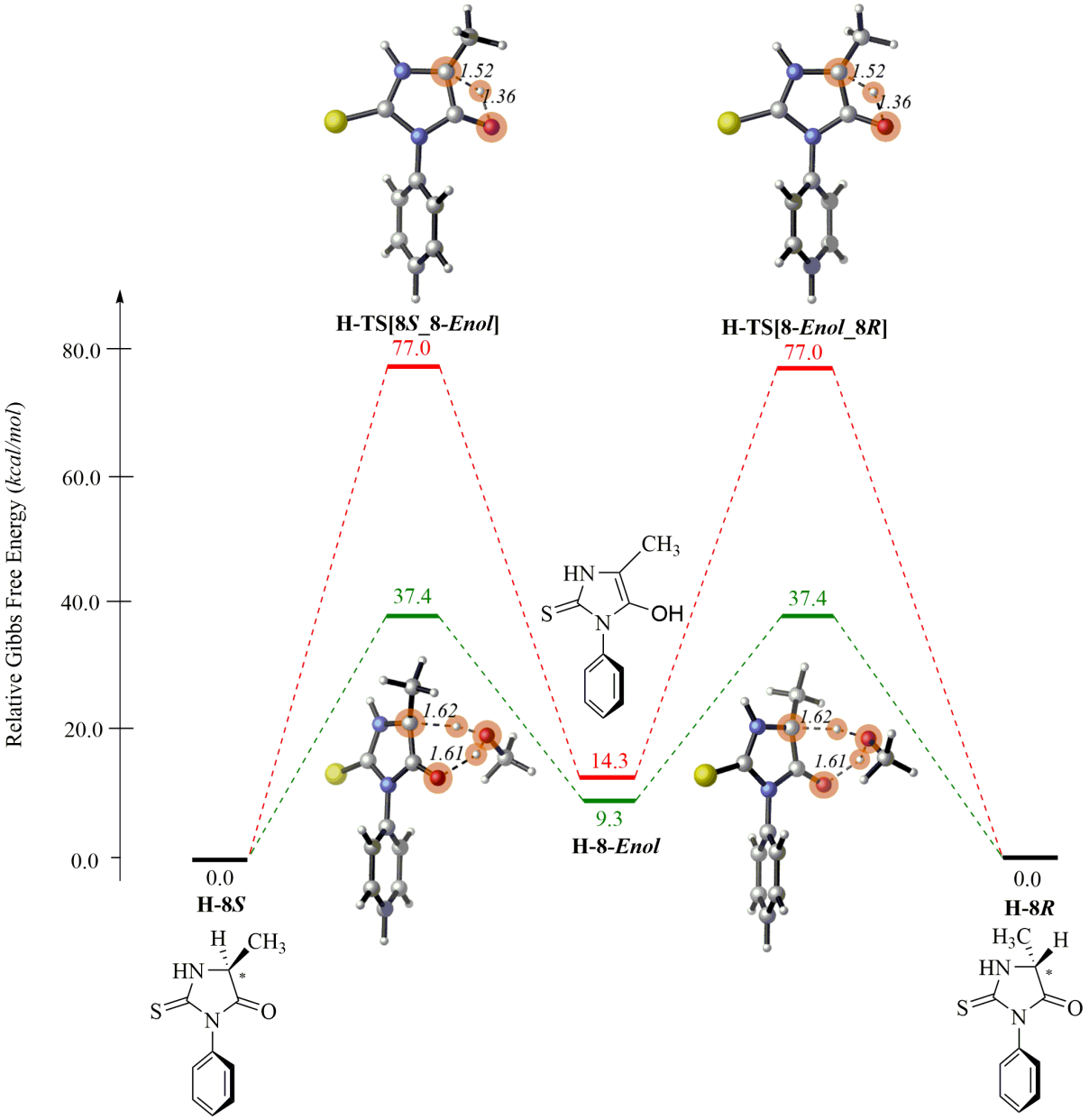


Figure 7.8: Free energy profile for the racemization of H-8*S* through enolate formation with (green lines) and without (red lines) a base catalyst (methanol).

88/12 according to the Boltzmann distribution function, which is in very good agreement with the reported experimental ratio (*S*/*R*:86/14).

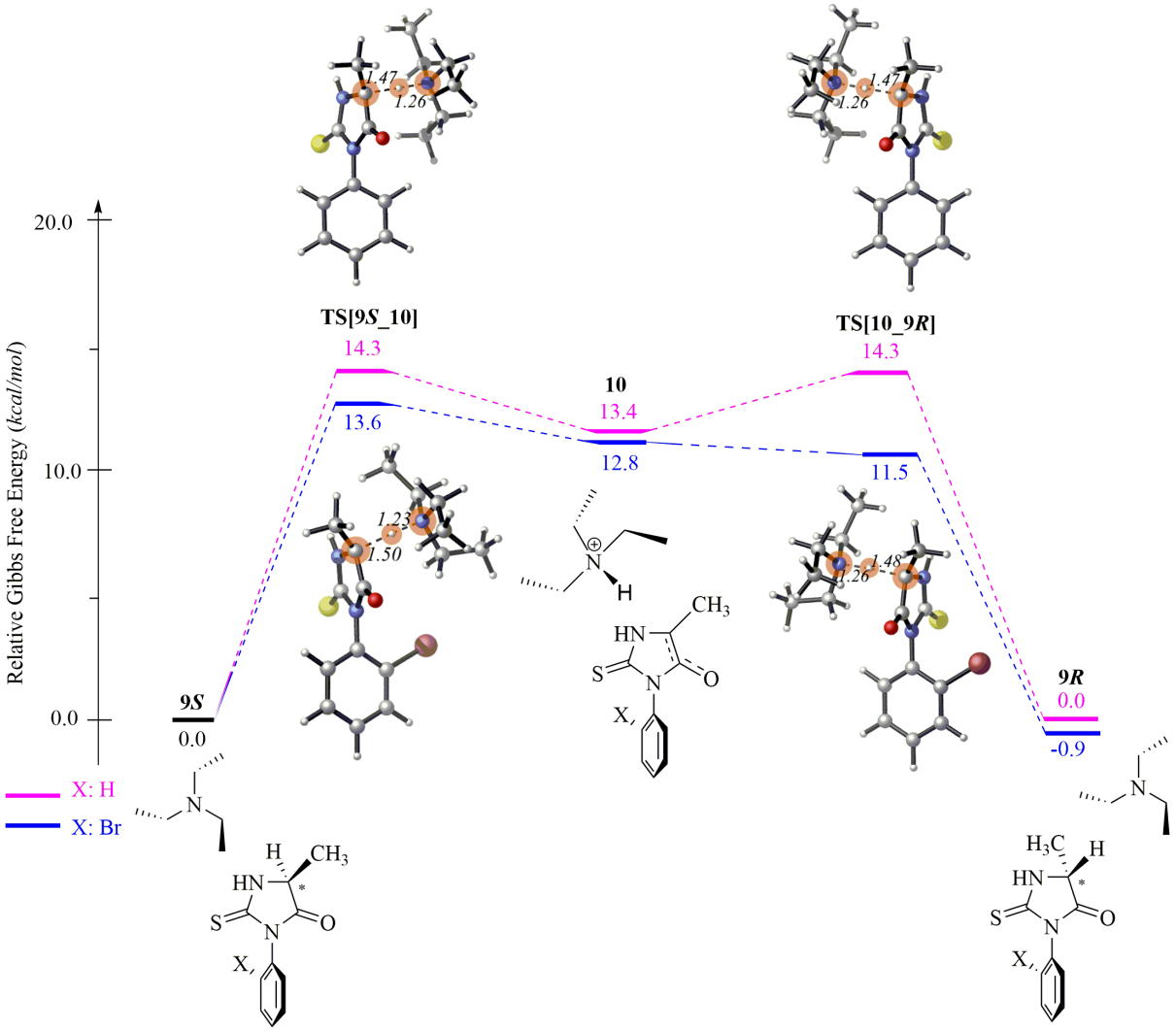


Figure 7.9: Triethylamine assisted racemization mechanism for the thiohydantoin derivatives.

Racemization Before Cyclization In this part of the study, we have evaluated several possible racemization mechanisms of 3*S*, 4*S* and 6*S* with and without TEA in order to understand whether racemization occurs before the cyclization as well. We have also discussed the formation of 8*S* and 8*R* by cyclization of 3*S*, 3*R*, 4*S* and 4*R*.

Note that Zhang *et al.* have synthesized hydantoins from enantiomerically pure amino amides and observed that triethylamine was not the only species responsible for the racemization [246]. They have used 1,1'-carbonyldiimidazole (CDI) instead of TEA and the reaction progressed with still some enantiomerization.

(i) Reactions of 3S

The intermediate 3*S* either cyclizes via TS[3*S*_8*S*] to give 8*S* or undergoes racemization via TS[3*S*_13] without any assistance or TS[11*S*_12] with the assistance of TEA, and then cyclizes to 8*R* (Figure 7.10).

The activation barriers for proton migration from the stereocenter (C_5) to the negatively charged arylamine nitrogen (N_3) via TS[3*S*_13] are 23.4 and 23.2 kcal/mol for compounds H and Br respectively giving the slightly stable intermediate 13 (Figure 7.11). Following the free rotation about the N_1-C_5 bond in intermediate 13, the arylamine nitrogen N_3 transfers the proton back to the sp^2 carbon (C_5) (TS[13_3*R*]) to give the enantiomer 3*R*.

The racemization of 3*S* may be achieved with the help of TEA through TS[11*S*_12] (Figure 7.10). 3*S* and TEA form the complex 11*S*, and this complex is 8.3 kcal/mol and 8.0 kcal/mol higher in energy as compared to the uncomplexed ligands for H-11*S* and Br-11*S* respectively due to the steric effects caused by TEA as it approaches to form the respective complexes (Figure 7.12). The complexation to H-11*S* and hydrogen abstraction in H-TS[11*S*_12] has a barrier of 23.0 kcal/mol, the formation of complex Br-11*S* and Br-TS[11*S*_12] requires 23.4 kcal/mol. The electron delocalization is weaker in intermediate 12, and this leads to a relatively unstable structure which lies slightly higher in energy with respect to the transition state structures. After the proton transfer from the protonated TEA to the intermediate 12 via TS[12_11*R*], 11*R* is formed as a complex of TEA. The isolated uncomplexed racemized product 3*R* further cyclizes to give 8*R*.

The two potential energy surfaces displayed in Figure 7.11 and Figure 7.12 indicate the formation of racemic mixtures for both H-3 and Br-3, since the barriers for the reverse and forward reactions are equal. In Br-3*R*, bromine is in antiperiplanar conformation with respect to methyl group attached to C_5 and Br-3*R* is 2.9 kcal/mol more stable than Br-3*R*-2 (Br atom in synperiplanar conformation) (Figure 7.13), due to the similar favorable interactions as observed in Br-3*S*. Thus, upon the cyclization of Br-3*R*, Br-8*R*M atropoisomer is expected to be formed.

3*S* and 3*R* can undergo cyclization via a concerted reaction mechanism: Following some conformational reorganizations, the nucleophilic attack of electron rich nitrogen atom (N_3) to the carbonyl carbon (C_4) (TS[3_8]) initiates the ring closure simultaneously promoting the hydrogen movement from the amidic nitrogen (N_1) to the methoxy oxygen, yielding the thiohydantoin ring (8) and methyl alcohol (Figure 7.10). This ring-closure transition state embodies a 5-endo-trig nucleophilic cyclization which has been considered as a geometrically disfavored process according to Baldwin's rules [247, 248, 249, 250]. There are a number of 5-endo-trig cyclizations which are reported in the recent literature [249]. In the case of cyclization of 3, there are no steric or electrostatic repulsions between N_3 and C_4. The carboxyl oxygen readily abstracts the hydrogen from N_1 and releases methanol as the ring closes. This concerted cyclization reaction requires 29.7 kcal/mol for the formation of the racemic H-8, 30.5 kcal/mol for Br-8*S*P and 30.2 kcal/mol for Br-8*R*M (Figure 7.14). Despite the relatively high activation barrier (compared to racemization of 3*S* or conversion of 3*S* to 4*S*) needed for cyclization of 3, there is a tremendous gain in thermodynamic stability in reaching the cyclized products; and this suggests a thermodynamically driven process. The exergonic character of this step can be attributed to the delocalization gained in the cyclic products.

The enol tautomer of 3*S* can be generated via the transfer of the C_5 proton to O_6. In dichloromethane, the relative energy of the enol tautomer (H-3-Enol) is found to be 22.0 kcal/mol higher with respect to the keto isomer (H-3*S*) (Figure 7.15), and its formation requires 76.2 kcal/mol indicating the presence of the keto isomer in the reaction medium only.

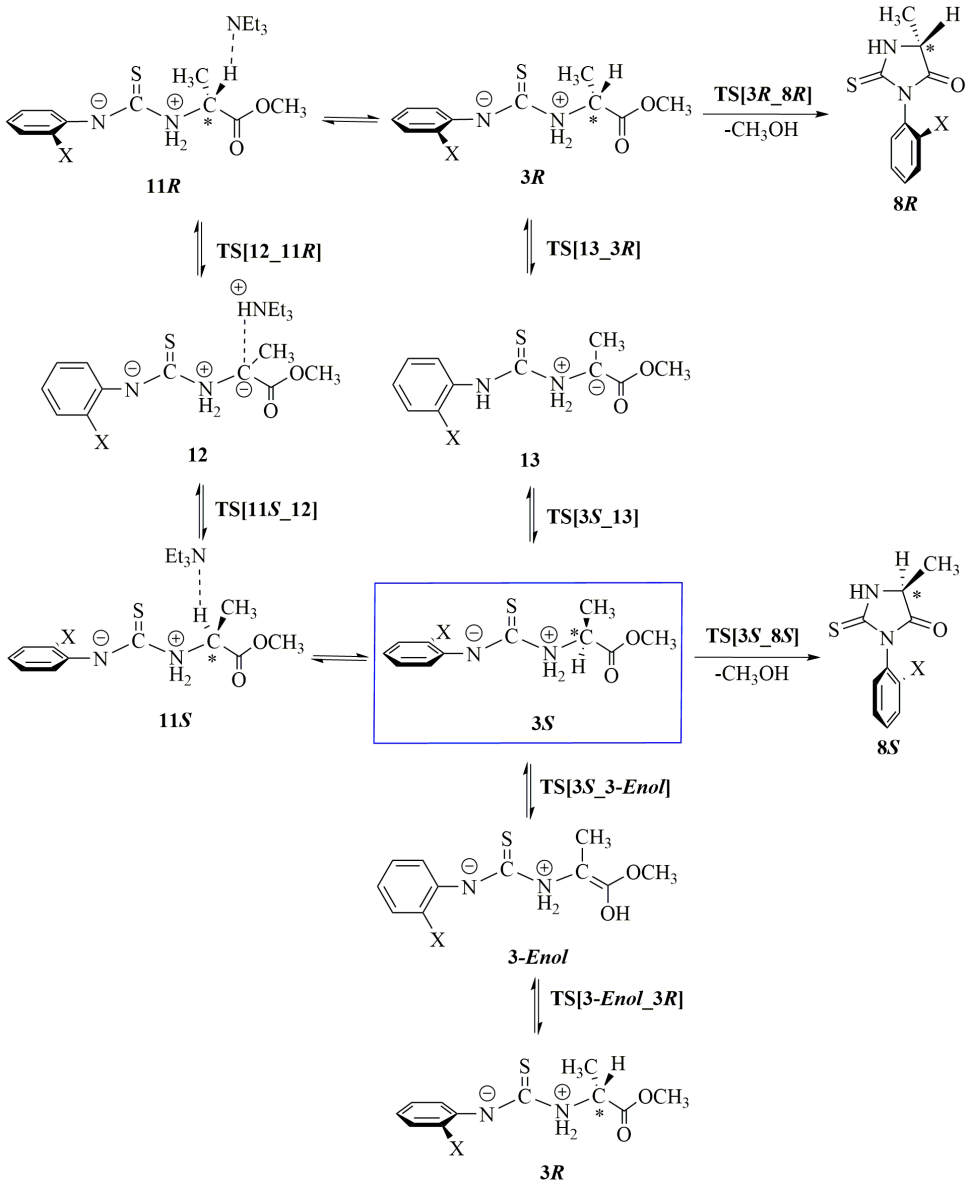


Figure 7.10: Alternative cyclization and racemization routes of 3*S* with and without the assistance of TEA.

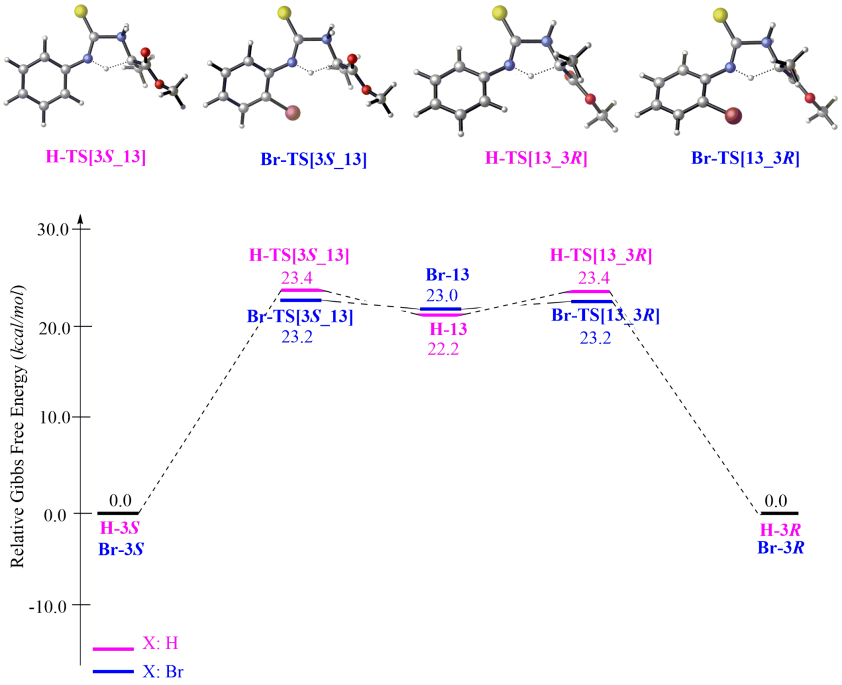


Figure 7.11: Free energy profile for 3*S* to 3*R* transformation.

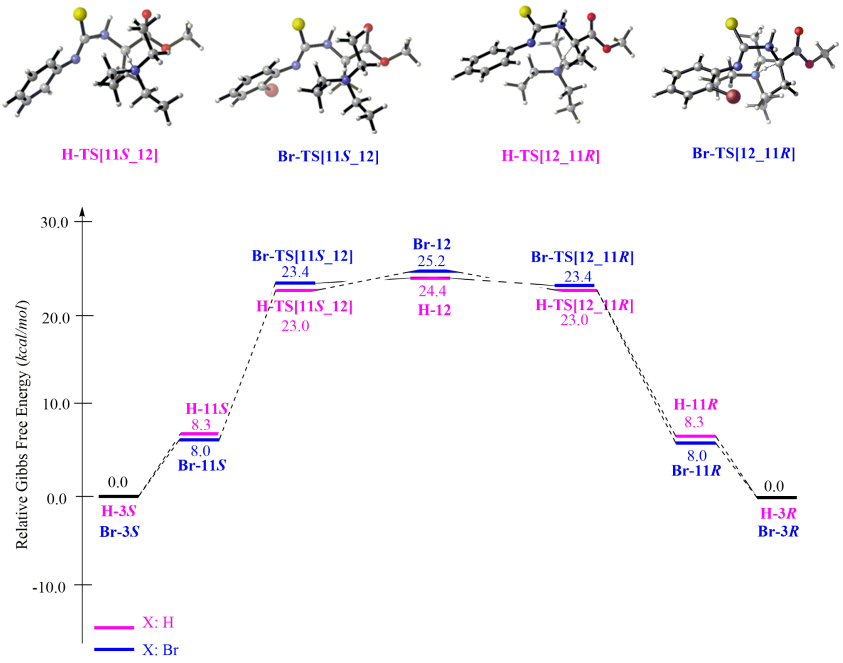


Figure 7.12: Free energy profile for 3*S* to 3*R* transformation with the help of TEA.

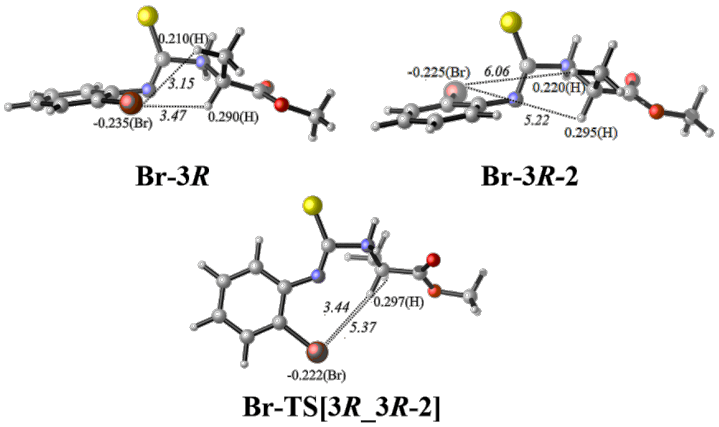


Figure 7.13: Charges (Mulliken Population Analysis) and distances (italics) for Br-3*R*, Br-3*R*-2 and the rotational transition state Br-TS[3*R*_3*R*-2].

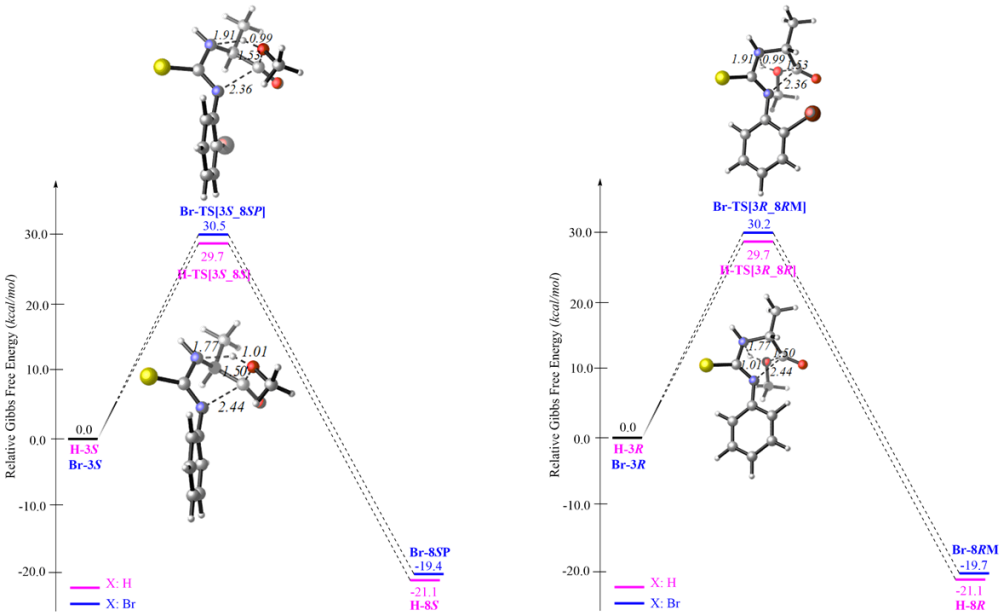


Figure 7.14: Free energy profile for cyclization from 3*S* (left) and 3*R* (right).

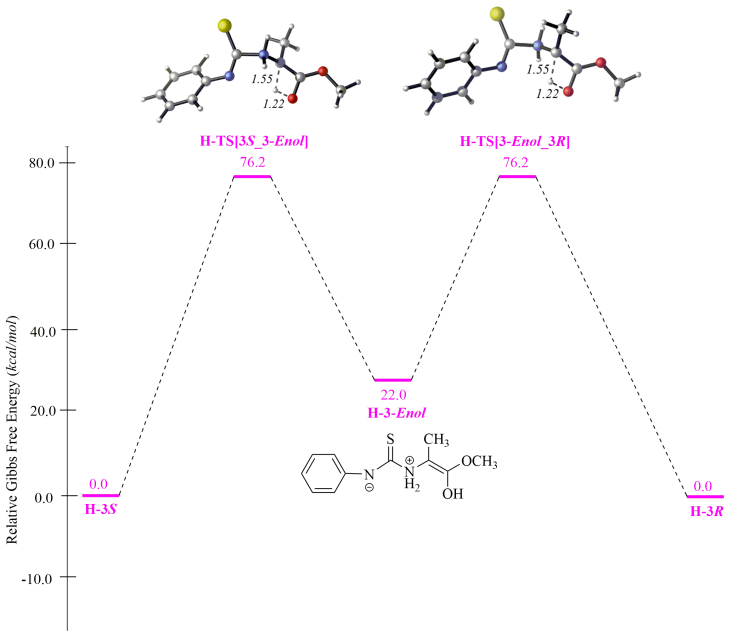


Figure 7.15: Free energy profile for the racemization of H-3*S* via keto-enol tautomerization.

(ii) Reactions of 4S

The intermediate 4*S* may either cyclize through TS[4*S*_8*S*] to give 8*S* or is assisted by TEA, it may enantiomerize via TS[14*S*_15] and then cyclize to 8*R* (Figure 7.16). The possibility of keto-enol tautomerizations of 4*S* is also evaluated for the observed racemizations.

TEA can form the complex 14*S* by interacting with the hydrogen atom of the stereocenter. The complexation to H-14*S* and hydrogen abstraction in H-TS[14*S*_15] overcomes a barrier of 25.3 kcal/mol, while the same barrier is 26.3 kcal/mol for Br-14*S* and Br-TS[14*S*_15] (Figure 7.17). In TS[15_14*R*] the protonated TEA transfers the proton to the sp^2 carbon (C_5) from the *Re* face of 15 to give the 14*R*. The potential energy surface (PES) in Figure 7.17 indicates the formation of racemic mixtures for both H-4 and Br-4 compounds, since reaching the summit of the potential energy surface requires equal energies in both forward and reverse reactions which is in agreement with experimental data for the formation of H-8 but does not satisfy the expectation for Br-8 for which the *S* isomer is dominant experimentally. On the other hand, as observed in Br-3*R*, bromine is in antiperiplanar conformation with respect to the C_5 methyl in Br-4*R*, which yields the *R*M atropoisomer upon the cyclization of Br-4*M* to Br-8*R*. Br-4*R*-2 conformer, in which bromine is in synperiplanar conformation, is 0.5 kcal/mol less stable than Br-4*R* (Figure 7.18).

On the other hand, 4*S* can undergo keto-enol tautomerization by the transfer of C_5 proton to O_6. The calculated barrier is 75.1 kcal/mol and the relative energy of the enol tautomer (H-4-Enol) is found to be 26.3 kcal/mol higher with respect to the keto isomer H-4*S* (Figure 7.19), indicating the presence of the keto isomer in the reaction medium only.

Structure 4 can undergo a concerted ring closure via the simultaneous hydrogen transfer from arylamine nitrogen (N_3) to the methoxy oxygen to release methyl alcohol with the attack

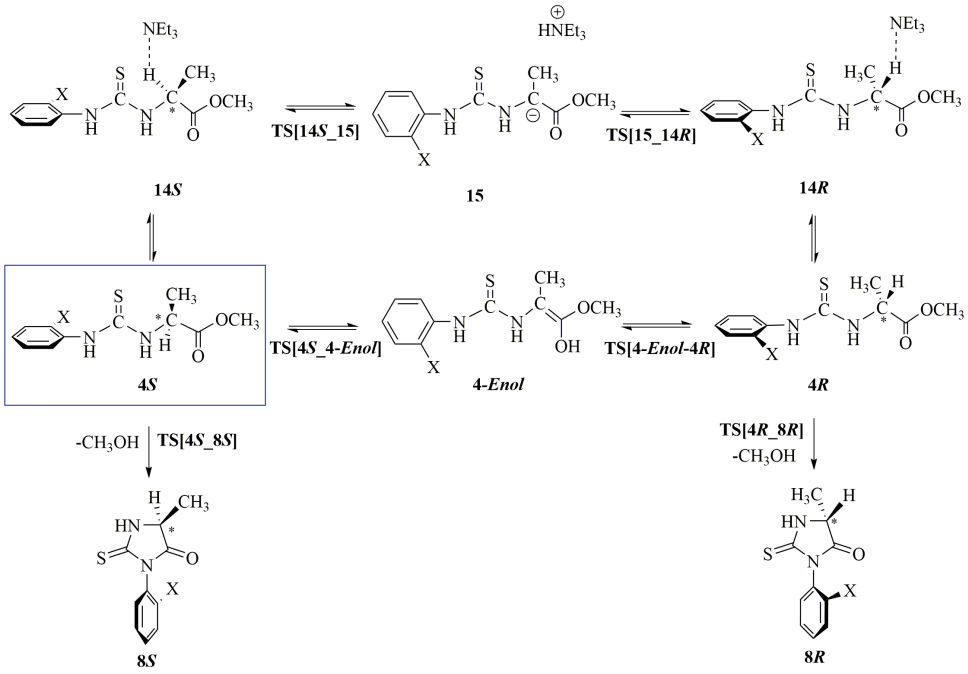


Figure 7.16: Transformations of 4*S* with and without the assistance of TEA.

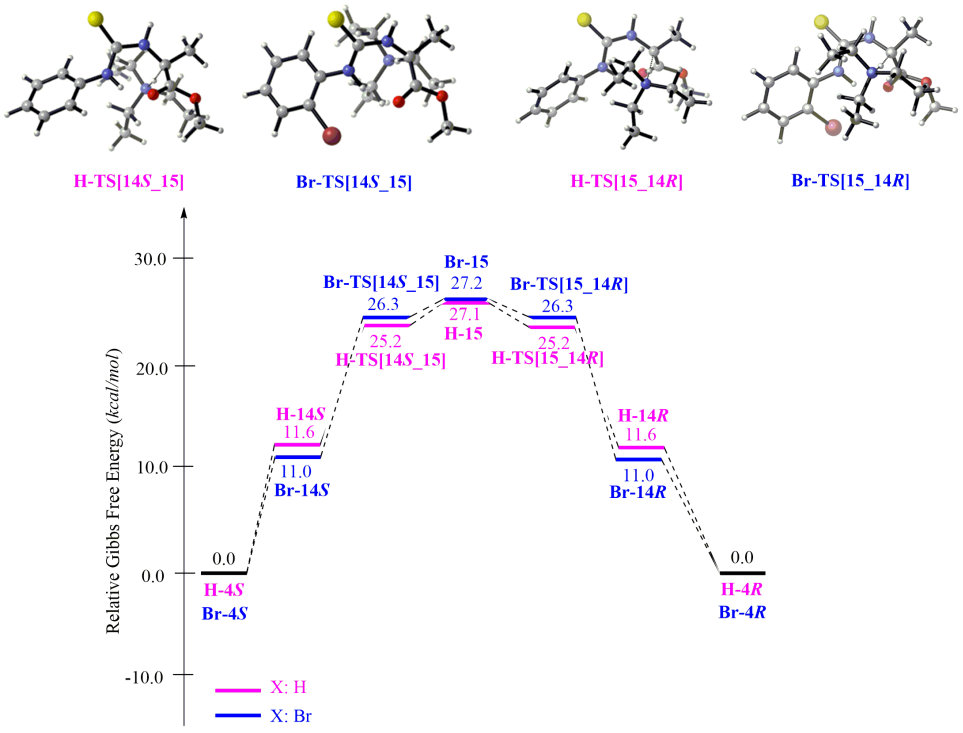


Figure 7.17: Free energy profile for 4*S* to 4*R* transformation

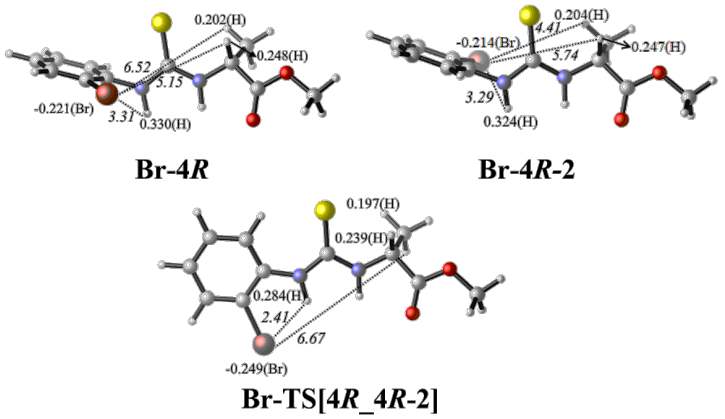


Figure 7.18: Charges (Mulliken Population Analysis) and distances (italics) for Br-4*R*, Br-4*R*-2 and the rotational transition state Br-TS[4*R*_4*R*-2].

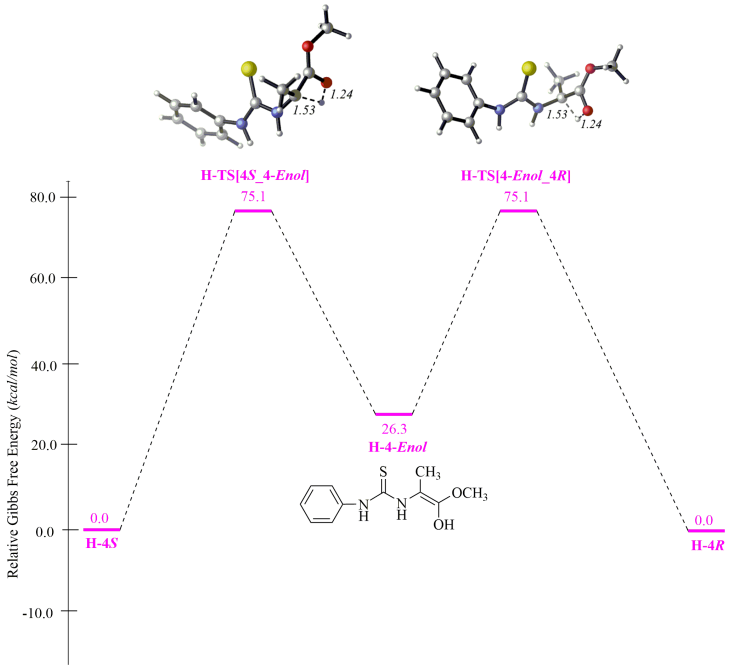


Figure 7.19: Free energy profile for the racemization of H-4*S* via keto-enol tautomerization.

of N_3 to C_4 (Figure 7.20). The transition state structure corresponding to the cyclization of 4*S* resembles the product; the activation barriers for this cyclization step are 49.6 kcal/mol for the formation of racemic mixture of H-8, 48.4 kcal/mol for Br-8*S*P and 48.4 kcal/mol for Br-8*R*M. These barriers are not attainable under the reaction conditions.

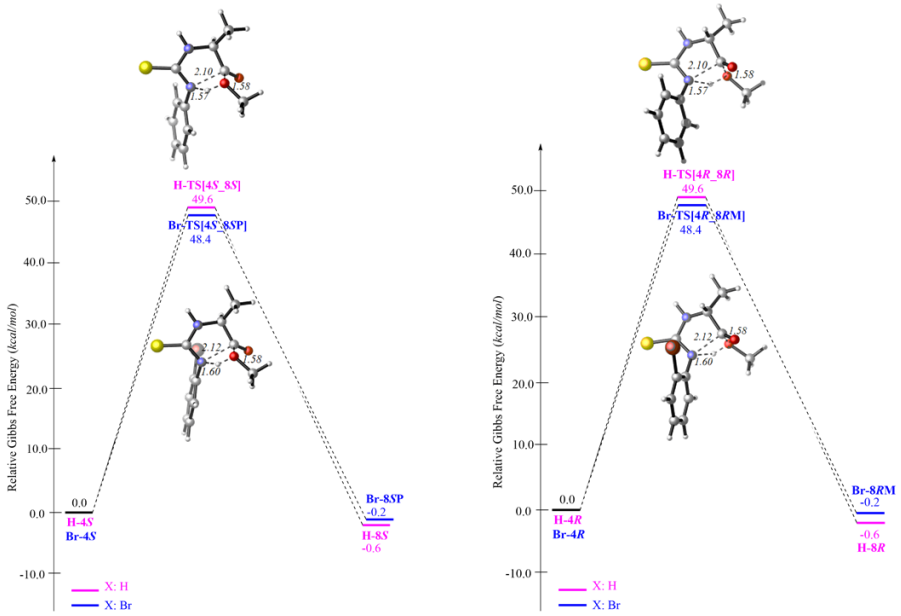


Figure 7.20: Free energy profile for cyclization from 4*S* (left) and 4*R* (right).

(iii) Reactions of 6S

The intermediate 6*S* either cyclizes to 8*S* in two-steps with the assistance of TEA as discussed in cyclization part or follows an independent enantiomerization via TS[6*S*_16], and then yields 8*R* by cyclization (Figure 7.21).

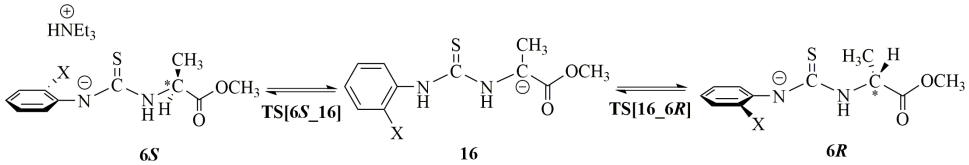


Figure 7.21: Main features for the racemization of 6*S*.

The electron rich N_3 abstracts the hydrogen atom from the stereocenter via TS[6*S*_16] with activation barriers of 21.2 kcal/mol for H and 22.6 kcal/mol for Br. After the free rotation about the N_1-C_5 bond in H-16 and Br-16, hydrogen at N_3 is transferred back to the sp^2 C_5 to give H-6*R* and Br-6*R*. The potential energy surface in Figure 7.22 indicates the formation of racemic mixtures for both H-6 and Br-6; the barriers for the reverse and forward reactions are the same. Moreover, as in the case of Br-3*R* and Br-4*R*, the bromine arranges itself in

the antiperiplanar conformation with respect to C_5 methyl, indicating that the M atropoisomer should be dominant in the final cyclized Br-8*R* product. Favorable interactions between Br and H atoms render Br-6*R* 1.8 kcal/mol more stable than its other conformer Br-6*R*-2 in which bromine is in synperiplanar conformation (Figure 7.23).

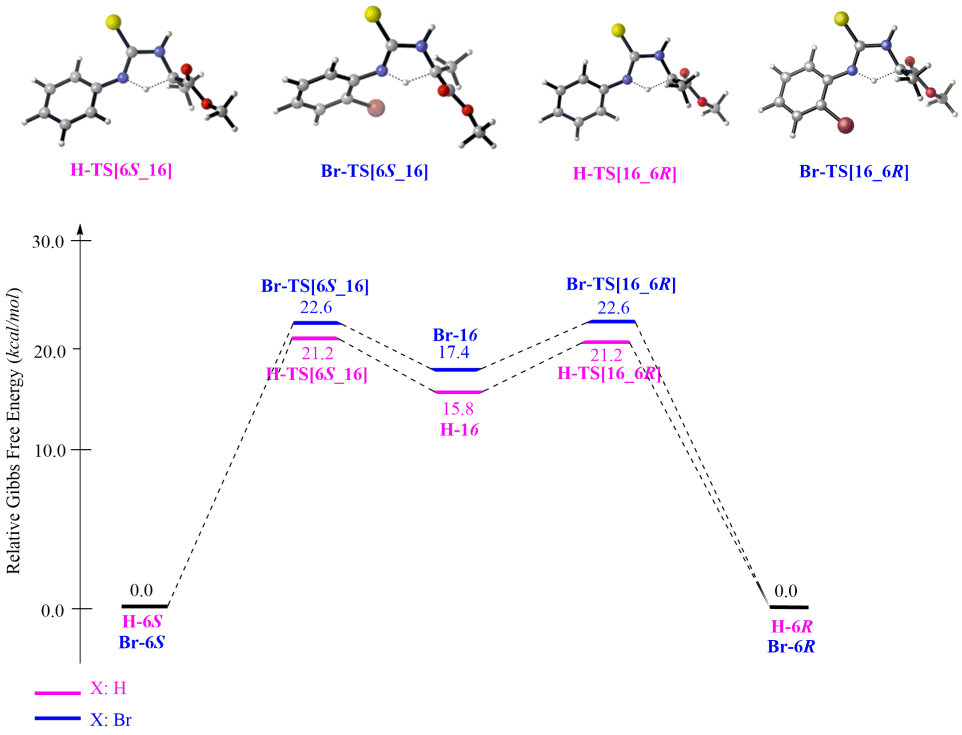


Figure 7.22: Free energy profile for 6*S* to 6*R* transformation.

Summary of the Results The reactions for the suggested mechanisms in this study are gathered in Figure 7.24. Note that in spite of the error margin of M06-2X, one of the most recent hybrid meta density functionals which is generally successful in modeling organic reactions[251], the proposed mechanism is in agreement with the experimental results of Sarigul and Dogan[22] who observed 86% *S* enantiomer dominance for Br-8 ($\Delta\Delta G^{\neq}$= -1.1 kcal/mol). As Br-3*S* forms in the medium, the most favorable path for cyclization follows the Br-3*S*→Br-4*S*→Br-6*S*→Br-8*S* route. The direct cyclizations of the intermediates (Br-3*S*→Br-8*S*P) and (Br-4*S*→Br-8*S*P) are quite demanding (Figure 7.24). Experimentalists have observed variable extents of racemization at C_5 of the heterocyclic ring, thus we have undertaken the task of clarifying computationally whether racemization occurs during the synthesis or after the thiohydantoin ring is formed. For this purpose four possible racemization mechanisms have been evaluated (Br-3*S*→Br-3*R*, Br-4*S*→Br-4*R*, Br-6*S*→Br-6*R*, Br-8*S*P→Br-8*R*P) and their Gibbs energy barriers have been evaluated. Although the barriers to overcome for enantiomerizations (Br-3*S*→Br-3*R*, Br-4*S*→Br-4*R*, Br-6*S*→Br-6*R*) would be reachable under the reaction conditions, the corresponding competing reactions (Br-3*S*→Br-4*S*→Br-6*S*→Br-8*S*P→Br-8*R*P) are highly preferred

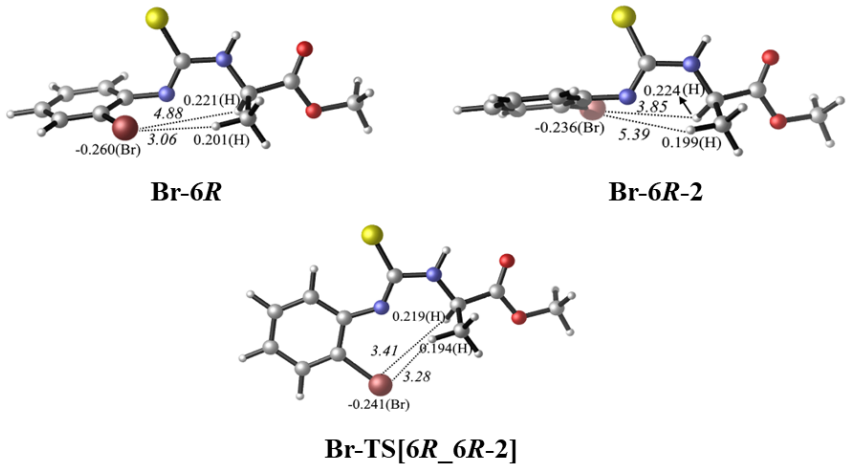


Figure 7.23: Charges (Mulliken Population Analysis) and distances (italics) for the lowest energy isomers of Br-6*R*, Br-6*R*-2 and the rotational transition state Br-TS[6*R*_6*R*-2].

due to their lower barriers; the route favored during the reaction is highlighted in Figure 7.24. It is also worth noting that if racemization occured before cyclization, the cyclized species Br-8*R* would have the M configuration in disagreement with experiment: calculations have shown that most stable conformers of Br-3*R*, Br-4*R* and Br-6*R* have Br in the antiperiplanar position with respect to the methyl group at C_5. Eventually, the preferred synperiplanar position of Br is kept starting from Br-3*S* to the final thiohydantoin formed, thus yielding atropoisomer Br-8*S*P in which bromine is transoid with respect to C_5 methyl. The rotation of bromoaryl ring, thus the transformation of Br-8*S*P to Br-8*S*M, has an experimentally determined barrier, 27.8 kcal/mol [22], and this relatively high barrier explains the fact that Br-8*S*P has 83% abundance whereas Br-8*S*M is only 3%. Calculations have shown that after racemization occurs at C_5 of Br-8*S*P, Br-8*R*P will be formed and the rotation of the bromoaryl ring (*R*P⇌*R*M) requires 27.5 kcal/mol to yield Br-8*R*M.

7.3.2 Computational Approach to the Elucidation of Stereoconversion of Thiohydantoin Derivatives

Recently, 5-Benzyl-3-aryl-2-thiohydantoin and 5-isobutyl-3-aryl-2-thiohydantoin derivatives were synthesized by using S-phenyl alanine methyl ester and S-leucine methyl ester hydrochloride with the corresponding aryl isothiocyanates in the presence of triethylamine by Dogan's group, following the procedure as described earlier [22]. When R_1 substituent in Figure 7.1 is benzyl group and X substituent is bromine atom, distribution of the isomers are reported as Benzyl-8*S*M:Benzyl-8*R*P:Benzyl-8*S*P:Benzyl-8*R*M=2:41:51:6. The stereoisomers of compounds were resolved and kept in ethanol to follow the kinetics of conversion of the stereoisomers and the energy barriers for racemization have been calculated. For the racemization of Benzyl-8*S*P compound the experimental barrier is 24.4 kcal/mol. In this part of the study, we have elucidated the mechanism of racemization of 2-thiohydantoins in ethanol environment.

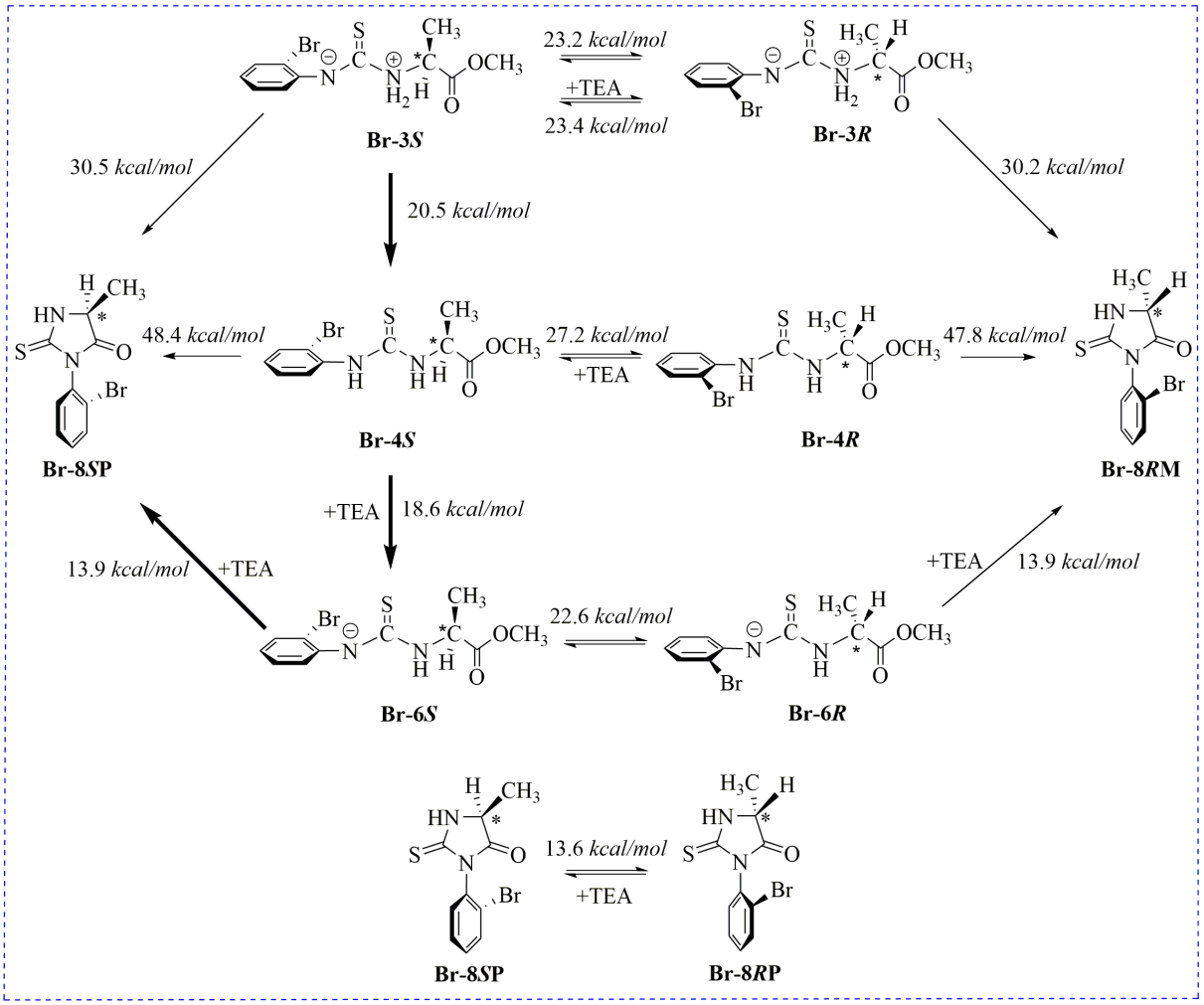


Figure 7.24: Summary of all pathways for the formation of Br-8*S*P and Br-8*R*M.

Keto-enol tautomerization is suggested to be one of the main mechanisms responsible from the racemization of cyclized products [142, 143, 144]. Therefore, in order to clarify whether the enolization is a viable process in racemization of thiohydantoins or not, we have modeled the intramolecular proton transfer from C_5 to O_6 with and without a base catalyst (Figure 7.25).

The data in Figure 7.25 demonstrates that the keto-enol tautomerization without a catalyst is not kinetically driven because of the very high activation barrier (79.4 kcal/mol, blue line in Figure 7.25). Thus we propose a proton relay racemization process, in which explicit ethanol molecules assist the intramolecular proton transfer from C_5 to O_6 forming the enol intermediate. We have modeled the entire reaction with one (Figure 7.25, red line) and two (Figure 7.25, green line) explicit ethanol molecules. The located transition states for the deprotonation/reprotonation at C_5 of the heterocyclic ring via one ethanol assistance embodies a six membered cyclic structure between Benzyl-8*S*P and ethanol molecule (Benzyl-TS[8*S*P_8-Enol] and Benzyl-TS[8-Enol_8*R*P]). Following a concerted bond reorganization in which double proton transfer occurs between C_5, ethanol and O_6, the enol intermediate (Benzyl-8-Enol) is formed. The activation energy for this deprotonation step is 35.9 kcal/mol. On the other hand, deprotonation by two ethanol molecules assistance has much lower activation barrier (22.4 kcal/mol), in which the transition state structures embody an eight-membered cyclic structure involving triple proton transfer. Shorter hydrogen bond lengths in two ethanol assisted mechanism indicates more stabilized transition states for both enolization (Benzyl-TS[8*S*P_8-Enol]) and ketonization (Benzyl-TS[8-Enol_8*R*P]). Reprotonation step in order to give *R* enantiomer requires 28.5 kcal/mol and 19.0 kcal/mol for one and two ethanol assisted mechanisms respectively. Suggested two ethanol assisted mechanism with 22.4 kcal/mol activation energy (rate determining step is the deprotonation step) is in very good agreement with the experimentally determined value.

Finally we have calculated the relative proportions for Benzyl-8 in terms of *S*/*R* isomer ratios according to the TEA assisted racemization mechanism suggested in the first part of this study, the obtained ratio of 56/44 is in very good agreement with the experimentally reported data which is 53/47 (Figure 7.26). Table 7.1 shows the computed NPA charges for the reactants and the transition states of deprotonation and reprotonation steps for Br-8*S*P and Benzyl-8*S*P which differ in their C_5 substituents; Br-8*S*P has a methyl group and Benzyl-8*S*P has a benzyl group attached. These substituents have direct impact on the extent of racemization at C_5 of the heterocyclic rings. Benzyl substituent turns out the ring more prone to racemization while methyl substituent suppresses the racemization. NPA charges on C_5 of Br-9*S*P (q_C=-0.164 e) and Benzyl-9*S*P (q_C=-0.150 e), the TEA complexes of Br-8*S*P and Benzyl-8*S*P, are closely related to the electron donating power of the substituents at C_5. Comparing the activation energies for deprotonation steps, replacing the methyl substituent with benzyl group lowers the barrier by 0.7 kcal/mol (Br-TS[9*S*P_10] in Figure 7.9 and Benzyl-TS[9*S*P_10] in Figure 7.26). This small barrier difference can be interpreted in terms of the C_5-H bond lengths which is 1.50 Å in Br-TS[9*S*P_10] and 1.51 Å in Benzyl-TS[9*S*P_10]. Slightly shorter bond length in Br-TS[9*S*P_10] is due to the fact that electron donating methyl group enhances the electron density between C_5 and H, thus intensifies the bond strength and the stronger bond leads to a higher activation energy (13.6 kcal/mol for Br-TS[9*S*P_10] and 12.9 kcal/mol for Benzyl-TS[9*S*P_10]). On the other hand, charges on the hydrogen atoms at stereocenter are 0.292 e and 0.298 e for Br-9*S*P and Benzyl-9*S*P respectively, and the relatively higher electron density

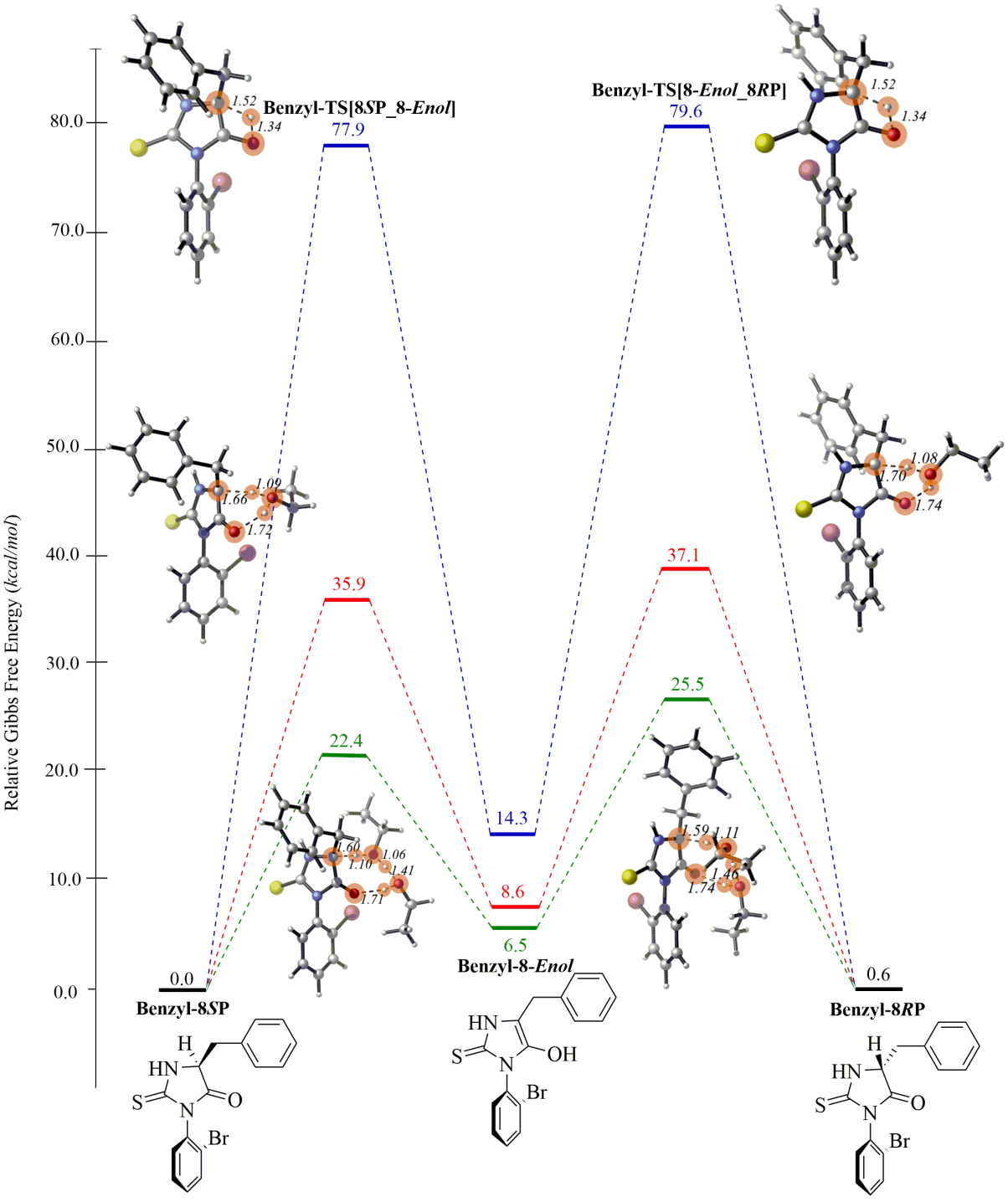


Figure 7.25: One ethanol (red line) and two ethanol molecules (green line) assisted racemization mechanism based on keto-enol tautomerization.

observed in H of Br-9*S*P accounts for the higher barrier in Br-TS[9*S*P_10]. In the reprotonation step, relative barrier heights are 11.5 kcal/mol for Br-TS[10_9*R*P] and 11.1 kcal/mol for Benzyl-TS[10_9*R*P]. C_5 at Br-TS[10_9*R*P] is richer in electron population (q_C=-0.252 e) compared to Benzyl-TS[10_9*R*P] (q_C=-0.244 e) and the transferring proton at Br-TS[10_9*R*P] has lower charge compared to Benzyl-TS[10_9*R*P], thus, as a result, a stronger electron density is observed between C_5 and H in Br-TS[10_9*R*P] with a shorter bond length 1.48 Å; where the same bond length is 1.50 Å in Benzyl-TS[10_9*R*P].

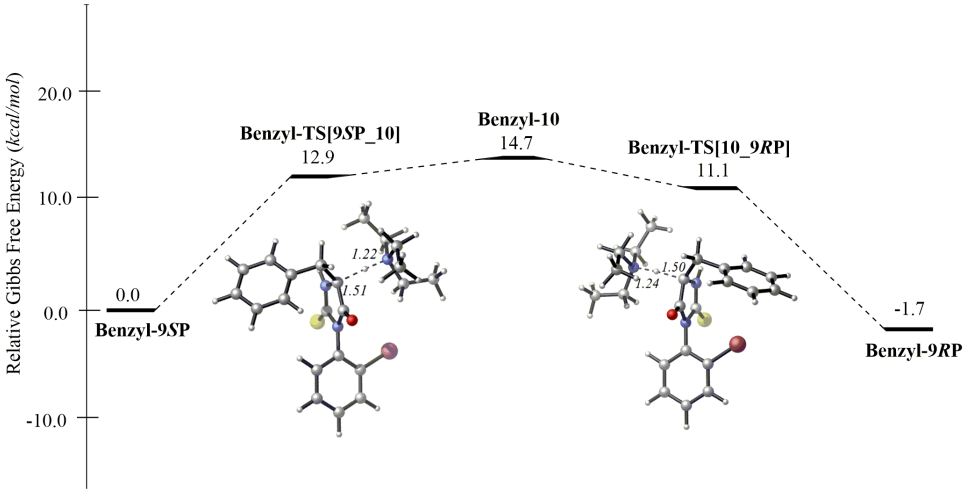


Figure 7.26: Racemization of Benzyl-8*S*P through TEA assisted mechanism.

Table 7.1: Calculated NPA atomic charges for the transition states of TEA assisted racemization at C_5 of Br-8*S*P and Benzyl-8*S*P (M06-2X/6-311+G(d,p) for C, H, O, N and S atoms; M06-2X/6-311++G(3df,3pd) for Br atom).

Structure	H	N1	C2	N3	C4	C5
Br-9SP	0.292	-0.622	0.329	-0.559	0.751	-0.164
Br-TS[9SP_10]	0.451	-0.594	0.327	-0.553	0.684	-0.261
Br-TS[10_9RP]	0.445	-0.595	0.325	-0.561	0.685	-0.252
Benzyl-9SP	0.298	-0.628	0.335	-0.568	0.755	-0.150
Benzyl-TS[9SP_10]	0.456	-0.596	0.330	-0.556	0.689	-0.250
Benzyl-TS[10_9RP]	0.453	-0.601	0.328	-0.564	0.697	-0.244

7.4 Conclusions

In this study, the racemization differences observed in axially chiral atropoisomeric and nonaxially chiral 2-thiohydantoins have been explored by quantum mechanical methods. We have suggested a plausible mechanism by modeling various possibilities for the formation of 2-thiohydantoins. The computational data shows that the most plausible pathway starts with the nucleophilic addition of the amino acid (2*S*) to the *o*-aryl isothiocyanate (1) yielding intermediate 3*S* which then follows the Br-3*S*→Br-4*S*→Br-6*S*→Br-8*S* route. Once 2-thiohydantoin ring (8*S*) is formed it can undergo racemization with triethyl amine which is already present in the reaction medium. Racemization via the formation of methanol assisted enol intermediate is not favorable due to very high activation barriers (37.4 kcal/mol). The proposed mechanisms corroborate the experimental findings regarding the dominance of the S isomer in the final product 8 when X:Br. When X:H on the other hand, the *S* and *R* configured intermediates are equally stable, thus racemic products are obtained. Overall, the atroposelectivity in the synthesis of axially chiral thiohydantoin derivatives can be attributed to the synperiplanar position of Br atom with respect to methyl group at C_5 in Br-3*S* which keeps its conformation through the intermediate states. Moreover, we have shown that the diversity in the extent of racemization at C_5 of 5-methyl-3-*o*-bromophenyl-2-thiohydantoin (Br-8) and 5-benzyl-3-*o*-bromophenyl-2-thiohydantoin (Benzyl-8) arises from the differences in the electron donating power of C_5 substituents. Additionally, DFT calculations were performed for the elucidation of racemization mechanism in ethanol, and we suggested a two-ethanol assisted enolization/ketonization process which is in good agreement with the experimentally determined activation barriers.

The computational methods utilized in this study can be used with confidence to predict the chirality in the atroposelective synthesis of axially chiral thiohydantoin derivatives.

Concluding Remarks

In this dissertation, different approaches have been used to get a deeper understanding of the characterization and rationalization of the ligands that can be regarded as drug candidates. Biological activity and selectivity of ligands have been identified and the ligand-protein interactions in NMDA receptor have been investigated.

Quantum chemical calculations were performed on several NMDAR-GluN1 subunit selective ligands which have antagonistic and partial agonistic activity. Several global and local molecular properties that could play a role in ligand binding to the GluN1 subunit of NMDA were examined and these descriptors were then used to discriminate the partial agonism-antagonism nature of the ligands. The HOMO energies of the partial agonists were calculated to be on average lower than those of the antagonists, whereas the reverse has been seen for the LUMO energies. Calculated chemical hardness values pointed out that the partial agonists are harder species than antagonists. Based on the hard and soft acids and bases principle, which states that hard acids prefer to bind to hard bases and soft acids to soft bases, this could indicate that the two types of ligands interact stronger or weaker with the residues on the binding sites of the receptor, depending on their hardness values.

In order to understand the differences between reactivity centers in ligand-receptor interactions of antagonists and partial agonists, local descriptors were applied to the nitrogen and oxygen atoms of the amide group. Since ligand-receptor recognition is mostly driven by interactions between partially charged atoms, the charge distribution on the amide group atoms were calculated. The results showed that the amide oxygen and nitrogen of the partial agonists are more negatively charged than those of the antagonists. This suggests stronger hydrogen bonding acceptor and donor ability of the partial agonists' amide group. These findings can be used in NMDAR targeted drug design strategies in order to predict the agonism-antagonism activity of drug candidates. Moreover, we have suggested a multiple linear regression model for calculating the intrinsic efficacies of the ligands and the predictions might be a guideline for experimentalists. It is expected that the quantum chemical calculations performed in this study will be tested and improved by theoreticians and experimentalists by expanding the application to the other iGluRs, namely AMPA and Kainate receptors.

Docking studies and molecular dynamics simulations have been performed in order to calculate the binding Gibbs free energies for the natural receptor's GluN1 subunit ligand binding domain-ligand complexes and obtained values have been correlated with their experimental efficacy and binding affinity (pK_i). Correlations have been checked, but poor linearities have been obtained. The results obtained in this study point out that the receptor's functionality is tightly linked to the other subunits including the ion channel. This should be taken into account when performing the simulations; GluN1 subunit should be modeled within the full NMDA receptor

to obtain a complete picture of the ligand activity. On the other hand, a good correlation has been found between the ligand's agonism-antagonism activity and the receptor's conformational states. By these correlations it is possible to discriminate between a ligand's agonism activity from antagonism activity. The preliminary results promisingly show that a mathematical model can be derived by extending the training set of ligands. For binding Gibbs free energy calculations of the NMDAR-ligand systems, GB^{HCT} and GB^{OBC} implicit solvation models can be used with confidence.

The solubility of a drug ligand is governed by the pK_a's of the acidic and basic sites within the molecule. When the ligand is in the target protein environment, its affinity, activity and efficacy is dependent on the pK_a's of the interacting residues of the active site. Regarding the successful applications of quantum mechanical charges as descriptors, an accurate protocol has been suggested for the fast prediction of pK_a's of ligands and amino acids bearing a carboxyl group. According to the suggested protocol, pK_a's are computed by using the equation derived from the linear regression of the experimental pK_a's with the maximum atomic charge on the oxygen atoms of the carboxylate fragment. Molecular dynamics simulations have been performed for a set of peptides and the prediction was found to be stable with respect to the geometrical changes. The protocol was also applied to the set of partial agonists and antagonists which have high affinity for NMDA receptor GluN1 subunit. With the suggested protocol, it is possible to predict the protonation states of the ionizable residues. The extension of the suggested protocol includes the application to the pK_{a} prediction of aspartate, glutamate and alanine within protein environment. In a next step, a new protocol will be suggested for pK_{a} prediction of a very important class of organic compounds, namely amines.

Stereoisomers of a chiral molecule may show different biological activities. A ligand can be stereoselective or regioselective depending on the differences in electronic and steric environment. For this reason, the demonstration of configurational stabilities of drug candidates and mechanisms by which they racemize are important issues. Therefore, in order to highlight the enantioselectivity differences observed in axially chiral atropoisomeric and nonaxially chiral 2-thiohydantoins, a plausible mechanism has been suggested by modeling various possibilities for the formation of 2-thiohydantoins. The observed racemization at C_5 of the heterocyclic ring was investigated; we have computationally clarified that the racemizaton occurs after the thiohydantoin ring is formed. Overall, it has been found that the reason for the observation of atroposelectivity in the synthesis of axially chiral thiohydantoin derivatives is the synperiplanar position of Br atom with respect to methyl group at C_5 position in the reactant, which keeps its conformation through the intermediate states formed. Moreover, we have shown that the diversity in the extent of racemization arises from the differences in the electron donating power of C_5 substituents. The computational methods utilized in this study can be used with confidence to predict the chirality in the atroposelective synthesis of axially chiral thiohydantoin derivatives.

Bibliography

[1] Furukawa, H. Structure and Function of Glutamate Receptor Amino Terminal Domains. *The Journal of Physiology* **2012**, *590*, 63–72.

[2] Pinheiro, P.; Mulle, C. Kainate Receptors. *Cell and Tissue Research* **2006**, *326*, 457–482.

[3] Traynelis, S. F.; Wollmuth, L. P.; McBain, C. J.; Menniti, F. S.; Vance, K. M.; Ogden, K. K.; Hansen, K. B.; Yuan, H.; Myers, S. J.; Dingledine, R. Glutamate Receptor Ion Channels: Structure, Regulation, and Function. *Pharmacological Reviews* **2010**, *62*, 405–496.

[4] Schmid, S. M.; Hollmann, M. To Gate or Not to Gate: Are the Delta Subunits in the Glutamate Receptor Family Functional Ion Channels? *Molecular Neurobiology* **2008**, *37*, 126–141.

[5] Kaye, S. L.; Sansom, M. S. P.; Biggin, P. C. Molecular Dynamics Simulations of the Ligand-Binding Domain of an N-Methyl-D-Aspartate Receptor. *Journal of Biological Chemistry* **2006**, *281*, 12736–12742.

[6] Lau, C. G.; Zukin, R. S. NMDA Receptor Trafficking in Synaptic Plasticity and Neuropsychiatric Disorders. *Nature Reviews. Neuroscience* **2007**, *8*, 413–26.

[7] Cull-Candy, S. G.; Leszkiewicz, D. N. Role of Distinct NMDA Receptor Subtypes at Central Synapses. *Science's STKE : Signal Transduction Knowledge Environment* **2004**, *2004*, re16.

[8] Parsons, C. G.; Danysz, W.; Quack, G. Glutamate in CNS Disorders As a Target For Drug Development: An Update. *Drug News & Perspectives* **1998**, *11*, 523–569.

[9] Danysz, W.; Parsons, C. G. Glycine and N-Methyl-D-Aspartate Receptors: Physiological Significance and Possible Therapeutic Applications. *Pharmacological reviews* **1998**, *50*, 597–664.

[10] Tikhonova, I. G.; Baskin, I. I.; Palyulin, V. A.; Zefirov, N. S. CoMFA and Homology-Based Models of the Glycine Binding Site of N-Methyl-D-Aspartate Receptor. *Journal of Medicinal Chemistry* **2003**, *46*, 1609–1616.

[11] Millan, M. J. N-Methyl-D-Aspartate Receptors as a Target for Improved Antipsychotic Agents: Novel Insights and Clinical Perspectives. *Psychopharmacology* **2005**, *179*, 30–53.

[12] Catarzi, D.; Colotta, V.; Varano, F.; Filacchioni, G.; Galli, A.; Costagli, C.; Vincenzo, C. Synthesis, Ionotropic Glutamate Receptor Binding Affinity, and Structure-Activity Relationships of a New Set of 4,5-Dihydro-8-heteroaryl-4-oxo-1,2,4-triazolo[1,5-a]quinoxaline-2-carboxylates Analogues of TQX-173. *Journal of Medicinal Chemistry* **2001**, *44*, 3157–3165.

[13] Nguyen, K. T.; Syed, S.; Urwyler, S.; Bertrand, S.; Bertrand, D.; Reymond, J. L. Discovery of NMDA Glycine Site Inhibitors from the Chemical Universe Database GDB. *ChemMedChem* **2008**, *3*, 1520–1524.

[14] Regan, M. C.; Romero-Hernandez, A.; Furukawa, H. A Structural Biology Perspective on NMDA Receptor Pharmacology and Function. *Current Opinion in Structural Biology* **2015**, *33*, 68–75.

[15] Newcomer, J. W.; Krystal, J. H. NMDA Receptor Regulation of Memory and Behavior in Humans. *Hippocampus* **2001**, *11*, 529–542.

[16] Ylilauri, M.; Pentikäinen, O. T. Structural Mechanism of N-Methyl-D-Aspartate Receptor Type 1 Partial Agonism. *PLoS ONE* **2012**, *7*, e47604.

[17] Blanke, M. L.; VanDongen, A. M. *Biology of the NMDA Receptor*; CRC Press/Taylor & Francis, 2009.

[18] Brunton, L. L.; Lazo, J. S.; Parker, K. L. *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 11th ed.; Mc.Graw-Hill Medical Pub.: NewYork, 2005.

[19] Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. Empirical Relationships Between Protein Structure and Carboxyl pKa Values in Proteins. *Proteins: Structure, Function and Genetics* **2002**, *48*, 388–403.

[20] Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Structure, Function and Genetics* **2005**, *61*, 704–721.

[21] Harris, T. K.; Turner, G. J. Structural Basis of Perturbed pKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* **2002**, *53*, 85–98.

[22] Sarigul, S.; Dogan, I. Atroposelective Synthesis of Axially Chiral Thiohydantoin Derivatives. *The Journal of Organic Chemistry* **2016**, *81*, 5895–5902.

[23] Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180.

[24] Cheng, Z.; Zhang, Y.; Fu, W. Predictive QSAR Models of 3-Acylamino-2-Aminopropionic Acid Derivatives as Partial Agonists of the Glycine Site on the NMDA Receptor. *Medicinal Chemistry Research* **2011**, *20*, 1235–1246.

[25] Zhu, X.; Cai, H.; Xu, Z.; Wang, Y.; Wang, H.; Zhang, A.; Zhu, W. Classification of 5-HT(1A) Receptor Agonists and Antagonists Using GA-SVM Method. *Acta Pharmacologica Sinica* **2011**, *32*, 1424–1430.

[26] Deng, C.; Luan, F.; Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M. N. D. S. Recent Advances on QSAR-Based Profiling of Agonist and Antagonist A3 Adenosine Receptor Ligands. *Current Topics in Medicinal Chemistry* **2013**, *13*, 1048–1068.

[27] Djeradi, H.; Rahmouni, A.; Cheriti, A. Antioxidant Activity of Flavonoids: A QSAR Modeling Using Fukui Indices Descriptors. *Journal of Molecular Modeling* **2014**, *20*, 2476.

[28] Yosa, J.; Blanco, M.; Acevedo, O.; Lareo, L. Molecular Orbital Differentiation of Agonist and Antagonist Activity in the Glycine B-iGluR-NMDA Receptor. *European Journal of Medicinal Chemistry* **2009**, *44*, 2960–2966.

[29] Di Fabio, R. et al. Substituted Indole-2-carboxylates as in Vivo Potent Antagonists Acting as the Strychnine-Insensitive Glycine Binding Site. *Journal of Medicinal Chemistry* **1997**, *40*, 841–850.

[30] Urwyler, S.; Floersheim, P.; Roy, B. L.; Koller, M. Drug Design, in Vitro Pharmacology, and Structure-Activity Relationships of 3-Acylamino-2-aminopropionic Acid Derivatives, a Novel Class of Partial Agonists at the Glycine Site on the N-Methyl-D-Aspartate (NMDA) Receptor Complex. *Journal of Medicinal Chemistry* **2009**, *52*, 5093–5107.

[31] Krishnamurty, S.; Pal, S. Intermolecular Reactivity Trends Using the Concept of Group Softness. *The Journal of Physical Chemistry A* **2000**, *104*, 7639–7645.

[32] Vijayaraj, R.; Subramanian, V.; Chattaraj, P. K. Comparison of Global Reactivity Descriptors Calculated Using Various Density Functionals: A QSAR Perspective. *Journal of Chemical Theory and Computation* **2009**, *5*, 2744–2753.

[33] Pearson, R. G. Hard and Soft Acids and Bases. *American Chemical Society* **1963**, *85*, 3533–3539.

[34] Pearson, R. G. Acids and Bases. *Science* **1966**, *151*, 172–177.

[35] Pearson, R. G. *Chemical Hardness*; Wiley-VCH, 1997; p 198.

[36] Padmanabhan, J.; Parthasarathi, R.; Elango, M.; Subramanian, V.; Krishnamoorthy, B.; Gutierrez-Oliva, S.; Toro-Labbe, A.; Roy, D. R.; Chattaraj, P. Multiphilic Descriptor for Chemical Reactivity and Selectivity. *Journal of Physical Chemistry A* **2007**, *111*, 9130–9138.

[37] Cai, S. X.; Zhou, Z. L.; Huang, J. C.; Whittemore, E. R.; Egbuwoku, Z. O.; Lu, Y.; Hawkinson, J. E.; Woodward, R. M.; Weber, E.; Keana, J. F. W. Synthesis and Structure-Activity Relationships of 1,2,3,4-Tetrahydroquinoline-2,3,4-trione 3-Oximes: Novel and Highly Potent Antagonists for NMDA Receptor Glycine Site. *Journal of Medicinal Chemistry* **1996**, *39*, 3248–3255.

[38] Joshi, K. A.; Patil, D. D.; Gejji, S. P. Molecular Electrostatic Potentials in Aromatic Substituted 4-hydroxyquino-2-lones: Glycine/NMDA Receptor Antagonists. *Journal of Molecular Modeling* **2009**, *15*, 383–390.

[39] Karakas, E.; Simorowski, N.; Furukawa, H. Subunit Arrangement and Phenylethanolamine Binding in GluN1/GluN2B NMDA Receptors. *Nature* **2011**, *475*, 249–253.

[40] Inanobe, A.; Furukawa, H.; Gouaux, E. Mechanism of Partial Agonist Action at the NR1 Subunit of NMDA Receptors. *Neuron* **2005**, *47*, 71–84.

[41] Armstrong, N.; Gouaux, E. Mechanisms for Activation and Antagonism of an AMPA-Sensitive Glutamate Receptor: Crystal Structures of the GluR2 Ligand Binding Core. *Neuron* **2000**, *28*, 165–181.

[42] Mayer, M. L. Crystal Structures of the GluR5 and GluR6 Ligand Binding Cores: Molecular Mechanisms Underlying Kainate Receptor Selectivity. *Neuron* **2005**, *45*, 539–552.

[43] Mayer, M. L.; Armstrong, N. Structure and Function of Glutamate Receptor Ion Channels. *Annual Review of Physiology* **2004**, *66*, 161–181.

[44] Mayer, M. L. Glutamate Receptors at Atomic Resolution. *Nature* **2006**, *440*, 456–462.

[45] Furukawa, H.; Gouaux, E. Mechanisms of Activation, Inhibition and Specificity: Crystal Structures of the NMDA Receptor NR1 Ligand-Binding Core. *The EMBO Journal* **2003**, *22*, 2873–2885.

[46] Kvist, T.; Bielefeldt, T. S.; Greenwood, J. R.; Mehrzad Tabrizi, F.; Hansen, K. B.; Gajhede, M.; Pickering, D. S.; Traynelis, S. F.; Kastrup, J. S.; Bräuner-Osborne, H. Crystal Structure and Pharmacological Characterization of a Novel N-Methyl-D-Aspartate (NMDA) Receptor Antagonist at the GluN1 Glycine Binding Site. *Journal of Biological Chemistry* **2013**, *288*, 33124–33135.

[47] Jin, R.; Banke, T. G.; Mayer, M. L.; Traynelis, S. F.; Gouaux, E. Structural Basis for Partial Agonist Action at Ionotropic Glutamate Receptors. *Nature Neuroscience* **2003**, *6*, 803–810.

[48] Dürr, K. L.; Chen, L.; Stein, R. A.; De Zorzi, R.; Folea, I. M.; Walz, T.; Mchaourab, H. S.; Gouaux, E. Structure and Dynamics of AMPA Receptor GluA2 in Resting, Pre-open, and Desensitized States. *Cell* **2014**, *158*, 778–792.

[49] Furukawa, H.; Singh, S. K.; Mancusso, R.; Gouaux, E. Subunit Arrangement and Function in NMDA receptors. *Nature* **2005**, *438*, 185–192.

[50] Vance, K. M.; Simorowski, N.; Traynelis, S. F.; Furukawa, H. Ligand-Specific Deactivation Time Course of GluN1/GluN2D NMDA Receptors. *Nature Communications* **2011**, *2:294*.

[51] Yao, Y.; Belcher, J.; Berger, A. J.; Mayer, M. L.; Lau, A. Y. Conformational Analysis of NMDA Receptor GluN1, GluN2, and GluN3 Ligand-Binding Domains Reveals Subtype-Specific Characteristics. *Structure* **2013**, *21*, 1788–1799.

[52] Hansen, K. B.; Tajima, N.; Risgaard, R.; Perszyk, R. E.; Jørgensen, L.; Vance, K. M.; Ogden, K. K.; Clausen, R. P.; Furukawa, H.; Traynelis, S. F. Structural Determinants of Agonist Ffficacy at the Glutamate Binding Site of N-methyl-D-aspartate Receptors. *Molecular Pharmacology* **2013**, *84*, 114–127.

[53] Yuriev, E.; Agostino, M.; Ramsland, P. A. Challenges and Advances in Computational Docking: 2009 in Review. *Journal of Molecular Recognition* **2011**, *24*, 149–164.

[54] Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer-Aided Drug Design* **2011**, *7*, 146–157.

[55] Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for Molecular Docking: A Review. *Biophysical Reviews* **2017**, *9*, 91–102.

[56] Zwanzig, R. W. High Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics* **1954**, *22*, 1420–1426.

[57] Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83*, 833–835.

[58] Aqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Engineering* **1994**, *7*, 385–391.

[59] Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research* **2000**, *33*, 889–897.

[60] Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *Journal of the American Chemical Society* **2001**, *123*, 5221–5230.

[61] Beveridge, D. L.; DiCapua, F. M. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annual Review of Biophysics and Biophysical Chemistry* **1989**, *18*, 431–492.

[62] Chipot, C.; Rozanska, X.; Dixit, S. B. Can Free Energy Calculations Be Fast and Accurate at the Same Time? Binding of Low-Affinity, Non-Peptide Inhibitors to the SH2 Domain of the SRC Protein. *Journal of Computer-Aided Molecular Design* **2005**, *19*, 765–770.

[63] Genheden, S.; Ryde, U. How to Obtain Statistically Converged MM/GBSA Results. *Journal of Computational Chemistry* **2010**, *31*, 837–846.

[64] Rastelli, G.; Degliesposti, G.; Del Rio, A.; Sgobba, M. Binding Estimation After Refinement, A New Automated Procedure for the Refinement and Rescoring of Docked Ligands in Virtual Screening. *Chemical Biology & Drug Design* **2009**, *73*, 283–286.

[65] Kongsted, J.; Ryde, U. An Improved Method to Predict the Entropy Term with the MM/PBSA Approach. *Journal of Computer-Aided Molecular Design* **2009**, *23*, 63–71.

[66] Dravid, S. M.; Burger, P. B.; Prakash, A.; Geballe, M. T.; Yadav, R.; Le, P.; Vellano, K.; Snyder, J. P.; Traynelis, S. F. Structural Determinants of D-cycloserine Efficacy at the NR1/NR2C NMDA Receptors. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* **2010**, *30*, 2741–2754.

[67] Postila, P. A.; Ylilauri, M.; Pentikäinen, O. T. Full and Partial Agonism of Ionotropic Glutamate Receptors Indicated by Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2011**, *51*, 1037–1047.

[68] Mamonova, T.; Speranskiy, K.; Kurnikova, M. Interplay Between Structural Rigidity and Electrostatic Interactions in the Ligand Binding Domain of GluR2. *Proteins* **2008**, *73*, 656–671.

[69] Dolino, D. M.; Cooper, D.; Ramaswamy, S.; Jaurich, H.; Landes, C. F.; Jayaraman, V. Structural Dynamics of the Glycine-Binding Domain of the N-Methyl-D-Aspartate Receptor. *The Journal of Biological Chemistry* **2015**, *290*, 797–804.

[70] Çifci, G.; Aviyente, V.; Akten, E. D.; Monard, G. Assessing Protein–Ligand Binding Modes with Computational Tools: The Ccase of PDE4B. *Journal of Computer-Aided Molecular Design* **2017**, *31*, 563–575.

[71] Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein Ionizable Groups: pKa Values and Their Contribution to Protein Stability and Solubility. *Journal of Biological Chemistry* **2009**, *284*, 13285–13289.

[72] Kim, J.; Mao, J.; Gunner, M. R. Are Acidic and Basic Groups in Buried Proteins Predicted to Be Ionized? *Journal of Molecular Biology* **2005**, *348*, 1283–1298.

[73] Ji, C.; Mei, Y.; Zhang, J. Z. Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp 20 in Thioredoxin. *Biophysical Journal* **2008**, *95*, 1080–1088.

[74] Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Garcia-Moreno E., B. Large Shifts in pKa Values of Lysine Residues Buried Inside a Protein. *Proceedings of the National Academy of Sciences* **2011**, *108*, 5260–5265.

[75] Li, H.; Robertson, A. D.; Jensen, J. H. The Determinants of Carboxyl pKa Values in Turkey Ovomucoid Third Domain. *Proteins: Structure, Function and Genetics* **2004**, *55*, 689–704.

[76] Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. pH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry* **1990**, *29*, 2403–2408.

[77] Schmidt, H.; Shah, G.; Sperling, L.; Rienstra, C. NMR Determination of Protein pKa Values in the Solid State. *Journal of Physical Chemistry Letters* **2010**, *1*, 1623–1628.

[78] Oksanen, E.; Chen, J. C.; Fisher, S. Z. Neutron Crystallography for the Study of Hydrogen Bonds in Macromolecules. *Molecules* **2017**, *22*, 1–26.

[79] Liptak, M. D.; Shields, G. C. Accurate pK_a Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods. *Journal of the American Chemical Society* **2001**, *123*, 7314–7319.

[80] Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. Absolute pK_a Determinations for Substituted Phenols. *Journal of the American Chemical Society* **2002**, *124*, 6421–6427.

[81] Rebollar-Zepeda, A. M.; Galano, A. First Principles Calculations of pKa Values of Amines in Aqueous Solution: Application to Neurotransmitters. *International Journal of Quantum Chemistry* **2012**, *112*, 3449–3460.

[82] Thapa, B.; Schlegel, H. B. Calculations of pKa's and Redox Potentials of Nucleobases with Explicit Waters and Polarizable Continuum Solvation. *Journal of Physical Chemistry A* **2015**, *119*, 5134–5144.

[83] Casasnovas, R.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. Theoretical pKa Calculations with Continuum Model Solvents, Alternative Protocols to Thermodynamic Cycles. *International Journal of Quantum Chemistry* **2014**, *114*, 1350–1363.

[84] Zhang, J.; Kleinoder, T.; Gasteiger, J. Prediction of pK_a Values for Aliphatic Carboxylic Acids and Alcohols with Empirical Atomic Charge Descriptors. *Journal of Chemical Information and Modeling* **2006**, *46*, 2256–2266.

[85] Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G. New and Original pKa Prediction Method Using Grid Molecular Interaction Fields. *Journal of Chemical Information and Modeling* **2007**, *47*, 2172–2181.

[86] Xing, L. and Glen, R. C., Novel Methods for the Prediction of logP, pKa, and logD. *Journal of Chemical Information and Computer Sciences* **2002**, *42*, 796–805.

[87] Xing, L.; Glen, R. C.; Clark, R. D., Predicting pKa by Molecular Tree Structured Fingerprints and PLS. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 870–879.

[88] Parthasarathi, R.; Padmanabhan, J.; Elango, M.; Chitra, K.; Subra-manian, V.; Chattaraj, P. K. pKa Prediction Using Group Philicity. *Journal of Physical Chemistry A* **2006**, *110*, 6540–6544.

[89] Tao, L.; Han, J.; Tao, F. M. Correlations and Predictions of Carboxylic Acid pKa Values Using Intermolecular Structure and Properties of Hydrogen-Bonded Complexes. *Journal of Physical Chemistry A* **2008**, *112*, 775–782.

[90] Abkowicz-Bienko, A. J. and Latajka, Z., Density Functional Study on Phenol Derivative-Ammonia Complexes in the Gas Phase. *Journal of Physical Chemistry A* **2000**, *104*, 1004–1008.

[91] Caballero-García, G.; Mondragón-Solórzano, G.; Torres-Cadena, R.; Díaz-García, M.; Sandoval-Lira, J.; Barroso-Flores, J. Calculation of Vs,Max and Its Use as a Descriptor for the Theoretical Calculation of pKa Values for Carboxylic Acids. *Molecules* **2019**, *24*.

[92] Grüber, C.; Buß, V. Quantum-Mechanically Calculated Properties for the Development of Quantitative Structure-Activity Relationships (QSAR'S). pKA-Values of Phenols and Aromatic and Aliphatic Carboxylic Acids. *Chemosphere* **1989**, *19*, 1595–1609.

[93] Soriano, E.; Cerdán, S.; Ballesteros, P. Computational Determination of pKa Values. A Comparison of Different Theoretical Approaches and a Novel Procedure. *Journal of Molecular Structure: THEOCHEM* **2004**, *684*, 121–128.

[94] Clarke, F. H.; Cahoon, N. M. Ionization Constants by Curve Fitting: Determination of Partition and Distribution Coefficients of Acids and Bases and Their Ions. *Journal of Pharmaceutical Sciences* **1987**, *76*, 611–620.

[95] Dixon, S. L.; Jurs, P. C. Estimation of pKa for Organic Oxyacids Using Calculated Atomic Charges. *Journal of Computational Chemistry* **1993**, *14*, 1460–1467.

[96] Citra, M. J. Estimating the pKa of Phenols, Carboxylic Acids and Alcohols from Semi-Empirical Quantum Chemical Methods. *Chemosphere* **1999**, *38*, 191–206.

[97] Varekova, R.S., Geidl, S., Ionescu, C.M., Skrehota, O., Kudera, M., Sehnal, D., Bouchal, T., Abagyan, R., Huber, H.J., Koca, J., Predicting pKa Values of Substituted Phenols from Atomic Charges: Comparison of Different Quantum Mechanical Methods and Charge Distribution Schemes. *Journal of Chemical Information and Modeling* **2011**, *51*, 1795–1806.

[98] Ugur, I.; Marion, A.; Parant, S.; Jensen, J. H.; Monard, G. Rationalization of the pK_a Values of Alcohols and Thiols Using Atomic Charge Descriptors and Its Application to the Prediction of Amino Acid pK_a's. *Journal of Chemical Information and Modeling* **2014**, *54*, 2200–2213.

[99] Haynes, W. M., Ed. *CRC Handbook of Chemistry and Physics*, 97th ed.; CRC Press, 2018.

[100] Zhang, S.; Baker, J.; Pulay, P. A Reliable and Efficient First Principles-Based Method for Predicting pK_a Values. 2. Organic Acids. *The Journal of Physical Chemistry A* **2010**, *114*, 432–442.

[101] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988**, *37*, 785–789.

[102] Becke, A. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Physical Review A, General Physics* **1988**, *38*, 3098–3100.

[103] Becke, A. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *The Journal of Chemical Physics* **1993**, *98*, 1372–1377.

[104] Handy, N. C.; Cohen, A. J. Left-Right Correlation Energy. *Molecular Physics* **2001**, *99*, 403–412.

[105] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865–3868.

[106] Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.

[107] Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Accounts of Chemical Research* **2008**, *41*, 157–167.

[108] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B* **2009**, *113*, 6378–6396.

[109] Scalmani, G.; Frisch, M. J. Continuous Surface Charge Polarizable Continuum Models of Solvation. I. General Formalism. *The Journal of Chemical Physics* **2010**, *132*, 114110.

[110] Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *Journal of Physical Chemistry A* **1998**, *102*, 1995–2001.

[111] Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics* **1955**, *23*, 1833–1840.

[112] Löwdin, P. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics* **1950**, *18*, 365–375.

[113] Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *The Journal of Chemical Physics* **1985**, *83*, 735–746.

[114] Fraile, J. M.; Lafuente, G.; Mayoral, J. A.; Pallarés, A. Synthesis and Reactivity of 5-Methylenehydantoins. *Tetrahedron* **2011**, *67*, 8639–8647.

[115] Al-Obaid, A.; El-Subbagh, H.; Khodair, A.; Elmazar, M. 5-Substituted-2-Thiohydantoin Analogs as a Novel Class of Antitumor Agents. *Anti-Cancer Drugs* **1996**, *7*, 873–880.

[116] El-Barbary, A. A.; Khodair, A. I.; Pedersen, E. B.; Nielsen, C. S-Glucosylated Hydantoins as New Antiviral Agents. *Journal of Medicinal Chemistry* **1994**, *37*, 73–77.

[117] Marton, J.; Enisz, J.; Hosztafi, S.; Timar, T. Preparation and Fungicidal Activity of 5-Substituted Hydantoins and Their 2-Thio Analogs. *Journal of Agricultural and Food Chemistry* **1993**, *41*, 148–152.

[118] Marx, J. V.; Richert, D. A.; Westerfeld, W. W. Peripheral Inhibition of Thyroxine by Thiohydantoins Derived from Amino Acids. *Journal of Medicinal Chemistry* **1970**, *13*, 1179–1181.

[119] Tompkins, J. E. 5,5-Diaryl-2-Thiohydantoins and 5,5-Diaryl N3-Substituted 2-Thiohydantoins as Potential Hypolipidemic Agents. *Journal of Medicinal Chemistry* **1986**, *29*, 855–859.

[120] Elwood, J.; Richert, D. A.; Westerfeld, W. A Comparison of Hypolipidemic Drugs in the Prevention of an Orotic Acid Fatty Liver. *Biochemical Pharmacology* **1972**, *21*, 1127–1134.

[121] Takahashi, A.; Matsuoka, H.; Ozawa, Y.; Uda, Y. Antimutagenic Properties of 3,5-Disubstituted 2-Thiohydantoins. *Journal of Agricultural and Food Chemistry* **1998**, *46*, 5037–5042.

[122] Froelich, E.; Fruehan, A.; Jackman, M.; Kirchner, F. K.; Alexander, E. J.; Archer, S. 5-heptyl-2-Thiohydantoin, a New Antitubercular Agent. *Journal of the American Chemical Society* **1954**, *76*, 3099–3100.

[123] Das, H. P.; Mahapatra, G. N. New 2-Thiohydantoins as Potential Pesticides. *Indian Journal of Chemistry. Section B* **1982**, *21*, 162–164.

[124] Pardasani, R. T.; Pardasani, P.; Ojha, C. K.; Sherry, D.; Chaturvedi, V.; Sharma, I. Syntheses of Indigoid Dye Precursors and Bioactive Compounds Via Condensation of 1,2- and 1,4-Diones with Thiohydantoins. *Phosphorus, Sulfur, and Silicon and the Related Elements* **2002**, *177*, 2435–2443.

[125] Curran, A. C. W. U.S. Patent 3 984 430, 1976.

[126] Mo, B.; Li, J.; Liang, S. Chemical Carboxy-Terminal Sequence Analysis of Peptides Using Acetyl Isothiocyanate. *Analytical Biochemistry* **1997**, *252*, 169–176.

[127] Kandil, S. S.; El-Hefnawy, G. B.; Baker, E. A. Thermal and Spectral Studies of 5-(Phenylazo)-2-Thiohydantoin and 5-(2- Hydroxyphenylazo)-2-Thiohydantoin Complexes of Cobalt(II), Nickel(II) and Copper(II). *Thermochimica Acta* **2004**, *414*, 105–113.

[128] Ichitani, M.; Kitoh, S.; Tanaka, K.; Fujinami, S.; Suda, M.; Honda, M.; Kuwae, A.; Hanai, K.; Kunimoto, K. Synthesis and Crystal Structure of (S)-5-Isopropyl-5-Methyl-2-Thiohydantoin. *European Journal of Chemistry* **2013**, *4*, 350–352.

[129] Metwally, M. A.; Abdel-Latif, E. Thiohydantoins: Synthetic Strategies and Chemical Reactions. *Journal of Sulfur Chemistry* **2012**, *33*, 229–257.

[130] Kawabata, T.; Kawakami, S.; Fuji, K. Enantioselective α-allylation of a phenylalanine derivative under the control of aggregation of a chiral nonracemic enolate. *Tetrahedron Letters* **2002**, *43*, 1465–1467.

[131] Aydeniz, Y.; Oğuz, F.; Yaman, A.; Konuklar, A. S.; Doğan, I.; Aviyente, V.; Klein, R. A. Barriers to Internal Rotation Around the C–N Bond in 3-(o-Aryl)-5-Methyl-Rhodanines Using NMR Spectroscopy and Computational Studies. Electron Density Topological Analysis of the Transition States. *Organic and Biomolecular Chemistry* **2004**, *2*, 2426–2436.

[132] Yildirim, A.; Konuklar, F. A. S.; Catak, S.; Van Speybroeck, V.; Waroquier, M.; Dogan, I.; Aviyente, V. Solvent-Catalyzed Ring-Chain-Ring Tautomerization in Axially Chiral Compounds. *Chemistry - A European Journal* **2012**, *18*, 12725–12732.

[133] Yilmaz, E. M.; Dogan, I. Axially Chiral N-(o-Aryl)-2-Thioxo-oxazolidine-4-one and Rhodanine Derivatives: Enantiomeric Separation and Determination of Racemization Barriers. *Tetrahedron: Asymmetry* **2008**, *19*, 2184–2191.

[134] Belot, V.; Farran, D.; Jean, M.; Albalat, M.; Vanthuyne, N.; Roussel, C. Steric Scale of Common Substituents from Rotational Barriers of N-(o-Substituted aryl)thiazoline-2-thione Atropisomers. *The Journal of Organic Chemistry* **2017**, *82*, 10188–10200.

[135] Bonne, D.; Rodriguez, J. Enantioselective Syntheses of Atropisomers Featuring a Five-Membered Ring. *Chemical Communications* **2017**, *53*, 12385–12393.

[136] Bonne, D.; Rodriguez, J. A Bird's Eye View of Atropisomers Featuring a Five-Membered Ring. *European Journal of Organic Chemistry* **2018**, *2018*, 2417–2431.

[137] Bringmann, G.; Price Mortimer, A. J.; Keller, P. A.; Gresser, M. J.; Garner, J.; Breuning, M. Atroposelective Synthesis of Axially Chiral Biaryl Compounds. *Angewandte Chemie International Edition* **2005**, *44*, 5384–5427.

[138] Erol, S.; Dogan, I. Axially Chiral 2-Arylimino-3-aryl-thiazolidine-4-one Derivatives: Enantiomeric Separation and Determination of Racemization Barriers by Chiral HPLC. *The Journal of Organic Chemistry* **2007**, *72*, 2494–2500.

[139] Kumarasamy, E.; Raghunathan, R.; Sibi, M. P.; Sivaguru, J. Nonbiaryl and Heterobiaryl Atropisomers: Molecular Templates with Promise for Atropselective Chemical Transformations. *Chemical Reviews* **2015**, *115*, 11239–11300.

[140] Renzi, P. Organocatalytic Synthesis of Axially Chiral Atropisomers. *Organic & Biomolecular Chemistry* **2017**, *15*, 4506–4516.

[141] Jacob, S. D.; Brooks, J. L.; Frontier, A. J. No Acid Required: 4π and 6π Electrocyclization Reactions of Dienyl Diketones for the Synthesis of Cyclopentenones and 2H-Pyrans. *The Journal of Organic Chemistry* **2014**, *79*, 10296–10302.

[142] Pem, B.; Vrček, V. Substituent Effects on the Stability of 1,4-Benzodiazepin-2-one Tautomers: A Density Functional Study. *International Journal of Quantum Chemistry* **2018**, *118*, e25523.

[143] Tian, C.; Xiu, P.; Meng, Y.; Zhao, W.; Wang, Z.; Zhou, R. Enantiomerization Mechanism of Thalidomide and the Role of Water and Hydroxide Ions. *Chemistry - A European Journal* **2012**, *18*, 14305–14313.

[144] Kawabata, T.; Majumdar, S.; Tsubaki, K.; Monguchi, D. Memory of Chirality in Intramolecular Conjugate Addition of Enolates: A Novel Access to Nitrogen Heterocycles with Contiguous Quaternary and Tertiary Stereocenters. *Organic & Biomolecular Chemistry* **2005**, *3*, 1609.

[145] Pople, J. A.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. *The Journal of Chemical Physics* **1965**, *43*, S136–S151.

[146] Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *The Journal of Chemical Physics* **1967**, *47*, 2026–2033.

[147] Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *The Journal of Chemical Physics* **1965**, *43*, S129–S135.

[148] Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *Journal of the American Chemical Society* **1977**, *99*, 4899–4907.

[149] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *Journal of the American Chemical Society* **1985**, *107*, 3902–3909.

[150] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *Journal of Computational Chemistry* **1989**, *10*, 209–220.

[151] Ceperley, D. M.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **1980**, *45*, 566–569.

[152] Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Eelectron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Canadian Journal of Physics* **1980**, *58*, 1200–1211.

[153] Perdew, J. P.; Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Physical Review B* **1981**, *23*, 5048–5079.

[154] Ortiz, G.; Ballone, P. Correlation Energy, Structure Factor, Radial Distribution Function, and Momentum Distribution of the Spin-Polarized Uniform Electron Gas. *Physical Review B* **1994**, *50*, 1391–1405.

[155] Cramer, C. J. *Essentials of Computational Chemistry : Theories and Models*; Wiley, 2004; p 596.

[156] Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chemical Physics Letters* **1989**, *157*, 200–206.

[157] Guidon, M.; Hutter, J.; VandeVondele, J. Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations. *Journal of Chemical Theory and Computation* **2010**, *6*, 2348–2364.

[158] Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.

[159] Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *The Journal of Chemical Physics* **2006**, *125*, 194101.

[160] Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **1971**, *54*, 724–728.

[161] Leach, A. R. *Molecular Modelling : Principles and Applications*; Prentice Hall, 2001; p 744.

[162] Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoretica Chimica Acta* **1977**, *44*, 129–138.

[163] Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *Journal of Chemical Theory and Computation* **2012**, *8*, 527–541.

[164] Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilizaion of AB Initio Molecular Potentials for the Prevision of Solvent Effects. *Chemical Physics* **1981**, *55*, 117–129.

[165] Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *Journal of Pjysical Chemistry A* **1998**, *102*, 1995–2001.

[166] Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications. *The Journal of Physical Chemistry B* **1997**, *101*, 10506–10517.

[167] Mennucci, B.; Tomasi, J. Continuum Solvation Models: A New Approach to the Problem of Solute's Charge Distribution and Cavity Boundaries. *The Journal of Chemical Physics* **1997**, *106*, 5151–5158.

[168] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.

[169] Verlet, L. Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967**, *159*, 98–103.

[170] Hockney,; W., R. The Potential Calculation and Some Applications. *Methods in Computational Physics* **1970**, *9*, 136–211.

[171] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76*, 637–649.

[172] Cerutti, D. S.; Case, D. A. Multi-Level Ewald: A Hybrid Multigrid / Fast Fourier Transform Approach to the Electrostatic Particle-Mesh Problem. *Journal of Chemical Theory and Computation* **2010**, *6*, 443–458.

[173] Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.

[174] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

[175] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 15–26.

[176] Ten Eyck, L. F.; Mandell, J.; Roberts, V. A.; Pique, M. E. Surveying Molecular Interactions with DOT. Proceedings of the 1995 ACM/IEEE conference on Supercomputing (CDROM) - Supercomputing '95. New York, USA, 1995; pp 22–es.

[177] Vakser, I. A. Evaluation of GRAMM Low-Resolution Docking Methodology on the Hemagglutinin-Antibody Complex. *Proteins* **1997**, *Suppl 1*, 226–230.

[178] Chen, R.; Weng, Z. Docking Unbound Proteins Using Shape Complementarity, Desolvation, and Electrostatics. *Proteins* **2002**, *47*, 281–294.

[179] Hart, T. N.; Read, R. J. A Multiple-Start Monte Carlo Docking Method. *Proteins: Structure, Function, and Genetics* **1992**, *13*, 206–222.

[180] Goodsell, D. S.; Lauble, H.; Stout, C. D.; Olson, A. J. Automated Docking in Crystallography: Analysis of the Substrates of Aconitase. *Proteins: Structure, Function, and Genetics* **1993**, *17*, 1–10.

[181] Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible Ligand Docking Using a Genetic Algorithm. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 113–130.

[182] Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology* **1997**, *267*, 727–748.

[183] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Journal of Computational Chemistry* **1998**, *19*, 1639–1662.

[184] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery* **2004**, *3*, 935–949.

[185] Sotriffer, C. *Virtual Screening : Principles, Challenges, and Practical Guidelines*; Wiley-VCH, 2011; p 519.

[186] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated Docking with Grid-Based Energy Evaluation. *Journal of Computational Chemistry* **1992**, *13*, 505–524.

[187] Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. *Journal of Medicinal Chemistry* **2003**, *46*, 499–511.

[188] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *Journal of Molecular Biology* **1996**, *261*, 470–489.

[189] Muegge, I. PMF Scoring Revisited. *Journal of Medicinal Chemistry* **2006**, *49*, 5895–5902.

[190] Velec, H. F. G.; Gohlke, H.; Klebe, G. DrugScore (CSD)-Knowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction. *Journal of Medicinal Chemistry* **2005**, *48*, 6296–6303.

[191] Shao, Y.; Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S.; et al., A. G. Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

[192] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871.

[193] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, A1133–A1138.

[194] Orio, M.; Pantazis, D. A.; Neese, F. Density Functional Theory. *Photosynthesis Research* **2009**, *102*, 443–453.

[195] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; et al., G. S. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT, 2009.

[196] Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *The Journal of Chemical Physics* **1997**, *107*, 3032–3041.

[197] Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *Journal of the American Chemical Society* **1983**, *105*, 7512–7516.

[198] Buhl, M.; Kaupp, M.; Malkina, O. L.; Malkin, V. G. The DFT Route to NMR Chemical Shifts. *Journal of Computational Chemistry* **1999**, *20*, 91–105.

[199] Ditchfield, R. Self-Consistent Perturbation Theory of Diamagnetism. *Molecular Physics* **1974**, *27*, 789–807.

[200] Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations. *Journal of the American Chemical Society* **1990**, *112*, 8251–8260.

[201] Miranda, C.; Escarti, F.; Lamarque, L.; J. R. Yunta, M.; Navarro, P.; Garcia-España, E.; Luisa, J. M. New 1H-Pyrazole-Containing Polyamine Receptors Able To Complex l-Glutamate in Water at Physiological pH Values. *Journal of the American Chemical Society* **2003**, *126*, 823–833.

[202] Bischoff, R.; Schlüter, H. Amino acids: Chemistry, Functionality and Selected Non-Enzymatic Post-Translational Modifications. *Journal of Proteomics* **2012**, *75*, 2275–2296.

[203] Moretti, L.; Pentikäinen, O. T.; Settimo, L.; Johnson, M. S. Model Structures of the N-Methyl-D-Aspartate Receptor Subunit NR1 Explain the Molecular Recognition of Agonist and Antagonist Ligands. *Journal of Structural Biology* **2004**, *145*, 205–215.

[204] Rios, A.; Amyes, T. L.; Richard, J. P. Formation and Stability of Organic Zwitterions in Aqueous Solution: Enolates of the Amino Acid Glycine and Its Derivatives. *Journal of American Chemical Society* **2000**, *122*, 9373–9385.

[205] Agarwal, A.; Taylor, E. W. 3-D QSAR for Intrinsic Activity of 5-HT1A Receptor Ligands by the Method of Comparative Molecular Field Analysis. *Journal of Computational Chemistry* **1993**, *14*, 237–245.

[206] Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34–40.

[207] Schapira, M.; Abagyan, R.; Totrov, M. Nuclear Hormone Receptor Targeted Virtual Screening. *Journal of Medicinal Chemistry* **2003**, *46*, 3045–3059.

[208] Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A Novel Method for the Shape-Directed Rapid Docking of Ligands to Protein Active Sites. *Journal of Molecular Graphics & Modelling* **2003**, *21*, 289–307.

[209] McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian Docking Functions. *Biopolymers* **2003**, *68*, 76–90.

[210] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47*, 1739–1749.

[211] Trott, O.; Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* **2010**, *31*, 455–461.

[212] Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in Docking Success Rates due to Dataset Preparation. *Journal of Computer-Aided Molecular Design* **2012**, *26*, 775–786.

[213] Zhao, H.; Caflisch, A. Discovery of ZAP70 Inhibitors by High-Throughput Docking into a Conformation of its Kinase Domain Generated by Molecular Dynamics. *Bioorganic & Medicinal Chemistry Letters* **2013**, *23*, 5721–5726.

[214] Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Computational Biology* **2014**, *10*.

[215] Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *Journal of Computational Chemistry* **2015**, *36*, 1132–1156.

[216] Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **2009**, *30*, 2785–2791.

[217] Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cruzeiro, T. C. V.; Darden, T.; et al., R. D. AMBER 2018. University of California, San Fransisco, 2018.

[218] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

[219] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

[220] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[221] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23*, 327–341.

[222] Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chemical Physics Letters* **1995**, *246*, 122–129.

[223] Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *Journal of Physical Chemistry* **1996**, *100*, 19824–19839.

[224] Tsui, V.; Case, D. A. Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275–291.

[225] 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.

[226] Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *Journal of Chemical Theory and Computation* **2007**, *3*, 156–169.

[227] Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *Journal of Chemical Theory and Computation* **2013**, *9*, 2020–2034.

[228] Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *Journal of Computational Chemistry* **1999**, *20*, 217–230.

[229] Zhao, Y.; Truhlar, D. G., Density Functionals with Broad Applicability in Chemistry. *Accounts of Chemical Research* **2008**, *41*, 157–167.

[230] Singh, U. C.; Kollman, P. A. An approach to Computing Electrostatic Charges for Molecules. *Journal of Computational Chemistry* **1984**, *5*, 129–145.

[231] Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *Journal of Computational Chemistry* **1990**, *11*, 361–373.

[232] Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A Summary of the Measured pKa Values of the Ionizable Groups in Folded Proteins. *Protein Science* **2008**, *18*, 247–251.

[233] Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pKa Values of the Ionizable Groups of Proteins. *Protein Science* **2006**, *15*, 1214–1218.

[234] Leeson, P. D.; Baker, R.; Carling, R. W.; Kulagowski, J. J.; Mawer, I. M.; Ridgill, M. P.; Rowley, M.; Smith, J. D.; Stansfield, I.; Stevenson, G. I.; Foster, A. C.; Kemp, J. A. Amino Acid Bioisosteres: Design of 2-Quinolone Derivatives as Glycine-Site N-Methyl-D-Aspartate Receptor Antagonists. *Bioorganic & Medicinal Chemistry Letters* **1993**, *3*, 299–304.

[235] Brown, W. H. *Organic chemistry*; Brooks/Cole Cengage Learning, 2009.

[236] Liu, L.-P.; Malhotra, D.; Paton, R. S.; Houk, K. N.; Hammond, G. B. The [4+2], not [2+2], Mechanism Occurs in the Gold-Catalyzed Intramolecular Oxygen Transfer Reaction of 2-Alkynyl-1,5-Diketones. *Angewandte Chemie International Edition* **2010**, *49*, 9132–9135.

[237] Cantillo, D.; Kappe, C. O. A Unified Mechanistic View on the Morita-Baylis-Hillman Reaction: Computational and Experimental Investigations. *The Journal of Organic Chemistry* **2010**, *75*, 8615–8626.

[238] Catak, S.; Hemelsoet, K.; Hermosilla, L.; Waroquier, M.; Van Speybroeck, V. Competitive Reactions of Organophosphorus Radicals on Coke Surfaces. *Chemistry - A European Journal* **2011**, *17*, 12027–12036.

[239] Winne, J. M.; Catak, S.; Waroquier, M.; Van Speybroeck, V. Scope and Mechanism of the (4+3) Cycloaddition Reaction of Furfuryl Cations. *Angewandte Chemie International Edition* **2011**, *50*, 11990–11993.

[240] Pluempanupat, W.; Abraham, M.; Brecker, L.; Wolschann, P.; Karpfen, A.; Arion, V. B.; Widhalm, M. Synthesis and Conformation of Chiral Biheteroaryls. *The Journal of Organic Chemistry* **2011**, *76*, 3222–3230.

[241] Degennaro, L.; Mansueto, R.; Carenza, E.; Rizzi, R.; Florio, S.; Pratt, L. M.; Luisi, R. Nitrogen Dynamics and Reactivity of Chiral Aziridines: Generation of Configurationally Stable Aziridinyllithium Compounds. *Chemistry - A European Journal* **2011**, *17*, 4992–5003.

[242] Mollet, K.; Catak, S.; Waroquier, M.; Van Speybroeck, V.; D'hooghe, M.; De Kimpe, N. Stereoselective Synthesis of cis -3,4-Disubstituted Piperidines through Ring Transformation of 2-(2-Mesyloxyethyl)azetidines. *The Journal of Organic Chemistry* **2011**, *76*, 8364–8375.

[243] Fukui, K. The Path of Chemical Reactions - the IRC Approach. *Accounts of Chemical Research* **1981**, *14*, 363–368.

[244] Cancès, E.; Mennucci, B. Comment on Reaction Field Treatment of Charge Penetration. *The Journal of Chemical Physics* **2001**, *114*, 4744.

[245] Kawabata, T.; Yahiro, K.; Fuji, K. Memory of Chirality: Enantioselective Alkylation Reactions at an Asymmetric Carbon Adjacent to a Carbonyl Group. *Journal of the American Chemical Society* **1991**, *113*, 9694–9696.

[246] Zhang, D.; Xing, X.; Cuny, G. D. Synthesis of Hydantoins from Enantiomerically Pure α-Amino Amides without Epimerization. *Journal of Organic Chemistry* **2006**, *71*, 1750–1753.

[247] Baldwin, J. E. Rules for Ring Closure. *Journal of the Chemical Society, Chemical Communications* **1976**, 734.

[248] Baldwin, J. E.; Cutting, J.; Dupont, W.; Kruse, L.; Silberman, L.; Thomas, R. C. 5-Endo-Trigonal Reactions: A Disfavoured Ring Closure. *Journal of the Chemical Society, Chemical Communications* **1976**, 736.

[249] Baldwin, J. E.; Thomas, R. C.; Kruse, L. I.; Silberman, L. Rules for Ring Closure: Ring Formation by Conjugate Addition of Oxygen Nucleophiles. *The Journal of Organic Chemistry* **1977**, *42*, 3846–3852.

[250] Chatgilialoglu, C.; Ferreri, C.; Guerra, M.; Timokhin, V.; Froudakis, G.; Gimisis, T. 5-Endo-trig Radical Cyclizations: Disfavored or Favored Processes? *Journal of American Chemical Society* **2002**, *124*, 10765–10772.

[251] Zhao, Y.; Truhlar, D. G. Density Functional Theory for Reaction Energies: Test of Meta and Hybrid Meta Functionals, Range-Separated Functionals, and Other High-Performance Functionals. *Journal of Chemical Theory and Computation* **2011**, *7*, 669–676.

Résumé

L'un des problèmes importants dans la conception de médicaments est l'identification de l'activité biologique des ligands en regard de leurs récepteurs. Le développement, la synthèse et les mesures d'activité des ligands revêtent une importance majeure dans la conception de médicaments. Cependant, les études expérimentales ne peuvent qu'être limitées: la synthèse de toutes les molécules potentiellement actives est par exemple irréaliste. Les études numériques pourraient donc apporter une aide précieuse aux études expérimentales, notamment dans le cas de la conception de ligands pour les récepteurs au glutamate comme le récepteur NMDA. En combinant les points forts des approches par dynamique moléculaire et par chimie quantique, il est possible d'établir une inspection, une caractérisation et une rationalisation plus ciblées des études de conception de médicaments. Dans cette thèse, des méthodes numériques ont été utilisées pour étudier les propriétés intrinsèques des molécules biologiquement actives responsables de la sélectivité. Les résultats de cette étude sont présentés en 4 chapitres. Dans les deux premiers chapitres, nous avons cherché à différencier les agonistes, les antagonistes et les agonistes partiels au récepteur NMDA en fonction de descripteurs tirés de la chimie quantique et d'énergies libres de liaison de Gibbs. Plusieurs propriétés moléculaires qui pourraient jouer un rôle dans la liaison du ligand à la sous-unité glycine GluN1 du récepteur NMDA ainsi que les énergies libres de liaison ont été utilisées pour établir un lien entre les efficacités et les affinités de liaison des ligands. La prédiction des constantes de dissociation acide d'acides aminés dans les protéines et les ligands nous permet d'avoir des informations sur l'affinité de liaison et l'efficacité des ligands envers leurs protéines cibles. Considérant l'importance des pK_{a}, nous avons exploré dans le chapitre suivant comment les charges atomiques des acides carboxyliques peuvent être liées à la prédiction de pK_{a} de ligands. Afin de mettre en lumière les origines de la stéréosélectivité de ligands biologiquement actifs, plusieurs voies mécanistiques ont été évalués dans le dernier chapitre pour les 2-thiohydantoïnes, qui sont de puissants antagonistes des récepteurs aux androgènes.

Mots-clés: Docking; Dynamique Moléculaire; DFT; pKa; NMDA

Abstract

One of the important issues in drug design is the identification of the biological activity of receptor ligands. Development, synthesis and activity measurements of ligands have a major importance in drug design. However, there are certain limits in experimental studies; synthesis of a large number of compounds to cover all the potentially active molecules is unrealistic. Computational studies could therefore provide a valuable aid to experimental studies on ligand design for glutamate receptors. By combining the strengths of Molecular Dynamics and Quantum Chemical approaches, a more focused inspection, characterisation and rationalization of the drug design studies is allowed to be established. In this dissertation, computational methods have been used to investigate the intrinsic properties of the biologically active molecules that cause the selectivity. The results of this study will be introduced in 4 chapters. In Chapters 4 and 5, we aimed to differentiate between agonists, antagonists and partial agonists based on Quantum Chemical descriptors and binding Gibbs free energies. Several molecular properties that could play a role in ligand binding to the glycine GluN1 subunit of NMDA and calculated binding Gibbs free energies were further used to provide a link between the efficacies and binding affinities of the ligands. Prediction of the acid dissociation constants of amino acids in proteins and ligands allows us to have information about the binding affinity and efficacy of the ligand to its target protein. Considering the significance of pK_a's, how atomic charges of carboxylic acids can be related to the prediction of pK_a of the ligands have been explored in Chapter 6. In order to shed light on the origins of the stereoselectivity of biologically active ligands, several mechanistic pathways have been evaluated for 2-thiohydantoins which are potent androgen receptor antagonists and the results are given in Chapter 7.

Keywords: Docking; Molecular Dynamics; DFT; pKa; NMDA