Universidade do Algarve

Departamento de Ciências Biomédicas e Medicina

Interaction between Isothiocyanates and Cytochromes P450: a Computational Docking Study

André Filipe da Conceição Marçal Alves Nunes

Master Thesis in Oncobiology – Molecular Mechanisms of Cancer

Supervisor: PhD Professor Paulo Martel University of Algarve

26 February 2016

Universidade do Algarve

Departamento de Ciências Biomédicas e Medicina

Interaction between Isothiocyanates and Cytochromes P450: a Computational Docking Study

André Filipe da Conceição Marçal Alves Nunes

Thesis submitted to the Universidade do Algarve in partial fulfilment of the requirements for the Master Degree in Oncobiology – Molecular Mechanisms of Cancer under the orientation of PhD Professor Paulo Martel (Universidade do Algarve).

26 February 2016

Interaction between Isothiocyanates and Cytochromes P450: a Computational Docking Study

Declaração de Autoria de Trabalho

" Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída."

André Filipe da Conceição Marçal Alves Nunes

Ass.:_____

Direitos de Cópia ou Copyright

Agradecimentos

Em primeiro lugar, gostaria de agradecer à minha família (Pai, Mãe, David e Avós) por todo o apoio e encorragamento que tenho recebido e desfrutado ao longo da minha vida. Sem vós e a vossa preciosa ajuda, chegar até onde cheguei hoje não teria sido possível. Agradeço, muito sinceramente do fundo do meu coração, por todo o carinho e apreço que tem tido por mim, pelo meu empenho e força de vontade em superar qualquer desafio. Espero continuar a surpreende-los por muitos e bons anos e que estejamos cá todos para o ver. Os meus maiores cumprimentos, beijinhos e abraços a todos vós pelo vosso amor incondicional. Muito obrigado.

Em segundo lugar, gostaria de agradecer ao professor e orientador desta minha tese de mestrado, o senhor professor Paulo Martel. Muito obrigado por me ter recebido e orientado neste meu percurso académico e por ter estado sempre disponível e prestável para esclarecer qualquer dúvida que eu pudesse ter. Muito obrigado.

Index

1. Abstract

Cancer is the second leading cause of death worldwide. Environmental and lifestyle factors play a crucial role in its development. Certain components of the diet, namely Isothiocyanates (ITCs), which are present in cruciferous vegetables, can act as cancer chemoprevention agents, which may reduce cancer development risk up to 50%.

ITCs are thought to act through inhibition of Phase I biotransformation enzymes, cytochromes P450 (CYPs), which are responsible for the bioactivation of many pre-carcinogenic compounds in the body. In this study we used computational (*in silico*) docking methods to investigate the interaction between ITCs and several important members of the CYP family who are known to be inhibited by this group of chemopreventive agents. Experimental inhibition constants (K_i) were compared with the estimates produce with the *in silico* docking algorithms.

In silico docking was performed using several ITCs (BITC, PEITC, PHITC, PPITC) as ligands and all the CYP files available in the PDB website for the following receptor molecules: CYP2A6, 2A13, 2B6, and 2C9. In order to investigate how different docking algorithms and scoring functions would affect the docking results, the softwares Autodock Vina (Vina) and AutoDock4 (AD4) were both used and compared to other results found in the literature. The relative performance of AD4 and Vina was assessed by *in silico* redocking the crystallographic ligands of the studied CYPs and comparing the results with the experimentally observed conformations

The best docked ligand structure found by both programs is, for most cases, very similar in position and conformation to the crystallographic structure, with Vina producing K_i values in closest agreement with experimental data. In terms of both rigid and flexible docking, Vina produced better results with more accurate K_i values when compared with AD4. Vina is better at flexible while AD4 is better at rigid docking. The docking of a ligand to its crystallographic partner as compared with docking to a different structure of the same CYP generally yields the second best K_i values.

ITCs were found to be good inhibitors of CYP2A6 and PHITC is a very good inhibitor of CYP2A13. High K_i values corresponding to weak inhibition were observed for CYP2B6, whereas very poor inhibition was observed for CYP2C9.

Keywords: Isothiocyanates; Cytochrome P450s; Docking; Inhibition

2. Resumo

O cancro é considerado a segunda causa de morte em todo o mundo. Os factores ambientais e o estilo de vida desempenham um papel crucial no desenvolvimento desta doença. Muitos estudos epidemiológicos mostram que uma dieta rica em frutas e vegetais é capaz de levar a uma diminuição do risco de se desenvolver cancro até 50%. Um dos elementos de uma dieta saudável que tem um papel especialmente relevante na prevenção do cancro são os vegetais crucíferos. Estes contêm compostos chamados glucosinolatos. Quando os glucosinolatos entram em contacto com a enzima mirosinase há a produção maioritária de isotiocianatos.

Os glucosinolatos conseguem prevenir o aparecimento e desenvolvimento de cancro através da regulação de vários mecanismos celulares, sendo de especial relevância a regulação do metabolismo de compostos xenobióticos pelas enzimas de biotransformação. Os isotiocianatos conseguem activar as enzimas de Fase II (destoxificação – conjugação) e inibir as enzimas de Fase I (destoxificação e bioactivação de carcinogénios – oxidação). Os citocromos P450 (CYPs) são os enzimas responsáveis pela maioria (70-80%) do metabolismo da Fase I. Estas versáteis monooxigenases são capazes de hidroxilar muitos compostos e estão presentes em maior abundância no retículo endoplasmático liso dos hepatócitos. As principais famílias de CYPs responsáveis pela carcinogénese são também aquelas que estão envolvidas no metabolismo de compostos exógenos (CYP1-2).

Os ITCs conseguem inibir, de forma reversível ou irreversível, a actividade catalítica de alguns CYPs, ligando-se ao centro activo destes enzimas.

O tempo de vida dos ITCs no organismo está dependente da actividade dos enzimas de Fase II. Estes são activados pelos ITCs, levando à sua posterior degradação pela via do ácido mercaptúrico e posterior excreção por via urinária (principal via) ou fecal.

Uma miríade de ITCs de origem natural e sintética são capazes de inibir a actividade dos CYPs. Após uma revisão da literatura disponível, foram escolhidos 4 ITCs [benzil-ITC (BITC); fenetil-ITC (PEITC); 6-fenilhexil-ITC (PHITC); 3-fenilpropil-ITC (PPITC)] e 4 CYPs (2A6; 2A13; 2B6; 2C9) para serem objecto de estudo nesta tese de mestrado.

A modelação molecular tem vindo a ganhar um maior destaque nos últimos anos devido à maior velocidade de processamento de dados por computadores pessoais. Neste tipo de estudos é necessário considerar diversos factores, tais como o tamanho e as propriedades electrónicas das moléculas. Várias abordagens teóricas e algoritmos de modelação podem ser usados quando se faz acoplamento (*docking*) molecular.

O método de *ab initio* tenta resolver a equação de Schrödinger através do uso da aproximação de Born–Oppenheimer assumindo o movimento dos núcleos e desprezando o movimento dos electrões. Por sua vez, o método de mecânica molecular tira partido das forças existentes entre as ligações atómicas numa molécula. Finalmente, o método de dinâmica molecular tenta resolver as equações Newton aplicadas ao movimento de moléculas.

Para os cálculos de *docking* molecular deste trabalho, utilizaram-se dois programas: AutoDock Vina (Vina) e AutoDock 4 (AD4). Em ambos os caos é usada uma função de pontuação (*scoring*) e um algoritmo de busca, de modo a tentar encontrat a conformação ideal de acoplamento de ligando e receptor. No entanto, os dois programas diferem na forma matemática da função de scoring e no mecanismo do algoritmo de pesquisa.

Os ITCs foram construídos utilizando o programa HyperChem, o qual providencia as ferramentas necessárias à criação de modelos moleculares num ambiente tridimensional. O processo de construção das moléculas envolve cálculos de mecânica quântica (*ab initio*) bem como de mecânica molecular (utilizando o campo de forças MM+), podendo os modelos gerados ser usados em diversos tipo de simulações moleculares. Este tipo de métodos permite prever o modo como a estrutura das moléculas simuladas responde a diversos tipos de perturbações.

Foi realizado o download dos ficheiros pdb dos CYPs de origem humana a partir do website do Protein Data Bank (PDB), o qual consiste numa grande base de dados para estruturas cristalográficas de macromoléculas biológicas e dos seus respectivos ligandos (quando presentes). A maioria destas estruturas foi obtida experimentalmente por cristalografia de raio-X, e em menor número por espectroscopia de Ressonância Magnética Nuclear (NMR). A primeira técnica de determinação de estrutura molecular baseia-se no uso de raios X que são incididos sobre um cristal da molécula cuja estrutura se pretende determinar levando à produção de padrões de difracção que podem ser usados

para resolver a estrutura molecular. A segunda técnica baseia-se na exposição da molécula em estudo a um campo magnético extremamente forte, o qual vai produzir a reorientação do spin nuclear de alguns tipos de átomos presentes na amostra, possibilitando assim a sua determinação estrutural.

O acoplamento molecular dos CYPs com os respectivos ligandos nativos bem como com os ITCs foi realizado através do uso do conjunto de programas AutoDock Tools (ADT), AutoGrid4/AutoDock4 (AD4) e AutoDock Vina (Vina). É através do emprego destes programas que se procurou encontrar a melhor energia de interação possível entre ligando e receptor. Esta tarefa é realizada através do uso de funções de *scoring*, as quais fornecem uma previsão da afinidade de ligação entre ligando e receptor, bem como da orientação e conformação relativa das duas moléculas. Aminimização de scoring é usada como objectivo na busca sistemática pela melhor conformação do ligando (flexível) no receptor (rígido ou flexível).

Os ficheiros pdb contendo as estruturas das moléculas em estudo têm de ser manualmente ajustados de modo a estarem aptos a serem utilizados no processo de *docking* molecular. As moléculas de água, os ligandos nativos e as cadeias suplementares da macromolécula são removidos através do programa PyMOL. No ADT é adicionado carga e resíduos flexíveis aos ficheiros do receptor e ligando, os quais são posteriormente gravados em formato pdbqt. A grelha de busca também é definida neste programa. De seguida correm-se os programas de *docking* molecular (AD4 e Vina) separadamente nos seus modos de *docking* rígido e flexível. O uso simultâneo dos dois programas e a comparação dos resultados obtidos fornecem uma forma de validação dos cálculos. Como forma adicional de validação dos cálculos, procedeu-se ao re-*docking* dos pares cristalográficos CYP-ligando nativo de forma a comparar os resultados com os dados obtidos experimentalmente.

Uma vez que o programa Vina é mais recente que o AD4, a necessidade de ter uma comparação mais equitativa levou a que se procedesse à alteração das cargas do grupo hémicos dos CYPs a acoplar pelo AD4 por cargas hémicas mais precisas, as quais foram fornecidas pelo trabalho de Shahrokh, K. et al.; 2012.

Uma revisão da literatura referente à inibição de CYPs por ITCs foi tida em conta por forma a se proceder à escolha dos CYPs a serem alvo de acoplamento molecular bem

como ter uma forma de comparar os resultados obtidos *in silico* com os obtidos experimentalmente.

Os valores experimentais de constante de inibição [Ki (μM)] de alguns ligandos foram usados como forma de validação dos resultados das simulações de *docking*. Enquanto o software AD4 fornece estimativas directas de Ki (sendo apenas necessária uma conversão de mM para μM), os valores de constantes de afinidade dados pelo software Vina têm de ser convertidos em valores de Ki.

Ambos os programas fornecem informação relativamente à raíz quadrada do desvio médio (RMSD) dos ligandos. O RMSD corresponde ao desvio médio da estrutura do ligando após o acoplamento face à sua posição cristalográfica original. Os valores de RMSD obtidos pelo programa AD4 foram os únicos utilizados uma vez que o Vina calcula o RMSD usando como referência não a estrutura original do mesmo mas sim melhor estrutura acoplada do ligando.

A análise dos resultados obtidos mostrou que as melhores estruturas acopladas por ambos os programas eram na sua grande maioria muito semelhantes às estruturas cristalográficas correspondentes. Também se constatou que a sua localização bem como orientação tridimensional eram as mais semelhantes com a estrutura cristalográfica. No entanto, o programa Vina foi capaz de gerar melhores resultados de Ki em relação aos encontrados na literatura quando comparado com os obtidos pelo AD4.

Relativamente aos modos de acoplamento rígido e flexível, verificou-se que o programa Vina conseguiu gerar os melhores resultados nas duas situações. No entanto, quando a análise é feita em termos de programa, verifica-se que o Vina é melhor em acoplamento flexível enquanto que o AD4 é melhor em acoplamento rígido.

Visto que existem várias estruturas cristalográficas do mesmo CYP, importa responder se o acoplamento de um ligando com o seu parceiro cristalográfico é capaz de gerar um melhor resultado do que com uma estrutura diferente mas do mesmo CYP. Em ambos os programas, alguns ligandos conseguem de facto ter um melhor valor de Ki quando acoplados com o seu parceiro cristalográfico. No entanto, maioritariamente isto só ocorre para o segundo melhor resultado. Apesar de o receptor sofrer algum ajuste ao acomodar o ligando, é possível que outros ligandos de estrutura parecida ocupar com alta afinidade o espaço tridimensional deixado vago pelo ligando original.

Uma vez que existem diferentes estruturas cristalográficas do mesmo par CYP-ligando, procedeu-se ao acoplamento do mesmo tipo de ligando mas com as outras estruturas referentes ao mesmo par de modo a ver se o programa AD4 conseguia descobrir alguma diferença. Tal não foi verificado, não havendo nenhuma diferença tanto ao nível do modo de acoplamento como de valores de Ki ou RMSD. Tal acontecimento deve-se possivelmente ao facto de que o processo de cristalização ter sido refinado até ao ponto em que certas variáveis, como quem ou quando os complexos foram cristalizados ter pouca ou nenhuma importância no produto final de cristalização.

De seguida procedeu-se ao acoplamento dos diversos CYPs com ligandos exógenos de forma a testar se os programas de acoplamento usados conseguiam distinguir um verdadeiro ligando de um não ligando. O que se verificou no nodo de acoplamento rígido foi que o único ligando escolhido capaz de exercer um controlo negativo sob o acoplamento nos CYPs 2A6/13 e 2B6 foi o 2QJ, facto que pode estar ligado à sua elevada complexidade estrutural. No caso do CYP2B6, como 2QJ é um ligando endógeno, o ligando IND foi aquele que acabou por funcionar como controlo negativo aceitável.

De um modo geral, não houve nenhum controlo negativo que se destacasse quando de procedeu ao acoplamento flexível. Tal pode dever-se ao facto de que no modo flexível, um outrora controlo negativo tenha depois espaço de manobra no interior do centro activo de modo a conseguir produzir um valor de Ki bastante inferior ao do produzido em modo rígido. O único ligando exógeno aceitável foi o 3QO quando este era acoplado ao CYP2A13 no programa Vina.

Ambos os programas de acoplamento usados são bons a acoplar ligandos que reúnam as seguintes características: moléculas não-polares, com um ou dois anéis (preferencialmente) aromáticos ou com uma estrutura em forma de cadeira e com poucas a nenhuma ligação rodável.

No caso dos ITCs e para ambos os programas, estes constituem bons inibidores de CYP2A6/2A13, inibidores moderados de CYP2B6 e maus inibidores de CYP2C9. PHITC é um excelente inibidor de CYP2A13.

Palavras-chave: Isotiocianatos; Citocromos P450; Acoplamento; Inibição

3. Introduction

3.1 What is Cancer?

Cancer is a complex and diversified disease of which not all cellular components are cancer cells (e.g. immune cells, endothelial cells, fibroblasts). Uncontrolled cellular proliferation in cancer cells, which is the most known characteristic of cancer, is caused by a progressive accumulation of multiple abnormalities (i.e. mutations) in the cells´ regulatory mechanisms. The several traits of cancer arise from six different types of alterations that happen during the cancer development process that leads to the formation of increasingly cancerous cells. These six cancer hallmarks are the following: self-sufficiency in growth signals, insensitivity to anti-growth signals, tissue invasion and metastasis, limitless replicative potential, sustained angiogenesis, and evading apoptosis. More recently, two new cancer hallmarks (deregulating cellular energetics, and avoiding immune destruction), as well as the corresponding enabling characteristics (tumour-promoting inflamation, and genome instability and mutation), were recognized as being involved in the carcinogenic process and, as such, they were added to the previously mentioned six (Cooper, G. M.; 2000)(Yates, L. R.; Campbell, P. J.; 2012)(Aktipis, C. A.; Nesse, R. M.; 2012)(Hanahan, D.; Weinberg, R. A.; 2000)(Hanahan, D.; Weinberg, R. A, 2011)(Tlsty, T. D.; Coussens, L. M.; 2006)(Bendall, S. C.; Nolan, G. P.; 2012)(Sudhakar, A.; 2009)(Mardis, E. R.; 2012)(Mendelsohn, J. et al.; 2008).

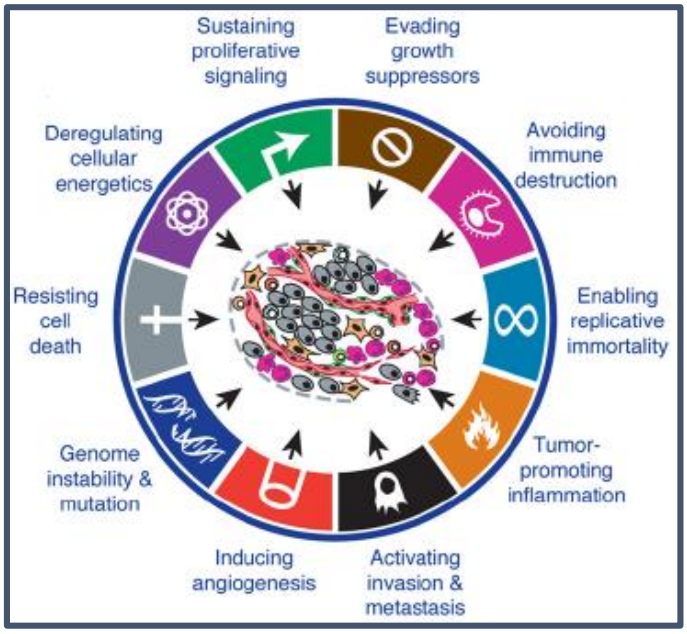


Fig.: 3.1 – The different cancer hallmarks and enabling characteristics that are crucial in the development of cancer (Hanahan, D; Weinberg, R. A.; 2011).

3.2 Distinction between Tumour and Cancer

Tumour is the same as cancer with regard to uncontrolled cellular proliferation. A benign tumour is one which does not invade any neighbourly tissue, or in other words, it neither penetrates nor crosses the basal membrane (i.e. membrane that surrounds the tumour cells). A malignant tumour, on the other hand, invades the basal membrane and the surrounding tissues and, with time, it can also eventually spread out to other tissues in the body through metastasis in the blood or lymph. Malignant tumours are the ones that should be properly known and referred to as cancers (Cooper, G. M.; 2000)(Mendelsohn, J. et al.; 2008).

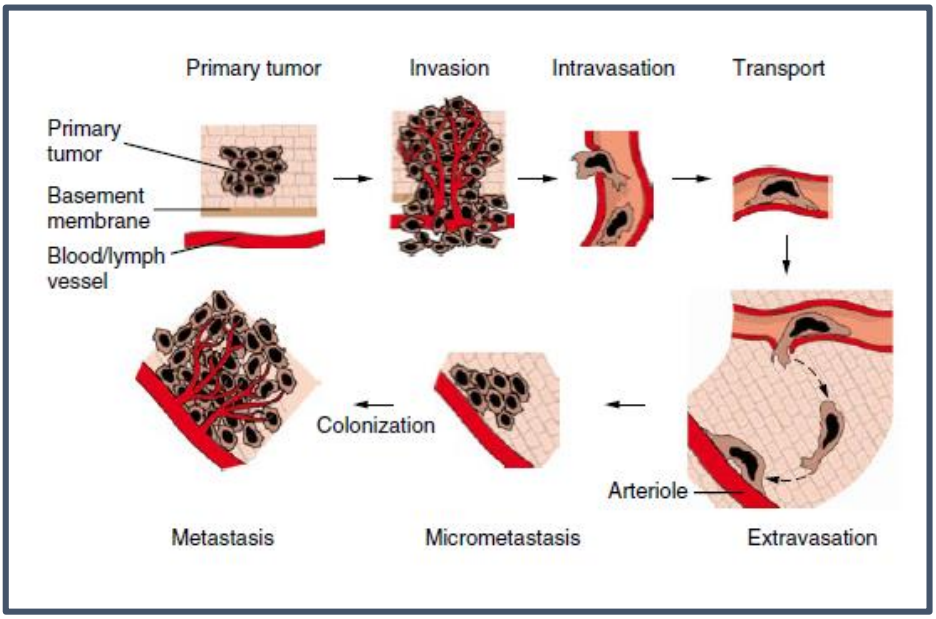


Fig.: 3.2 – A tumour is confined by the basement membrane until the last is ruptured by the first and invasion of the surround tissues occurs. Cancer cells can then procced to invade other tissues and organs in the body by using the linfatic system or the blood circulatory system (Mendelsohn, J. et al.; 2008).

3.3 Cancer Statistics Worldwide

3.3.1 Incidence

Cancer may affect up to at least 24 million people in the world, with an anual incidence rate of 12.7 million new cases of the disease diagnosed every year. Cancer incidence in the globe is expected to increase steadily over the next few years (Baskar, R. et al.; 2012)(Li, W. W. et al.; 2012)(Anand, P. et al.; 2008)(Steward, W. P.; Brown, K.; 2013)(see Bibliography website section for references).

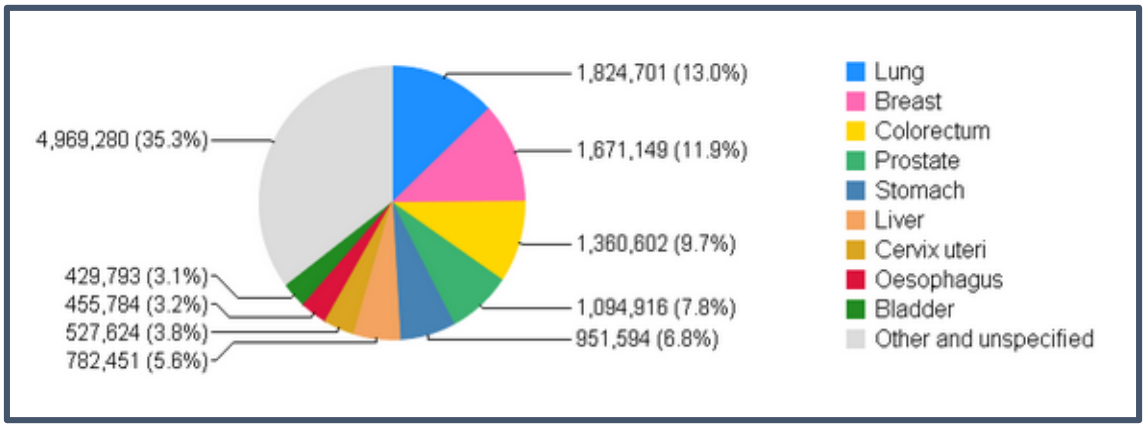


Fig.: 3.3 – Incidence of different types of cancer in the world, with lung being the one with the highest incidence affecting 13.0% of all cancer incidence cases (see Bibliography website section for references).

3.3.2 Mortality

Cardiovascular diseases are the first leading cause of death worldwide; the second been cancer. Every year, an estimated 30 million people will die of chronic diseases. 3.9 million (13%) of all these people died because of cancer. Some recently data pointed out to the estimated cancer death cases to be as high as 7.6 million deaths annually (Senthilkumar, K.; Kim, S.; 2013)(Thomas, F. et al.; 2012)(Baskar, R. et al.; 2012)(Li, W. W. et al.; 2012)(Rath, M.; 2001)(Li, W. W. et al.; 2012)(Baskar, R. et al.; 2012)(Sudhakar, A.; 2009)(see Bibliography website section for references).

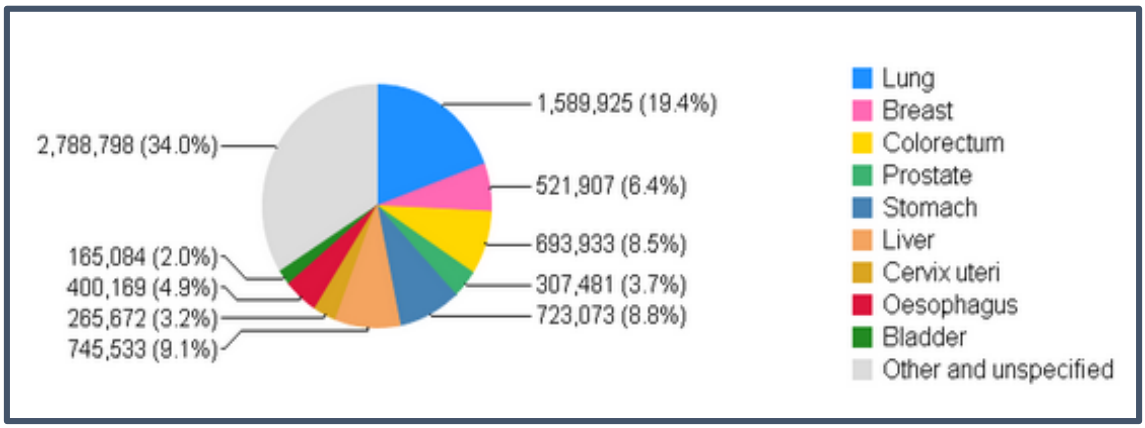


Fig.: 3.4 – Mortality of different types of cancer in the world, with lung being the one with the highest mortality affecting 19.4% of all cancer patients (see Bibliography website section for references).

3.4 Cancer Statistics by Gender

3.4.1 Incidence

It was estimated that in the year 2012, 14.1 million people would develop cancer. Of all this new cases of cancer, 53% of them would be in men and 47% would be in women. Lung cancer was the most usually diagnosed cancer in men and it affected 16.8% of the new cases of cancer in men in 2012. Breast cancer, on the other hand, was the most usually diagnosed cancer in women, affecting 25.1% of all the new cases of cancer in women for the same year (see Bibliography website section for references).

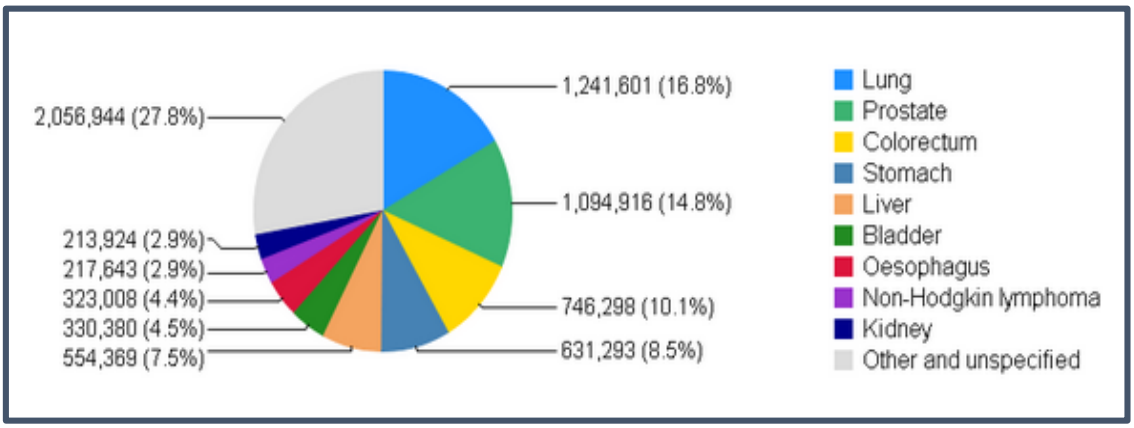


Fig.: 3.5 – Incidence of different types of cancer in men, with lung being the one with the highest incidence (see Bibliography website section for references).

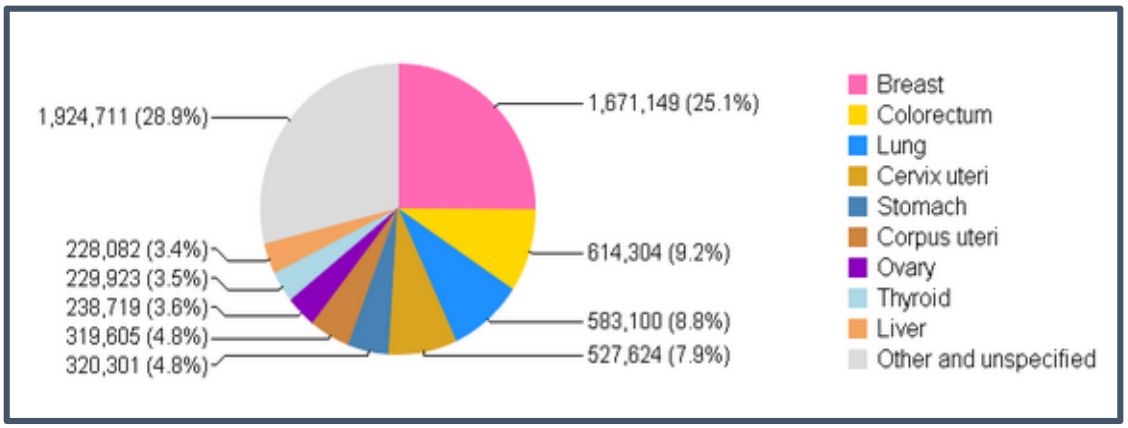


Fig.: 3.6 – Incidence of different types of cancer in women, with breast being the one with the highest incidence (see Bibliography website section for references).

3.4.2 Mortality

In terms of cancer deaths around the globe, estimates were made that pointed out to 8.2 million people dying in 2012 because of this disease. 57% of these cancer deaths would affect men and 43% would affect women. Lung cancer is the main cancer related cause of death in men resulting in 23.6% of all cancer mortality in men, while breast cancer is the cancer type that causes more cancer deaths in women resulting in 14.7% of all cancer mortality in women (see Bibliography website section for references).

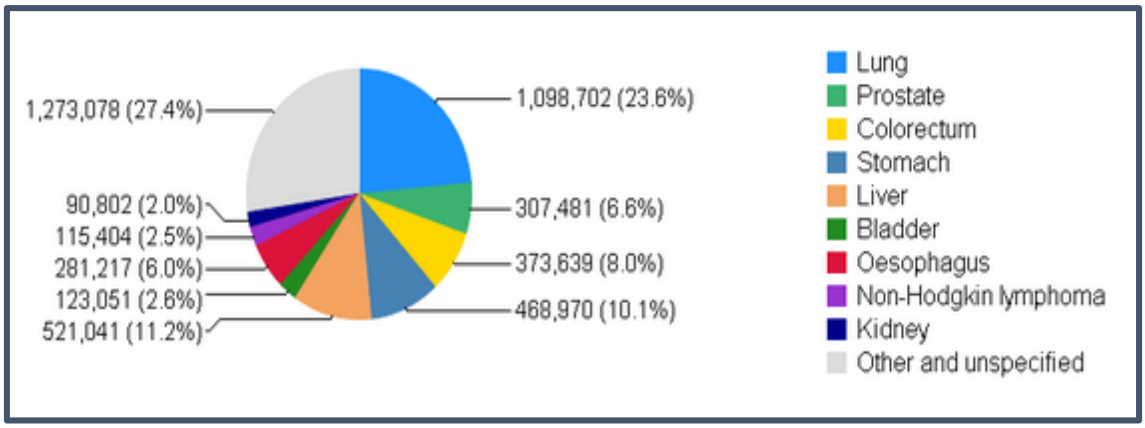


Fig.: 3.7 – Mortality of different types of cancer in men, with lung being the one with the highest mortality (see Bibliography website section for references).

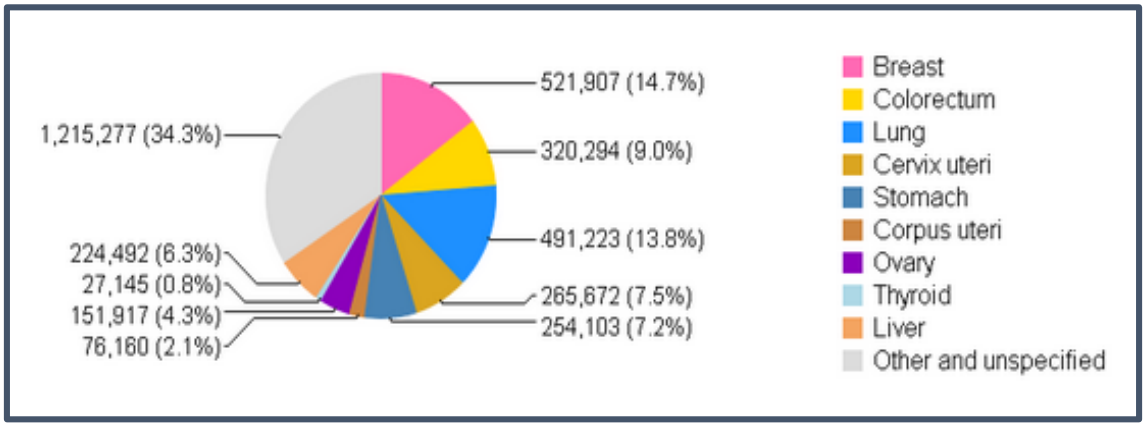


Fig.: 3.8 – Mortality of different types of cancer in women, with breast being the one with the highest mortality (see Bibliography website section for references).

3.5 Development of Cancer

Cancer development is generally a long process resulting from an accumulation of mutations over a period of time, which could take as many as 20 years or even more until invasion of nearby tissues happens. This fact explains why cancer commonly develops in old-aged people. As time goes by, tumour cells are being constantly selected for an increasingly cancerous phenotype (i.e. capacity for proliferation, survival, invasion, and metastasis)(Cooper, G. M.; 2000)(Sporn, M. B.; 1996)(Sudhakar, A.; 2009)(Vogelstein, B. et al.; 2013)(see Bibliography website section for references).

Initially, a cell suffers a mutation that is not repaired and which affects the cells´ regulatory mechanisms. This first mutational event that leads to an abnormal proliferation rate of a single cell is called tumour initiation. As time passes, that cell will start to divide more frequently than its counterparts. At some point, these altered cells suffer yet another mutational event that turns those even more aberrant compared with the normal cells in the tissue. This continued accumulation of defects is called tumour progression. Tumour cells are selected towards their capacity to proliferate, survive, invade and metastasise in the same way as Darwinian natural selection occurs in living beings (i.e. somatic evolution – cancer cell formation is favoured – as opposed to organismal evolution – beings capable of maintaining their genome integrity across all of its cells). This clonal selection for progressively more cancerous cells is a result of a plethora of different factors (e.g. chemo or radiotherapy treatment, microenvironment factors) that influence the selection of a certain type of clone to the detriment of the others existing in the tumour. It is this heterogeneous cell population in cancer that is responsible for relapses after a certain treatment has being done in a patient. One additional problem in terms of cancer treatment is the fact that cancer has intra- and inter-heterogeneity (Yates, L. R.; Campbell, P. J.; 2012)(Cooper, G. M.; 2000)(Thomas, F, et al.; 2012)(DeVita, Jr., V. T.; Chu, E.; 2008)(Gatenby, R. A.; Gillies, R. J.; 2004)(Aktipis, C. A., Nesse, R. M.; 2012)(Bendall, S. C.; Nolan, G. P.; 2012)(Sudhakar, A.; 2009)(Mardis, E. R.; 2012)(Vogelstein, B. et al.; 2013)(Mendelsohn, J. et al.; 2008).

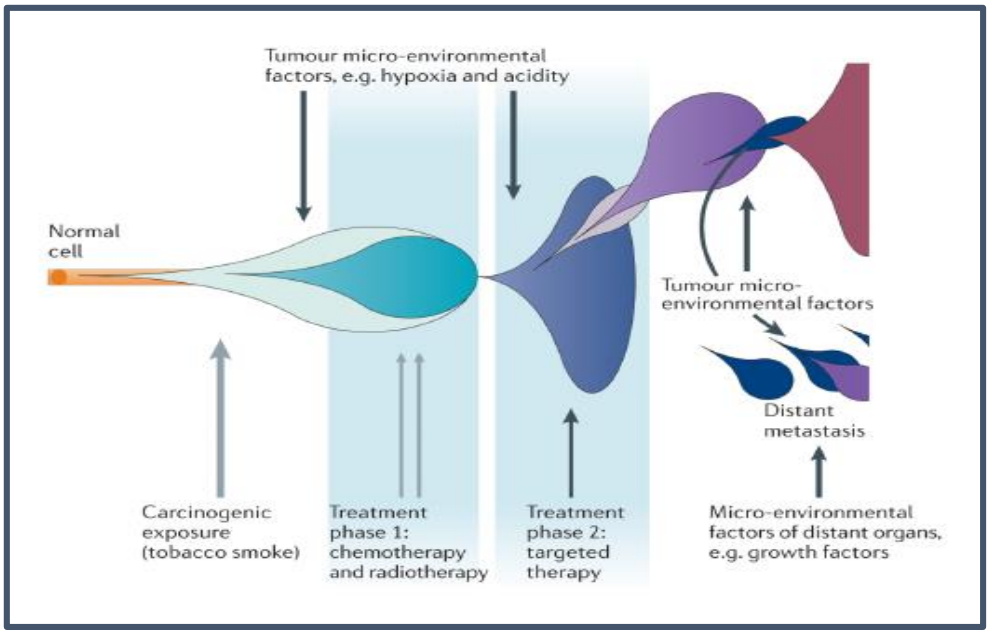


Fig.: 3.9 – The evolutionary procees by which a single mutated cell can lead to the formation of a heterogenous cancer mass (Yates, L. R.; Campbell, P. J.; 2012).

3.6 Cancer Treatment

The treatment of cancer consists in three main lines of approach: surgery, radiotherapy, and chemotherapy. Surgery was already practiced as far back as in ancient Egypt, were the very first description of cancer was made and discovered in an ancient papyrus dated from 3000 B.C., the Edwin Smith papyrus. In it, it is described what nowadays is considered to be a breast cancer and the fact that, at the time, there was no treatment available to cure this disease; cutting it off the body could only delay the progression of the aforementioned ailment. Radiotherapy consists on the use of ionizing radiation, more specifically X-rays to try and kill cancer cells by destroying their DNA. X-rays and their use as a means to fight against cancer were introduced in 1895, by Wilhelm Conrad Röntgen in Germany. Both surgery and radiotherapy treatment modalities were the dominat types of treatment given to cancer pacients until the 1960s decade, inclusively. Unfortunately, the cure rates resulting from the application of these treatments had plateaued at about 33% due to the existence of undetected micrometastases. Chemotherapy was, therefore, the path to follow next, the best cure rates obtained for combined chemotherapy provided to advanced cancer patients. The term chemotherapy was coined in the beginnings of the 1900s´decade in Germany by Paul Ehrlich. These new kind of treatment was based on the development of new and improved

drugs that could be used to treat diseases, one of them of course being cancer (DeVita, Jr., V. T.; Chu, E.; 2008)(Sudhakar, A.; 2009)(Papavramidou, N. et all; 2010)(Hajdu, S. I.;. 2004)(Baskar, R. et al.; 2012)(Fairchild, A. et al.; 2008)(Mendelsohn, J. et al.; 2008)(see Bibliography website section for references).

3.7 Risk Factors for Cancer

Chronic diseases, including cancer, are mostly caused by environmental and lifestyle factors. These factors are responsible for 90-95% of all cancers, with genetics having a scarce contribution (5-10%) for the development of cancer (Anand, P., et al.; 2008)(Senthilkumar, K.; Kim, S; 2013)(Brawley, O. W.; 2011).

By changing a person´s lifestyle it is possible to drastically decrease the risk of ever developing cancer. The most avoidable risk factors for the prevention of cancer are the following: diet (30-35%), tobacco (25-30%), infections (15-20%), obesity (10-20%), and alcohol (4-6%) (Anand, P., et al.; 2008) (Danaei, G.; 2005)(Brawley, O. W.; 2011).

3.8 A New Perspective about Cancer

All the methods previously mentioned in this manuscript are used to treat cancer, most of the time when the disease has already developed in the body or when it is at an advanced state. Early detection of cancer helps, but inevitably, the majority of cancer patients only receive their diagnosis when the cancer has already advanced to a point where little hope for cure exists. Cancer prevention, on the other hand, deals with trying to diminish the incidence and mortality due to cancer by altering the modifiable risk factors that increase the cancer risk. This is done mainly by analysing the findings from different sources (e.g. epidemiology studies, basic research and clinical trials) that will then be used as a way to help to improve (i.e reduce) the cancer burden. Cancer chemoprevention was originally defined by Michael B. Sporn, in 1976, as the chronic usage of chemical agents (natural or synthetic) to try and reverse, suppress, delay or even prevent the emergence or further progression of an existing cancer. Dietary components have a key importance in the chemoprevention of cancer, specially the consumption of fruits and vegetables (Steinmetz, K. A.; Potter, J. D.; 1996)(Liu, R. H.; 2004)(Wargovich, M. J. et al.; 2001)(Steward, W. P.; Brown, K.; 2013)(Mettlin, C; 1997)(Tsao, A. S. et al.; 2004)(Umar, A. Et al.; 2012) (Li, W. W. et al.; 2012)(Wolf, C. R.; 2001)(see Bibliography website section for references).

3.9 Relation between Diet and Disease

Many epidemiologic studies have shown that a diet with a high abundance and variety of vegetables and fruits is associated with a reduction in the incidence, risk and development of several types of cancer and cardiovascular diseases. In some cases, a high consumption of fruits and vegetables has led to a decrease in the risk of developing cancer up to 50% (Manchali, S.; Murthy, K. N. C.; Patil, B. S.; 2012)(Talalay, P.; Fahey, J. W.; 2001)(Moiseeva, E. P.; Manson, M. M.; 2009)(Castro, I. M. et al.; 2008) (Rao, C. V.; 2013)(Totusek, J. et al.; 2011)(Cartea,M. E. et al.; 2011)(Liu, R. H.; 2004).

3.10 Cruciferous Plants – A Special Case

Among all the variety of vegetables that are eaten in a normal and healthy diet, cruciferous (also called Brassica) vegetables seem to be especially effective as cancer preventive agents against several types of malignancies, even to the extent of their consumption being reversely associated with cancer incidence at multiple different sites (e.g. lung, breast, colorectal and prostate). Colorectal cancer is a prime example of how the intake of this kind of vegetables in the diet can have a positive effect in disease prevention (Talalay, P.; Fahey, J. W.; 2001) (Wallig, M. A. et al.; 2005)(Rao, C. V.; 2013)(Shapiro, T. A. et al., 2001)(Seow, A. et al.; 2002)(Totusek, J. et al.; 2011)(James, D. et al.; 2012)(Gasper, A. V. et al.; 2007)(Epplein,M. et al.; 2009)(Kurilich, A. C.; 1999)(Scott, O. et al.; 2012)(Higdon, J. V., et al., 2007)(Wu, X. et al.; 2009)(Ishida, M. et al.; 2014)(Cartea,M. E. et al.; 2011).

3.11 Importance of The Brassicaceae Family in General and *B. oleracea* in Particular

The Brassicaceae family includes several species of crops (e.g. leaf vegetables, root vegetables, oilseed and condiments) that are cultivated in many parts of the globe because of their high nutritional value and agricultural economic importance. The majority of the species of cruciferous vegetables that are consumed all over the world belong to the *Brassica* genus (e.g *B. oleracea*, *B. rapa*, *B. napus*, *B. carinata*, *B. nigra* and *B. juncea*). *B. oleracea* is the species of cruciferous vegetables that is most commonly seen in a healthy diet. This species is very diverse, morphologically, and it comprises multiple different varieties (sub species) of cultivars, including kale, collard greens, cabbage, Brussels sprouts, broccoli, kohlrabi, and cauliflower. (Scott, O. et al.; 2012)(Park, M. et al.; 2013)(Rosa, E. A. S.; Rodrigues, A. S.; 2001)(Sotelo, T. et al.;

2014)(Cartea,M. E. et al.; 2011)(Yu, J. et al.; 2013)(Parkin, I. A. P.; 2014)(Wu, X. et al.; 2009)(Sotelo, T. et al.; 2014)(Ishida, M. et al.; 2014).

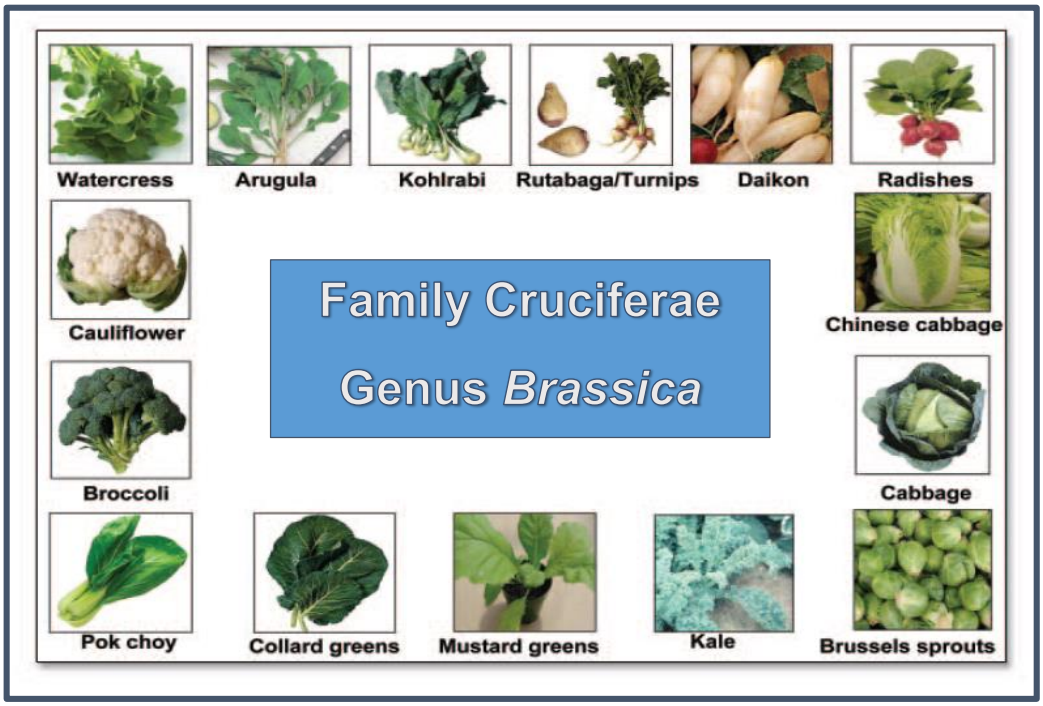


Fig.: 3.10 – Different aspects of vegetables belonging to the Brassica genus (e.g. *B. oleracea*) (Aggarwal, B. B.; Ichikawa, H.; 2005).

3.12 How Can Cruciferous Vegetables Prevent Cancer?

Vegetable and fruit consumption is responsible for the intake of many different types of biologically active compounds [e.g. vitamins (C and E, for example), minerals, fibre, antioxidants, phenols, carotenoids, polyphenols, folate, isoflavonoids, indoles and others] that are thought to be responsible for the cancer-preventive properties of vegetables (Scott, O. et al.; 2012)(Higdon, J. V., et al., 2007)(Totusek, J. et al.; 2011)(Gratacós-Cubarsí, M. et al.; 2010)(Park, M. et al.; 2013)(Kapusta-Duch, J.; Leszczynska, T.; 2013)(Liu, R. H.; 2004)(Kim, M. K.; Park, J. H. Y.; 2009).

Phytochemicals ("phyto" means plant in the Greek language), or secondary metabolites, considered to be non-essential nutrients, are compounds that are present in plant-derived foods. These compounds are able to perform a myriad of different functions: they are able to modulate the activity of biotransformation enzymes, scavenge free radicals, alleviate inflammation, stimulate immune functions, inhibit malignant transformation, and regulate the growth of cancer cells. It is by doing all of this functions

that they provide a way by which the consumption of cruciferous vegetables may contribute to the reduction of the risk of developing cancer (Hennig, K. et al.; 2014)(Lampe, J. W.; Peterson, S.; 2002)(Kapusta-Duch, J.; Leszczynska, T.; 2013)(Cartea, M. E. et al.; 2011)(Talalay, P.; Fahey, J. W.; 2001)(Liu, R. H.; 2004).

Beyond the nutritional value of cruciferous vegetables, these also have several secondary metabolites with health improvement benefits as it was previously mentioned (e.g. flavonoids, anthrocyanins, coumarins, carotenoids, antioxidant enzymes, terpenes and others). Of particular interest is the fact that these vegetables have high levels of glucosinolates, which are thought to be responsible for the chemopreventive effect provided by the consumption of this kind of vegetables. This type of phytochemical compound exists solely in 15 plant families belonging to the Capparales order, of which members of the Brassicaceae family possess the highest concentration (Wu, X. et al.; 2009) (Castro, I. M. et al.; 2008)(Rao, C. V.; 2013)(Molina-Vargas, L. F.; 2013)(Epplein M. et al.; 2009)(Ishida, M. et al.; 2014)(Park, M. et al.; 2013)(Manchali,S. et al.; 2012)(Seow; A. et al.; 2002).

3.13 What are Secondary Metabolites?

Secondary metabolites are usually small molecules (molecular weight less than 3,000 Da) that are synthesized by a natural source, like plants, animals and microorganisms. These compounds were named like this because of two reasons: firstly, they are not synthesized by the general metabolic pathways; and secondly, they have no direct role in the growth, development and reproduction of the organism. Instead, these secondary metabolites are used as a means of defence against the habitat´s conditions in which the organism lives (Bhatnagar, I.; Kim, S.; 2010) (Martins, A; Vieira, H.; Gaspar, H; Santos, S; 2014).

3.14 What are Glucosinolates?

As mentioned before, glucosinolates (GSLs, GSs or GLs), also known as mustard oils glucosides, constitute a class of secondary metabolites that is very abundant in cruciferous vegetables (corresponding to approximately 1% of their dry weight), namely in the Brassicaceae family (Ratzka, A. et al.; 2002)(Kliebenstein D. J. et al.; 2005)(Park, M. et al.; 2013)(Ishida M. et al.;2014)(Sotelo, T. et al.; 2014)(James, D. et al.; 2012)(Agrawal, A. A.; Kurashige, N. S.; 2003)(Talalay, P.; Fahey, J. W.; 2001)(Marca, M. L. et al.; 2012)(Zhang, J. et al.; 2015).

3.15 Chemical Structure of GSLs

GSLs are β-thioglucoside *N*-hydroxysulfates with the molecular formula R-C(=N-O-SO_3^-)-*S*-glucose. These compounds (anionic thioglucosides) are sulphur- and nitrogen-based molecules derived from glucose and an α-amino acid precursor. GSLs are made up of three moieties: a β-D-thioglucose, a sulfonated oxime and a variable aglycone (R) side chain moietie (Ishida M. et al.;2014)(Wu, X. et al.; 2009)(Kim, Y. B. et al.; 2013)(Gratacós-Cubarsí, M. et al.; 2010)(Ratzka, A. et al.; 2002)(James, D. et al.; 2012)(Redovnikovic, I. R. et al.; 2008)(Lambrix, V. et al.; 2001).

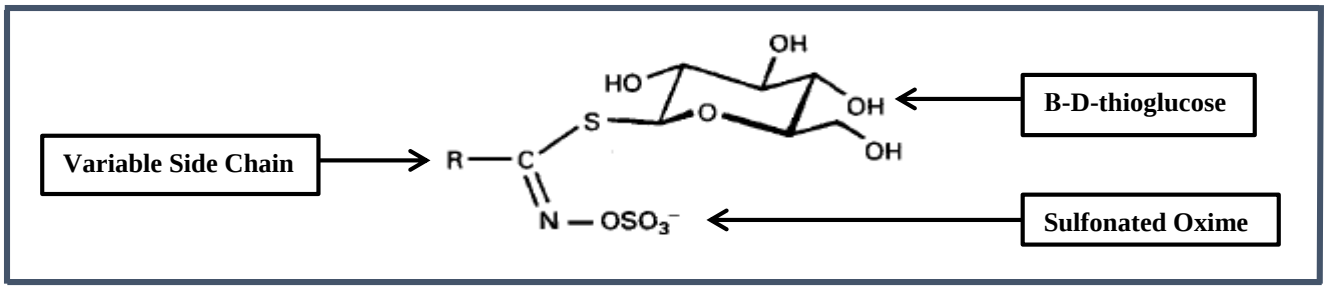


Fig.: 3.11 – The intact and general structure of a GSL (Nugon-Baudon, L.; Rabot, S.; 1994).

3.16 Classification of GSLs

Up to date, approximately 200 different types of GSLs have been isolate from plants and identified so far. These natural products (NPs) are classified in accordance with the structure of the specific amino acid precursor that originated the GSL. As such, GSLs can be classified into three different groups: aliphatic, indolic, and aromatic. The aliphatic GSLs have a side chain that derived from alanine, leucine, isoleucine, valine and methionine (Met); while the aromatic GSLs have a side chain that derived from phenylalanine and tyrosine. On the other hand, all indolic GSLs R group derived from tryptophan. In Brassica crops, the majority of GSLs described had a side chain that was synthesized from Met (Ishida M. et al.;2014)(Kim, Y. B. et al.; 2013)(Sotelo, T. et al.; 2014)(Rosa, E. A. S.; Rodrigues, A. S.; 2001)(Kliebenstein, D. J. et al.; 2005).

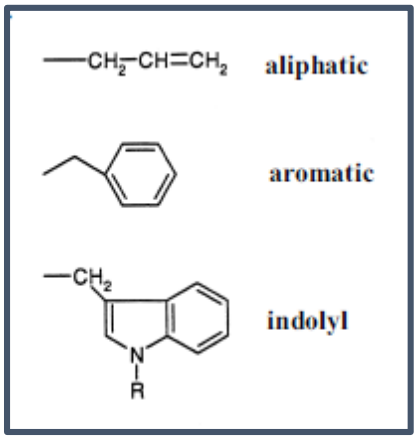


Fig.: 3.12 – The three groups of GSLs according to their derived amino acid side chain (Borgen, B. H.; 2002).

3.17 How Abundant Are GSLs In Cruciferous Vegetables?

The different types of GSLs present in plants and their respective concentrations are greatly influenced by several factors: genotype (i.e. different species and species variety), environmental (e.g. climate, soil fertility), developmental [i.e. different parts/tissues of the same plant and different developmental stages (i.e. sprouts can have 20-50 times more GSLs compared with mature market-stage plants) can constrain both the levels and types of GSLs present], and agronomic factors (i.e. cultivation conditions – fertilization, harvest time, plant position, soil type –, post-harvest storage, etc.)(Shapiro, T. A. et al.; 2001)(Sotelo, T. et al.; 2014)(Ishida, M. et al.; 2014)(Gratacós-Cubarsí, M. et al.; 2010)(Nugon-Baudon, L.; Rabot, S.; 1994)(Sarıkamış, G. et al.; 2009).

3.18 Are Raw and Cooked Vegetables Different?

Brassica vegetables are normally consumed after being cooked. This process invariably alters the GSL content of these vegetables, leading to a lower intake of GSLs from the diet. The act of cooking the vegetables itself can lead to a drastic decrease of GSLs content between 30-60%. During the cooking process GSLs, which are water soluble molecules, are able to leach into the cooking water. Another fact that may contribute to the decrease in GSLs content in cooked vegetables is thermal degradation, without intervention of degradation enzymes. Both processes, leaching and thermal degradation will obviously depend upon several factors: the GSL type (i.e. chemical structure), the specific vegetable that is used (i.e. concentration and variety of GSLs), the size of the vegetable pieces, the temperature and cooking time, as well as the pH of the

cooking liquid. These two processes are possible explanation for the observed GSL content drops in cruciferous vegetables after they have been cooked (McNaughton, S. A.; Marks, G. C.; 2003)(Hennig, K. et al.; 2014)(Nugon-Baudon, L.; Rabot, S.; 1994)(Park, M. et al.;2013)(Rungapamestry, V. et al.; 2007).

Interestingly, different food processing techniques have different impacts on the GSL content of Brassica vegetables. Chopping, cooking, steaming, and microwaving, which are considered as being domestic processing techniques performed at home have a much higher impact than freezing, fermenting and hot packing, which are techniques considered to be of industrial level. The reasons for this remarkable difference remain unclear (Park, M. et al.; 2013).

Another fact that may help explain the lower quantity of GSLs in cooked cruciferous vegetables, in comparison with uncooked ones, is the catalytic activity of the myrosinase enzyme (Nugon-Baudon, L.; Rabot, S.; 1994)(Yoshigae, Y. et al.; 2013)(Hecht, S. S.; 1999)(Rungapamestry, V. et al.; 2007).

3.19 Myrosinase

The enzyme commonly called myrosinase (MYR), also called β-thioglucoside glucohydrolase (EC 3.2.3. l), belongs to a class of enzymes termed β-thioglucosidases. MYR is a dimeric glycoprotein with a molecular weight between 124-150 kDa. This highly glycosylated enzyme (i.e. the carbohydrates correspond to 20% of its molecular weight) is responsible for hydrolysing GSLs in order to produce an array of different compounds. Several MYR isoenzymes, which are enzymes with differences in their amino acid sequence but that are able to catalyse the same enzymatic reaction, have been discovered (Ratzka, A. et al.; 2002)(Kliebenstein, D. J. et al; 2005) (Nugon-Baudon, L.; Rabot, S.; 1994)(Benn, M.; 1977)(Ishida M. et al.; 2014) (Rask, L. et al.; 2000)(Botti, M. G. et al.; 1995)(Redovnikovic, I. R. et al.; 2008)(Bones, A. M.; Rossiter, J. T.; 1996)(Eriksson, S. et al.; 2002)(Halkier, B. A.; Gershenzon, J.; 2006).

3.20 MYR and GSLs Compartmentalisation

MYR coexists with GSLs in the same plant, although they are both physically separated, with GSLs being located inside vacuoles in the cytoplasm of cells from various plant tissues, while MYR is located and stored within special cellular structures called myrosin granules that are exclusively present in a specific type of plant cell, called

idioblast or myrosin cell, or even toxic mine, that is scattered throughout the plant´s tissues. GSLs are not found in idioblasts, and likewise MYR is not found in non-idioblast cells. MYR expression can also be found on the cell wall external surface (Wink, M.; 2010)(Benn, M.; 1977)(Bridges, M. et al.; 2002)(Zhang, Y.; 2010) (Nugon-Baudon, L.; Rabot, S.; 1994))(Wu, X. et al.; 2009)(Andréasson, E. et al.; 2001)(Rask, L. et al.; 2000)(Ahuja, I. et al.; 2015)(Hara, M. et al.; 2000)(Redovnikovic, I. R. et al.; 2008)(Marca, M. L. et al; 2012).

3.21 The Myrosinase-Glucosinolate System

The myrosinase-glucosinolate (MYR-GSL) system, also known as the mustard oil bomb consists on a defence mechanism employed by cruciferous plants to defend themselves against generalist insects, herbivores and phytopathogens (i.e. plant disease agents – fungus, bacteria, and nematodes), as well as functioning as regulators of the interactions of the plant itself with specialist insects. The two main players in this system are GSLs and MYR. This system works by putting in contact both GSLs and MYR but only in the case where there is tissue disruption (Molina-Vargas, L.F.; 2013)(Ratzka, A. et al.; 2002)(Textor, S.; Gershenzon, J.; 2009)(Ishida M. et al.; 2014)(Bones, A. M.; Rossiter, J. T.; 1996)(Textor, S.; Gershenzon, J.; 2009)(Grubb, C. D.; Abel, S.; 2006).

3.22 How Does It All Work?

It is upon tissue damage, that can be caused by either chewing, preparation for cooking (e.g. slicing), or simply plant injury caused by herbivore insects, that both components – GSLs and MYR – are able to meet and initiate the enzymatic activity that leads to the production of volatile GSLs hydrolysis products (Agrawal, A. A.; Kurashige, N. S.; 2003) (Hennig, K. et al.; 2014)(Nugon-Baudon, L.; Rabot, S.; 1994)(Andréasson, E. et al.; 2001)(Wu, X. et al.; 2009)(Glendening, T. M.; Poulton, J. E.; 1988).

3.23 GSL-MYR Reaction Process

When MYR is put into contact with GSLs, the first will hydrolyse the second on the thioglucoside bond. This process leads to the release of the β-D-glucose moiety from the substrate, which subsequently leads to the formation of an unstable aglycone (sometimes also referred to as an aglucone) intermediate, with the molecular formula R-C(-SH)=N-O-SO_3^- , that is often referred to as thiohydroximate-*O*-sulphonate. This intermediate then undergoes a spontaneous rearrangement, which leads to the release of

a sulphate anion (SO_4^{2-}) and the production of a single compound from a plethora of GSL degradation products – thiocyanates, isothiocyanates, and nitriles, among others (Ishida M. et al.; 2014)(Wu, X. et al.; 2009)(Nugon-Baudon, L.; Rabot, S.; 1994)(Bones, A. M.; Rossiter, J. T.; 1996)(Verhoeven, D. T. H. et al.; 1997)(Eriksson, S. et al.; 2002)(Ratzka, A. et al.; 2002)(Castro, I.M. et al.; 2008)(Botti, M. G. et al.; 1995)(Andréasson, E. et al.; 2001).

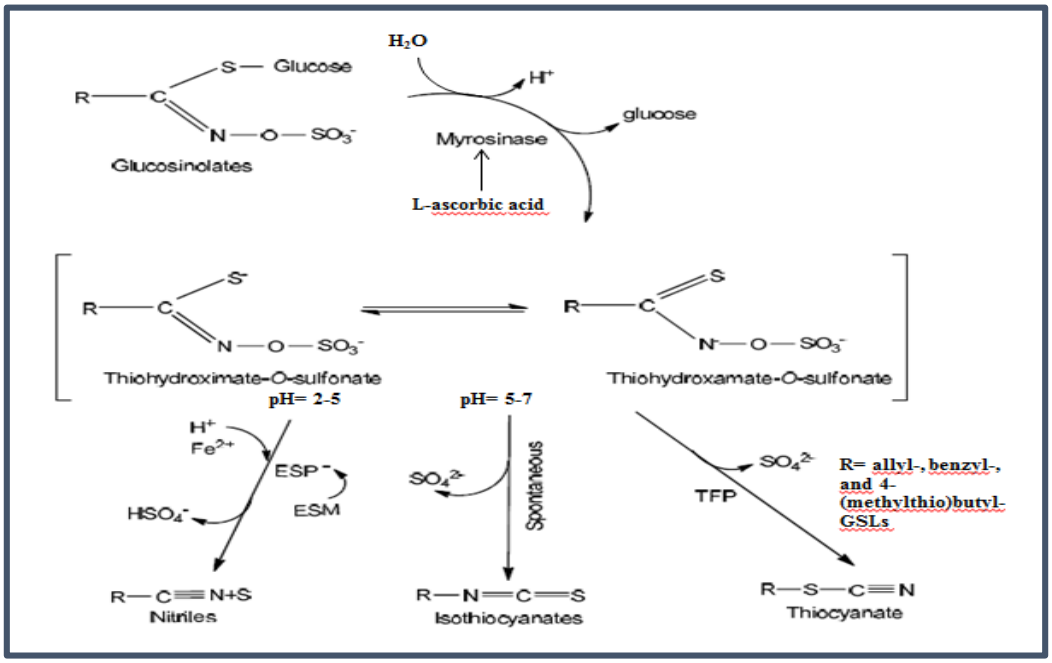


Fig.: 3.13 – The GSL-MYR reaction process, which leads to the formation of several GSLs degradation products. ESP – epithospecifier protein; ESM – epithiospecifier modifier; TFP – thiocyanate forming proteins; R-variable side chain (Zhang, Y; Talalay, P.; 1994)(Wu, X. et al.; 2009)(Tong, Z.; 2007)(Borgen, B. H.; 2002)(Redovnikovic, I. R. et al.; 2008)(Rask, L. et al.; 2000)(Nugon-Baudon, L.; Rabot, S.; 1994)(Lambrix, V. et al.; 2001)(Hayes, J. D. et al.; 2008)(Alnsour, M.; 2013).

3.24 Isothiocyanates – What are They?

As metioned before, isothiocyanates (ITCs), also called mustard oils, are one of several end-products of the reactions catalyzed by MYR. When the GSL has an aliphatic side chain and the reaction medium has a neutral pH or a high temperature and there is no ESP, the formation of ITCs is favoured over the production of other hydrolysis end-products. ITCs are the predominant product formed by the MYR-GSL system, closely followed by nitriles as the second main end products. It is the hydrolysis products, mainly ITCS, rather than the GSLs themselves that are the biological active compounds responsible for the chemoprotective effect conferred by the consumption of Brassica

vegetables. Additionally, these compounds are the ones responsible for the pungent odour and spicy or sharp, some even say acrid flavour that is so characteristic of the *B. oleracea* species as well as of some other cruciferous vegetables. The term mustard oil originated from the flavour of mustard seeds as well as from the oil extracted from them, which are rich sources of a type of ITC named allyl ITC, or more commonly known as just oil of mustard or mustard oil (Nugon-Baudon, L.; Rabot, S.; 1994)(Wu, X. et al.; 2009)(Lambrix, V. et al.; 2001)(Wink, M.; 2010)(Kliebenstein, D. J. et al.; 2005)(Seow, A. et al.; 2002)(Park, M. et al.; 2013)(Hennig, K. et al.; 2014)(Agrawal, A. A.; Kurashige, N. S.; 2003)(Lampe, J. W.; Peterson, S.; 2002)(Ishida, M. et al.; 2014)(Zhang, Y. et al.; 2006)(Molina-Vargas, L. F.; 2013)(Talalay, P.; Fahey, J. W.; 2001)(Kraker, J. de; Gershenzon, J.; 2011)(Ishida, M. et al.; 2014)(Zhang, Y.; Talalay, P.; 1994)(Remaud, G. S. et al.; 1997)(Kimball, E. S. et al.; 2006)(Jordt, S. et al.; 2004).

3.25 ITCs´ Chemical Structure

ITCs constitute a family of small compounds with the molecular formula R-N=C=S. This kind of molecules are characterized by having an –N=C=S group and the fact that the side chain (R) can only be an alkyl or an aryl group. (Wu, X. et al.; 2009)(Zhang, Y, et al.; 2006)(Tang, L.; Zhang, Y.; 2004)(Talalay, P.; Fahey, J. W.; 2001)(Zhang, Y.; Talalay, P.; 1994).

3.26 Cooked Vegetables, Uncooked Ones and MYR – What Should I Eat?

ITCs are formed by MYR, as it was previously explained in this manuscript. However, the enzyme MYR is inactivated by high temperatures (i.e. heat), which happens when the vegetables are cooked; consequently, the quantity of ITCs that will be available from cooked vegetables after they have been ingested will be much smaller compared with uncooked ones. Nevertheless, GSLs from the diet can still be converted into ITCs. This happens because the bacteria from the microflora present in the gastrointestinal system have MYR-like activity (bacterial β-thioglucosidase), which results in the formation of ITCs, albeit the rate at which it happens is 3-10 times lower than that of plant MYR. This process is responsible for the conversion of about 40% of the GSLs that escape plant MYR. Another noteworthy way by which these GSLs can be converted into ITCs is by a thioglucosidase activity present in animal tissues (Hennig, K. et al.; 2014)(Talalay, P.; Fahey, J. W.; 2001)(Hecht, S. S.; 1999)(Getahun, S. M.; Chung, F.; 1999)(Zhang, Y.; 2010)(Wu, X. et al.; 2009)(Navarro, S. L. et al.; 2011).

3.27 How Are ITCs Chemopreventive Against Cancer?

ITCs are capable of being chemopreventive against cancer because they are able to regulate several cellular mechanisms associated with cancer, including: modulation of Phase I and Phase II detoxification enzymes, induction of apoptosis, inhibition of cell proliferation, inhibition of angiogenesis, modulation of epigenetic mechanisms, activation of cell cycle arrest, among many others (Tang, L.; Zhang, Y.; 2005)(Yang, G. et al.; 2010)(Tang, L.; Zhang, Y.; 2004)(Zhang, Y. et al.; 2006)(Yoshigae, Y. et al.; 2013)(Sarıkamış, G. et al.; 2009)(Rao, C. V.; 2013)(Watson, R. R.; Preedy, V. R.; 2010)(Navarro, S. L. et al.; 2011).

The main biological activity of ITCs is the regulation of carcinogen metabolism by the biotransformation enzymes. This task is accomplished firstly by the induction of detoxification enzymes (Phase II), and secondly by the repression of xenobiotic (carcinogen) activation enzymes (Phase I). This is the primary process by which dietary ITCs are able to be chemoprotective against cancer (Scott, O. et al.; 2012)(Epplein, M. et al.; 2009)(Higdon, J. V.; et al.; 2007)(Wallig, M. A. et al.; 2005)(Yang, G. et al.; 2010)(Seow, A. et al.; 2002)(Castro, I. M. et al.; 2008)(Park, M. et al.; 2013)(Gasper, A. V. et al.; 2007)(Talalay, P.; Fahey, J. W.; 2001)(Totusek, J. et al.; 2011)(Hecht, S. S.; 1999)(Hecht, S. S.; 2004).

3.28 Compounds and Their Biotransformation Enzymes

The organism is always and inevitably exposed to a myriad of different compounds, both exogenous (e.g. drugs, carcinogens, pesticides, pollutants, secondary plant metabolites, toxins) and endogenous (e.g. steroids, prostaglandins, bile acids), which are capable of exerting beneficial or harmful effects on the organism depending on the compound. The exogenous or foreign molecules to the organism are often designated by the term xenobiotics. All of these compounds (exo- and endogenous alike) are metabolized by different ubiquitous enzymes known collectively as biotransformation enzymes, drug metabolizing enzymes (DMEs) or xenobiotic metabolizing enzymes (XMEs). These enzymes are mostly expressed in the endoplasmic reticulum and in the cytoplasm of various human tissues (e.g. skin, lungs, nasal mucosa, eyes, kidneys, gastrointestinal tract). Most of the xenobiotic metabolism occurs primarilly in the liver because of several important factors: firstly, the liver consists in the largest internal organ in the human body; secondly, it is strategically well located inside the body, close to the

gastrointestinal tract as well as other nearby important organs; and thirdly and finally, it is considered to be the organ that has the highest concentration of XMEs. Though the drug metabolism performed by DMEs is important, it probably only accounts for less than 1% of all the functions these enzymes carry out in the human body (Jancova, P. et al.; 2010)(Gaikovitch, E. A.; 2003)(Penner, N. et al.; 2012)(Brandon, E. F. A. et al.; 2003)(Nebert, D.W.; 1997)(Nebert, D. W.; Dalton, T. P.; 2006)(Yang, C. S. et al.; 1992)(Irigaray, P.; Belpomme, D.; 2009)(Xu, C. et al.; 2005)(Rose, R. L.; Hodgson, E.; 2004)(Sturgill, M. G.; Lambert, G. H.; 1997)(Ennulat, D. et al.; 2010)(Brandon, E. F. A. et al.; 2003).

3.29 The Problem with Eliminating a Previously Absorbed Compound

If an exogenous compound is absorbed by the organism because of its lipophilicity (i.e. can bind to lipid membranes and be transported around the body in the blood by lipoproteins because it is a lipid-soluble compound), it will end up becoming a problem when it time comes to be excreted from the body. The biotransformation process is responsible for turning an otherwise lipophilic compound into a more hydrophilic (i.e. polar or water-soluble) compound, in order to make it more easily excreted in the urine or bile. The excretion rote is dependent on several crucial factors of the compound to be excreted (e.g. chemical structure, molecular size, molecular weight, and polarity) and in some characteristics of the liver itself (e.g. active transport sites). The biotransformation process is usually separated into two phases: Phase I and Phase II (Higdon, J. V. et al.; 2007)(Rose, R. L.; Hodgson, E.; 2004)(Brandon, E. F. A. et al.; 2003)(Penner, N. et al.; 2012)(Gaikovitch, E. A.; 2003)(Sturgill, M. G; Lambert, G. H.; 1997)(Rollins, D. E.; Klaassen, C. D.; 1979)(Klaassen, C. D.; 1975)(Mutlib, A. E. et al.; 2000).

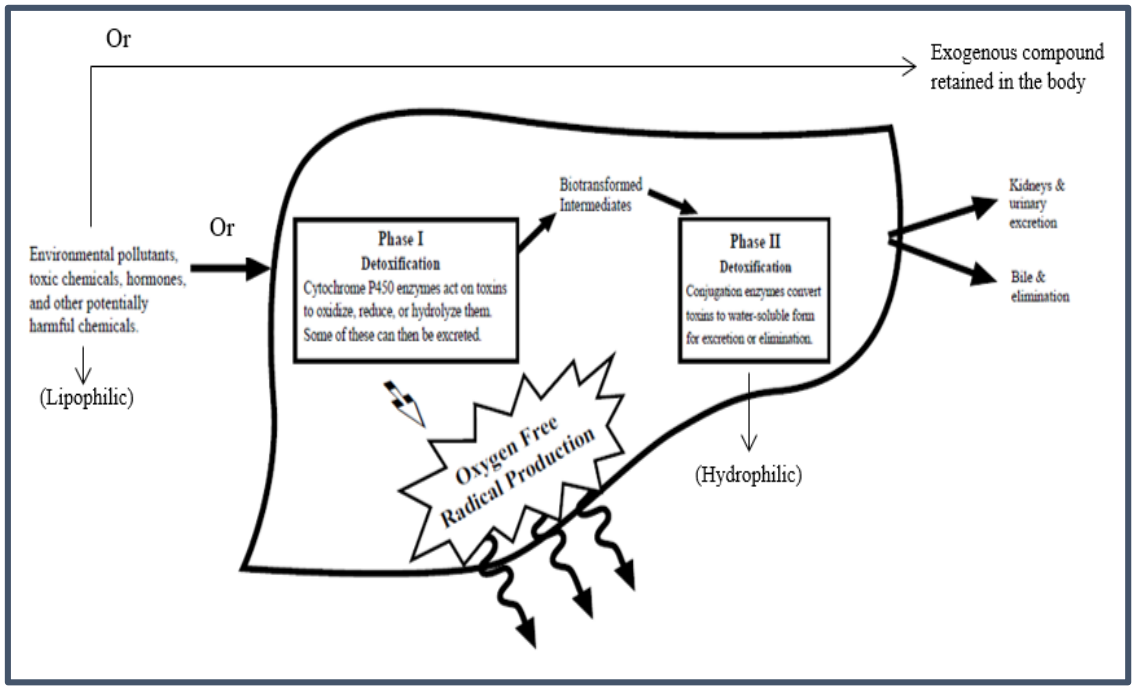


Fig.: 3.14 – Importance of the biotransformation process for the removal of exogenous substances from the body (Guttmacher, A. E.; 2003)(Percival, M.; 1997).

3.30 Biotransformation Enzymatic Reactions and the Corresponding Enzymes

Xenobiotics can be metabolized by Phase I (PI) reactions, also called functionalization reactions (e.g. oxidation, reduction, and hydrolysis) and by Phase II (PII) reactions (conjugation). Phase I reactions are responsible for exposing or adding a functional group (e.g. -OH, -NH_2, -SH or COOH) to the parent compound resulting in a mild increase in its polarity. On the other hand, Phase II reactions are responsible for conjugating endogenous cofactors/ligants (i.e. water-soluble derivatives) with functional groups of the substrate in order to increase its hydrophilicity substantially, when compared with Phase I reactions. A compound can be metabolized by both biotransformation phases, one after the other (i.e. Phase I followed by Phase II), or by solely one of them, without the need of a phase precedence over the other. Though in general terms, it is more common for Phase II reactions to happen after Phase I. Phase II reactions occur much faster than Phase I reactions and as such, Phase I reactions are the limiting step in the detoxification process (i.e. excretion) (Nebert, D. W.; 1997)(Jancova, P. et al.; 2010)(Gerhauser, C. et al.; 1997)(Brandon, E. F. A. et al.; 2003)(Penner, N. et al.; 2012)(Gaikovitch, E. A.; 2003)(Wilkening, S. et al.; 2003)(Higdon, J. V. et al.;

2007)(Talalay, P; Fahey, J. W.; 2001)(Wilkening, S. et al.; 2003)(James, D. et al.; 2012)(Percival, M.; 1997)(Wolf, C. R.; 2001).

3.31 What is the Importance of Biotransformation?

Rough estimates pointed out to the fact that nearly ¼ of all carcinogenic compounds are directly toxic in their natural; unmetabolize state, while the ¾ remaining of them are metabolically activated from pro-carcinogens to carcinogenic substances. Phase I enzymes are involved in both detoxification and carcinogen activation (i.e. bioactivation), albeit the last is not a common event. Bioactivation is achieved by increasing the polarity of a parent compound (in this case, a pro-carcinogen), which leads to the formation of a considerably more reactive compound [reactive oxygenated intermediate (ROI or ROM), which is the activated carcinogen] in comparison. Phase II enzymes are generally involved in xenobiotic or carcinogen detoxification through conjugation reactions in order to complete the detoxification cycle (Gaikovitch, E. A.; 2003)(Nebert, D. W.; 1997)(Brandon, E. F. A. et al.; 2003)(Gerhauser, C. et al.; 1997)(Hecht, S. S.; 2004)(Jancova, P. et al.; 2010)(Irigaray, P.; Belpomme, D.; 2010)(Higdon, J. V. et al.; 2007)(Nakajima, M. et al.; 2001)(Wilkening, S. et al.; 2003)(Nebert, D. W.; 1997)(Nebert, D. W.; Dalton, T. P.; 2006)(Rodriguez-Antona, C; Ingelman-Sundberg, M.; 2006)(Yoshigae, Y. et al.; 2013).

3.32 Fundamentals of Cytochromes P450

Cytochromes P450 (cytochromes P_{450}, CYPs, P450s or CYP450s) are responsible for 70-80% of all Phase I xenobiotic metabolism, which makes them the main DMEs responsible for catalysing the reactions that can lead to carcinogen activation. CYPs exist in almost all the tissues in the body (e.g. small intestine, lungs, kidneys, placenta) but they are primarily expressed in the smooth endoplasmic reticulum of hepatocytes (i.e. liver cells), followed by mitochondrial expression. CYPs are a superfamily of heme-thiolate enzymes responsible for catalysing oxidation reactions mostly, and sometimes reduction and hydrolysis reactions. In addition to xenobiotic metabolism, CYPs are also involved in several other metabolic pathways (e.g. endobiotics metabolism), including: steroid, bile acids, corticosteroids and cholesterol biosynthesis, retinoic acids hydroxilation, synthesis and degradation of biogenic amines, metabolism of eicosanoids, vitamin D_3 synthesis and metabolism, fatty acid metabolism (prostacyclins, thromboxane A_2), and maintenance of calcium homeostasis (Guengerich, F. P.; 2008)(Gaikovitch, E. A.; 2003)(Irigaray, P.;

Belpomme, D.; 2010)(Penner, N. et al.; 2012)(Jancova, P. et al.; 2010)(Rodriguez-Antona, C.; Ingelman-Sundberg, M.; 2006)(McKinnon, R. A. et al.; 2008)(Lynch, T.; Price, A.; 2007)(Sridhar, J. et al.; 2012)(Badyal, D. K.; Dadhich, A. P.; 2001)(Monga, S. P. S.; Cagle, P. T.; 2010)(Nassar, A. F. et al.; 2009)(Mi, L. et al.; 2011)(Talalay, P.; Fahey, J. W.; 2001)(Nebert, D. W.; Dalton, T. P.; 2006).

3.33 Human CYP Families Responsible For Carcinogenesis

CYPs are versatile monooxigenases (i.e. their principal action is to perform hydroxylations), though they can also act as dioxygenases and hydrolases. These Phase I enzymes are responsible for many reactions that take place in the body. They have many substrates that can be catalysed upon, though some CYPs also possess a certain degree of substrate specificity. In the human genome, 57 active *CYP* genes and 58 pseudogenes are known to exist, of which five of them (i.e. *CYP1A2*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, and *CYP3A5*) are held responsible for nearly 90-95% of all the drugs that are metabolized in the liver. *CYP* families *CYP1-3/4* are considered the ones involved in drug metabolism, with *CYP1-2* being associated with exogenous carcinogen activation, while the remaining *CYP* families are associated with catalysing enzymatic reactions upon endogenous substrates, which may lead to endogenous tumour promotion. Indeed, CYPs seem to be a major player in terms of xenobiotic metabolism and carcinogen activation, with genes that can count up to 1% of the human coding genome (Guengerich, F. P.; 2008)(Irigaray, P.; Belpomme, D.; 2010)(Penner, N. et al.; 2012)(Jancova, P. et al.; 2010)(Kelly, S. L. et al.; 2006)(Rodriguez-Antona, C.; Ingelman-Sundberg, M.; 2006)(Lynch, T.; Price, A.; 2007)(Sridhar, J. et al.; 2012)(Monga, S. P. S.; Cagle, P. T.; 2010)(Nassar, A. F. et al.; 2009)(Jushchyshyn, M. I. et al.; 2003)(Ingelman-Sundberg, M.; Rodriguez-Antona, C.; 2005).

3.34 How Do CYPs Work?

Since CYPs are monooxigenases, they use oxygen in their reaction mechanism to oxidise their substrates. The oxidative cycle (or redox cycling) performed by CYPs is a rather complex process with many players involved such as: CYP enzyme, nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FAD)-containing P450 reductase, among others (Gaikovitch, E. A.; 2003)(Guengerich, F. P.; 2008)(Penner, N. et al.; 2012)(Yang, C. S. et al.; 1992)(Monga, S. P. S.; Cagle, P. T.; 2010)(Nassar, A. F. et al.; 2009).

An inactive CYP has its iron atom (Fe) of the heme group (i.e CYP active site) in the ferric state (Fe^{3+}). When the CYP binds a ligand (e.g. O_2 or CO), the Fe atom is reduced to the ferrous state (Fe^{2+}). After ligand binding occurs, the CYP enzyme can now activate molecular oxygen (O_2), with the electrons provided by NADPH (electron donor) via a FAD-containing P450 reductase (electron bridge – also simply known as NDAPH-P450 reductase or by its other name NADPH oxidoreductase). One atom of the molecular oxygen is incorporated in the CYP´s substrate (RH), converting it in the oxidized form (ROH) while the other oxygen is converted to H_2O using reducing equivalents donated by NADPH. Afterwards, the CYP enzyme releases the now oxidised substrate (ROH) and enters the resting (inactive) state where the heme Fe atom in its oxidised state (Fe^{3+}). NADPH-P450 reductase is required, because CYP enzymes are unable to directly bind NADPH; as such NADPH oxidoreductases help in the CYP activity by transferring the electrons necessary for the oxidative reaction from NADPH to the CYP. Moreover, there is another important enzyme that has a crucial role to perform in the CYP oxidative cycle, Cytochrome b5. This flavoprotein is also involved in the transference of electrons, but its main role is to increase the reduction rate and the substrate binding affinity of the CYP enzyme. With respect to mitochondrial CYP enzymes, ferrodoxin and ferrodoxin reductase are two other enzymes that are important in the electron transference process that takes place in the mytochrondria (Gaikovitch, E. A.; 2003)(Guengerich, F. P.; 2008)(Penner, N. et al.; 2012)(Monga, S. P. S.; Cagle, P. T.; 2010)(Nassar, A. F. et al.; 2009).

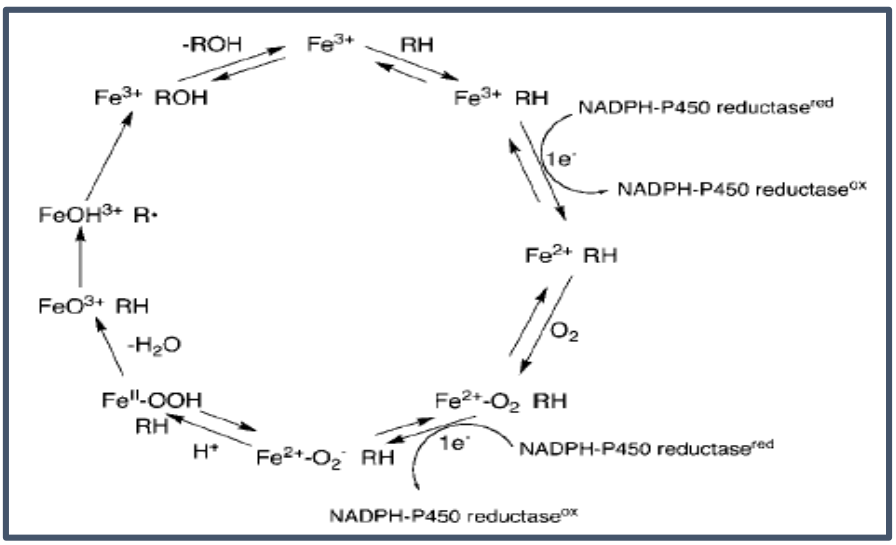


Fig.: 3.15 – Catalytic mechanism (i.e. oxidative cycle) of the heme group of CYP enzymes. RH – substrate; ROH – oxidised substrate; Fe^{3+} – iron in the ferric state; Fe^{2+} – iron in the ferrous state; red – reduced; ox – oxidised (Guengerich, F. P.; 2008).

3.35 ITCs and CYPs

There are compounds present in the diet capable of binding to the active site of CYPs and serve as substrates or inhibitors of these enzymes. ITCs, which are dietary molecules, are capable of modulating CYP activity through several distinct mechanisms depending on the CYP isoform and the specific ITC in question. In general ITCs inhibit CYPs, though there has been some ITCs reported to be capable of inducing some of these enzymes. A plausible explanation for this seems to be the fact that some *CYP* genes have the xenobiotic response element (XRE) in their gene regulatory regions, which works in combination with the aryl-hydrocarbon receptor (AhR), which in turn is activated by binding ITCs. Additionally, some ITCs are what is called bifunctional inducers (e.g. they are capable of inducing both Phase I and II enzymes). Nevertheless, the consensus is that ITCs are primarily CYP inhibitors. ITC-mediated inhibition of CYPs works in the catalytic activity unit of these enzymes in a multitude of different ways with some of them being reversible and others irreversible, namely: direct inhibition of the CYP catalytic activity by binding of ITC to crucial residues (e.g. Cys) in the enzyme´s active site that ends up interfering with the binding of the substrate and O_2, covalent protein modification (e.g. NADPH), competitive inhibition, non-competitive inhibition, uncompetitive inhibition or inhibition by a mechanism-based inactivator. Noteworthy is the fact that the –N=C=S group, characteristic of ITCs, seems to be essential for the inhibition process (Yang, C. S. et al.; 1992)(Zhang, Y.; Talalay, P.; 1994)(Hecht, S. S.; 1999)(Spitz, M. R. et al.; 2000)(Nakajima, M. et al.; 2001)(Jushchyshyn, M. I. et al.; 2003)(Wallig, M. A. et al.; 2005)(Gasper, A. V. et al.; 2007)(Peterson, S. et al.; 2009)(Wu, X. et al.; 2009)(Mi, L. et al.; 2011)(Navarro, S. L. et al.; 2011)(Zhang, Y.; 2012)(Yoshigae, Y. et al.; 2013)(Hecht, S. S.; 2004)(Marca, M. L. et al.; 2012).

3.36 A More Detailed Approach To Some ITCs-CYPs Inhibition Processes

A diverse array of synthetic and natural ITCs is able to inhibit several different CYP enzymes through different mechanisms, which results in cancer chemoprevention. This fact is, however, dependent on the specific type of ITC used, the CYP isoform responsible for carcinogen activation, the animal species as well as the target tissue, and the specific carcinogen that was employed in the animal experimental treatment (Mori, Y. et al.; 2005)(Wu, X. et al.; 2009)(Marca, M. L. et al.; 2012)(Nakajima, M. et al.; 2001)(Conaway, C. C. et al.; 1996)(Yoxall, V. et al.; 2005)(Zhang, Y.; 2012)(Molina-Vargas, L. F.; 2013).

3.37 CYP inhibition by Benzyl-ITC

Benzyl isothiocyanate, also known by the name benzyl-ITC (BITC) is a natural alkyl-aryl aromatic ITC that is especially abundant in garden cress plants in the form of the precursor GSL glucotropeolin. BITC is capable of inhibiting several rat CYP enzymes (CYP1A1, 1A2, 2B1, and 2E1) in *in vivo* rat liver microsomes as well as other human CYP proteins (2B6, and 2D6) and a rabbit CYP enzyme (2E1) by using a mechanism-based inhibition process. CYP2B1 is inactivated by BITC in a time- and concentration-dependent manner by binding and covalently modifying the CYP protein. Both human CYPs 2A6 and 2A13 enzymes are inhibited by this ITC (Marca, M. L. et al.; 2012)(Zhang, Y.; 2012)(Munday, R. et al.; 2008)(Nakajima, M. et al.; 2001)(Rao, C. V. et al.; 1995)(Cheung, K. L.; Kong, A.; 2009)(Wu, X. et al.; 2009)(Yang, X. F.; Zeng, F. D.; 2006)(Smith, T. J. et al.; 1990)(Stoner, G. D. et al.; 1998)(Molina-Vargas, L. F.; 2013)(Goosen, T. C. et al.; 2001)(Zhang, Y. et al.; 2003)(Schlicht, K. E.; 2007)(Moreno, R. L. et al.; 1999)(Moreno, R. L. et al.; 2001)(Goosen, T. C. et al.; 2000)(Bendich, A.; Deckelbaum, R. J.; 2005).

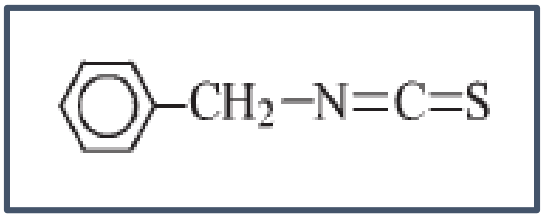


Fig.: 3.16 – Molecular structure of BITC (Tang, L.; Zhang, Y.; 2004)

3.38 CYP inhibition by Phenethyl-ITC

Phenethyl isothiocyanate, also termed phenethyl-ITC (PEITC) is also a natural alkyl-aryl aromatic ITC like the previously mentioned BITC. PEITC is commonly found in high quantity in watercress plants as the GSL gluconasturtin. PEITC is capable of inhibiting through multiple inhibition processes (i.e. competitive, non-competitive, mixed-type of competitive and non-competitive, and mechanism-based inhibition) a great multitude of human CYP isoenzymes (i.e. e.g. CYP1A2, 2A6, 2B6, 2C9, 2C19, 2D6, 2E1, and 3A4) that were, at the time, expressed in *in vivo* microsomes from baculovirus-infected cells. PEITC is also capable of inactivating CYP2B1 by protein modification in a time- and concentration dependent manner and direct inhibition of CYPs 1A1, 1A2, and 2B1 in a dose-dependent way in rat liver microsomes. With regard to CYP1A2 inhibition,

PEITC is capable of binding to the CYP binding site and work as a competitive inhibitor; it is also capable of being metabolized by the CYP1A2 into a reactive intermediate metabolite that will then proceed to react with the enzyme, resulting in its inactivation by a mechanism called suicide inhibition. The rat CYPs 2E1 (inhibited by competitive inhibition in rat liver microsomes) and 2A3 as well as the human CYP2A13 enzyme are also inhibited by this ITC (Zhang, Y.; 2012)(Munday, R. et al.; 2008)(Nakajima, M. et al.; 2001)(Marca, M. L. et al.; 2012)(Conaway, C. C. et al.; 1996)(Rao, C. V. et al.; 1995)(Cheung, K. L.; Kong, A.; 2009)(Wu, X. et al.; 2009)(Yang, X. F.; Zeng, F. D.; 2006)(Smith, T. J. et al.; 1990)(Barcelo, S. et al.; 1996)(Smith, T. J. et al.; 1996)(Stoner, G. D. et al.; 1998)(Molina-Vargas, L. F.; 2013)(Conaway, C. C. et al; 1999)(Goosen, T. C. et al.; 2001)(Zhang, Y. et al.; 2003)(Schlicht, K. E.; 2007)(Morris, C. R. et al.; 2004)(Thapliyal, R.; Maru, G. B.; 2001)(Moreno, R. L. et al.; 2001)(Ishizaki, H. et al.; 1990)(Wildman, R. E. C. et al.; 2002)(Ioannides, C.; 2008).

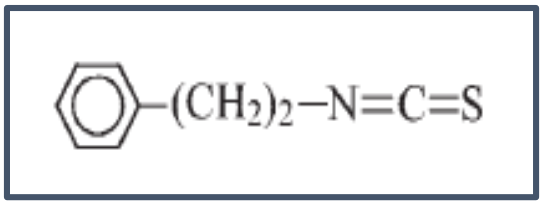


Fig.: 3.17 – Molecular structure of PEITC (Tang, L.; Zhang, Y.; 2004)

3.39 CYP inhibition by 3-phenylpropyl-ITC And 6-phenyhexyl-ITC

Both 3-phenylpropyl isothiocyanate, also known as 3-phenylpropyl-ITC or simply as phenylpropyl (PPITC), and 6-phenylhexyl isothiocyanate, also termed 6-phenyhexyl-ITC or merely phenylhexyl (PHITC), are synthetic alkyl-aryl aromatic ITCs. These two compounds are known inhibitors of CYP2B1 (by competitive inhibition) and of human CYPs 2A6 and 2A13, with PPITC being capable of inhibiting them (i.e. the two last mentioned CYPs) by binding to the active sites of both of these enzymes. PPITC and PHITC are both longer alkyl chain homologues of the previously mentioned compound PEITC. The reason as to why they were initially synthesized to have longer alkyl chains might have to do with the fact that the length of the alkyl chain of the phenylalkyl moiety (Ph-[CH_2]n-NCS, where n=0-6, 8, and 10) of ITCs constitutes an important factor in determining the inhibitory effect of the ITC (i.e. the longer the alkyl chain, the higher the inhibitory potency of the ITC). One such example is the fact that PHITC is nearly 50-100

times more powerful at inhibiting tumorigenesis than PEITC. This event might be explained if one considers the fact that the substrate access channel of the CYP enzyme active site may simply accommodate ITCs with variable alkyl lengths up to a certain point and that it may also have an optimal substrate binding preference for ITCs with a specific number of alkyl carbons. Another consequence of some ITCs having a longer alkyl chain moiety than others is the fact that it increases the ITC´s lipophilicity and additionally results in a mild chemical reactivity of the ITC towards GST (i.e. there is less ITC being eliminated, which translates into its cellular accumulation in the body), which might explain why PHITC is a better inhibitory agent than PEITC. Unfortunately, this effect has only been observed for ITCs with an alkyl chain length between C_8-C_{10}, the inhibitory effect declining for longer phenyl moieties. Furthermore, the length of the alkyl chain in ITCs is not essential for their inhibitory activity, but is rather responsible for its enhancement (Samaha, H. S. et al.; 1997)(Nishikawa, A. et al.; 1996)(Stoner, G. D. et al.; 1998)(Smith, T. J. et al.; 1990)(Rao, C. V. et al.; 1995)(Marca, M. L. et al.; 2012)(Conaway, C. C. et al.; 1996)(Wu, X. et al.; 2009)(Smith, T. J. et al.; 1990)(Stoner, G. D. et al.; 1998)(Conaway, C. C. et al.; 1999)(Nishikawa, A. et al.; 1996)(Schlicht, K. E.; 2007)(Meskin, M. S. et al.; 2002)(Weymarn, L. B. von et al.; 2007)(Goosen, T. C. et al.; 2000).

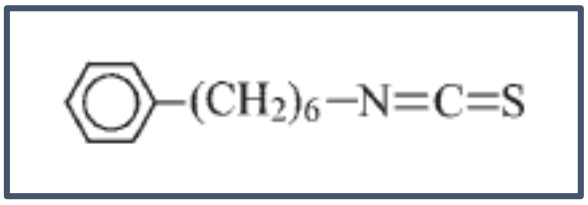


Fig.: 3.18 – Molecular structure of PHITC (Tang, L.; Zhang, Y.; 2004)

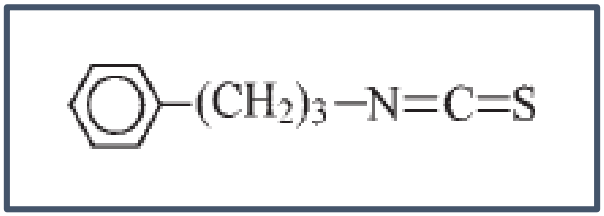


Fig.: 3.19 – Molecular structure of PPITC (Tang, L.; Zhang, Y.; 2004)

4. Materials and Methods

4.1 Introduction

A growing interest in *in silico* molecular modelling methods, and their widespread used has been seen in recent years, largely due to the increasing processing speed of digital computers and major devolopments in methodology.

In order to correctly model the interaction of a small ligand molecule with a macromolecule, certain aspects must be considered: the size of the molecules, the type of interactions (e.g. hydrogen bonds) between the molecules (ligand and receptor), the stabilities of both molecules and their electronic properties. All of these factors are to be taken into account when performing molecular modelling. This in turn allows the development of algorithms that are capable of correlating the chemical properties of certain molecules with their structures and, in the case of this work, to dock ligand-receptor molecule pairs, producing docking results that are accurate when compared with experimental data.

Various theoretical approaches can be used in the modelling of molecules: the more rigourous, but slow, *ab initio* or molecular quantum chemical method; the semi-empirical quantum chemical method; and the empirical force-field calculation method, more commonly known as molecular mechanics. For the modelling of large molecules (e.g. proteins and other biological macromolecules), the last method is the most frequently used one due to its simplicity and computationa efficiency. Molecular mechanics methodos can be used for various types of calculations, including energy minimization, conformational searching, molecular dynamics and receptor-ligand docking.

Force-field based molecular mechanics methods are often described as "classical" since they produce classical trajectories for point charge/mass atoms. *Ab* initio and semi-empirical methods are quantum methods yielding molecular wave functions of the system.

4.2 The Quantum Approach – *Ab Initio*

The *ab initio* calculation methods are quantum-based, using the Schrödinger equation to compute the energies and corresponding wave functions of a quantum system (e.g. molecule) containting electrons and nuclei. The *ab initio* method will use a basis set

for each molecule (i.e. finite set of atomic orbitals or basis functions centred in the nucleus of each atom that composes the molecule at hand to describe molecular orbitals that emcompass all the electrons in an atom) in order to solve the Schrödinger equation for that said molecule.

$$\left[\hat{T} + \hat{V}(\vec{r})\right]\psi(\vec{r}) = E\psi(\vec{r})$$

Fig.: 4.1 – The Schrödinger equation. The wave function (Ψ) is dependente on the position of $\vec{r}$. E is the energy of a quantum state described by a wave function Ψ , T and V are the kinetic and potential energy operators and their sum is generally represented as the single H (Hamiltonian operator).

$$\Psi_M = \sum_i c_i \psi_{Ai}$$

Fig.: 4.2 – Molecular orbital wave function (ΨM), ψ_{Ai} are the atomic orbital functions and *ci* are the weights of every atomic orbital function in the molecular function.

As an exact analytical solution of the Schrödinger equation for polyelectronic atoms does not exist, several approximate methdos were devised resulting in calculations that can be lengthy and difficult. The Born–Oppenheimer approximation postulates that electronic and nuclear degrees of freedom can be separated, so that the molecular wave function can be solved for a given set of nuclear coordinates. The simples and oldest *ab* initio method is the Hartree-Fock [or self-consistent field (SCF)] approximation.

$$\Psi_{\text{total}} = \psi_{\text{electronic}} \times \psi_{\text{nuclear}}$$

$$E_{\text{total}} = E_{\text{electronic}} + E_{\text{vibrational}} + E_{\text{rotational}} + E_{\text{nuclear}}$$

Fig.: 4.3 – In the Born–Oppenheimer approximation, the wave function of the molecule can be separated into its two components (electronic and nuclear), while also separating the vibrational and rotational components (see Bibliography website section for references).

Using the Born–Oppenheimer approximation, the Schrödinger equation is solved by assuming the nuclei are in fixed spatial positions. The Hartree-Fock (SCF) methods

with an initial guess solution of the Schroedinger eqation and progressively refines it until convergence (self-consistency) is attained.

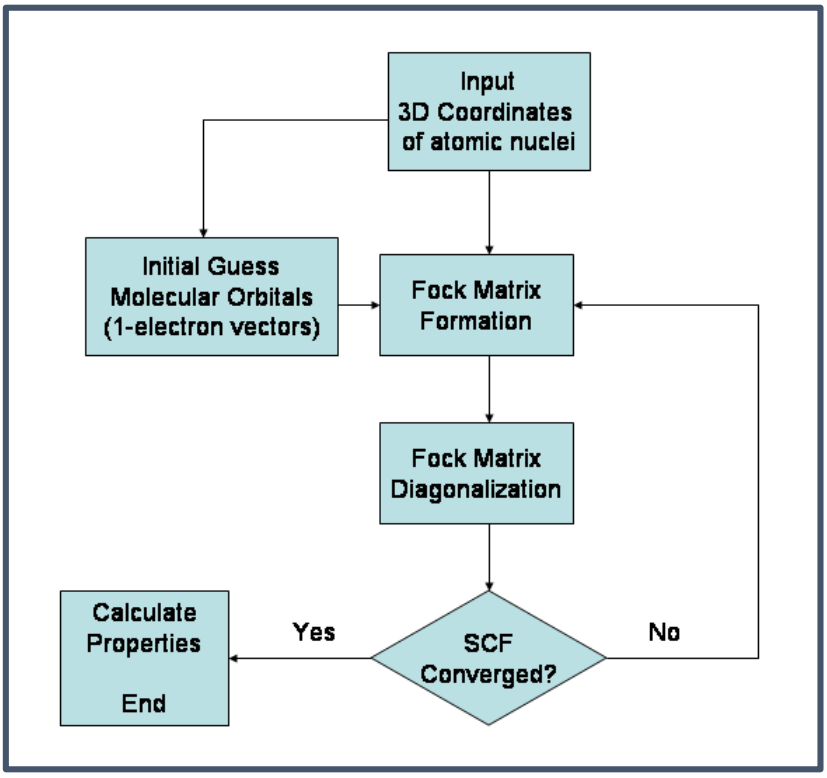


Fig.: 4.4 – The Hartree-Fock method algorithm (see Bibliography website section for references).

Generally, the ab initio calculation method is used when accurate values for molecular properties or parameters are required, or when studying mechanisms involving bond breaking/formation (e.g. catalysis), electron transfer, excited states, etc. In spite of this cases, molecular mechanics is mosty frequently employed when modelling due to its speed and simplicity. .

4.3 The Classical Approach – Molecular Mechanics

Contrary to *ab initio*, molecular mechanics (MM) is based on classical mechanics and it uses the atomic coordinates, charges and masses instead of atomic orbitals when modelling. MM also relies on the Born–Oppenheimer approximation but with a crucial difference: while quantum methods look for the electronic structure while keeping the nuclei fixed in space, the MM method will consider nuclear movements and treat the corresponding electrons only in an indirect fashion. This happens because the electrons and the nucleus of atoms are seen as aggregate point mass/charges, while the bonds

connecting them are seen as “springs” (harmonic approximation). In other words, MM will treat the molecule as if it was merely a collection of atoms bounded by elastic forces (bending, torsional energy, van der Waals energy, and electrostatics). These forces are are calculated from a series of potential energy functions comprising the molecular forcefield, and can be used to determine the lowest energy conformation of a molecule.

$$U = \Sigma\, U(r) + \Sigma\, U(\theta) + \Sigma\, U(\phi) + \Sigma\, U(d) + \ldots$$

Fig.: 4.5 – Potential molecular energy or molecular forcefield (U) corresponding to the sum of various energetic contributions, or potentials [bond length (r), bond angles (θ), dihedral angles (φ) and non-bonded interactions (d)] (Coelho, Lilian W. et al.; 1999).

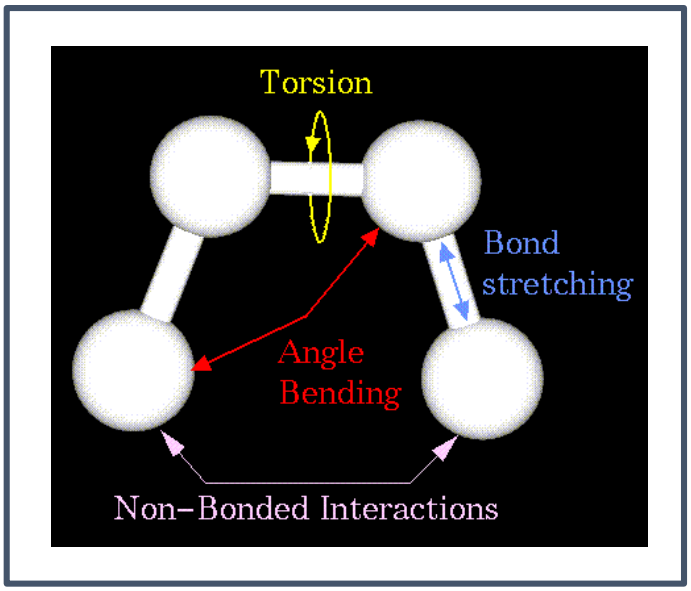


Fig.: 4.6 – The different forces that contribute to the potential energy of a molecule (see Bibliography website section for references).

The term force-field corresponds to the set of energy functions and parameters used in the MM calculation, or in other words, it is a potential energy function used to calculate the potential energy of a molecular conformation as the sum of different individual terms. These force-fields contain multiple parameters describing different types of bonded and nonbonded interactions (i.e. atomic distances, bond angles, electrostatic interactions, van der Waals potentials), which can be derived from experimental work or quantum mechanical calculation.

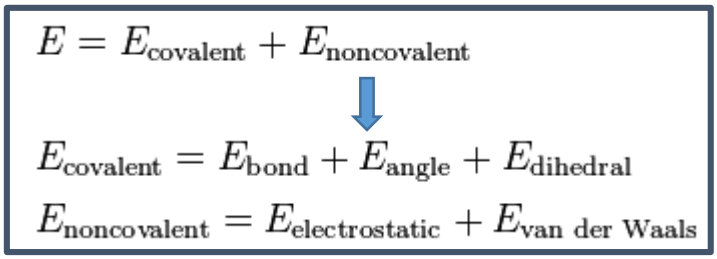


Fig.: 4.7 – The covalent and non-covalent potentials that make up the force-field can be decomposed into simpler terms (see Bibliography website section for references).

These force-field parameters remain fairly constant for the same type of atoms, bonds and chemical functions, which means that the same parameters can be used in many different strucutres as long as the atom types and chemical groups remain the same. Nowadays, there are many force-fields available for use in MM, with ever increasing terms to correctly describe a molecular structure. These force-fields usually differ in the form and parameters of the potential energy terms and in the type of system they are best suitedo to.

4.4 The Classical Approach – Molecular Dynamics

Molecular dynamics (MD) is used to simulate the movement of atoms and molecules by integrating the equations of motion of the system under the particular forcefield, allowing a glimpse of possible fluctuations and conformational changes that happen to a molecule or the trajectory of atoms in space during a fixed time window. This method allows the study of complex and dynamic biological processes (e.g. protein stabilitie, protein folding, conformational changes). By solving the Newtonian equations of motion of the system, it is possible to predicte how a molecule might move in space when the system is perturbed. Atoms move and interact with one another through the forces applied to them. These forces are calculated from the potential as described by the MM molecular forcefield.

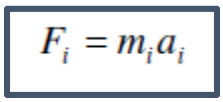


Fig.: 4.8 – The Newton equation of motion, where *Fi* corresponds to the force that leads to an atom´s acceleration (*ai*). The atom´s mass is *mi*. *Fi* can also be derived from the already mentioned potential energy (*U*) with respect to the atom´s coordenates.

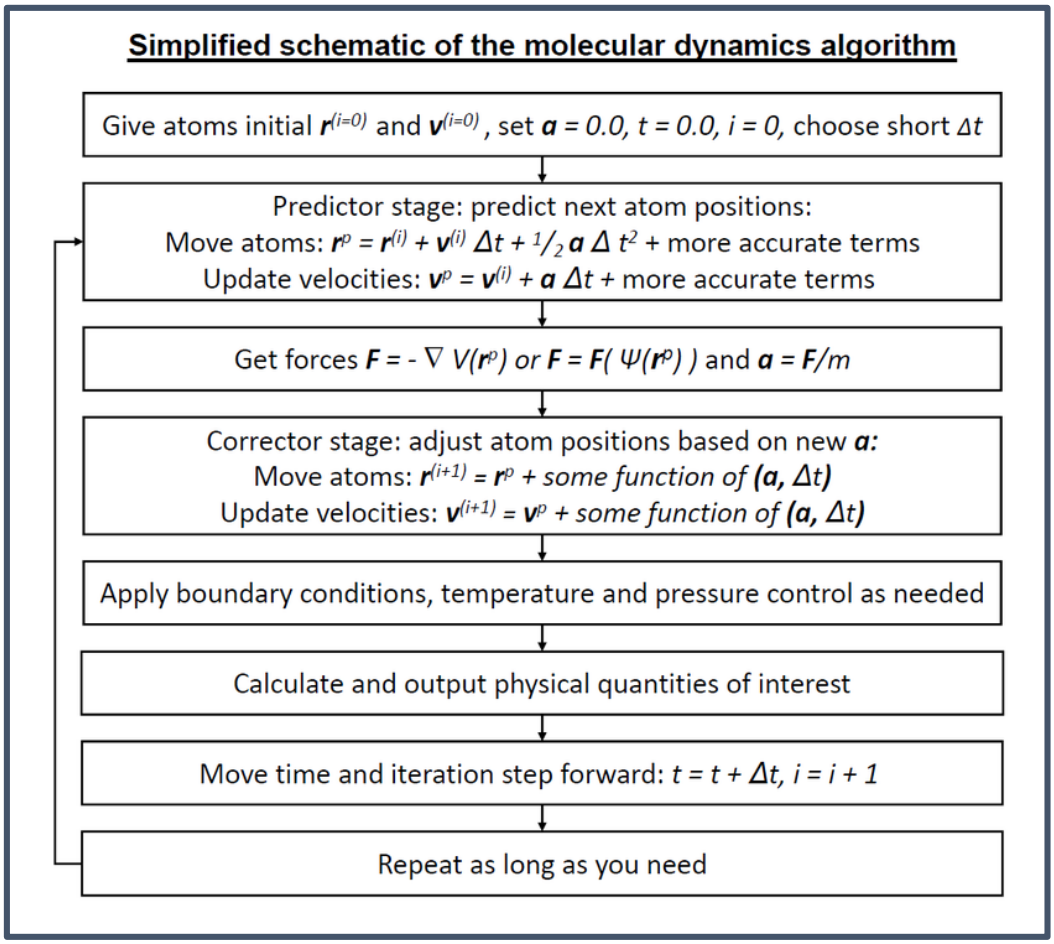


Fig.: 4.9 – A simplified example of a MD algorithm (see Bibliography website section for references).

An exact analytical solution to the Newton equations of motion is only possible for two-particle systems. For anything beyond that, only approximate solutions can be obtained for the atomic movements.Various numerical methods have been developed to integrate Newton´s laws by assuming that the inicial atom positions, velocities, and accelerations are know.

$$r(t+\Delta t) = r(t) + \frac{dr}{dt}\Delta t + \frac{d^2 r}{dt^2}\frac{\Delta t^2}{2} + \ldots$$

Fig.: 4.10 – An approximate numerical solution for more than two atoms systems can be obtained using a Taylor series expansion. Its solution is dependente on knowing the initial position of the atoms *r(t)*, their speed *dr/dt*, the corresponding atom´s acceleration *d2r/dt2* as well as other important factors.

Accuracy of the solutions depends on the time timestep used – smaller time steps provide beter solutions but lengthier calculations. For biomolecular simulations time steps of the order of the femtosecond (10^{-15}s) are normally used, and total simulations times range from pico- (1ps = 10^{-12} s) to nanosecond (10^{-9} s), though microsecond simulations are becoming commonplace. Small timesteps and long simluations times will result in extremely long simulations that can run for weeks or months on a desktop computer.

4.5 Docking Program Methodologies

Both programs used [AutoDock Vina (Vina) and AutoDock4 (AD4)] work in a similar way in the sense that they use a scoring function and a search algorithm in order to dock the ligand to its target macromolecule. The main difference lies in the way in which they operate and in the forms of both the scoring function and the search algorithm. Scoring functions are used to predict the energy of a certain ligand conformation and the choice of a particular function heaviuly depends on the intended speed and/or accuracy of the calculation. On another hand, the search algorithm will try and find the best (having the lowest value of the scoring function) docked pose of the ligand in the docking site. Autodock uses MM-type forcefields in order to dock a molecule to its target receptor.

$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

$$W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+$$

$$W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)+$$

$$W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+$$

$$W_{sol}\sum_{i,j}(S_i V_j + S_j V_i)e^{(-r_{ij}^2/2\sigma^2)}+$$

$$W_{tor}N_{tor}$$

Fig.: 4.11 – The scoring function of AD4, which takes into account the dispersion/repulsion forces, electrostatic forces, influence of hydrogen bonds, desolvation and torsional entropy of the ligand (see Bibliography website section for references).

$$\Delta G_{binding} = \Delta G_{gauss} + \Delta G_{repulsion} + \Delta G_{hbond} + \Delta G_{hydrophobic} + \Delta G_{tors}$$

ΔG_{gauss}
Attractive term for dispersion, two gaussian functions

$\Delta G_{repulsion}$
Square of the distance if closer than a threshold value

ΔG_{hbond}
Ramp function - also used for interactions with metal ions

$\Delta G_{hydrophobic}$
Ramp function

ΔG_{tors}
Proportional to the number of rotatable bonds

Fig.: 4.12 – Scoring function of Vina, which uses several terms common to AD4, while at the same time not using others (e.g. electrostatics) (see Bibliography website section for references).

As for the search algorithm, while Vina uses a systematic algorithm, AD4 uses a stochastic one. The main differences between them are the following: though the systematic algorithm is exhaustive and deterministic it is only adequate for docking small ligands. The Stochastic algorithm, on the other hand, is random and, therefore the outcomes will varie but it is feasible for larger atom systems.

The iterated local search algorithm is a way that was devised in order to get better results by doing a random perturbation to the system and start the local search again but from another point in its neighbourhood. This will allow the generation of better results than the ones produced from repeated random attempts.

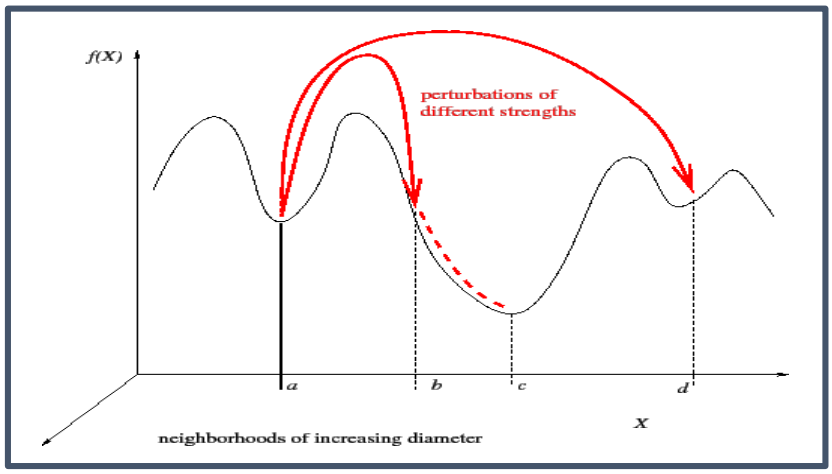


Fig.: 4.13 – The iterated local search algoritm works by doing perturbations to the system and then trying to find a new local minimum (see Bibliography website section for references).

The genetic algorithm used by Autodock will find the best solutions by following a Darwinian evolutionary path (i.e. it looks for new solutions by randomly mutating and crossing over previous solutions and selecting mutants according to their "fitness").

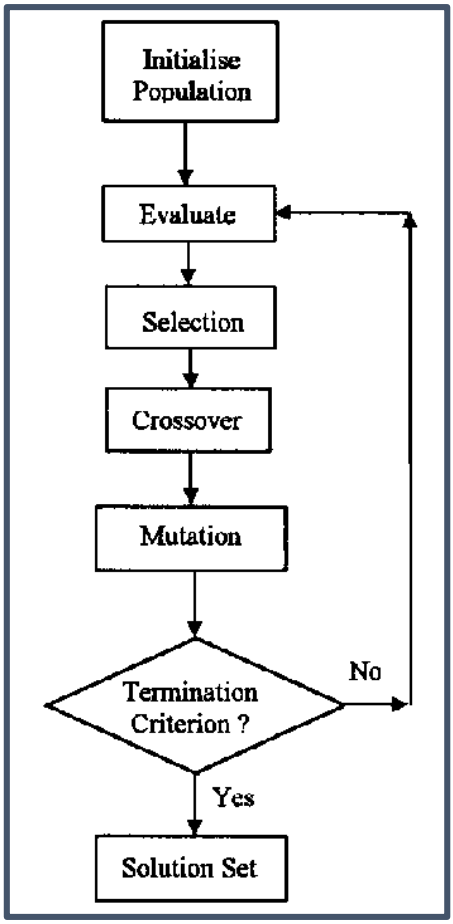


Fig.: 4.14 – The genetic algorithm works by applying to the system the same evolutionary principles that govern biological systems (see Bibliography website section).

4.6 Molecular Construction of the ITC molecules

Models for the previously mentioned ITC molecules (BITC, PEITC, PHITC, and PPITC) were created in a molecular modelling environment provided by the program "HyperChem".

HyperChem is an easy, useful and userfriendly software for building many kinds of molecules, which can then be visualized in a tridimensional (3D) fashion by the user. This program is also capable of performing several other different tasks related to the molecule at hand.

The ITCs 3D molecular structures were drawn in HyperChem, with the different atoms and bond types connecting them. Structures were finished by adding hydrogens (H) to them with the option "model build", which ensures chemically correct structures. After this step, electrostatic point charges for all atoms belonging to the ITC structures were calculated using an *ab initio* quantum mechanical method, (Hatree-Fock single point calculation with medium complexity basis set. The molecule was then setup for the molecular mechanics forcefield MM+ and a 10ps molecular dynamics simulation was ran at 300K. This was done to ascertain the molecular dynamics behaviour (i.e. fluctuations and conformational changes) of the ITC structures. Furthermore, this simulation step also enables the user to see how the atoms that make up the molecule act when they are allowed to move and interact with one another in the structure and in a physical environment. Finally, a geometry optimization (energy minimization) was done on the end structure of the MD simulation, to produce a stereochemically correct molecular structure.

4.7 Getting the CYP Molecules

The human CYP molecules (CYP2A6, 2A13, 2B6, and 2C9) crystallographic structures needed for the *in silico* molecular docking experiment were downloaded from the Protein Data Bank (PDB) website (http://www.rcsb.org/pdb/home/home.do). This website consists of a database of structures for biological macromolecules and their respective ligands (if present), obtained using mostly two experimental methods: X-ray crystalography and Nuclear Magnetic Resonance (NMR) spectroscopy (see Annex 7.1 and 7.2).

4.8 X-ray Crystalography

The majority of the protein strucutres stored in the PDB website were determined by X-ray crystallography. This technique employs X-ray diffraction to determine the atom positions in a crystal sample. Firstly, crystals of our macromolecule must be produced. Afterwards, an X-ray beam is shined on the crystal, resulting in multiple diffracted beams along specific directions, producing a pattern that depends on arrangement of atoms in the crystal lattice. It is this diffraction pattern, which is unique for each protein that allows the determination of its structure. After data analysis, structure determination, refinement and validation, the final structure is uploaded to the PDB website.

4.9 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is the second most frequently employed technique used to determine protein structure. As opposed to what happens in X-ray crystallography technique, in which a crystalized form of the protein is needed to determine its molecular structure; the NMR protein structure determination can be done simply with an aqueous solution of our sample. This fact is very important, since not all proteins can be crystalized easily or at all (e.g. membrane proteins). This technique works by taking advantage of the effect of a strong magnetic field upon the nuclear spin of certain types of atoms. The spin will tend to orient under the magnetic field along, emiting or absorbing electromagnetic radiation at a precise frequency (i.e. resonance frequency). Unfortunately, this technique only works for atoms that have nonzero nuclear spin (only particular isotopic forms of certain elements have this property). The term spin is associated with the orientation state [+½ (or α) and -½ (or β)] that charged subatomic particles or even atomic nuclei have when they are immersed in a magnetic field. In NMR spectroscopy procedure, by convention, the magnetic field orientation is aligned with α, which results in an energy difference between both spin states that is dependent on the strength of the said magnetic field. These energy differences, which are affected by the presence, and distance from, nearby atoms in a protein, as well as the strength of the applied magnetic field, are what creates the detectable NMR signals. These signals are then analysed to produce a list of data that contains the atoms´ location and distance between them in order to build the molecular structure model of the molecules under study.

4.10 Docking Procedure

In order to perform the docking of the CYP molecules with the ITC ligands and the native ligands in the CYPS´ PDB files, the AutoDock Tolls (ADT) suite of programs (available here: mgltools.scripps.edu/downloads), AutoGrid4/AutoDock4 (AD4) (available here: http://autodock.scripps.edu.) and the AutoDock Vina (or just simply called Vina) (available here: http://vina.scripps.edu) programs were used, all of which were developed at The Scripps Research Institute, California (USA). The purpose in using these programmes is to search for the best conformation and energy for the interaction between ligand and receptor molecule. This is done by using a scoring function to try and predict the strength of the non-covalent interaction (i.e. binding affinity) between the two aforementioned molecules. This task is accomplished by

computing and using both the charges of the previously mentioned molecules as well as a systematic search for any chemical conformation of the ligand (flexible) in order to find its best binding geometries and locations in the receptor molecule [rigid (by default) or flexible]. The reason as to why the native ligands were also used in the docking procedure, apart from the ITCs themselves, was to compare the docking predictions made by both softwares with the existing positioning and location of the native ligands in their experimental complexes with CYP molecules. Comparison between theory and experiment is the best way to assess the performance and reliability of the docking procedures.

4.11 ADT – Preparing the Ligand and Receptor for Docking

Autodock Tools (ADT) is a suite of programs used to prepare both the receptor protein as well as the ligand for the docking procedure (available here: mgltools.scripps.edu/downloads). In this software program, both the protein to be used for docking, as well as the ligand, are corrected for any type of error (e.g. broken chains) that the PDB files downloaded from the PDB website may contain as well as being prepared for the docking procedure.

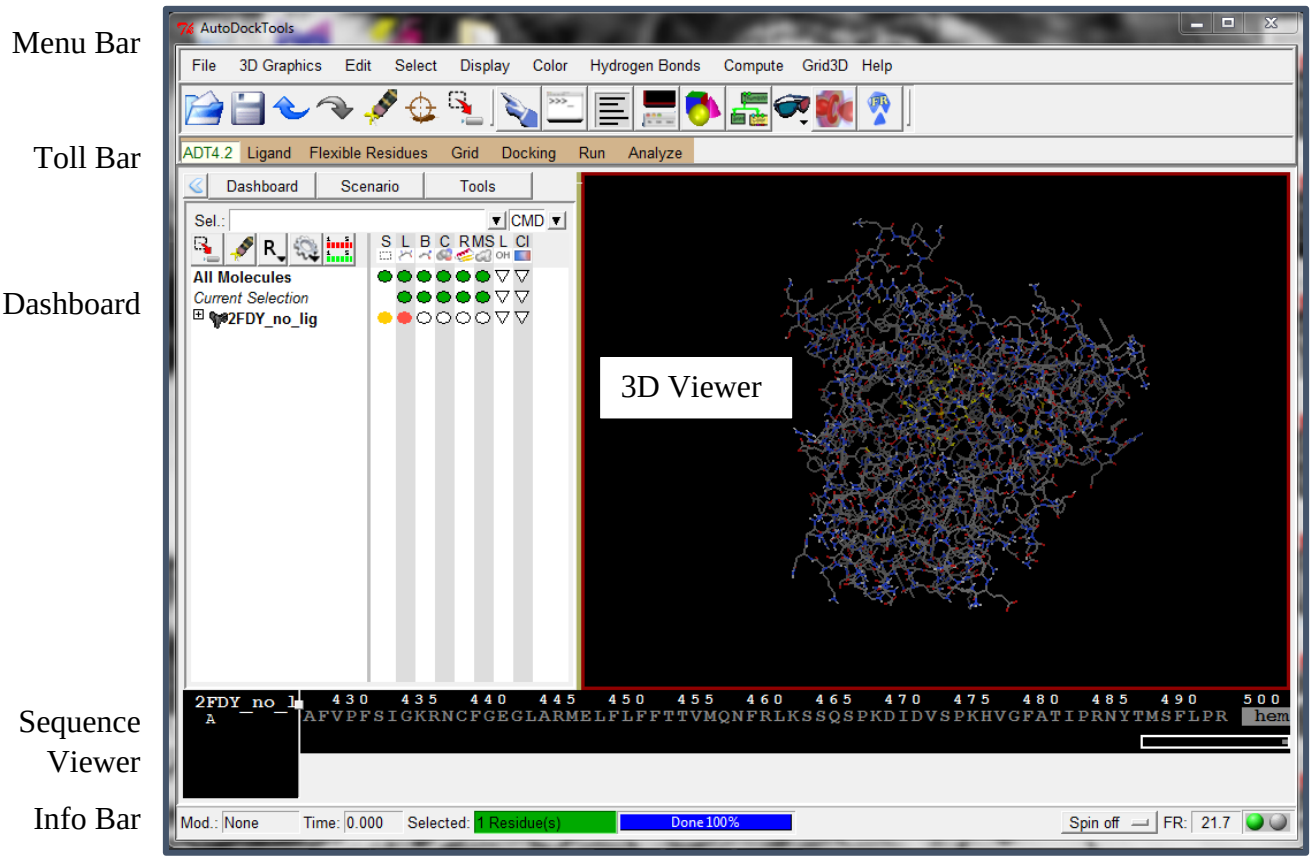


Fig.: 4.15 – The graphical user interface (GUI) of ADT, with the heme group coloured in faint yellow in the centre of the 2FDY crystallographic receptor molecule of CYP2A6.

In order to load the molecule into the software program, one needs to choose the "File" option in the Menu Bar and then select "Read Molecule", followed by selection of the protein PDB molecular file. Since Protein Databank Files normally lack hydrogens, these must be added to the model. Polar hydrogens are of crucial importance for the docking process since they impact the interactions that happen between ligand and receptor molecule. This task is done by clicking in the option "Edit" in the Menu Bar and then selecting "Hydrogens", followed by "Add" and the option "Polar Only". Another pivotal factor to bear in mind is the fact that this program will automatically compile and add charges to a molecule every time its PDB file is open; another way to perform this task is to the user to add the charges to the molecular file by choosing that option (Edit – Charges – and then chose one of the charge options available) from the program´s Menu Bar. Unfortunatelly, ADT does not assign charges to metal ions, which constitutes a problem since these charges are taken into account when performing the actual docking. A simple way to deal with this problem is to open the protein PDBQT file, which will be latter explained in this manuscript, in a text editor (WordPad was used in this case) and manually alter the charge of the iron atom from 0.000 to +3.000.

In order to input the ligand molecule, the user needs to select the option "Ligand" form the Tool Bar, followed by "Input", "Open" and then selecting the ligand file that is to be used. In terms of atomic torsions (free bond rotations) that are present in the ligand molecule, the user can see and choose the torsion root, i.e. the central atom in the ligand to be used as the root of the molecule, by going to "Ligand", "Torsion Tree", "Detect Root". ADT also allows the user to see the number as well as the type of bonds that exist in the ligand molecule. This can be done by choosing the following options from the Toll Bar: "Ligand", "Torsion Tree", "Choose Torsion" for the type of torsion bonds available; and "Ligand", "Torsion Tree", "Set Number of Torsions…" for the number of active bonds in the ligand molecule. The different types of bonds in the ligand are differentiated from one another by their colour (i.e green – rotatable; magenta – non-rotatable; and red – unrotatable). The rotatable bonds can be altered manually by the user in the window "Torsion Count" and then pressing "Done" when finished. Finally, the user saves the

ligand molecule in the PDBQT format by choosing “Ligand”, “Output”, “Save as PDBQT …” and naming the PDBQT ligand file.

The Grid Menu in the Toll Bar is used o choose both the atom types (by choosing the ligand and the receptor molecule) as well as the Grid Box Parameters. The atom types are chosen by selecting “Grid”, “Macromolecule”, “Select …” and by doing this the receptor molecule is chosen as being the receptor and is saved in PDBQT format after it has been named. For the continuation of the atom type selection, the ligand needs to be choosen as well. For that, the user just needs to select “Grid”, “Set Map Types”; “Choose Ligand”. Now for the Grid Box parameters (i.e. the 3D space in the receptor molecule were the conformational search of the ligand will be performed in order to assess both the ligand´s conformation and location in the receptor) can be chosen from the Tool Bar by choosing the option “Grid”; “Grid Box”, followed by the appearende of the “Grid Options” painel. Since the aim of this master thesis project was to study the inhibition potential (i.e. binding affinity) of the ligands (i.e. ITCs) with the receptor molecules (i.e. CYP enzymes), it follows that the 3D search was centered in the active site (heme group) of these proteins. This task is accomplished by selecting the protein in the Dashboard and then following the protein aminoacid sequence in the Sequence Viewer until the heme group appears in sight, then it is just a matter of selecting it, which in turn coulours it in faint yellow in the 3D Viewer, and trying to place the grid box in such a way as to fully encompass the heme group and active site cavity. Hereafter it is of crucial importance to choose the values for the Grid Box Parameters. The values chosen for the Grid Box in the Grid Options window are of crucial importance for the docking procedure and they will be used latter in both Vina and AD4 programs, which are the ones that will actually compute the docking calculation and, therefore be responsible for the docking. In ADT, this step consists in merely a visual interpretation of the search parameters where, afterwards, Vina and AD4 will try to do the calculations for the docking. In the Grid Options window, the user selects the number of grid points in every dimension (X =22; Y = 22; Z = 22), which corresponds to the scope of the 3D search that will be performed by both docking programs. It is also necessary to choose the exact location where such a search will be carried out, and for that the user needs to centre the Grid Box with the heme group, which is different for every crystallographic structure. Since ADT measures the search space in points instead of angstrom, the spacing was changed to 1 angstrom. This way every search point corresponds to one angstrom. After all the adjustments of the Grid

Box, the user needs to choose the option “File” in the “Grid Options” windown and select “Close saving current”.

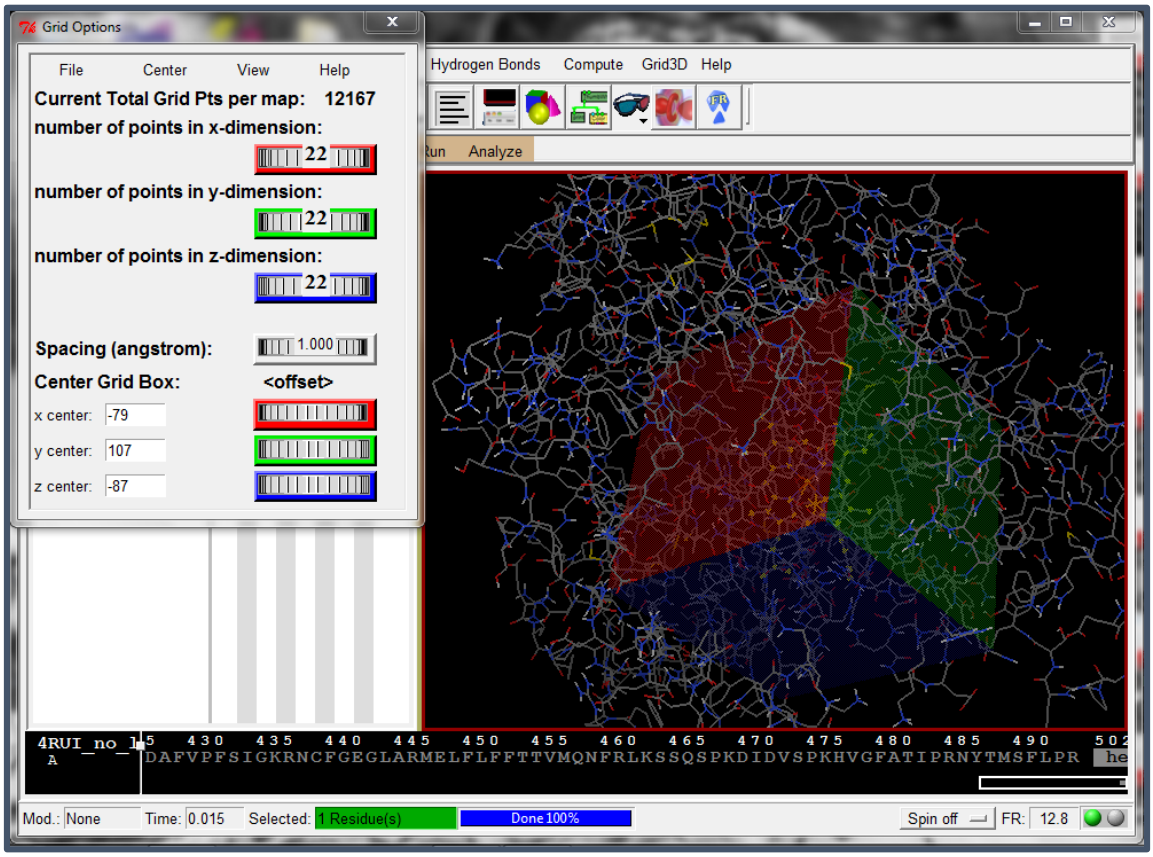


Fig.: 4.16 – The grid box is centred in the heme group by altering the XYZ values for each dimension as well as the number of points in each one.

Table: 4.1 – A list of the different grid box centre values for each crystallographic structure according to its type of CYP.

CYP	PDB Code	Grid Box Centre		
		X	Y	Z
2A6	1Z10	52	85	59
	1Z11	52	85	58
	2FDU	52	82	58
	2FDV	52	84	58
	2FDW	52	84	57
	2FDY	53	85	57
	2PG5	53	85	58
	2PG6	52	83	58
	2PG7	52	83	57
	3EBS	-19	5	-6.5
	3T3Q	-39	33	-50

	3T3R	-2	61	-58
	4EJJ	-7	-15	-5
	4RUI	-79	107	-87
2A13	2P85	23	-37	16
	3T3S	16	8	43
	4EJG	16	-13	37
	4EJH	27	-10	30
	4EJI	61	-23	26
2B6	3IBD	20	15	23
	3QOA	-47	13	-19
	3QU8	-14	4	4
	3UA5	13	-2	45
	4I91	47	-13	-19
	4RQL	8	36	21
	4RRT	4	-11	-14
2C9	1OG2	14	69	21
	1OG5	-25	79	30
	1R9O	7	25	-4
	4NZ2	-53	-46	-21

Following the completion of this step and having both the receptor and ligand files saved in the PDBQT format (i.e. includes all the atoms´ type and charge), the Vina docking procedure can now finally begin.

4.12 Vina Docking Procedure

This program will be responsible for performing the docking calculations, by means of a scoring function, between ligand and receptor, yielding binding affinity energies for the different conformations of the ligand-receptor complex. Additionally, Vina also presents the user with the most favourable conformations and locations of the ligand in the receptor. After Vina has been downloaded and successfully installed into one´s computer; in order to operate it, firstly the user needs to open the command line, which is located in the Accessories folder. After the command line has been open, the user only needs to get access to the usage summary of instructions on how to run Vina by typing its location in the respective file, followed by pressing the ENTER key on the keyboard. Alternatively, one just needs to type “vina --help” in the command line for the instructions to appear. To actually run the program Vina in an easy way, it is best to just place all the files of both the receptor and ligand in the PDBQT format, as well as a config file in the same folder. The purpose of using a config file is to further facilitate the process

of running Vina. This config file is written with the data from the previously used program ADT, more precisely the Grid Box parameter values, as well as the identification of the files that belong to the receptor and ligand molecules. If the user chooses to do so, there is another way of doing this instead of using a config file, which would be for the user to write down all of the config file data into the command line. Hereafter, when all the files needed to run Vina properly are assembled together in one folder, the user just needs to type in the command line the following: “the location of vina” --config “name of config file” --log “name of log file”, and then press the ENTRER key. After this has been done, Vina will then compute and give out the results in the form of two files: an out PDBQT file and a log text file. The log file contains all the information about the binding affinity energies, while the out file will have the data concerning the ligand position and conformational shape in the receptor molecule.

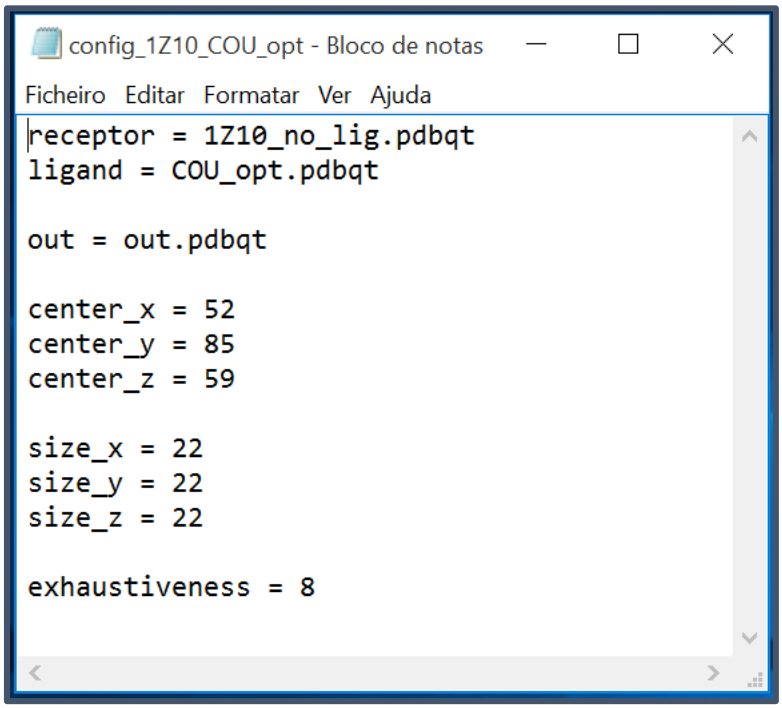


Fig.: 4.17 – Example of a Vina config file.

One particular crucial aspect of Vina is the exhaustiveness level, which corresponds to the number of runs (i.e. the number of docking calculations) the program will try to perform in order to dock the ligand with the receptor after there has been a slightly alteration of the conformation of the ligand or the receptor (if flexible residues have been chosen). Of important notice is the fact that the exhaustiveness level can also

be manipulated, hence the following methodology was adopted when dealling with poor docking results: when performing the docking, the exhaustiveness level should first be the default value (i.e. 8); but when the last result in a set of them has an energy affinity lower than -5.0 kacl/mol, then the docking should be repeated but with an exhaustiveness level equal to 10. Afterwards and if there is still a result with an energy level lower than the one required, then the exhaustiveness level should be changed to 50. This is done in order to ascertain if the Vina´s docking algorithm is capable of finding better solutions by increasing the number of runs.

The Vina program just requires three simple things to work properly: the molecular structure of the ligand and receptor molecules used in the docking process, and the parameters of the Grid Box from ADT. Neither receptor protein nor ligand atom charge assignment is needed for it to work. Nevertheless, and since Vina was designed and built to work with the file format used for AutoDock 4 structure files, the input and output files must be in the PDBQT format. As such, though Vina does not need molecules with charge to work because it attributes charges according to the molecular structure of the molecules used for docking it still needs the structural files of those molecules in the PDBQT format (i.e. atom type and charge).

4.13 PyMOL – The Molecule Viewer Tool

PyMOL is a powerful but user friendly program tool used to visualize molecules, from proteins to small ligands. PyMOL is made up of two windows: a command (top window, know as PyMOL Tcl/Tk GU or external graphics user interface, GUI) and a graphics window (main window, known as PyMOL Viewer). The last one is where the molecular visualization takes place. The main goal in using this software program was to view the results of the docking process performed by Vina (i.e. the out PDBQT file). Even so, this program was also used to prepare the receptor protein for docking by performing several tasks, namely: elimination of water molecules, removal of any additional chains of the same protein, and deletion of the ligands that were crystalized together with the protein. This preparation for docking was achieved through several different commands that can be written in both PyMOL windows since both of them have a command line. To eliminate the water molecules, the user can choose to write in the command line the following: “remove HOH/”. Alternatively, the user can simply go to the “Object Menu Panel” located in the graphics window, choose the menu “A” (from action) and then chose the option “remove waters”. To remove any surplus amino acid chain from the PDB file,

one just has to write in the command line the following: “remove not chain a”. Finally, to delete any ligands that may come with the protein downloaded from the PDB site, the user can write, once again in the command line, “remove <name that the ligand has in the PDB website>”. Another way to do this is for the user to select the option “S” from the “Mouse Menu” at the right inferior corner of the graphics window. Alternatively, the user can go to the command window, choose the menu “Display”, followed by the option “Sequence” in order for the program PyMOL to show the sequence. This option enables the user to see the sequence of amino acids from a protein as well as its ligand(s), with the last one(s) placed generally at the end of the sequence. After scrolling down the sequence, the user just needs to select the ligand(s) with the right mouse button and then select the option “Remove” to delete the ligand(s). The elimination of the native ligand(s) step is very important for the docking process, since the presence of other ligand(s) other than the ones that are to be docked will inevitable interfere once Vina tries to dock the protein-native ligand complex with our desired ligand(s). Needtheless to say that before the removal of the water molecules and the deletion of the ligand(s) that may exist in the pdb file, one must write in the command line the following: “set retain_order, 1”. This command allows for the maintenance of the sequence order in the pdb file, which is vital for the latter use of the program AutoDock4 (AD4), since for it to work properly, it needs the input files to have the same exact order as they had when they were originaly extracted form the pdb website.

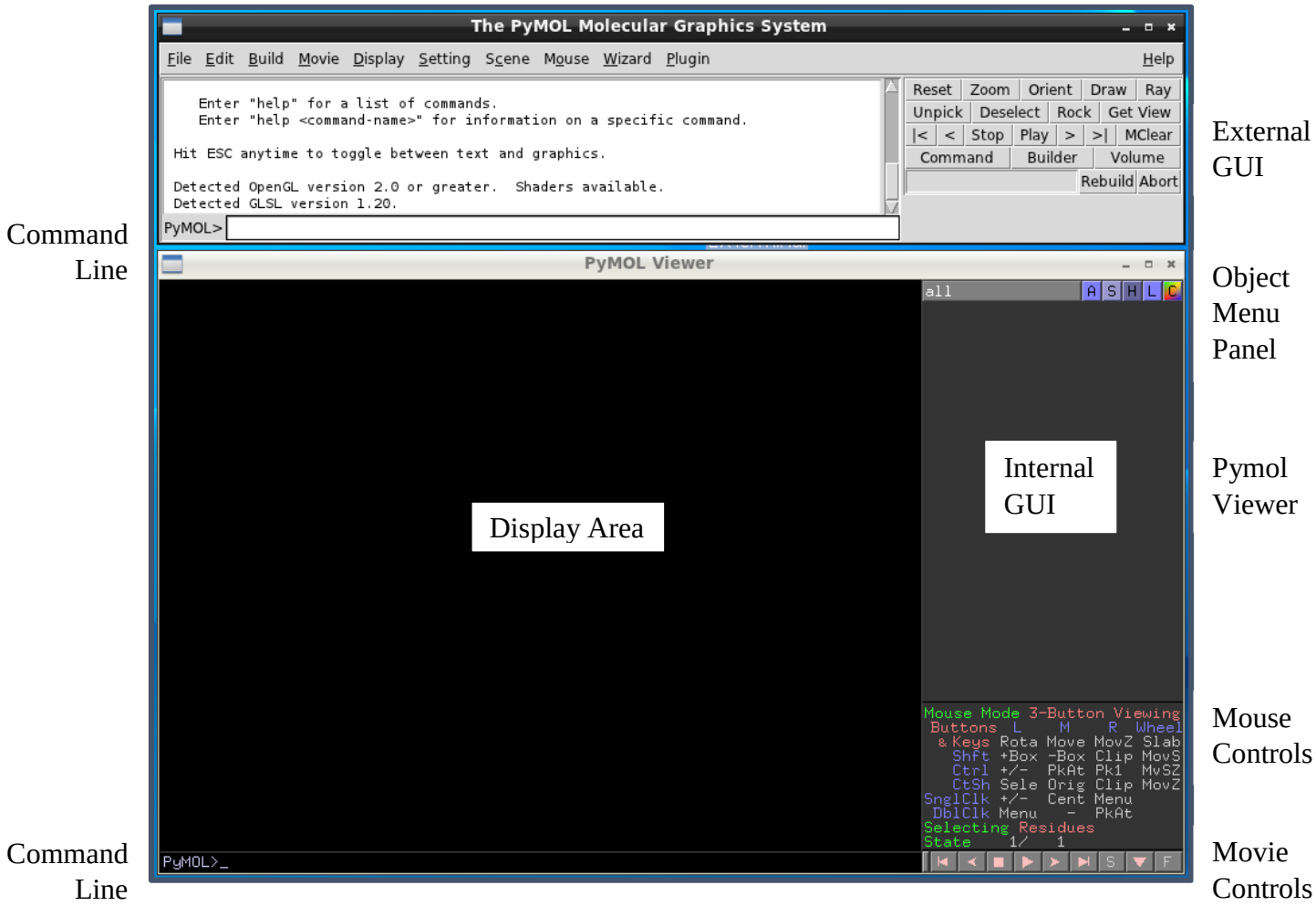


Fig.: 4.18 – The molecular visualisation program Pymol consists of an external GUI, smaller window, as well as the Pymol Viewer, where molecule display occurs.

4.14 AutoGrid4/AD4 – Old and Outdated, but Useful and Practical

AutoGrid4 and AD4 are two programs that need to be used one after the other, respectively, in order to produce docking output results. The first program is necessary for the generation and setting of several pre-calculated interaction energy maps of the receptor molecule which are then used by AD4 for the purpose of performing the actual docking.

Autodock 4 (AD4) is an older software developed by the same research team that created Vina. The usage of this software for the docking procedure is closely tied to the fact that it is imperative to have strong evidence that the outcome predictions given out by Vina, regarding the docking of the ITCs molecules into the CYP enzymes, are both accurate and correct. This can be accomplished in two ways: one way, already mentioned

before, is to perform docking runs of the existing ligands in the CYPs molecular PDB files with those same CYPs in order to assertain if the obtained result predicitons are in aggreament with both the natural positioning and location of such ligands in their corresponding CYP molecules. A second way to further strengthen our predictions is to use an alternative docking program and procced to work with the same dockings that were performed in Vina. Since the program AD4 was made by the same developers that made Vina, it was though that this program would make the ideal choice for the task of producing an aditional set of for comparison with those obtained with Vina, thus having a further way in which to judge if the ITCs are well docked with the CYP molecules.

The docking procedure is parcially the same as with Vina. Firstly, it is necessary to have both the ligand and receptor files in the PDB format. Afterwards both files are subjected to the same protocol that was used with the ADT suite of programs in order to have both molecular files in the PDBQT format. From here on out, the process largely differs from the one used in Vina. In the first place, it is crucial to prepare the AutoGrid Parameter File (GPF), which as the name implies, it is a file that is used by AutoGrid4 to create several different types of maps that take into account several factors (desolvation; electrostatics; and the types of atoms in both the ligand and receptor molecules). To do this, the user just needs to select the menu “Grid” in the Tool Bar and follow the same protocol that was mentioned before for this particular menu, since the map types are particularly dependent on the atom types that make up the ligand and the receptor. Afterwardds, the user merely has to save the GPF file by choosing “Grid”, “Output”, “Save GPF…”, name it and select “Save”. In order to start the AutoGrid4 program, the user has to choose the option “Run” located at the right corner of the Tool Bar, followed by “Run AutoGrid…”. Then it is only necessary to provide the location of the AutoGrid4 program as well as for the GPF file as press the “Launch” button.When AutoGrid4 has reached its conclusion and stopped, it is then necessary to procced to the preparation of the AD4 Parameter File (DPF), which works in a similar manner to the GPF in the way that it consists in a file that is going to be used by AD4 for the purpose of performing the docking of the ligand into the receptor. For the creation of the DPF file the user needs to choose the option menu “Docking”, followed by “Macromolecule”, “Set Rigid Filename”, and then choose the receptor file. After having choosen the gridmaps of the corresponding macromolecule it is now essential to do the same with the ligand (i.e. another set of parameters that can be altered if the user sees fit, in this case all parameters

were left in default), for which the user needs to choose “Docking”; “Ligand”, “Choose…”, “Select Ligand”, and “Accept”. Next, it is necessary to set the search method as well as the number of evaluations per each run. For this to be executed, one just merely choses the menu “Docking”, “Search Parameters”, “Genetic Algorithm…”, “Accept”. As in the previous case of the ligand parameters, the values for the number of evaluations in the DPF file were left untouched. Afterwards, it was time for the Docking Parameters (i.e. another set of parameters which the user can alter freely), which were also left in the default values (“Docking”, “Docking Parameters…”, “Close”). Finally, to save the DPF file the user needs to choose “Docking”, “Output”, “Lamarckian GA…”, name it and press “Save”. This last part allows the saving of the file that will be used by the AD4 program and which contains both docking parameters set by the user as well as instructions for a Genetic Algorithm Local Search, also called GA-LS, which consists in a heuristic yet very effective local search that tries to find solutions (i.e. high performance regions) in a vast and complex search space; in other words, what it tries to do is to find a solution when classical methods fail to even find one or are to slow to provide an answer.

4.15 Modification of the Heme Group Partial Charges

AD4 is expected to perform worse than Vina at protein-ligand docking due to the use of an older, less optimized algorithm; in an attempt to somewhat improve the AD4 results, more accurate partial charges for the heme group were supplied. In another study (Shahrokh, K. et al.; 2012), it has been reported that adding heme partial charges results in an vast improvement of the docking results both in terms of the number of solutions found and closeness to experimental data. Bearing that fact in mind, the approach taken in order to improve AD4 results was to replace the heme partial charges given by the program mentioned earlier by a set of heme partial charges that are more robust and consistent with molecular docking simulations (Shahrokh, K. et al.; 2012).

The docking partial charges assignment set chosen was the one from the full heme model (F-HM) IC6 for several reasons. Firstly, the IC*(4 or 6) species, a penta coordinate ferric species, is the one that corresponds to the heme state in which the dockings were and are supposed to be made (i.e. the Fe^{3+} state), because that is the state in which the ligand first interacts with the CYP molecule active site. Secondly, IC6 and IC4 show differences with the partial charges and spin density assignment and though the iron atom remains with very similar values in both species; notwithstanding, IC6 represents a closer agreement than IC4 with the mean spin densities after reorientation of the heme and its

comparison with the original optimization. Thirdly, since some of the other species docking parameter sets had oxygen (bound and unbound) present in the heme group, it was decided no to use such parameter sets because it would affect the docking performance (i.e. the programms are sensitive to the presence or absence of oxygen). Fourtly and final, the use of a fully extended heme group parameter set instead of the truncaded (T-HM) version of it yielded a better representation of the electrostatics potential of the heme group when compared with the T-HM, hence the choosing of the T-HM over the other would have impared the results of the docking simulations (Shahrokh, K. et al.; 2012).

After choosing the F-HM IC6 file as the ideal one to use to replace the heme partial charges in the CYP molecules, a way to actually change those charges was in order. For that, a small script in phython was created that was used to replace the heme charges in the CYP molecules extracted from the PDB website with the F-HM IC6 set of atomic charges.

```
import sys


# Mol2 file with HEM charges
mol_file = open("HEM_IC6_penta.mol2", "r")
# Pdbqt file with original hem charges
pdbqt_file = open("1Z10_clean.pdbqt", "r")
# Pdbqt file with new charges
pdbqt_new_file = open("1Z10_new_charges.pdbqt","w")

# Read mol2 charges
atm_charges =  {}
for a in mol_file.readlines():
    if a[0] == ' ' and len(a.split()) == 10:
        fields=a.split()
        atm_charges[fields[1]]= "%6.3f" % float(fields[8])

mol_file.close()

# Replace charges in pdbqt file, creating a new file
for a in pdbqt_file.readlines():
    if a[17:20] == 'HEM' :
        newline =
a[0:70]+str(atm_charges[a[12:16].strip()])+a[76:]
        pdbqt_new_file.write(newline)
    else :
        pdbqt_new_file.write(a)


# Close all files
pdbqt_file.close()
pdbqt_new_file.close()
```

Fig.: 4.19 – The python script that enabled the changing of the heme partial charges of the CYP molecules.

4.16 Creating Flexible and Rigid Files

Up until now, only rigid docking has been described in this manuscript. As for flexible docking, what actually becomes flexible are the handpicked amino acid residues from the rigid receptor; since the ligand was made flexible in the very beginning of the docking process. Truthfully, the ligand has always been flexible, even in rigid docking, and the terms rigid/flexible, simply refers to the presence or absence of flexible groups in the receptor.

In chossing the residues of the receptor protein to be made into flexible ones, the user has two approaches that can be used. Either choose the lowest value for the distance calculation between the atoms of the ligand and those from the receptor molecule until sufficient atoms are found to make up to 3 residues (minimum number of flexible residues to be used for the flexible dockings) or more (depending on technical constraints); or choose a distance value that has enough space to accommodate 3 or more residues but that has residues that are deemed appropriate, and as such, that are worthy to become flexible in the flexible residues file to be used in this type of docking procedure. In this work, residues from the receptor that were closest to the ligand were preferable, and therefore took precedence when there was the need to choose flexible residues. Nevertheless, if other residues seemed to be more appropriated to be made flexible, in order to achieve a better docking result, then some of those were also chosen so as to have, at the end, 3 flexible residues in the flexible residues file.

To create flexible residues files for both Vina and AD4, one needs first to use both Pymol (used to see in a graphical way which residues are to be made flexible) and ADT (used to actually create the flexible residues files with the residues that were chosen in the previous programm). In Pymol, the user needs to load both the result file of the docking [Vina (pdbqt format) or AD4 (dlg format)] and the receptor protein file (pdb or pdbqt format can both be used but just one at a time). After this, in the Pymol comamd line window, the user has to type the following command, “split_states, <name given to the result file used>” in order to have all the possible results given for a specific ligand-receptor docking separated. This step will allow a better visualization of the possible residues to take into account when chosing the flexible residues for that particular set of results in the ligand-receptor result file. Afterwards, the distance between the atoms of the ligand and the receptor is measured through this command: “distance <name to give to the result of this command>, <receptor file name>, <result file name>, <distance value

in angstrom to be measured between both molecules (ex.: 4.0) >". Now, it is time to select those atoms in the receptor molecule that are located at the specified distance form the result file and to complete the residues to which the atoms selected belong to. In order to achieve this, first the user needs to write the command to select those atoms ("select <name to give to the result of this command>, <receptor file name> within <distance value in angstrom to be measured between the molecules used> of <result file name>"), and then the user has to go to the "select result name" in the object control panel located in the graphical user interface window, click with the right mouse button on top of the "select result name" and select "Modify", "Complete", "Residues". Finally, the user just needs to type one last command: "iterate <name of the select result name> and name CA, print resn, resi". This command will simultaneously give the names and sequence numbers of the selected residues in the receptor protein. Afterwards, the user simply has to choose the most appropriate residues to be made into flexible ones in ADT.

In the ADT program, and before all else is done, the user needs to load the receptor protein in pdbqt format by going to the "Flexible Resiudes" menu located in the Tool Bar and choose the option "Input", "Choose Macromolecule…". After this step has benn carried out, it is time to select the residues by clicking in the "Select" menu in the Menu Bar, followed y the option "Select From String" and then proceed to add manually all the residues that are to be made into flexible ones; afterwards click dismiss. Now the user has to open the "Flexible Residues" menu once again and select "Choose Torsions in Currently Selected Residues…". The following process is identical to the one applied to the ligand, in the sense that the user just needs to choose which bonds are to become rotatable by chosing between different types of bond torsions; next the user selects "Close". Once again, the user needs to go to the "Flexible Residues" menu, but now select "Output", "Save Flexible PDBQT…" to have the flexible residues file. Since the program Autogrid4 needs to calculate grids (for AD4) without having moving residues, there is the need to save a rigid receptor protein but without the said moving residues; for that the user merely has to do the same procedure for saving the flexible residues file with the important exception that instead of chosing "Save Flexible PDBQT" it is necessary to choose "Save Rigid PDBQT" in order to save the said rigid receptor file.

4.17 Flexible Docking in Vina and AD4

After the formation of the rigid and flexible files of the receptor protein, it is time to perform the flexible docking. This type of docking protocol consists in fairly the same one that has already been used for the rigid docking, albeit with some small but very meaningful changes. When doing flexible docking it is very important not to use the entire receptor when asked for the macromolecule, but rather only the rigid portion of the receptor without the moving residues. This alteration to the docking protocol is implemented to prevent the docking programs from trying to dock a ligand with a receptor that has the same residues in a moving and unmovable states, simultaneouly. Since some of the flexible residues might get out of the search space, it is also necessary to increase the number of points in each dimension (X = 50, Y = 50, Z = 50) so as to be able to accommodate the docking movement of such residues.

In the program Vina, it is necessary to alter the config file in order for it to accommodate the flexible and rigid file options. The rest of the Vina protocol remains exactly the same as the one used for rigid docking.

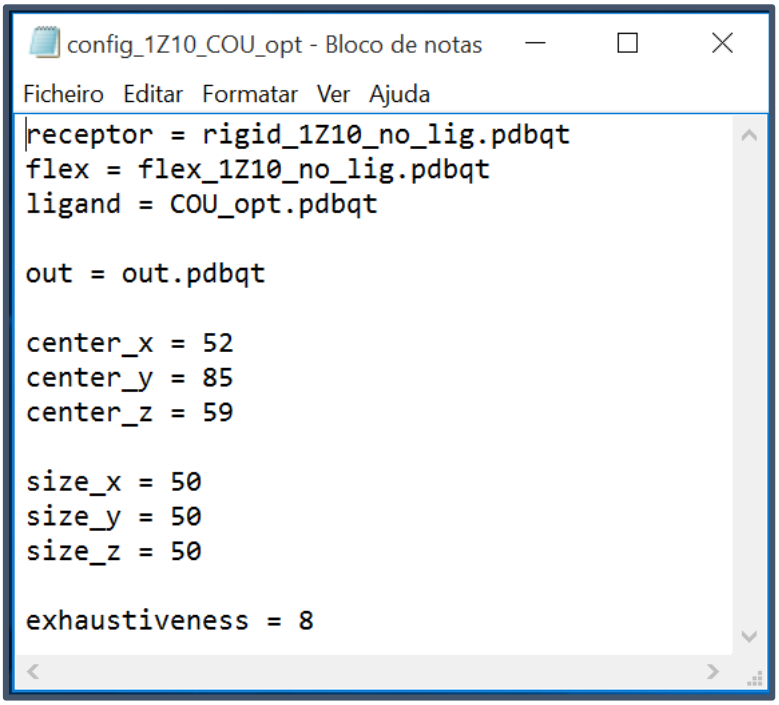


Fig.: 4.20 – Vina config file for flexible docking. The only changes to the rigid docking config file are related to the receptor and flexible sites.

In terms of flexible docking with the program AD4, there are also some small but crucial differences compared with rigid docking. Firstly, when creating the GPF file for the creation of the different grid maps, it is necessary to choose the rigid protion of the receptor by selecting the menu "Grid" in the Toll Bar, followed by the option "Macromolecule", "Choose…" and selecting the file with the rigid portion of the receptor instead of the fully rigid receptor. Also, there is the need to specify the flexible residues file for the flexible docking, which is done by selecting "Grid", "Set Map Types", "Choose FlexRes…" and selecting the appropriate file containing said residues. The 3D search space values are also incremented to 50 for each dimension because of the reasons mentioned before. This is carried out in the "Grid Options" window by changing the values manually after the next brief set of steps has been cleared: select "Grid", "Grid Box…". The remaining of the protocol for the creation of the grid maps that are going to be used in the actual docking is the same for both rigid and flexible docking. The final part that differs from the rigid docking concerns the creation of the DPF file to be used in the flexible docking with AD4. In order to create a DPF file that includes flexible residues, the same procedure of selecting the rigid and flexible portion of the receptor is necessary; for that, the next set of steps must be carried out: "Docking", "Macromolecule", "Set Rigid Filename…" and select the file providing the rigid portion of the receptor molecule. For the specification of the flexible part, the steps are as follows: "Docking", "Macromolecule", "Set the Flexible Residues Filename…" and then choose the file providing the flexible residues to the docking. As for the rest of the process, it is the same as the one previously used in the rigid docking.

4.18 Meaningful Data Extraction and Result Compilation

An initial scanning of the available literature about ITCs-mediated inhibition of CYP enzymes revelead that several of those proteins had their activity inhibited in the presence of such ligand molecules. Furthermore, several articles even provided Ki values obtained from real live CYP-ITC binding inhibition assays, which were compiled and latter used as both a basis for choosing which ITCs-CYPs molecules to dock in a virtual environment as well as a way to further compare the results obtained from the docking experiment with experimental values.

In the docking studies, a parallel was drawn between the type of results given out by Vina [binding affinity energy and root-mean-square deviation (RMSD)] and AD4 [binding affinity, inhibition constant (Ki) and RMSD)] that would enable not only the

comparison between both programs but also between those set of results and the live experimental-driven (i.e. empirical) data found in the literature.

The binding affinity energy is inversely correlated with how strong the affinity of a certain ligand is towards a specific receptor protein; the more negative the binding energy is, the stronger the affinity of that protein-ligand complex.

The RMSD of a particular and specific ligand corresponds to a measurement value that emcompasses the existing difference between the atom positions of a docking predicted structure and its originally observed (i.e. lab-driven) position in a crystal; the more deviated a predicted structure is from the reference structure, the higher the RMSD for that structure will be.

$$\mathrm{RMSD}_{ab} = \max(\mathrm{RMSD}'_{ab}, \mathrm{RMSD}'_{ba}),$$

Where

$$\mathrm{RMSD}'_{ab} = \sqrt{\frac{1}{N}\sum_{i}\min_{j} r_{ij}^2},$$

Fig.: 4.21 – Formula to calculate the RMSD value between two identical structures (*a*, *b*); *N* stands for heavy atoms in structure *a*, while *min* correspondes to the minimum of all the atoms (*j*) in structure *b* that belong to the same atom type (*i*) in structure *a* (Trott, O.; Olson, A. J.; 2010).

The program Vina assumes the best docked ligand structure is the actual position of the ligand, and in doing so, it assigns the first docked ligand structure a RMSD value of 0. Vina will subsequently calculate the RMSD for the remaning ligand structures using the first one as reference. As such, the only means to have usable RMSD values for the best ligand structure is to use the program AD4. To actually use the RMSD values computed by AD4, the coordenates of both crystallographic structures to be compared have to be the same or at least they have to be very close together. This last aspect is essential in order to have an estimate as real as possible of how differente the ligand position would be if it was docked to another receptor molecule because AD4 uses the ligand's original coordenates when calculating the RMSD. The alignment process is

worked out in PyMOL with the receptor molecules in the PDB format by following the next set of commands: load both molecules (“load <name of the crystallographic structure>)” to which a calculated estimate of RMSD is to be achieved one after the other; remove water molecules (“remove HOH/”) and additional molecular chains (“remove not chain a”); align the polypeptide chains (“align <name of the crystallographic strcture to be aligned>, <name of the fixed crystallographic structure>,”); and finaly save the resulting alignment (save <name of the aligned crystallographic structure>, <name of the crystallographic strcture to be aligned>”). This last step ensures that the saved file is has the first crystallographic strcture but with coordenates that have been aligned with the second structure.

The Ki values for specific complexes can be empirically obtained in a lab setting by running specific assays on those same complexes for which the Ki is to be known. Another way of doing this is to obtain it by running a docking program with exactly the same ligand and receptor protein. Of course the estimated Ki value generated by virtual docking will always have to be supported by concrete factual evidence provided by the experimental data. Accurately predicting a Ki through docking alone is quite difficult since many aspects need to be considered when estimating such a measurment (e.g. static receptor and absence of water molecules).

While Vina binding affinities can be used to assess the docking performance, RMSD values produced by this program are unusable since they are calculated relative to the best docking mode and not the original ligand conformation (experimental conformation in the PDB file). Since both Vina´s and AD4 set of results were to be compared with the results provided by the articles in the current literature, it was thought to be more practical to convert the affinity binding energies into K_i values by using the same formula used by AD4 and then procced to compare all the results available at hand.

```
// equilibrium:   E  +  I  <=>    EI
// binding:      E  +  I  ->    EI       K(binding),     Kb
// dissociation:     EI    ->  E  +  I     K(dissociation), Kd
//
//                      1
//        K(binding) = ---------------
//                   K(dissociation)
// so:
//      ln K(binding) = -ln K(dissociation)
//            ln Kb = -ln Kd
// Ki = dissociation constant of the enzyme-inhibitor complex = Kd
//      [E][I]
// Ki = ------
//       [EI]
// so:
//            ln Kb = -ln Ki
// deltaG(binding)    = -R*T*ln Kb
// deltaG(inhibition) =  R*T*ln Ki
//
// Binding and Inhibition occur in opposite directions, so we
// lose the minus-sign:  deltaG = R*T*lnKi,  _not_ -R*T*lnKi
// => deltaG/(R*T) = lnKi
// => Ki = exp(deltaG/(R*T))
```

Fig.: 4.22 – Converting binding free energy into Ki (see Bibliography website section for references)

From both the formula shown above and the relevant information mentioned previously, one can conclude that a more stable ligand-receptor protein will lead to a more negative affinity binding energy (i.e. deltaG), which in turn will lead to a smaller Ki value, and vice-versa.

5. Results and Discussion

The purpose of this work was to study how effective the binding, and therefore, the inhibition of CYP enzymes by ITCs actually is and, at the same time, to work out an assessment and comparative evaluation of the docking programs used. The present section is organized as a series of fundamental questions about the work that are each presented and answered based on the relevant results of the work.

5.1 Are the best docked structures close to the experimental docked structures?

After carefully viewing and analyzing both the out.pdbqt and log files from Vina and the .dlg files from AD4, one can conclude that, in the great majority of the cases, the orientation and placement of the best docked ligand structure are quite similar to those of the experimental crystallographic structure (see Fig. 5.1). However, the best docked structure is not always the one with the best (most negative) affinity energy – at the expense of a less negative ΔG, the RMSD (relative to the crystal structure) of that structure will turn out to be smaller than the one for the other docked structure, which means that it will be closer, in terms of space positioning, to the crystallographic ligand structure. Nevertheless, even the best docked structures usually present a small displacement relative to the reference crystallographic structure. An explanation to this aspect may lie with either the serach algorithm (further improvement and refinment may lead to more accurate docking results) or the chosen docking parameters. Furthermore, despite the fact that the programs used have different docking algorithms, they were both capable of finding several best solutions of which simultaneously had the best affinity energy value and the 3D molecular orientation close to the experimental docked structure. Finally, the mean Ki values of the in-silico docked ligands obtained with Vina were, for the most part, in close agreement with the ones found in literature. As for the set of results from AD4 there were some large differences when compared with the Ki values obtained from experimental docking assays. Also, the number of results per set for each docking run was much higher in Vina than in AD4. This difference in the results can be explained by considering the fact that Vina has a more up to date search algorithm and docking function than AD4. Vina also uses a different set of docking parameters, which possibily influences the docking results, leading it to generate results that more closely resemble

the crystallographic strucutres and their corresponding Ki values than its counterpart (see Annex 7.3, 7.4, and the Supplementary Material).

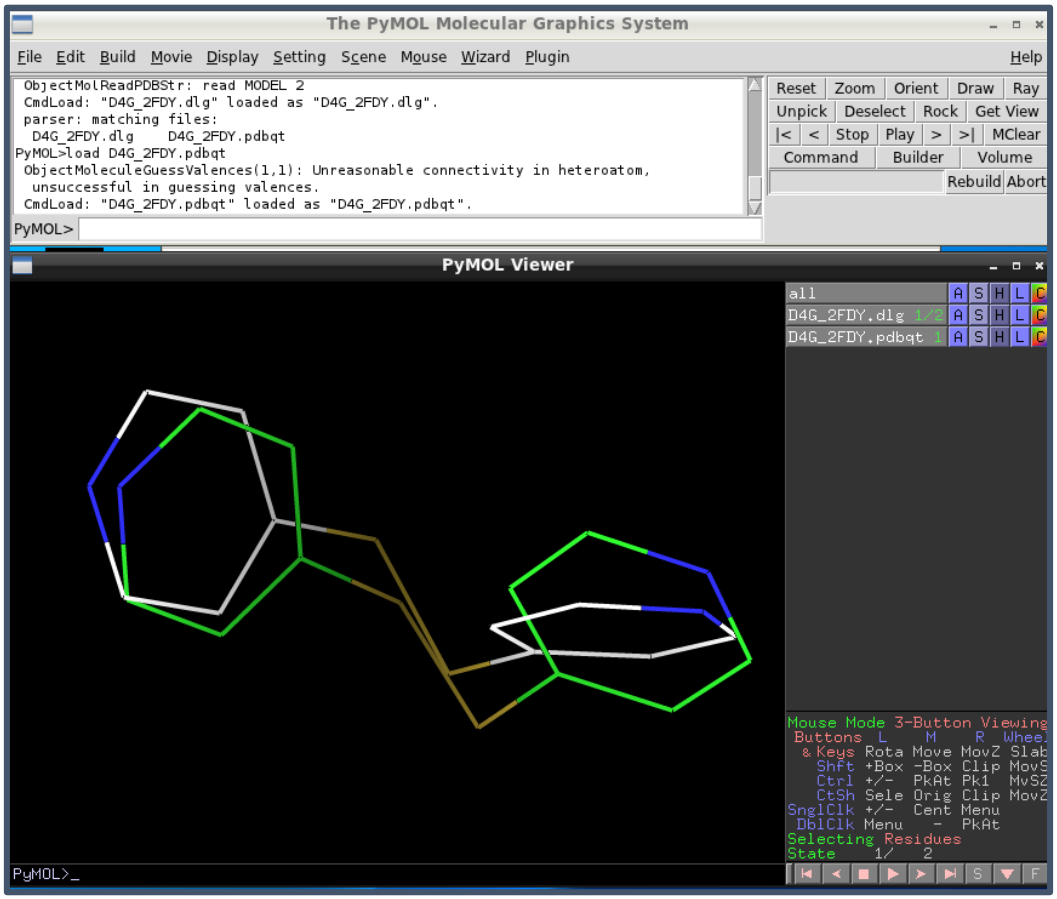


Fig.: 5.1 – The final result of the docking of the ligand D4G with its receptor file (CYP2A6 2FDY). As one can see, the best docked structure has a 3D conformation that is very close to the corresponding crystal ligand.

5.2 Which program is better for rigid docking?

Analysis of the Ki values obtained for the rigid docking using both programs revealed that Vina produces better results (i.e. lower Ki values) for the docked ligand than AD4. Vina was also capable of generating docking results that are more similar to the experimental data compared with AD4. This happens throughout all the ligand-receptor complexes that were produced and evaluated for their estimated Ki values. Nevertheless, there are a few exceptions; not all the results from Vina have the lowest Ki values when compared with AD4 [e.g. D4G-2FDY(CYP2A6)], though that may result from the fact that Vina works with an non-deterministic algorithm, a random number generator is used in the algorithm and the software may not always find the best solution in the set number

of iterations of the method. While this means that Vina will not always get the right answer, i.e. the best docked mode when compared with AD4, an increase in the exaustiveness of the search will normally provide a better solution. In general, Vina will normally find both quicker and better results for a ligand-protein complex docking prediction when compared with AD4. This is to be expected since Vina is endowed with a different and more up to date algorithm with increased accuracy and rapid response, allowing it to find better results than its predecessor, statistical speaking (see Fig. 5.2 and the Supplementary Material).

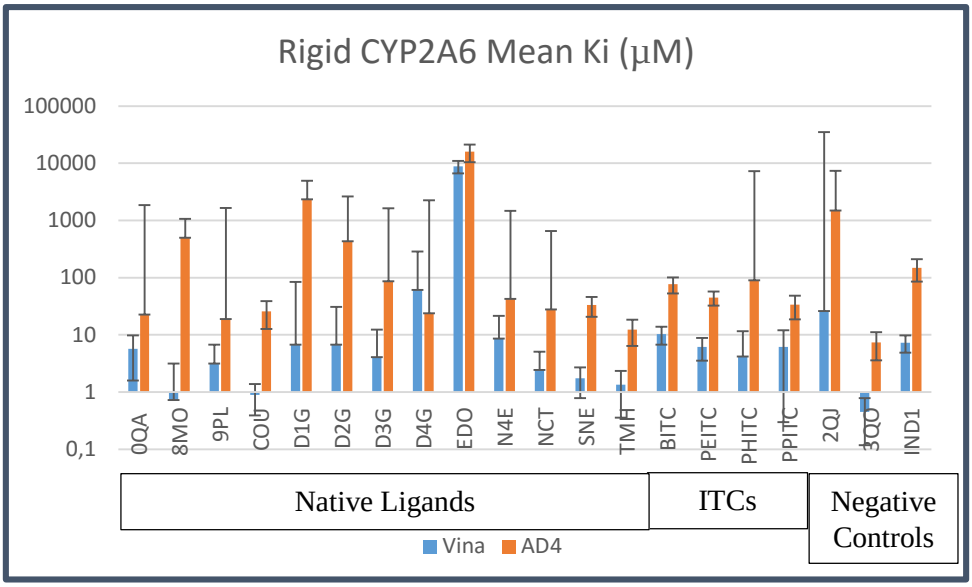


Fig.: 5.2 – Comparison between Vina and AD4 Mean Ki value results for rigid mode using CYP2A6 receptor protein against ITCs and native ligands.

5.3 Which program is better for flexible docking?

As for the flexible docking, Vina still manages to produce the best docking results. Once again, this is explained by the fact that Vina uses a different and more up to date algorithm with increased accuracy and rapid response. The results in flexible mode derived from Vina are better than the ones from rigid mode because there is the selection of flexible residues around the docked ligand structure, which will alter the ligand binding site properties. This will result in a more adequate conformational space for the ligand to dock into the heme group of the protein, which will lead to a better output result for the docking calculation. Additionaly, the difference observed between both programs is greater in flexible than in rigid mode (see Fig. 5.3 and the Supplementary Material).

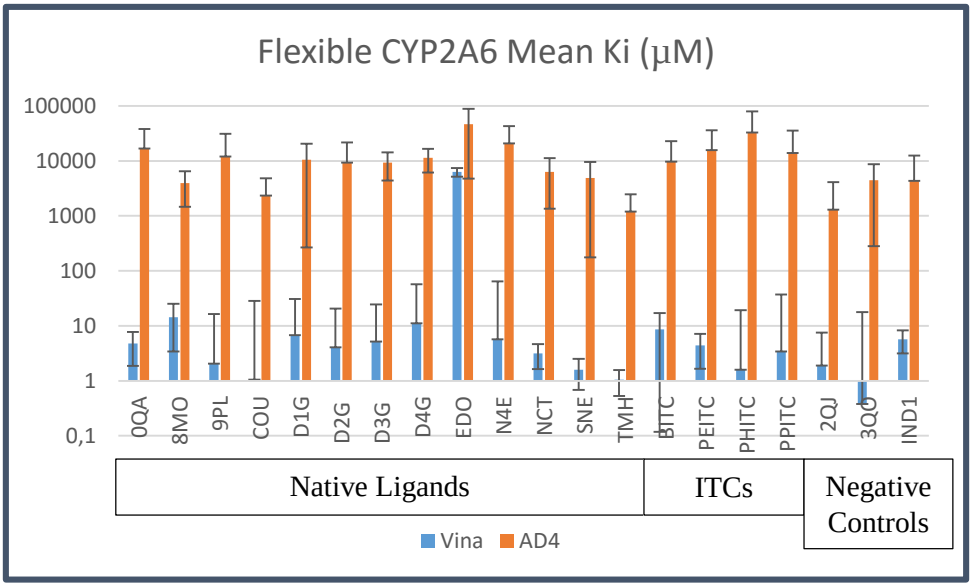


Fig.: 5.3 – Comparison between Vina and AD4 Mean Ki value results for flexible mode using CYP2A6 receptor protein against ITCs and native ligands.

5.4 Are results generally better for flexible docking?

After comparing the flexible and rigid docking results from both programs, a distinction is clearly visible between flexible docking and rigid docking. In Vina, flexible docking is only marginally better at producing lower Ki values from the predicted docking poses of the ligand when compared with rigid docking. In the case of AD4, however, the flexible docking results were actually worse, albeit only slightly more than the ones from rigid docking. As for the AD4´s RMSD values, there is a noticeable difference between flexible and rigid docking, with the last producing lower docking RMSD values for most of the native ligands. This may be due to the different docking algorithms employed by the two programs together with the altered fitting of the ligand into the receptor protein, brought about by choosing flexible residues in the macromolecule. Another possible explanation that may account for why this difference in Ki values is greater in AD4 than in Vina is the first´s energy function, which is used to estimate Ki. Finding the closest conformation (low RMSD) is not the same as finding the the best binder (low Ki). Calibration errors in the energy function may explain the observed discrepance. There is also the problem of the high RMSD values calculated by AD4 in flexible docking mode, since they should not be as high as they are because the receptor proteins only have about 30-40 angstrom in diameter. A possible explanation may lie with the fact that perhaps the program AD4 has some kind of error that keeps it from correctly calculating the RMSD

for the docked ligands when the said docking has been performed in flexible mode (see Figs. 5.4, 5.5, 5.6 and the Supplementary Material).

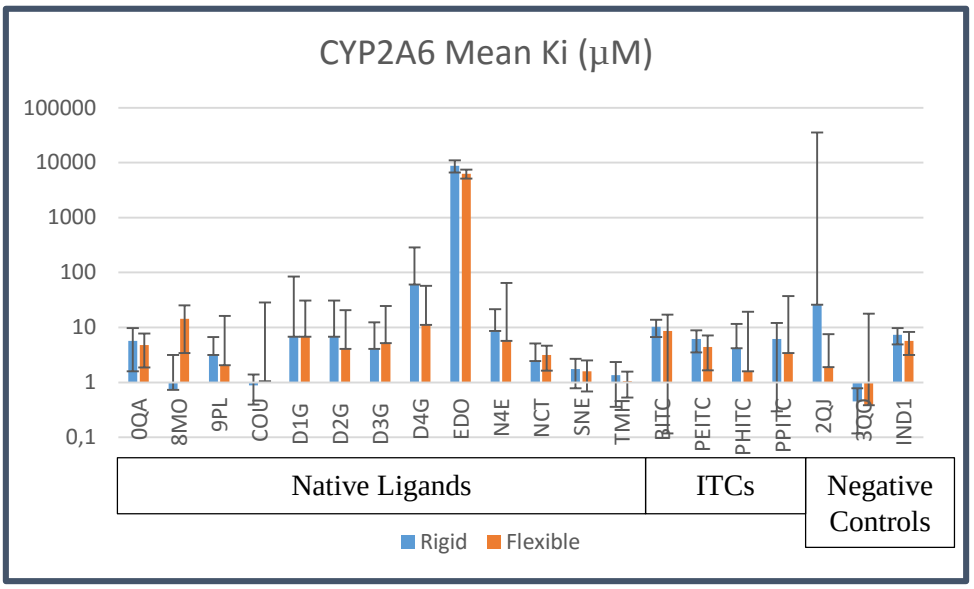


Fig.: 5.4 – Comparison between mean Ki values obtained for the rigid and flexible docking modes in Vina, using CYP2A6 receptor protein and ITCs and native ligands.

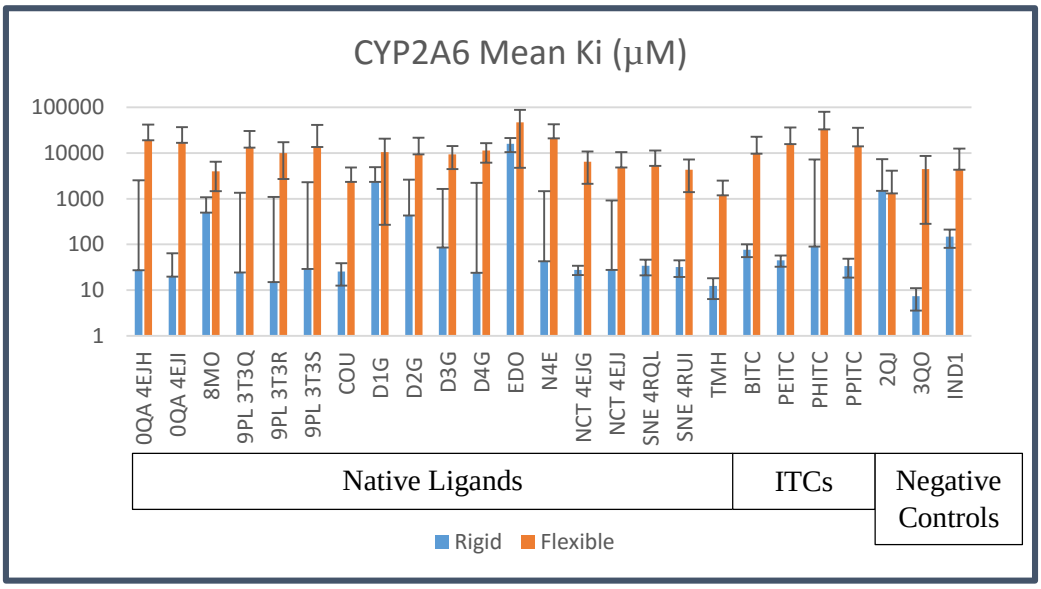


Fig.: 5.5 – Comparison between mean Ki values obtained for the rigid and flexible docking modes in AD4, using CYP2A6 receptor protein and ITCs and native ligands.

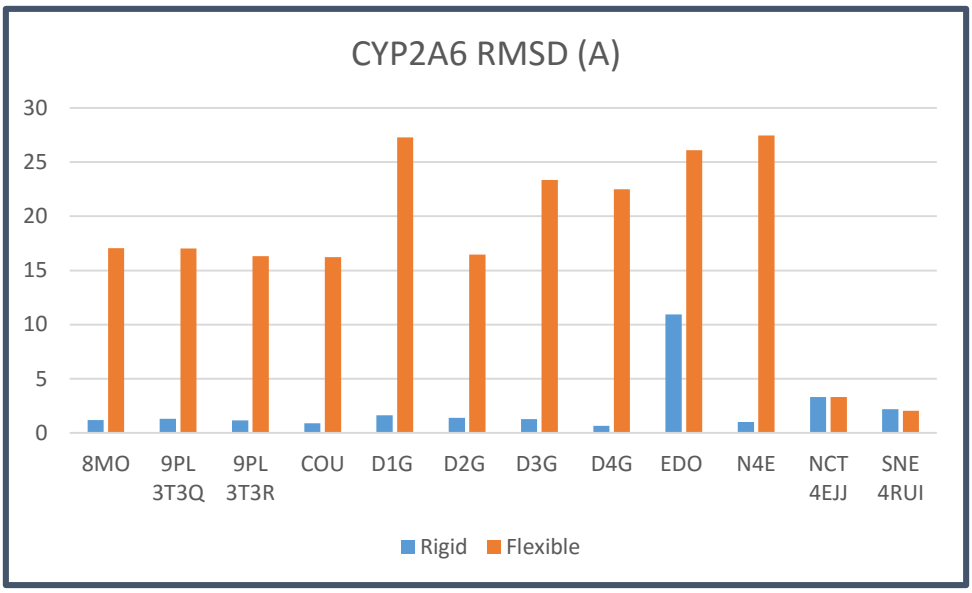


Fig.: 5.6 – Comparison between RMSD values obtained for the rigid and flexible docking modes in AD4, using CYP2A6 receptor protein and the respective native ligands.

5.5 Are the results better when docking a ligand to its crystallographic partner, as compared to docking to a different structure of the same CYP?

Since there were several crystallographic structures of the same CYP molecule deposited in the PDB website, one interesting fact to know is whether or not the native ligands achieve a better docked state when docked to its crystallographic partner as opposed to other existing crystallographic structures of the same CYP molecule. For both programs, some few ligands do produce a better Ki when docked to its crystallographic partner, but for the majority of them, docking of the native ligand to its crystallographic partner does produce the lower Ki value available, though it is generally the second best result. In AD4, the rigid RMSD values are better when it is the result of the native ligand docking with its crystallographic partner. In terms of AD4 flexible RMSD values, the same pattern is also generally observed, but to a much lesser extent. This happens because in rigid docking the receptor molecule has adapated in order to accommodate the partner ligand and as such it will be more difficult to dock a non-partner ligand to an already induced fitted receptor. In flexible docking, on the other hand, the receptor can slightly adapte to accommodate a non-partner ligand, which may improve the docking results for a non-partner ligand. Though it is true that the receptor undergoes induced fitting in order to accommodate the ligand, perhaps some other crystallographic structure of the same CYP molecule offers a better 3D space in the heme group for other ligands other than its

own. Additionally, in terms of AD4 flexible RMSD values, it is noteworthy to mention that the program AD4 has a possible error that leads it to incorrectly calculate the RMSD value for ligands docked in flexible docking mode (see Annex 7.1 and the Supplementary Material).

5.6 How different are the docking results for different crystallographic structures of the same CYP-ligand complex?

Provided there were a few crystallographic structures in the PDB database that shared the same ligand-protein complex, it was interesting to try and compare how the same ligands would fare if they were to be docked to each other´s crystallographic receptor partner. This was performed with the AD4 program since this was the only one that provided useable RMSD values, which were a pivotal aspect of this comparison. The said docking experiment was done with the ligands 9PL-3T3Q (CYP2A6) and 9PL-3T3R (CYP2A6) and the respective CYP2A6 receptor crystallographic strcutures; and 0QA-4EJH (CYP2A13) and 0QA-4EJI (CYP2A13) and the respective CYP2A13 receptor crystallographic strucutres. Unsurprisingly, flexible docking of these particular ligands continues to be worse than rigid docking in terms of their Ki values. In rigid mode, the docking of 9PL-3T3Q (CYP2A6) with its crystallographic partner is sligly better than with its homonym receptor 3T3R (CYP2A6). The same patter happens for the flexible docking of 9PL_3T3R. The flexible docking of 9PL-3T3Q (CYP2A6) with its crystallographic partner is a little bit worse than with 3T3R (CYP2A6). On the other hand, the flexible docking of 9PL-3T3R (CYP2A6) with its crystallographic partner yields a better Ki value than with the non-partner receptor molecule. In the case of CYP2A13, the rigid docking of 0QA-4EJH (CYP2A13) with its receptor crystallographic partner yields a little bit worse than with the crystallographic receptor 4EJI (CYP2A13), contrary to what happens for the rigid docking of 0QA-4EJI (CYP2A13). The flexible docking of CYP2A13 crystallographic partners yields the lowest Ki value when compared with docking between flexible non-crystallographic partners. Additionally, the rigid docking of 9PL-3T3R (CYP2A6) with its crystallographic partner yields the closest Ki result to the experimental data, while in the case of CYP2A13 was the 0QA-4EJH (CYP2A13) with rigid 4EJH (CYP2A13) that was closest to the experimental Ki. In general, the Ki values obtained do not varie much when compared within the same docking mode.

As for the RMSD values obtained from the docking, the flexible mode yields the highest (i.e. worst) RMSD values. The rigid docking of 9PL-3T3Q (CYP2A6) with its crystallographic partner results in a slightly worse RMSD value than with 3T3R (CYP2A6), while the docking of 9PL-3T3R with its crystallographic partner yields a better RMSD value than with 3T3Q (CYP2A6). The flexible docking between crystallographic partners of CYP2A6 yields the lowest RMSD value for that docking mode, albeit being more pronounced in the docking of 9PL-3T3R (CYP2A6). As for the CYP2A13, the docking between crystallographic partners yields a higher RMSD value for both docking modes, with rigid mode having better RMSD values than flexible mode for the docking between crystallographic partners. As for non-crystallographic partners, rigid docking results in RMSD values slightly better than the rigid docking of crystallographic partners. In flexible mode, the same patter observed for non-crystallographic partner rigid docking is also present, albeit being much more pronounced than in rigid mode.

By analyzing the data for the same CYP-ligand crystallographic complex docking, it is visible that rigid docking mode of AD4 is clearly better at finding lower RMSD and more accurate Ki values (i.e more closely resemble the experimental Ki values found in literature) than AD4 flexible docking mode. The difference between flexible and rigid RMSD values is perhaps due to the possible error in the way which AD4 calculates the RMSD for ligands docked in flexible mode. Nevertheless, more complexes of this type have to be available in order to make a statistically viable statement about wether or not AD4 is capable of establish a distinction between the docking of different crystallographic structures of the same CYP-ligand complex (see Figs. 5.7, 5.8, Annex 7.1 and the Supplementary Material).

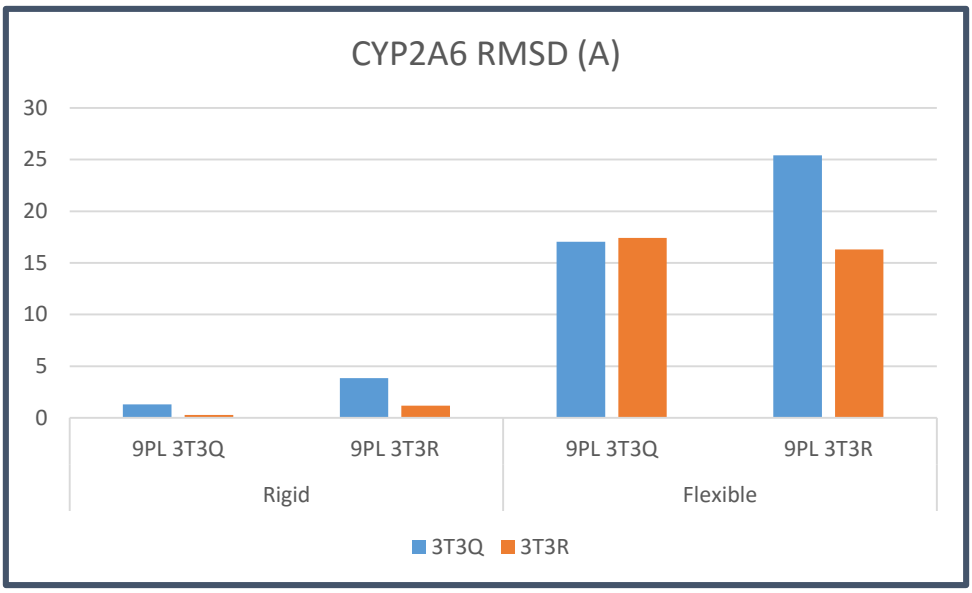


Fig.: 5.7 - Comparison between the RMSD values obtained for the same CYP2A6-9PL complex of different crystallographic structures in both rigid and flexible docking modes in AD4.

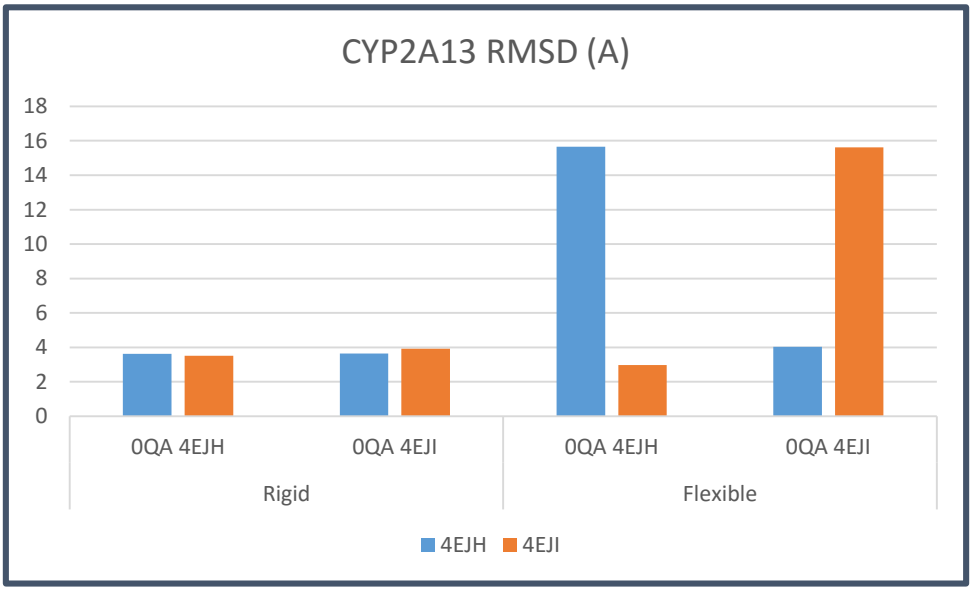


Fig.: 5.8 - Comparison between the RMSD values obtained for the same CYP2A13-0QA complex of different crystallographic structures in both rigid and flexible docking modes in AD4.

5.7 How do different ligands rank when docked to an apo structure of its CYP partner?

There are three apo proteins (i.e. a receptor protein that was crystallized without ligands) in the list of CYPs procured from the PDB protein library. These CYPs (2PG6 and 2PG7 from CYP2A6; and 1OG2 from CYP2C9) are receptor molecules in which induced fitting to accommodate ligands did not take place. As such, they constitute a further source of information on how the docking programs might predict the binding and inhibition of a receptor by a foreign ligand without having to crystalize the said ligand-receptor complex of interest (see Annex 7.1).

By analyzing the Vina Ki values of both crystallographic structures of CYP2A6, it is visible that the best docked ligand structures are 9PL-3T3Q (CYP2A6), NCT-4EJJ (CYP2A6), and SNE-4RUI (CYP2A6). In the case of AD4, the same ligands also managed to score the best Ki values when docked to both 2PG6 and 2PG7. However in the Vina docking with 2PG7, COU-1Z10 (CYP2A6) is the one with the best Ki value in both types of docking, followed by SNE-4RUI (CYP2A6), 9PL-3T3Q (CYP2A6), and NCT-4EJJ (CYP2A6). In AD4 rigid mode, 9PL-3T3R (CYP2A6), 9PL-3T3Q (CYP2A6), NCT-4EJJ (CYP2A6) and SNE-4RUI (CYP2A6) are the ligands that managed to produce the best Ki values, while in flexible mode it was the ligands 8MO-1Z11 (CYP2A6), D1G-2FDU (CYP2A6), 9PL3T3R (CYP2A6), and NCT-4EJJ (CYP2A6). As for the RMSD values, the ligands D1G-2FDU (CYP2A6), D2G-2FDV (CYP2A6), D3G-2FDW (CYP2A6), D4G-2FDY (CYP2A6), 8MO-1Z11 (CYP2A6), COU-1Z10 (CYP2A6), and EDO-2PG5 (CYP2A6) are the ones with the lowest RMSD values in both flexible vs rigid modes as well as in the two CYP2A6 crystallographic apo proteins. Once again the RMSD values for AD4 flexible mode are too high, though the rigid RMSD is high as well. This is possibly explained by an error in the way which AD4 calculates the RMSD (see Figs. 5.9, 5.10, 5.11 and the Supplementary Material).

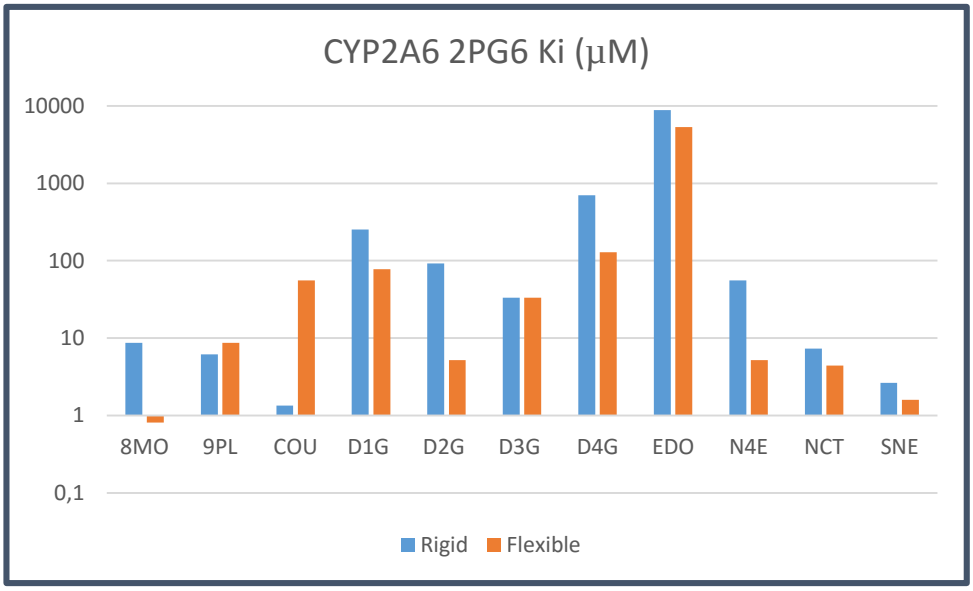


Fig.: 5.9 – Comparison between Ki values obtained for the rigid and flexible docking modes in Vina, using CYP2A6 2PG6 receptor protein and respective CYP2A6 native ligands.

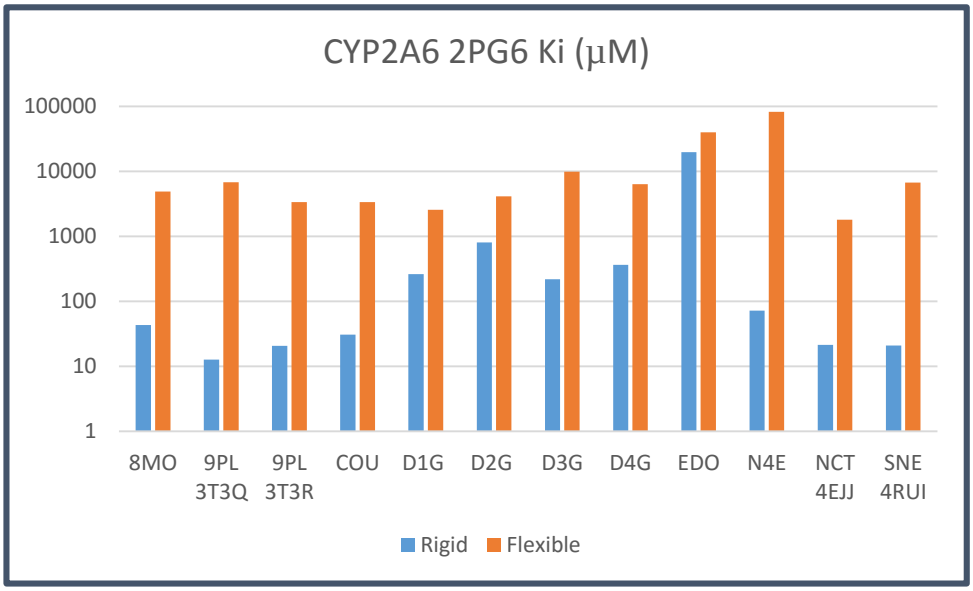


Fig.: 5.10 – Comparison between Ki values obtained for the rigid and flexible docking modes in AD4, using CYP2A6 2PG6 receptor protein and respective CYP2A6 native ligands.

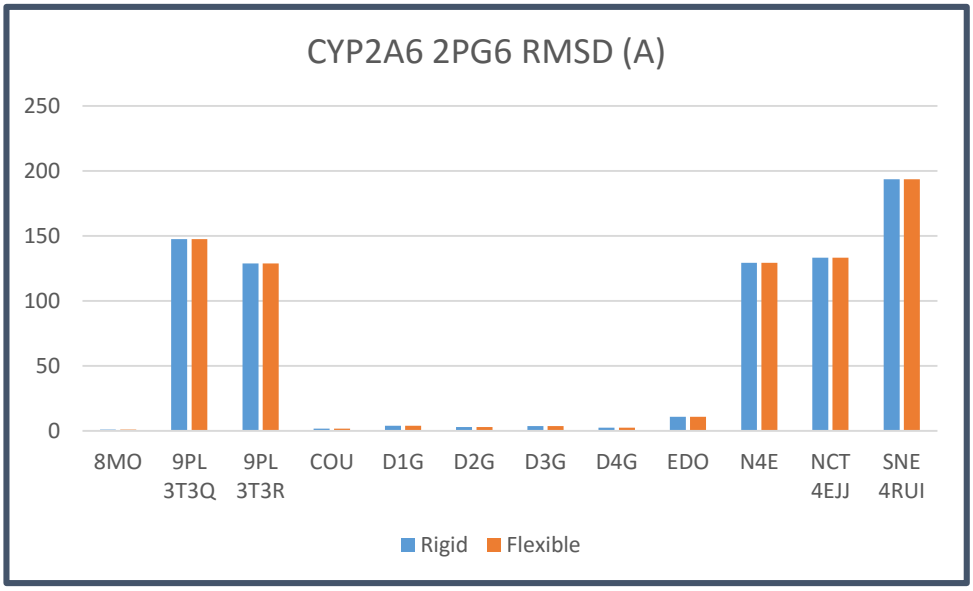


Fig.: 5.11 – Comparison between RMSD values obtained for the rigid and flexible docking modes in AD4, using CYP2A6 2PG6 receptor protein and respective CYP2A6 native ligands.

In the only apo protein available for CYP2C9, the ligands 2QJ-4NZ2 (CYP2C9) was the one with the smallest Ki, followed by SWF-1OG5 (CYP2C9) and FLP-1R9O (CYP2C9) in rigid docking in both programs. In Vina flexible docking, 2QJ-4NZ2 (CYP2C9) had the smallest Ki followed by the ligand FLP-1R9O (CYP2C9) and SWF-1OG5 (CYP2C9). In AD4 flexible docking, FLP-1R9O (CYP2C9) had the smallest Ki of the three CYP2C9 native ligands, followed by SWF-1OG5 (CYP2C9) and 2QJ-4NZ2 (CYP2C9). In the case of the RMSD, however, 2QJ-4NZ2 (CYP2C9) was the ligand with the worst RMSD value of all the CYP2C9 ligands, SWF-1OG5 (CYP2C9) managed to have the lowest RMSD values of the ligand set. In this case, the RMSD values for flexible docking are also too high because of a possible error in the program AD4 that leads it to miscalculate the RMSD (see Supplementary Material).

It seems that the ligand pocket located right above the heme group of the CYP2A6 has a better affinity for ligands that are apolar, with metil groups and with a carbon ring in its structure. In the case of CYP2C9, it would seem it has an affinity for ligands with 2 to 3 ring structures and with both methyl and oxygen groups (see Annex 7.2 and the Supplementary Material).

5.8 Are results worst when docking a ligand to a non-partner CYP?

In order to assess if the docking programs were capable of distinguishing a true ligand from a experimentally known non-ligand molecule when performing the docking, it was of crucial importance to choose negative controls for the *in-silico* docking experiments carried out in this work. These non-partner ligands were chosen from the complete set of crystallographic ligands used in this work. Since these non-partner ligands were to be docked with the CYP molecules studied in this work, it was necessary that two conditions were fulfilled in order the molecules to be selected for the negative control docking runs. Four ligands were selected, one from every CYP molecule and the conditions were (1) each ligand showed inhibitory activity towards the CYP it was bound to in the PDB structure, and none other and (2) if more than one ligand could filled the previous condition, than the one choosen would be the ligand with the best docked Ki value for that particular CYP. This second condition would, in theory, minimize the possibility that the chosen ligand could inhibit other receptor molecules, and in particular, the other CYP molecules used in this work. As such, the four selected ligands for the role of negative controls were the following: D4G-2FDY (CYP2A6); IND-2P85 (CYP2A13); 3QO-3QOA (CYP2B6); and 2QJ-4NZ2 (CYP2C9) (see Annex 7.1, 7.2 and the Supplementary Material).

As it can be seen in rigid docking, the non-partner ligand 2QJ-4NZ2 (CYP2C9) seems to be the only one that is actually working as a negative control at all. In the case of rigid CYP2A6, the endogenous ligands EDO-2PG5 (CYP2A6) and D1G-2FDU (CYP2A6) achieved a worse docking Ki value than any of the controls used in the docking against this type of receptor, though 2QJ-4NZ2 (CYP2C9) managed to have a worse Ki than all the other ligands in the case of AD4 rigid docking. In the case of flexible AD4 docking, the controls performed a little better than the native ligands. In rigid Vina docking, 2QJ-4NZ2 (CYP2C9) is the only negative controls that works as intended, while in flexible Vina docking, the negative controls performed better than the native ligands (see Fig. 5.2 – 5.5).

In the case of CYP2A13, the control ligand 2QJ-4NZ2 (CYP2C9) stands, once again, above all the other ligands as being the worst ligand to dock with this CYP in both docking programs, albeit with a higher proeminence in AD4. The other two ligands,

however, are also incapable of performing bad enough to be considered negative controls (see Supplementary Material).

In CYP2B6, it is the same story all over again for the other two controls other than 2QJ-4NZ2 (CYP2C9), with one noticeable exception, 06X-3UA5 (CYP2B6). This ligand managed to achieve really bad docking results when compared with all the other ligands, endogenous and non-partner alike. This event may have something to do with the higher structural complexity of the 06X-3UA5 (CYP2B6) ligand when compared with 2QJ-4NZ2 (CYP2C9), which might hamper down its docking, more so than what happens with 2QJ-4NZ2 (CYP2C9) (see Supplementary Material).

Finally, in the case of CYP2C9, IND-2P85 (CYP2A13) was the only negative control ligand that managed to act like a proper negative control (see Supplementary Material).

Nonetheless, if 2QJ-4NZ2 (CYP2C9) seemed to be a fine negative control in rigid docking, on the other hand, it was supplanted by many other ligands in flexible docking. Though flexible docking produces better results than its rigid counterpart, in Vina at least; this also means that making receptor residues flexible may increase the chances of having a bad ligand fitting better in the receptor molecule, thus giving out better Ki values than would be expected for what should have been a negative control. In regards to the other negative controls used, once in flexible docking, they were not capable of standing out as negative controls, even to the point where they were undistinguishable from the other ligands. Perhaps the only praiseworthy reference for the negative controls in flexible mode was of 3QO-3QOA (CYP2B6) when docked to CYP2A13 in Vina, which managed to stand out a little apart from the background of the other ligands, both endogenous and negative controls alike (see Fig. 5.2 – 5.5 and the Supplementary Material).

In terms of RMSD values, the negative controls were not easily set apart from the rest of the ligands in both flexible and rigid modes. In CYP2A6, several native ligands had similar docking results to the negative controls. In CYP2A13, only 8MO-1Z11 (CYP2A6), 9PL-3T3Q (CYP2A6) and 9PL-3T3R (CYP2A6) had RMSD values as high as the negative controls, though the rest of the native ligands did not perform much better. In CYP2B6, D2G-2FDV (CYP2A6), D3G-2FDW (CYP2A6) and SNE-4RUI (CYP2A6) had RMSD equaly bad as the negative controls 2QJ-4NZ2 (CYP2C9) and D4G-2FDY (CYP2A6). In CYP2C9, the negative controls had a RMSD very close to the native

ligands' RMSD. A possible explanation for the AD4 flexible RMSD values to be so elevated compared with AD4 rigid RMSD may be due to the presence of an error that affects the way in which AD4 calculaes the RMSD for ligands docked in flexible mode (see Supplementary Material).

Despite the fact that negative control ligands were carefuly selected based on the aforementioned reasoning, one must consider the fact that they belong to the same group with the rest of the ligands used in this work. As such, since they were capable of inhibiting CYPs, albeit different types of CYPS nonetheless, perhaps some of them were simply too structurally similar or chemically related to actually serve as usable docking negative controls.

5.9 Are AD4 or Vina better or worse at docking specific chemical classes?

With the exception that Vina can easily find better docking solutions than AD4, both of these programs seem to be able to best dock ligands with a few shared characteristics among themselves: non-polar molecules, with one to two (preferable) aromatic ring structures or a chair ring structure and with few to nil rotatable bonds. All of these characteristics make up the kind of structure the best docked molecules have in common with one another in both docking programs (see Annex 7.2 and the Supplementary Material).

5.10 How good fare the ITCs when compared with the native ligands?

The performance of the ITCs´docking status with the intended CYPs is not the best there is but it is also not the worst. In the CYP2A6 docking, there are some remarkably good results from Vina Ki values when compared with the native ligands. However, in AD4, the docking results seem more unstable and the ITCs are in agreament with the other native ligands in the sense that it does not seem to be any result that is capable of standing out from the rest (see Figs. 5.2 – 5.5).

In the case of CYP2A13, Vina Ki values from the ITCs are in aggreament with the other ligand docking results, except PHITC, which had very good docking results with this particular type of CYP. As for the AD4 docking performance of this type of ligands, the picture is similar to CYP2A6 (See Supplementary Material).

The docking with CYP2B6 seems to separate the ligands into three categories: the good, the bad, and the middle ones. The docking results for the ITCs fall in this last

category, indicating that they are just mild inhibitors of this type of CYP. In the program AD4, however, there does not seem to be any kind of distinct and fine separation between the docking results of the several native and ITC ligands (See Supplementary Material).

Lastly but not least, CYP2C9 docking results from Vina clearly dictate that ITCs are very poor inhibitor ligands of this type of CYP, achieving the worst Ki values from all the ligands used in the in-silico docking experiment. In AD4, the docking difference still is that ITCs have performed poorly but it is not so transparent as the one from Vina, with some native ligands having equally or worst docking results than ITCs (See Supplementary Material).

In the case of the mean RMSD values from AD4 for the docking, since the ITCs used in the docking were created in HyperChem, their RMSD values are not to be considered since they were, obviously, not compared with a structure naturally found on those receptor CYPs but rather from the 3D initial coordinates produced by the molecular drawing software. Nonetheless, there doesn't seem to be any difference from those RMSD values when compared with the ones from the docking of the native ligands (See Supplementary Material).

5.11 Is there a correlation between the experimental and the calculated docking results?

For each ligand, the calculated Ki that is closest to the experimental value was selected, seeking to maximize the expected correlation between experiment and calculation. Though it can be said that Vina has slightly more dots on the X = Y line than AD4, there is not a clear correlation between the results from either program and the experimental data found in the literature. It is clear, however, that AD4 computed values have a strong tendency to be below the X=Y line, meaning that calculated values are almost always larger than the experimental ones (this is likely to be a calibration problem with the scoring function in AD4). It may be possible that parameter problems with either the scoring function or the search algoritm are contributing to the discrepancy observed between the calculated and experimental Ki results. There is also the possibility that some other unknown factor is influencing the calculated docking results and seting them further apart from the experimental ones. A more thourgh research is needed in order to elucidate what is causing this difference (see Figs. 5.12, 5.13, Annex 7.3, 7.4, and the Supplementary Material).

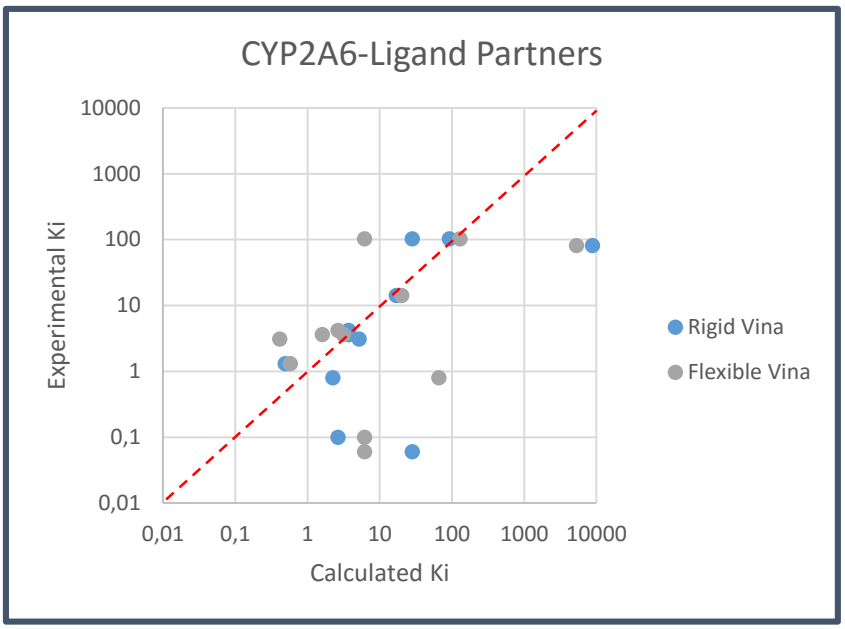


Fig.: 5.12 – Relationship between the experimental and Vina calculated Ki docking results for rigid and flexible mode.

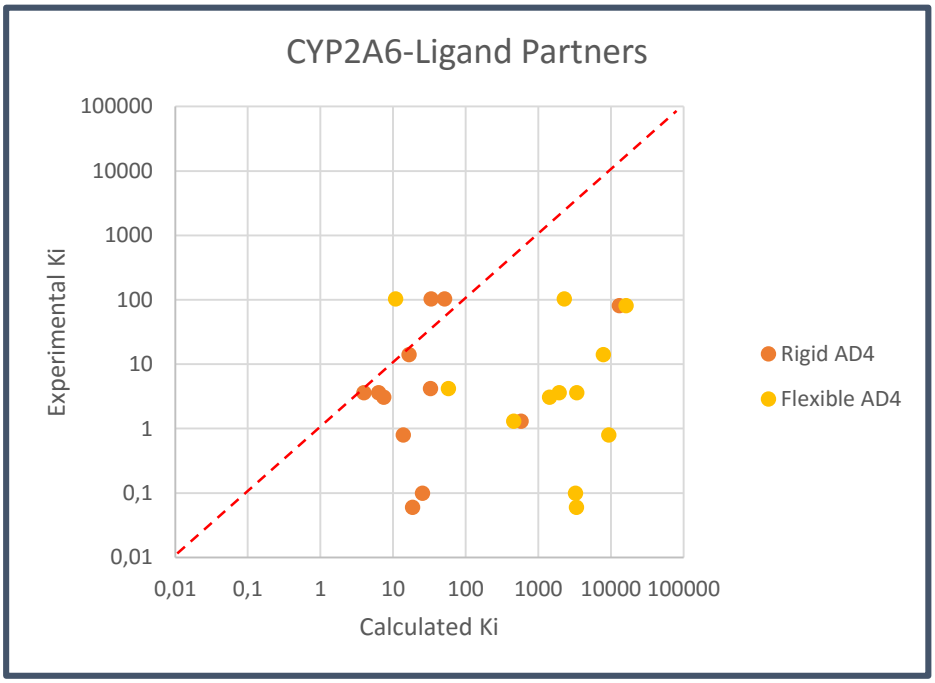


Fig.: 5.13 – Relationship between the experimental and AD4 calculated Ki docking results for rigid and flexible mode.

6. Bibliography

6.1 Articles

- Abdul E. Mutlib, John Shockcor, Robert Espina, Nilsa Graciani, Alici Du, and Liang-Shang Gan; Disposition of Glutathione Conjugates in Rats by a Novel Glutamic Acid Pathway: Characterization of Unique Peptide Conjugates by Liquid Chromatography/Mass Spectrometry and Liquid Chromatography/NMR; The Journal of Pharmacology And Experimental Therapeutics; 294(2): 735-745; 2000

- Adeline Seow, Jian-Min Yuan, Can-Lan Sun, David Van Den Berg, Hin-Peng Lee, and Mimi C. Yu; Dietary isothiocyanates, glutathione S-transferase polymorphisms and colorectal cancer risk in the Singapore Chinese Health Study; Carcinogenesis; 23(12): 2055-2061; 2002

- Akiyoshi Nishikawa, Fumio Furukawa, Shinichiro Ikezaki, Zen-yo Tanakamaru, Fung-Lung Chung, Michihito Takahashi, and Yuzo Hayashi; Chemopreventive Effects Of3-Phenylpropyl Isothiocyanate On Hamster Lung Tumorigenesis Initiated With N-Nitrosobis(2-oxopropyl)amine; Jpn. J. Cancer Res.; 87: 122-126; February 1996

- Akulapalli Sudhakar; History of Cancer, Ancient and Modern Treatment Methods; J Cancer Sci Ther.; 1(2): 1-4; 1 December 2009

- Alysa Fairchild, Kristin Harris, Elizabeth Barnes, Rebecca Wong, Stephen Lutz, Andrea Bezjak, Patrick Cheung, and Edward Chow; Palliative Thoracic Radiotherapy for Lung Cancer: A Systematic Review; 26(24): 4001-4011; 20 August 2008

- Amy V. Gasper, Maria Traka, James R. Bacon, Julie A. Smith, Moira A. Taylor, Christopher J. Hawkey, David A. Barrett, and Richard F. Mithen; Consuming Broccoli Does Not Induce Genes Associated with Xenobiotic Metabolism and Cell Cycle Control in Human Gastric Mucosa; J. Nutr.; 137: 1718-1724; 2007

- Ana Martins, Helena Vieira, Helena Gaspar, and Susana Santos; Marketed Marine Natural Products in the Pharmaceutical and Cosmeceutical Industries: Tips for Success; Mar. Drugs; 12: 1066-1101; 2014

- Andreas Ratzka, Heiko Vogel, Daniel J. Kliebenstein, Thomas Mitchell-Olds, and Juergen Kroymann; Disarming the mustard oil bomb; PNAS; 99(17): 11223-11228; 20 August 2002

- Anne C. Kurilich, Grace J. Tsau, Allan Brown, Lenora Howard, Barbara P. Klein, Elizabeth H. Jeffery, Mosbah Kushad, Mathew A. Wallig, and John A. Juvik; Carotene, Tocopherol, and Ascorbate Contents in Subspecies of Brassica oleracea; J. Agric. Food Chem.; 47: 1576-1581; 1999

- Anne S. Tsao, Edward S. Kim, and Waun Ki Hong; Chemoprevention of Cancer; CA Cancer J Clin; 54: 150-180; 2004

- Anurag A. Agrawal, and Nile S. Kurashige; A Role For Isothiocyanates In Plant Resistance Against The Specialist Herbivore Pieris rapae; Journal of Chemical Ecology; 29(6): 1403-1415; June 2003

- Asad Umar, Barbara K. Dunn, and Peter Greenwald; Future directions in cancer prevention; Nature Reviews Cancer; 12: 835-848; December 2012

- Atle M. Bones, and John T. Rossiter; The myrosinase-glucosinolate system, its organisation and biochemistry; Physiologia Plantarum; 97: 194-208; 1996

- Barbara Ann Halkier, and Jonathan Gershenzon; Biology and Biochemistry of Glucosinolates; Annu. Rev. Plant Biol.; 57: 303–33; 2006

- Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz Jr., and Kenneth W. Kinzler; Cancer Genome Landscapes; Science; 339(6127): 1546–1558; 29 March 2013

- Bharat B. Aggarwal, and Haruyo Ichikawa; Molecular Targets and Anticancer Potential of Indole-3-Carbinol and Its Derivatives; Cell Cycle; 4(9): 1201-1215; September 2005

- C. Athena Aktipis, and Randolph M. Nesse; Evolutionary foundations for cancer biology; Evolutionary Applications; 6: 144-159; 2013

- C. Clifford Conaway, Ding Jiao, and Fung-Lung Chung; Inhibition of rat liver cytochrome P450 isozymes by isothiocyanates and their conjugates: a structure-activity relationship study; Carcinogenesis; 17(11): 2423-2427; 1996

- C. Clifford Conaway, Ding Jiao, Toshiyuki Kohri, Leonard Liebes, and Fung-Lung Chung; Disposition And Pharmacokinetics Of Phenethyl Isothiocyanate And 6-Phenylhexyl Isothiocyanate In F344 Rats; Drug Metabolism And Disposition; 27(1): 13-20; 1999

- C. D. Klaassen; Biliary excretion of xenobiotics; CRC Crit Rev Toxicol.; 4(1): 1-30; October 1975

- C. Douglas Grubb, and Steffen Abel; Glucosinolate metabolism and its control; Trends in Plant Science; 11(2): 89-100; February 2006

- C. Mettlin; Chemoprevention: will it work?; Int J Cancer.; 10: 18-21; 1997

- C. Rodriguez-Antona, and M. Ingelman-Sundberg; Cytochrome P450 pharmacogenetics and cancer; Oncogene; 25: 1679-1691; 2006

- C. Roland Wolf; Chemoprevention: Increased potential to bear fruit; PNAS; 98(6): 2941-2943; 13 March 2001

- Carlos M. R Sant'Anna; Métodos de Modelagem Molecular para Estudo e Planejamento de Compostos Bioativos: Uma Introdução; Rev. Virtual Quim.; 1 (1): 49-57; 2009

- Changcheng Zhou, Emma-Jane Poulton, Felix Grün, Theo K. Bammler, Bruce Blumberg, Kenneth E. Thummel, and David L. Eaton; The dietary isothiocyanate, sulforaphane is an antagonist of the human steroid and xenobiotic nuclear receptor (SXR); Molecular Pharmacology; 71(1): 220-229; 2007

- Changjiang Xu, Christina Yong-Tao Li, and Ah-Ng Tony Kong; Induction of phase I, II and III drug metabolism/transport by xenobiotics; Archives of Pharmacal Research; 28(3): 249-268; March 2005

- Chantey R. Morris, Sheng C. Chen, Lin Zhou, Lawrence M. Schopfer, Xinxin Ding, and Sidney S. Mirvish; Inhibition by Allyl Sulfides and Phenethyl Isothiocyanate of Methyl-n-pentylnitrosamine Depentylation by Rat Esophageal Microsomes, Human and Rat CYP2E1, and Rat CYP2A3; Nutrition and Cancer; 48(1): 54-63; 2004

- Chinthalapally V. Rao, Abraham Rivenson, Barbara Simi, Edith Zang, Rachid Hamid, Gary J. Kelloff, Vernon Steele, and Bandaru S. Reddy; Enhancement of Experimental Colon Carcinogenesis by Dietary 6-Phenylhexyl Isothiocyanate; Cancer Research; 55: 4311-4318; 1 October 1995

- Chinthalapally V. Rao; Benzyl Isothiocyanate: Double Trouble for Breast Cancer; Cancer Prev Res; 6(8): 760-763; 2013

- Chung S. Yang, John F. Brady, and Jun-Yan Hong; Dietary effects on cytochromes P450, xenobiotic metabolism, and toxicity; FASEB J.; 6: 737-744; 1992

- Clarissa Gerhäuser, Min You, Jinfang Liu, Robert M. Moriarty, Michael Hawthorne, Rajendra G. Mehta, Richard C. Moon, and John M. Pezzuto; Cancer Chemopreventive Potential of Sulforamate, a Novel Analogue of Sulforaphane That Induces Phase 2 Drug-metabolizing Enzymes; Cancer Research; 57: 272-278; 15 January 1997

- D. E. Rollins, and C. D. Klaassen; Biliary excretion of drugs in man; Clin Pharmacokinet; 4(5): 368-379; September-October 1979

- D. K. Badyal, and A. P. Dadhich; Cytochrome P450 And Drug Interactions; Indian Journal of Pharmacology; 33: 248-259; 2001

- Dan J. Kliebenstein, Juergen Kroymann, and Thomas Mitchell-Olds; The glucosinolate-myrosinase system in an ecological and evolutionary context; Current Opinion in Plant Biology; 8: 264-271; 2005

- Daniel W. Nebert, and Timothy P. Dalton; The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis; Nature Reviews Cancer; 6: 947-960; December 2006

- Daniel W. Nebert; Polymorphisms in Drug-Metabolizing Enzymes: What Is Their Clinical Relevance and Why Do They Exist?; Am. J. Hum. Genet.; 60: 265-271; 1997

- Daniela Ennulat, Dana Walker, Frances Clemo, Michal Magid-Slav, David Ledieu, Mark Graham, Suzanne Botts, and Laura Boone; Effects of Hepatic Drug-metabolizing Enzyme Induction on Clinical Pathology Parameters in Animals and Man; Toxicologic Pathology; 38: 810-828; 2010

- Don James, Sridevi Devaraj, Prasad Bellur, Shantala Lakkanna, John Vicini, and Sekhar Boddupalli; Novel concepts of broccoli sulforaphanes and disease: induction of phase II antioxidant and detoxification enzymes by enhanced-glucoraphanin broccoli; Nutrition Reviews; 70(11): 654-665; 2012

- Dorette T. H. Verhoeven, Hans Verhagen, R. Alexandra Goldbohm, Piet A. van den Brandt, and Geert van Poppel; A review of mechanisms underlying anticarcinogenicity by brassica vegetables; Chemico-Biological Interactions; 103: 79-129; 1997

- Douglas Hanahan, and Robert A. Weinberg; Hallmarks of Cancer: The Next Generation; Cell; 144: 646-674; 4 March 2011

- Douglas Hanahan, and Robert A. Weinberg; The Hallmarks of Cancer; Cell; 100: 57–70; 7 January 2000

- Eduardo A. S. Rosa, and Ana S. Rodrigues; Total and Individual Glucosinolate Content in 11 Broccoli Cultivars Grown in Early and Late Seasons; HortScience; 36(1): 56-59; 2001

- Edward S. Kimball, Craig R. Schneider, Nathaniel H. Wallace, and Pamela J. Hornby; Agonists of cannabinoid receptor 1 and 2 inhibit experimental colitis induced by oil of mustard and by dextran sulfate sodium; Am J Physiol Gastrointest Liver Physiol; 291: G364-G371; 2006

- Elaine R. Mardis; Genome Sequencing and Cancer; Curr Opin Genet Dev.; 22(3): 245-250; June 2012

- Elena P. Moiseeva, and Margaret M. Manson; Dietary Chemopreventive Phytochemicals: Too Little or Too Much?; Cancer Prev Res; 2(7): 611-616

- Erik Andréasson, Lise Bolt Jørgensen, Anna-Stina Höglund, Lars Rask, and Johan Meijer; Different Myrosinase and Idioblast Distribution in Arabidopsis and Brassica napus; Plant Physiology; 127: 1750-1763; December 2001

- Esther F. A. Brandon, Christiaan D. Raap, Irma Meijerman, Jos H. Beijnen, and Jan H. M. Schellens; An update on in vitro test methods in human hepatic drug biotransformation research: pros and cons; Toxicology and Applied Pharmacology; 189: 233-246; 2003

- F. Peter Guengerich; Cytochrome P450 and Chemical Toxicology; Chem. Res. Toxicol.; 21: 70-83; 2008

- Fangxing Yanga, Shulin Zhuanga, Chao Zhangb, Heping Daib, and Weiping Liu; Sulforaphane inhibits CYP1A1 activity and promotes genotoxicity induced by 2,3,7,8-tetrachlorodibenzo-p-dioxin in vitro; Toxicology and Applied Pharmacology; 269(3): 226-232; 15 June 2013

- Frédéric Thomas, Daniel Fisher, Philippe Fort, Jean-Pierre Marie, Simon Daoust, Benjamin Roche, Christoph Grunau, Céline Cosseau, Guillaume Mitta, Stephen Baghdiguian, François Rousset, Patrice Lassus, Eric Assenat, Damien Grégoire, Dorothée Missé, Alexander Lorz, Frédérique Billy, William Vainchenker, François Delhommeau, Serge Koscielny, Raphael Itzykson, Ruoping Tang, Fanny Fava, Annabelle Ballesta, Thomas Lepoutre, Liliana Krasinska, Vjekoslav Dulic, Peggy Raynaud, Philippe Blache, Corinne Quittau-Prevostel, Emmanuel Vignal, Hélène Trauchessec, Benoit Perthame, Jean Clairambault, Vitali Volpert, Eric Solary, Urszula Hibner, and Michael E. Hochberg; Applying ecological and evolutionary theory to cancer. A long and winding road; Evolutionary Applications; 6: 1-10; 2012

- Gary D. Stoner, Charlotte Adams, Laura A. Kresty, Shantu G. Amin, Dhimant Desai, Stephen S. Hecht, Sharon E. Murphy, and Mark A. Morse; Inhibition of N´-nitrosonornicotine-induced esophageal tumorigenesis by 3-phenylpropyl isothiocyanate; Carcinogenesis; 19(12): 2139-2143; 1998

- Gérald S. Remaud, Yves-Loïc Martin, Gilles G. Martin, Norbert Naulet, and Gérard J. Martin; Authentication of Mustard Oils by Combined Stable Isotope Analysis (SNIF-NMR and IRMS); J. Agric. Food Chem.; 45(5): 1844-1848; 1997

- Gölge Sarıkamış, Ahmet Balkaya, and Ruhsar Yanmaz; Glucosinolates within a collection of white head cabbages (Brassica oleracea var. capitata sub.var. alba) from Turkey; African Journal of Biotechnology; 8(19): 5046-5052; 5 October 2009

- Gong Yang, Yu-Tang Gao, Xiao-Ou Shu, Qiuyin Cai, Guo-Liang Li, Hong-Lan Li, Bu-Tian Ji, Nathaniel Rothman, Marcin Dyba, Yong-Bing Xiang, Fung-Lung Chung, Wong-Ho Chow, and Wei Zheng; Isothiocyanate exposure, glutathione S-transferase polymorphisms, and colorectal cancer risk; Am J Clin Nutr; 91: 704-711; 2010

- Goodarz Danaei, Stephen Vander Hoorn, Alan D. Lopez, Christopher J. L. Murray, Majid Ezzati, and the Comparative Risk Assessment collaborating group (Cancers); Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors; Lancet; 366: 1784–1793; 2005

- H. Ishizaki, J. F. Brady, S. M. Ning, and C. S. Yang; Effect of phenethyl isothiocyanate on microsomal N-nitrosodimethylamine metabolism and other monooxygenase activities; Xenobiotica; 20(3): 255-264; 1990

- Hanan S. Samaha, Gary J. Kelloff, Vernon Steele, Chinthalapally V. Rao, and Bandaru S. Reddy; Modulation of Apoptosis by Sulindac, Curcumin, Phenylethyl-3-methylcaffeate, and 6-Phenyihexyl Isothiocyanate: Apoptotic Index as a Biomarker in Colon Cancer Chemoprevention and Promotion; Cancer Research; 57: 1301-1305; 1 April 1997

- Ira Bhatnagar, and Se-Kwon Kim; Immense Essence of Excellence: Marine Microbial Bioactive Compounds; Mar. Drugs; 8: 2673-2701; 2010

- Ishita Ahuja, Nicole Marie van Dam, Per Winge, Marianne Trælnes, Aysel Heydarova, Jens Rohloff, Mette Langaas, and Atle Magnar Bones; Plant defence responses in oilseed rape MINELESS plants after attack by the cabbage moth Mamestra brassicae; Journal of Experimental Botany, 66(2): 579–592; 2015

- Isobel A. P. Parkin, Chushin Koh, Haibao Tang, Stephen J. Robinson, Sateesh Kagale, Wayne E. Clarke, Chris D. Town, John Nixon, Vivek Krishnakumar, Shelby L. Bidwell, France Denoeud, Harry Belcram, Matthew G. Links, Jérémy Just, Carling Clarke, Tricia Bender, Terry Huebert, Annaliese S. Mason, J. Chris Pires, Guy Barker, Jonathan Moore, Peter G. Walley, Sahana Manoli, Jacqueline Batley, David Edwards, Matthew N. Nelson, Xiyin Wang, Andrew H. Paterson, Graham King, Ian Bancroft, Boulos Chalhoub, and Andrew G. Sharpe; Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea; Genome Biology; 15: 1-18; 2014

- Ivana Radojčić Redovniković, Tatjana Glivetić, Karmela Delonga, and Jasna Vorkapić-Furač; Glucosinolates and their potential role in plant; Periodicum Biologorum; 110(4): 297-309; 2008

- Izabela Miranda de Castro, Marianna Ramos dos Anjos, and Elba dos Santos de Oliveira; Dterminação De Isotiocianato de Benzila Em Carica papaya Utilizando Cromatografia Gasosa Com Detectores Selectivos; Quim. Nova; 31(8): 1953-1959; 2008

- Jane V. Higdon, Barbara Delage, David E. Williams, and Roderick H. Dashwood; Cruciferous Vegetables and Human Cancer Risk: Epidemiologic Evidence and Mechanistic Basis; Pharmacol Res.; 55(3): 224-236; March 2007

- Jan-Willem de Kraker, and Jonathan Gershenzon; From Amino Acid to Glucosinolate Biosynthesis: Protein Sequence Changes in the Evolution of Methylthioalkylmalate Synthase in Arabidopsis; The Plant Cell; 23: 38-53; January 2011

- Jayalakshmi Sridhar, Jiawang Liu, Maryam Foroozesh, and Cheryl L. Klein Stevens; Insights on Cytochrome P450 Enzymes and Inhibitors Obtained Through QSAR Studies; Molecules; 17: 9283-9305; 2012

- Jifang Zhang, Xiaobo Wang, Feng Cheng, Jian Wu, Jianli Liang, Wencai Yang, and Xiaowu Wang1; Lineage-specific evolution of Methylthioalkylmalate synthases (MAMs) involved in glucosinolates biosynthesis; Frontiers In Plant Science; 6(18): 1-13; February 2015

- Jingyin Yu, Meixia Zhao, Xiaowu Wang, Chaobo Tong, Shunmou Huang, Sadia Tehrim, Yumei Liu, Wei Hua, and Shengyi Liu; Bolbase: a comprehensive genomics database for Brassica oleracea; BMC Genomics; 14(664): 1-7; 2013

- Jiří Totušek, Jan Tříska, Danuše Lefnerová, Jan Strohalm, Naděžda Vrchotová, Ondřej Zendulka, Jiřina Průchová, Jana Chaloupková, Pavla Novotná, and Milan Houška; Contents of Sulforaphane and Total Isothiocyanates, Antimutagenic Activity, and Inhibition of Clastogenicity in Pulp Juices from Cruciferous Plants; Czech J. Food Sci.; 29(5): 548-556; 2011

- Joanna Kapusta-Duch, and Teresa Leszczyńska; Comparison of Vitamin C and β-Carotene in Cruciferous Vegetables Grown in Diversified Ecological Conditions; Pol. J. Environ. Stud.; 22(1): 167-173; 2013

- Johanna W. Lampe, and Sabrina Peterson; Brassica, Biotransformation and Cancer Risk: Genetic Polymorphisms Alter the Preventive Effects of Cruciferous Vegetables; J. Nutr.; 132: 2991-2994; 2002.

- John D. Hayes, Michael O. Kelleher, and Ian M. Eggleston; The cancer chemopreventive actions of phytochemicals derived from glucosinolates; Eur J Nutr; 47: 73-88; 2008

- John O. Miners, and Donald J. Birkett; Cytochrome P4502C9: an enzyme of major importance in human drug metabolism; Br J Clin Pharmacol; 45: 525-538; 1998

- K. Hennig, R. Verkerk, M. A. J. S. van Boekel, M. Dekker, and G. Bonnema; Food science meets plant science: A case study on improved nutritional quality by breeding for glucosinolate retention during food processing; Trends in Food Science & Technology; 35: 61-68; 2014

- Ka Lung Cheung, and Ah-Ng Kong; Molecular Targets of Dietary Phenethyl Isothiocyanate and Sulforaphane for Cancer Chemoprevention; The AAPS Journal; 12(1): 87-97; March 2010

- Kalimuthu Senthilkumar, and Se-Kwon Kim; Marine Invertebrate Natural Products for Anti-Inflammatory and Chronic Diseases; Evidence-Based Complementary and Alternative Medicine; 2013: 1-10; 2013

- Karine Mahéo, Fabrice Morel, Sophie Langouët, Harry Kramer, Eric Le Ferrec, Brian Ketterer, and André Guillouzo; Inhibition of Cytochromes P-450 and Induction of Glutathione S-Transferases by Sulforaphane in Primary Human and Rat Hepatocytes; Cancer Research; 57: 3649-3652; 1 September 1997

- Kristi A. Steinmetz, and John D. Potter; Vegetables, Fruit, and Cancer Prevention; J Am Diet Assoc.; 96(10): 1027-1039; October 1996

- Kumars Shahrokh, Anita Orendt, Garold Yost, and Tom Cheatham; Quantum mechanically derived AMBER-compatible heme parameters for various states of

the cytochrome P450 catalytic cycle; J. Comput. Chem.; 33(2): 119-133; 15 January 2012

- Lars Rask, Erik Andréasson, Barbara Ekbom, Susanna Eriksson, Bo Pontoppidan, and Johan Meijer; Myrosinase: gene family evolution and herbivore defense in Brassicaceae; Plant Molecular Biology; 42: 93–113; 2000

- Li Tang, and Yuesheng Zhang; Dietary Isothiocyanates Inhibit the Growth of Human Bladder Carcinoma Cells; J. Nutr.; 134: 2004-2010; 2004

- Li Tang, and Yuesheng Zhang; Dietary Isothiocyanates Inhibit the Growth of Human Bladder Carcinoma Cells; J. Nutr.; 134: 2004-2010; 2004

- Li Tang, and Yuesheng Zhang; Mitochondria are the primary target in isothiocyanate-induced apoptosis in human bladder cancer cells; Mol Cancer Ther; 4(8): 1250-1259; 2005

- Lilian W. Coelho, Georgia M. A. Junqueira, Juan O. M. Herrera, Sérgio de P. Machado, and Bianca da C. Machado; Aplicação de Mecânica Molecular em Química Inorgânica; Química Nova; 22(3): 396-404; 1999

- Linda B. von Weymarn, Jamie A. Chun, Gabriel A. Knudsen, and Paul F. Hollenberg; Effects of Eleven Isothiocyanates on P450 2A6- and 2A13-Catalyzed Coumarin 7-Hydroxylation; Chem. Res. Toxicol.; 20(9): 1252–1259; 2007

- Lionelle Nugon-Baudon, and Sylvie Rabot; Glucosinolates And Glucosinolate derivatives: Implications For Protection Against Chemical Carcinogenesis; Nutrition Research Reviews; 7: 205-231; 1994

- Lixin Mi, Anthony J.Di Pasqua, and Fung-Lung Chung; Proteins as binding targets of isothiocyanates in cancer prevention; Carcinogenesis; 32(10): 1405-1413; 2011

- Lixin Mi, Zhen Xiao, Timothy D. Veenstra, and Fung-Lung Chung; Proteomic Identification of Binding Targets of Isothiocyanates: A Perspective on Techniques; J Proteomics; 74(7): 1036-1044; 10 June 2011

- Lucy R. Yates, and Peter J. Campbell; Evolution of the cancer genome; Nat Rev Genet.; 13(11): 795-806; November 2012

- Luis Frederico Molina-Vargas; Mechanism of action of isothiocyanates. A review; Agronomia Colombiana; 31(1): 68-75; 2013

- M. Gratacós-Cubarsí, A. Ribas-Agustí, J. A. García-Regueiro, and M. Castellari; Simultaneous Evaluation of Intact Glucosinolates and Phenolic Compounds by UPLC-DAD-MS/MS in Brassica oleracea L. var. botrytis; Food Chem; 121(1): 257-263; 2010

- M. Grazia Botti, Malcolm G. Taylor, and Nigel P. Botting; Studies on the Mechanism of Myrosinase Investigation Of The Effect Of Glycosyl Acceptors On

Enzyme Activity; The Journal of Biological Chemistry; 270(35): 20530-20535; 1995

- M. La Marca, P. Beffy, C, Della Croce, P. G. Gervasi, R. Iori, E. Puccinelli, and V. Longo; Strucutral influence of isothiocyanates on expression of cytochrome P450, phase II enzymes, and activation of Nrf2 in primary rat hepatocytes; Food and Chemical Toxicology; 50: 2822-2830; 2012

- Magnus Ingelman-Sundberg, and Cristina Rodriguez-Antona; Pharmacogenetics of drug-metabolizing enzymes: implications for a safer and more effective drug therapy; Phil. Trans. R. Soc. B; 360: 1563–1570; 2005

- Marc G. Sturgill, and George H. Lambert; Xenobiotic-induced hepatotoxicity: mechanisms of liver injury and methods of monitoring hepatic function; Clinical Chemistry; 43(8): 1512–1526; 1997

- Margaret R. Spitz, Cherie M. Duphorne, Michelle A. Detry, Patricia C. Pillow, Christopher I. Amos, Lei lei, Mariza de Andrade, Xiangjun Gu, Waun K. Hong, and Xifeng Wu.; Dietary Intake of Isothiocyanates: Evidence of a Joint Effect with Glutathione S-Transferase Polymorphisms in Lung Cancer Risk; Cancer Epidemiology, Biomarkers & Prevention; 9:1017-1020; October 2000

- María Elena Cartea, Marta Francisco, Pilar Soengas, and Pablo Velasco; Phenolic Compounds in Brassica Vegetables; Molecules; 16: 251-280; 2011

- Mark Percival; Phytonutrients And Detoxification; Clinical Nutrition Insights; 5(2): 1-4; 1997

- Markus Piotrowski, Andreas Schemenewitz, Anna Lopukhina, Axel Müller, Tim Janowitz, Elmar W. Weiler, and Claudia Oecking; Desulfoglucosinolate Sulfotransferases from Arabidopsis Thaliana Catalyze the Final Step in the Biosynthesis of the Glucosinolate Core Structure; The Journal of Biological Chemistry; 279(49): 50717-50725; 3 December 2004

- Masahiko Ishida, Masakazu Hara, Nobuko Fukino, Tomohiro Kakizaki, and Yasujiro Morimitsu; Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables; Breeding Science; 64: 48-59; 2014

- Masakazu Hara, Yoko Fujii, Yuri Sasada, and Toru Kuboi; cDNA Cloning of Radish (Raphanus sativus) Myrosinase and Tissue-Specific Expression in Root; Plant Cell Physiol.; 41(10): 1102-1109; 2000

- Matthew A. Wallig, Kathleen M. Heinz-Taheny, Donna L. Epps, and Tamara Gossman; Synergy among Phytochemicals within Crucifers: Does It Translate into Chemoprotection?; J. Nutr.; 135: 2972S-2977S; 2005

- Matthew Bridges, Alexandra M. E. Jones, Atle M. Bones, Chris Hodgson, Rosemary Cole, Elspeth Bartlet, Roger Wallsgrove, Vassiliki K. Karapapa, Nigel Watts, and John T. Rossiter; Spatial organization of the glucosinolate–myrosinase

system in brassica specialist aphids is similar to that of the host plant; Proc. R. Soc. Lond. B; 269: 187-191; 2002

- Me-Hea Park, Mariadhas Valan Arasu, No-Young Park, Yun-Jong Choi, Sang-Won Lee, Naif Abdullah Al-Dhabi, Jung Bong Kim, and Sun-Ju Kim; Variation of glucoraphanin and glucobrassicin: anticancer components in Brassica during processing; Food Sci. Technol; 33(4): 624-631; October-December 2013

- Meira Epplein, Lynne R. Wilkens, Maarit Tiirikainen, Marcin Dyba, Fung-Lung Chung, Marc T. Goodman, Suzanne P. Murphy, Brian E. Henderson, Laurence N. Kolonel, and Loïc Le Marchand; Urinary Isothiocyanates; Glutathione S-Transferase M1, T1, and P1 Polymorphisms; and Risk of Colorectal Cancer: The Multiethnic Cohort Study; Cancer Epidemiol Biomarkers Prev; 18(1): 314-320; January 2009

- Melinda C. Myzak, and Roderick H. Dashwood; Chemoprotection by sulforaphane: Keep one eye beyond Keap1; Cancer Lett.; 233(2): 208-218; 28 February 2006

- Mi Kyung Kim, and Jung Han Yoon Park; Cruciferous vegetable intake and the risk of human cancer: epidemiological evidence; Proceedings of the Nutrition Society; 68: 103-110; 2009

- Michael B. Sporn; The war on cancer; Lancet; 347: 1377-1381; 1996

- Michael Benn; Glucosinolates; Pure & Appl. Chem.; 49: 197-210; 1977

- Michael J. Wargovich, Cynthia Woods, Destiny M. Hollis, and Mary E. Zander; Herbals, Cancer Prevention and Health; J. Nutr.; 131: 3034S-3036S; 2001

- Miki Nakajima, Ryoko Yoshida, Noriaki Shimada, Hiroshi Yamazaki, and Tsuyoshi Yokoi; Inhibition And Inactivation Of Human Cytochrome P450 Isoforms By Phenethyl Isothiocyanate; Drug Metabolism And Disposition; 29(8): 1110-1113; 2001

- Monica I. Jushchyshyn, Ute M. Kent, and Paul F. Hollenberg; The Mechanism-Based Inactivation Of Human Cytochrome P450 2B6 By Phencyclidine; Drug Metabolism And Disposition; 31(1): 46-52; 2003

- Niki Papavramidou, Theodossis Papavramidis, and Thespis Demetriou; Ancient Greek and Greco–Roman Methods in Modern Surgical Treatment of Cancer; Ann Surg Oncol; 17: 665-667; 2010

- Oleg Trott, and Arthur J. Olson; AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading; J Comput Chem.; 31(2): 455-461; 30 January 2010

- Ori Scott, Elaine Galicia-Connolly, Denise Adams, Soleil Surette, Sunita Vohra, and Jerome Y. Yager; The Safety of Cruciferous Plants in Humans: A Systematic Review; Journal of Biomedicine and Biotechnology; 2012: 1-29; 2012

- Otis W. Brawley; Avoidable Cancer Deaths Globally; CA Cancer J Clin; 61: 67-68; 2011

- Paul Talalay, and Jed W. Fahey; Phytochemicals from Cruciferous Plants Protect against Cancer by Modulating Carcinogen Metabolism; J. Nutr.; 131: 3027S-3033S; 2001

- Petra Jancova, Pavel Anzenbacher, and Eva Anzenbacherova; Phase II Drug Metabolizing Enzymes; Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.; 154(2): 103-116; June 2010

- Philippe Irigaray, and Dominique Belpomme; Basic properties and molecular mechanisms of exogenous chemical carcinogens; Carcinogenesis; 31(2): 135-148; 2010

- Preetha Anand, Ajaikumar B. Kunnumakara, Chitra Sundaram, Kuzhuvelil B. Harikumar, Sheeja T. Tharakan, Oiki S. Lai, Bokyung Sung, and Bharat B. Aggarwal; Cancer is a Preventable Disease that Requires Major Lifestyle Changes; Pharmaceutical Research; 25(9): 2097-2116; September 2008

- R. Thapliyal, and G.B. Maru; Inhibition of cytochrome P450 isozymes by curcumins in vitro and in vivo; Food and Chemical Toxicology; 39(6): 541–547; June 2001

- Rajamanickam Baskar, Kuo Ann Lee, Richard Yeo, and Kheng-Wei Yeoh; Cancer and Radiation Therapy: Current Advances and Future Directions; International Journal of Medical Sciences; 9(3): 193-199; 2012

- Rex Munday, Paulette Mhawech-Fauceglia, Christine M. Munday, Joseph D. Paonessa, Li Tang, John S. Munday, Carolyn Lister, Paula Wilson, Jed W. Fahey, Warren Davis, and Yuesheng Zhang; Inhibition of Urinary Bladder Carcinogenesis by Broccoli Sprouts; Cancer Res; 68(5): 1593-1600; 2008

- Rex Munday, Yuesheng Zhang, Christine M. Munday, Meghana V. Bapardekar, and Joseph D. Paonessa; Structure-activity relationships and organ specificity in the induction of GST and NQO1 by alkyl-aryl isothiocyanates; Pharm Res.; 25(9): 2164-2170.; September 2008

- Richard Weinshilboum; Inheritance and Drug Response; N Engl J Med; 348(6): 529-537; 6 February 2003

- Robert A. Gatenby, and Robert J. Gillies; Why Do Cancers Have High Aerobic Glycolysis?; Nature Reviews Cancer; 4: 891-899; November 2004

- Rosa L. Moreno, Theunis Goosen, Ute M. Kent, Fung-Lung Chung, and Paul F. Hollenberg; Differential Effects of Naturally Occurring Isothiocyanates on the Activities of Cytochrome P450 2E1 and the Mutant P450 2E1 T303A; Archives of Biochemistry and Biophysics; 391(1): 99-110; 1 July 2001

- Rosa L. Moreno, Ute M. Kent, Kimberly Hodge, and Paul F. Hollenberg; Inactivation of Cytochrome P450 2E1 by Benzyl Isothiocyanate; Chem. Res. Toxicol.; 12(7): 582-587; 1999

- Ross A. McKinnon, Michael J. Sorich, and Michael B. Ward; Cytochrome P450 Part 1: Multiplicity and Function; J Pharm Pract Res; 38: 55-7; 2008

- Rui Hai Liu; Potential Synergy of Phytochemicals in Cancer Prevention: Mechanism of Action; J. Nutr.; 134: 3479S-3485S; 2004

- S. A. McNaughton, and G. C. Marks; Development of a food composition database for the estimation of dietary intakes of glucosinolates, the biologically active constituents of cruciferous vegetables; British Journal of Nutrition; 90: 687-697; 2003

- S. L. Kelly, D. C. Lamb, and D. E. Kelly; 8th International Symposium on Cytochrome P450 Biodiversity and Biotechnology; Biochemical Society Transactions; 34(6): 1159-1160; 2006

- Sabrina Peterson, Yvonne Schwarz, Shuying S. Li, Lin Li, Irena B. King, Chu Chen, David L. Eaton, John D. Potter, and Johanna W. Lampe; CYP1A2, GSTM1, and GSTT1 Polymorphisms and Diet Effects on CYP1A2 Activity in a Crossover Feeding Trial; Cancer Epidemiol Biomarkers Prev; 18(11): 3118-3125; November 2009

- Sandi L. Navarro, Fei Li, and Johanna W. Lampe; Mechanisms of Action of Isothiocyanates in Cancer Chemoprevention: An Update; Food Funct.; 2(10): 579-587; 14 October 2011

- Sean C. Bendall, and Garry P. Nolan; From single cells to deep phenotypes in cancer; Nature Biotechnology; 30(7): 639-647; July 2012

- Serkadis M. Getahun, and Fung-Lung Chung; Conversion of Glucosinolates to Isothiocyanates in Human after Ingestion of Cooked Watercress; Cancer Epidemiology, Biomarkers & Prevention; 8: 447-451; May 1999

- Shivapriya Manchali, Kotamballi N. Chidambara Murthy, and Bhimanagouda S. Patil; Crucial facts about health benefits of popular cruciferous vegetables; Journal of Functional Foods; 4(1): 94-106; January 2012

- Silvia Barcelo, John M. Gardiner, Andreas Gescher, and J. Kevin Chipmanw; CYP2El-mediated mechanism of anti-genotoxicity of the broccoli constituent sulforaphane; Carcinogenesis; 17(2): 277-282; 1996

- Stefan Wilkening, Frank Stahl, and Augustinus Bader; Comparison Of Primary Human Hepatocytes And Hepatoma Cell Line HEPG2 With Regard To Their Biotransformation Properties; Drug Metabolism And Disposition; 31(8): 1035-1042; 2003

- Stephen S. Hecht; Chemoprevention of Cancer by Isothiocyanates, Modifiers of Carcinogen Metabolism; J. Nutr.; 129: 768S-774S; 1999

- Steven I. Hajdu; A Note From History: Landmarks in History of Cancer, Part 1; Cancer; 117: 1097-1102; 2011

- Steven I. Hajdu; Greco-Roman Thought about Cancer; Cancer; 100(10): 2048-2051; 15 May 2004

- Susanna Eriksson, Erik Andréasson, Barbara Ekbom, Georg Granér, Bo Pontoppidan, Jan Taipalensuu, Jiaming Zhang, Lars Rask, and Johan Meijer; Complex Formation of Myrosinase Isoenzymes in Oilseed Rape Seeds Are Dependent on the Presence of Myrosinase-Binding Proteins; Plant Physiology; 129: 1592-1599; August 2002

- Susanne Textor, and Jonathan Gershenzon; Herbivore induction of the glucosinolate-myrosinase defense system: major trends, biochemical bases and ecological significance; Phytochem Rev; 8: 149-170; 2009

- Sven-Eric Jordt, Diana M. Bautista, Huai-hu Chuang, David D. McKemy, Peter M. Zygmunt, Edward D. Högestätt, Ian D. Meng, and David Julius; Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1; Nature; 427: 260-265; 15 January 2004

- Tamara Sotelo, Pilar Soengas, Pablo Velasco, Víctor M. Rodríguez, and María Elena Cartea; Identification of Metabolic QTLs and Candidate Genes for Glucosinolate Synthesis in Brassica oleracea Leaves, Seeds and Flower Buds; PLOS ONE; 9(3): 1-11; March 2014

- Thea D. Tlsty, and Lisa M. Coussens; Tumor Stroma and Regulation of Cancer Development; Annu. Rev. Pathol. Mech. Dis.; 1: 119–50; 2006

- Theresa A. Shapiro, Jed W. Fahey, Kristina L. Wade, Katherine K. Stephenson, and Paul Talalay; Chemoprotective Glucosinolates and Isothiocyanates of Broccoli Sprouts: Metabolism and Excretion in Humans; Cancer Epidemiology, Biomarkers & Prevention; 10: 501-508; May 2001

- Theresa J. Smith, Zuyu Guo, F. Peter Guengerich, and Chung S. Yang; Metabolism of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) by human cytochrome P450 1A2 and its inhibition by phenethyl isothiocyanate; Carcinogenesis; 17(4): 809-813; April 1996

- Theresa J. Smith, Zuyu Guo, Paul E. Thomas, Fung-Lung Chung, Mark A. Morse, Karin Elkind, and Chung S. Yang; Metabolism of 4-(Methylnitrosamino)-l-(3-pyridyl)-l-butanone in Mouse Lung Microsomes and Its Inhibition by Isothiocyanates; Cancer Research; 50: 6817-6822; 1 November 1990

- Theunis C. Goosen, Danielle E. Mills, and Paul F. Hollenberg; Effects of Benzyl Isothiocyanate on Rat and Human Cytochromes P450: Identification of

Metabolites Formed by P450 2B1; The Journal OF Pharmacology And Experimental Therapeutics; 296(1): 198-206; 2001

- Theunis C. Goosen, Ute M. Kent, Linda Brand, and Paul F. Hollenberg; Inactivation of Cytochrome P450 2B1 by Benzyl Isothiocyanate, a Chemopreventative Agent from Cruciferous Vegetables; Chem. Res. Toxicol.; 13(12): 1349-1359; 2000

- Thomas M. Glendening, and Jonathan E. Poulton; Glucosinolate Biosynthesis: Sulfation of Desulfobenzylglucosinolate By Cell-Free Extracts Of Cress (Lepidium Sativum L.) Seedlings; Plant Physiol.; 86: 319-321; 1988

- Tom Lynch, and Amy Price; The Effect of Cytochrome P450 Metabolism on Drug Response, Interactions, and Adverse Effects; Am Fam Physician; 76: 391-6; 2007

- Vanessa Rungapamestry, Alan J. Duncan, Zoë Fuller, and Brian Ratcliffe; Effect of cooking brassica vegetables on the subsequent hydrolysis and metabolic fate of glucosinolates; Proceedings of the Nutrition Society; 66: 69-81; 2007

- Victoria Yoxall, Peter Kentish, Nick Coldham, Nikolai Kuhnert, Maurice J. Sauer, and Costas Ioannides; Modulation of hepatic cytochromes P450 and phase II enzymes by dietary doses of sulforaphane in rats: Implications for its chemopreventive activity; Int. J. Cancer; 117: 356-362; 2005

- Vincent T. DeVita, Jr., and Edward Chu; A History of Cancer Chemotherapy; Cancer Res; 68(21): 8643–53; 2008

- Virginia Lambrix, Michael Reichelt, Thomas Mitchell-Olds, Daniel J. Kliebenstein, and Jonathan Gershenzon; The Arabidopsis Epithiospecifier Protein Promotes the Hydrolysis of Glucosinolates to Nitriles and Influences Trichoplusia ni Herbivory; The Plant Cell; 13: 2793-2807; December 2001

- W. P. Steward, and K. Brown; Cancer chemoprevention: a rapidly evolving field; British Journal of Cancer; 109: 1-7; 2013

- William W. Li, Vincent W. Li, Michelle Hutnik, and Albert S. Chiou; Tumor Angiogenesis as a Target for Dietary Cancer Prevention; Journal of Oncology; 2012: 1-23; 2012

- William W. Li, Vincent W. Li, Michelle Hutnik, and Albert S. Chiou; Tumor Angiogenesis as a Target for Dietary Cancer Prevention; Journal of Oncology; 2012: 1-23; 2012

- Xiang Wu, Qing-hua Zhou, and Ke Xu; Are isothiocyanates potential anti-cancer drugs?; Acta Pharmacol Sin; 30(5): 501-512; May 2009

- Xiu Fen Yang, and Fan Dian Zeng; Environmental factors, cytochrome P-450 polymorphisms, cytochrome P-450 inhibitors and lung cancer; Asian Journal of Pharmacodynamics and Pharmacokinetics; 6(3): 193-195; 2006

- Yasushi Yoshigae, Chitra Sridar, Ute M. Kent, and Paul F. Hollenberg; The Inactivation of Human CYP2E1 by Phenethyl Isothiocyanate, a Naturally Occurring Chemopreventive Agent, and Its Oxidative Bioactivation; Drug Metabolism and Disposition; 41: 858-869; April 2013

- Yeon Bok Kim, Xiaohua Li, Sun-Ju Kim, Haeng Hoon Kim, Jeongyeo Lee, HyeRan Kim, and Sang Un Park; MYB Transcription Factors Regulate Glucosinolate Biosynthesis in Different Organs of Chinese Cabbage (Brassica rapa ssp. pekinensis); Molecules; 18: 8682-8695; 2013

- Yoshimasa Nakamura, and Noriyuki Miyoshi; Electrophiles in Foods: The Current Status of Isothiocyanates and Their Chemical Biology; Biosci. Biotechnol. Biochem.; 74(2): 242-255; 2010

- Yuesheng Zhang, and Paul Talalay, Anticarcinogenic Activities of Organic Isothiocyanates. Chemistry and Metabolism; Cancer Research (Suppl); 54: 1976s-1981s; 1 April 1994

- Yuesheng Zhang, and Paul Talalay; Mechanism of Differential Potencies of Isothiocyanates as Inducers of Anticarcinogenic Phase 2 Enzymes; Cancer Research; 58: 4632-4639; 15 October 1998

- Yuesheng Zhang, Li Tang, and Veronica Gonzalez; Selected isothiocyanates rapidly induce growth inhibition of cancer cells; Molecular Cancer Therapeutics; 2: 1045-1052; 2003

- Yuesheng Zhang, Song Yao, and Jun Li; Vegetable-derived isothiocyanates: anti-proliferative activity and mechanism of action; Proceedings of the Nutrition Society; 65: 68-75; 2006

- Yuesheng Zhang; Allyl isothiocyanate as a cancer chemopreventive phytochemical; Mol Nutr Food Res; 54(1): 1-16; January 2010

- Yuesheng Zhang; The molecular basis that unifies the metabolism, cellular uptake and chemopreventive activities of dietary isothiocyanates; Carcinogenesis; 33(1): 2-9; 2012

- Yuesheng Zhang; The molecular basis that unifies the metabolism, cellular uptake and chemopreventive activities of dietary isothiocyanates; Carcinogenesis; 33(1): 2-9; 2012

- Yukio Mori, Akihiro Koide, Kenjiro Tatematsu, Shigeyuki Sugie, and Hideki Mori; Effects of a-naphthyl isothiocyanate and a heterocyclic amine, PhIP, on cytochrome P-450, mutagenic activation of various carcinogens and glucuronidation in rat liver; Mutagenesis; 20(1): 15-22; 2005

6.2 Books

- Awad, Atif B.; and Bradford, Peter G. – Nutrition and Cancer Prevention. In Nutrition and Disease Prevention. First Edition. New York: CRC Press, 2005. ISBN-10: 0849339456. Book 4. 640 Pages

- Bendich, Adrianne; and Deckelbaum, Richhard J. – Preventive Nutrition: The Comprehensive Guide for Health Professionals. Third Edition. Totowa (New Jersey): Humana Press, Inc., 2005. ISBN: 1-59259-880-3. 984 Pages

- Cooper, Geoffrey M. – The Cell: A Molecular Approach. Second Edition. Boston University: Sunderland (MA): Sinauer Associates, 2000. ISBN: 10: 0-87893-106-6. 689 Pages

- Hecht, Stephen S. – 2 Chemoprevention by Isothiocyanates. In Cancer Chemoprevention: Promising Cancer Chemoprevention Agents. Illustrated Edition. Totowa, New Jersey: G. J. Kelloff, E. T. Hawk, and C. C. Sigman (Humana Press Inc.), 2004. ISBN: 9781588290762. Volume 1. 21-35 15 Pages

- Ioannides, Costas – Cytochromes P450: Role in the Metabolism and Toxicity of Drugs. Royal Society of Chemistry, Cambridge, UK: RSC Publishing, 2008. ISBN: 978-0-85404-274-6. 521 Pages

- Mendelsohn, John; Howley, Peter M.; Israel, Mark A.; Gray, Joe W.; Thompson, Craig B. – The Molecular Basis of Cancer. Third Edition. Philadelphia: Saunders Elsevier Inc., 2008. ISBN: 978–1–4160–3703–3. 727 Pages

- Meskin, Mark S.; Bidlack, Wayne R.; Davies, Audra J.; Omaye, Stanley T. – Phytochemicals in Nutrition and Health. Boca Raton, Florida: CRC Press, 2002. ISBN: 1-58716-083-8. 224 Pages

- Monga, Satdarshan P. S.; and Cagle, Philip T. – Molecular Pathology of Liver Diseases. In Molecular Pathology Library. New York: Springer Science & Business Media, 2010. ISBN: 978-1-4419-7107-4. Volume 5. 954 Pages

- Nassar, Ala F.; Hollenberg, Paul F.; and Scatina, JoAnn – Drug Metabolism Handbook: Concepts and Applications. Hoboken, New Jersey: John Wiley & Sons, Inc., 2009. ISBN: 978-0-470-11803-0. 800 Pages

- Penner, Natalia; Woodward, Caroline; and Prakash, Chandra – Drug Metabolizing Enzymes And Biotransformation Reactions. In «ADME-Enabling Technologies in Drug Design and Development». First Edition. New Jersey: John Wiley & Sons, Inc., 2012. ISBN: 978-0-470-54278-1. Appendix. 545-565 21 Page(s)

- Rath, Matthias – Cellular Health Series: Cancer. First Edition. Santa Clara: MR Publishing, Inc., 2001. CA 95054. 40 Pages

- Rose, Randy L.; and Hodgson, Ernest – Metabolism of Toxicants. In « A Textbook of Modern Toxicology». Third Edition. Hoboken, New Jersey: Ernest Hodgson (John Wiley & Sons, Inc.), 2004. ISBN: 9780471265085. Chapter 7. 111-148 38 Pages

- Watson, Ronald Ross; and Preedy, Victor R. – Bioactive Foods and Extracts: Cancer Treatment and Prevention. New York: CRC Press, 2010. ISBN: 9781439816196. 663 Pages

- Wildman, Robert E.C.; Wildman, Robert; and Wallace, Taylor C. – Handbook of Nutraceuticals and Functional Foods. Third Edition. Boca Raton (Florida): CRC Press, 2002. ISBN 9780849387340. 568 Pages

- Wink, Michael – Annual Plant Reviews: Biochemistry of Plant Secondary Metabolism. Second Edition. Oxford, United Kingdom: Wiley-Blackwell, 2010. ISBN: 978-1-4051-8397-0. Volume 40. 464 Pages

6.3 Thesis

- Alnsour, Mohammad; Influence of exogenous factors on glucosinolate accumulation in horseradish (Armoracia rusticana Gaertn., Mey. & Scherb.). Technischen Universität Carolo-Wilhelmina zu Braunschweig, 2013. Doctor in Science Thesis. 191 Pages

- Borgen, Birgit Hafeld; Functional Analysis of Plant Idioblasts (Myrosin Cells) and their role in Defense, Development and Growth. Norwegian University of Science and Technology (NTNU), 2002. Doctor Scientiarum Thesis. 158 Pages

- Gaikovitch, Elena A.; Genotyping of the polymorphic drug metabolizing enzymes cytochrome P450 2D6 and 1A1, and N-acetyltransferase 2 in a Russian sample. Humboldt-Universität zu Berlin, 2003. Doctor in Medicine Thesis. 92 Pages

- Schlicht, Kari Elizabeth; Human Cytochrome P450-mediated Metabolism of Nicotine and Tobacco-specific Carcinogens: N-nitrosonornicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. University of Minnesota, 2007. Doctor of Philosophy Thesis. 122 Pages

- Tong, Zhan; Isothiocyanates Derivatives As Anti-Cancer Agents. National University Of Singapore, 2007. Master Thesis. 81 Pages

6.4 Website Section

- http://en.wikipedia.org/wiki/Edwin_Smith_Papyrus#mediaviewer/File:Edwin_Smith_Papyrus_v2.jpg, accessed on 22 February 2015

- http://globocan.iarc.fr/Pages/fact_sheets_population.aspx, accessed on 23 February 2015

- http://www.cancerresearchuk.org/cancer-info/cancerstats/world/incidence/, accessed on 18 February 2015

- http://www.cancerresearchuk.org/cancer-info/cancerstats/world/mortality/, accessed on 18 February 2015
- http://publications.cancerresearchuk.org/downloads/Product/CS_KF_WORLDWIDE.pdf, accessed on 18 February 2015
- http://www.wcrf.org/int/cancer-facts-figures/worldwide-data, accessed on 18 February 2015
- http://publications.cancerresearchuk.org/downloads/product/CS_REPORT_WORLD.pdf , accessed on 18 February 2015
- http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/radiotherapy/about/what-radiotherapy-is, accessed on 18 February 2015
- http://www.cancerresearchuk.org/cancer-info/healthyliving/introducingcancerprevention/?script=true, accessed on 18 February 2015
- http://globocan.iarc.fr/Pages/fact_sheets_population.aspx, accessed on 18 February 2015
- http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-027766.pdf, accessed on 18 February 2015
- http://www.ons.gov.uk/ons/dcp171778_367563.pdf, accessed on 18 February 2015
- http://www.rcsb.org/pdb/home/home.do, accessed on 06 March 2015
- http://www.rcsb.org/pdb/home/home.do, accessed on 25 March 2015
- http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm, accessed on 20 April 2015
- http://autodock.scripps.edu/faqs-help/faq/how-autodock-4-converts-binding-energy-kcal-mol-into-ki, accessed on 27 October 2015
- https://en.wikipedia.org/wiki/Born%E2%80%93Oppenheimer_approximation, accessed on 11 December 2015
- https://en.wikipedia.org/wiki/Hartree%E2%80%93Fock_method, accessed on 11 December 2015
- http://c125.chem.ucla.edu/NIH/MolMechanics.htm, accessed on 11 December 2015

- https://en.wikipedia.org/wiki/Molecular_mechanics, accessed on 11 December 2015
- https://en.wikipedia.org/wiki/Molecular_dynamics, accessed on 11 December 2015
- http://www.bch.cuhk.edu.hk/croucher11/tutorials/day1_autodock_presentation.pdf, accessed on 11 December 2015
- https://en.wikipedia.org/wiki/Iterated_local_search, accessed on 11 December 2015
- http://www.ecs.umass.edu/mie/labs/mda/mechanism/papers/genetic.html, accessed on 11 December 2015
- http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm, accessed on 6 January 2016

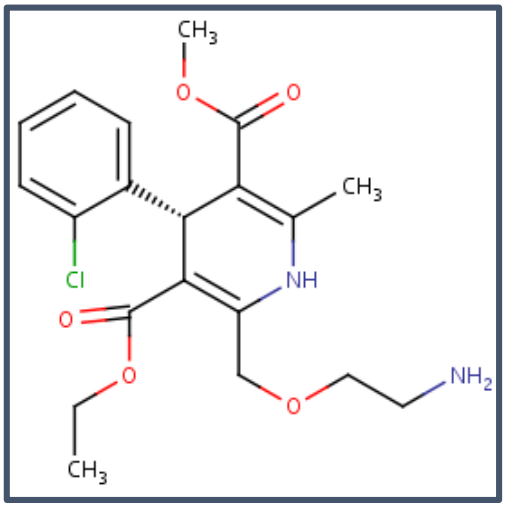
7. Annexes

7.1 Receptor-Ligand Complex Identification

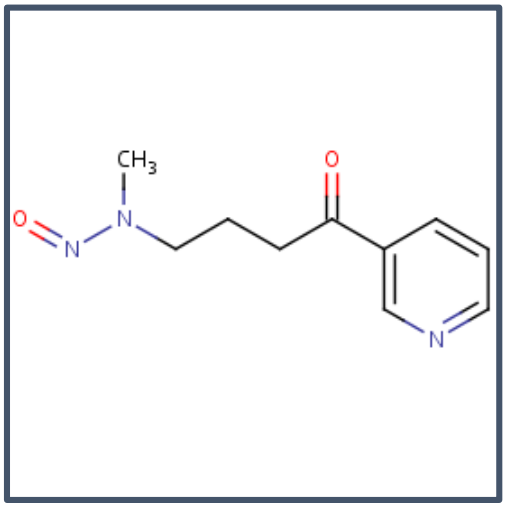
CYPs		Native Ligands			
Name	**PDB Code**	**Scientific Name**	**Common Name**	**PDB Code**	**Molecular Formula**
2A6	1Z10	2H-1-benzopyran-2-one	Coumarin	COU	$C_9 H_6 O_2$
	1Z11	9-methoxy-7H-furo[3,2-G][1]benzopyran-7-one	Methoxsalen	8MO	$C_{12} H_8 O_4$
	2FDU	N,N-dimethyl(5-(pyridin-3-yl)furan-2-yl)methanamine		D1G or Nic2c	$C_{12} H_{14} N_2 O$
	2FDV	N-methyl(5-(pyridin-3-yl)furan-2-yl)methanamine		D2G or Nic2b	$C_{11} H_{12} N_2 O$
	2FDW	(5-(pyridin-3-yl)furan-2-yl)methanamine		D3G or Nic2a	$C_{10} H_{10} N_2 O$
	2FDY	4,4'-dipyridyl disulfide	Adrithiol	D4G	$C_{10} H_8 N_2 S_2$
	2PG5	1,2-ethanediol	Ethylene Glycol	EDO	$C_2 H_6 O_2$
	2PG6				
	2PG7				
	3EBS	N-(4-ethoxyphenyl)acetamide	Phenacetin	N4E	$C_{10} H_{13} N O_2$
	3T3Q	(3S,4R)-3-ethyl-4-[(1-methyl-1H-imidazol- 5-yl)methyl]dihydrofuran-2(3H)-one	Pilocarpine	9PL	$C_{11} H_{16} N_2 O_2$
	3T3R	(3S,4R)-3-ethyl-4-[(1-methyl-1H-imidazol- 5-yl)methyl]dihydrofuran-2(3H)-one	Pilocarpine	9PL	$C_{11} H_{16} N_2 O_2$
	4EJJ	(S)-3-(1-methylpyrrolidin-2-yl)pyridine	Nicotine	NCT	$C_{10} H_{14} N_2$
	4RUI	(1S,5S)-4-methylidene-1-(propan-2-yl)bicyclo[3.1.0]hexane	Sabinene	SNE	$C_{10} H_{16}$
	2P85		indole	IND	$C_8 H_7 N$
	3T3S	(3S,4R)-3-ethyl-4-[(1-methyl-1H-imidazol- 5-yl)methyl]dihydrofuran-2(3H)-one	Pilocarpine	9PL	$C_{11} H_{16} N_2 O_2$

2A13	4EJG	(S)-3-(1-methylpyrrolidin-2-yl)pyridine	Nicotine	NCT	C_{10} H_{14} N_2
	4EJH	4-[methyl(nitroso)amino]-1-(pyridin-3-yl)butan-1-one		0QA or NNK	C_{10} H_{13} N_3 O_2
	4EJI	4-[methyl(nitroso)amino]-1-(pyridin-3-yl)butan-1-one		0QA or NNK	C_{10} H_{13} N_3 O_2
2B6	3IBD	4-(4-chlorophenyl)imidazole		CPZ	C_9 H_7 Cl N_2
	3QOA	4-benzylpyridine		3QO	C_{12} H_{11} N
	3QU8	4-(4-nitrobenzyl)pyridine		3QU	C_{12} H_{10} N_2 O_2
	3UA5	O3-ethyl O5-methyl 2-(2-azanylethoxymethyl)-4-(2-chlorophenyl)-6-methyl-1,4-dihydropyridine- 3,5-dicarboxylate	Amlodipine	06X	C_{20} H_{25} Cl N_2 O_5
	4I91	(+)-3,6,6-trimethylbicyclo[3.1.1]hept-2-ene	(+)-alpha-Pinene	TMH	C_{10} H_{16}
	4RQL	(1S,5S)-4-methylidene-1-(propan-2-yl)bicyclo[3.1.0]hexane	Sabinene	SNE	C_{10} H_{16}
	4RRT	(1R,6S)-3,7,7-trimethylbicyclo[4.1.0]hept- 3-ene	(+)-3-carene	3V4	C_{10} H_{16}
2C9	1OG2				
	1OG5		S-warfarin	SWF	C_{19} H_{16} O_4
	1R9O		Flurbiprofen	FLP	C_{15} H_{13} F O_2
	4NZ2	(2R)-N-{4-[(3-bromophenyl)sulfonyl]-2-chlorophenyl}- 3,3,3-trifluoro-2-hydroxy-2-methylpropanamide		2QJ	C_{16} H_{12} Br Cl F_3 N O_4 S

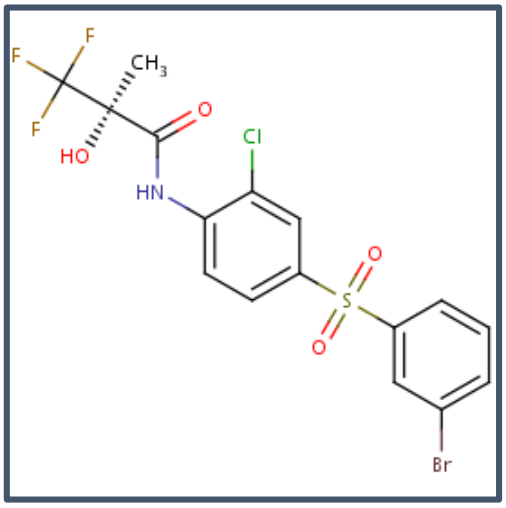
7.2 Native Ligands Molecular Structure



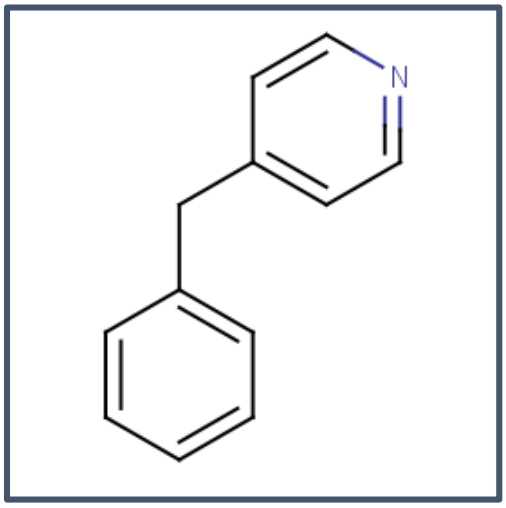
06X



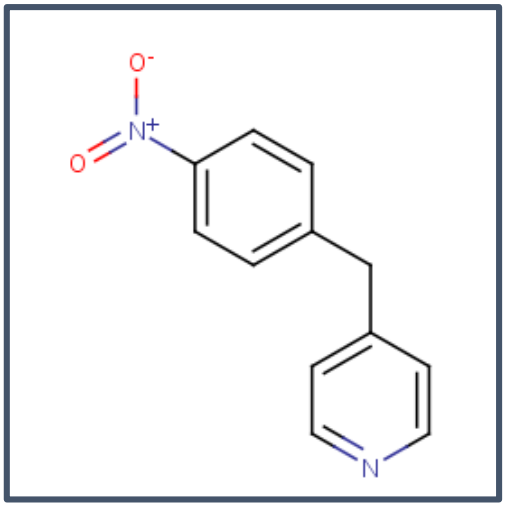
0QA or NNK



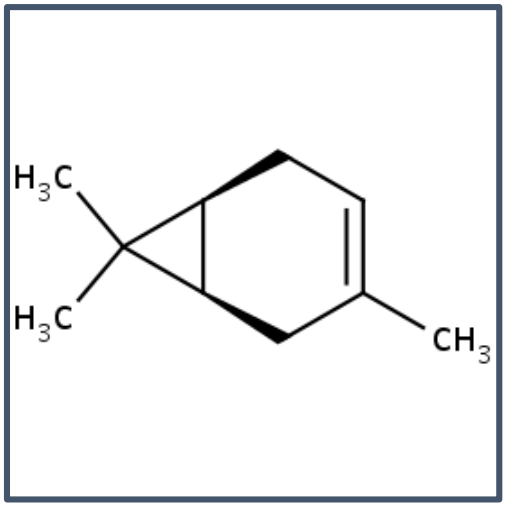
2QJ



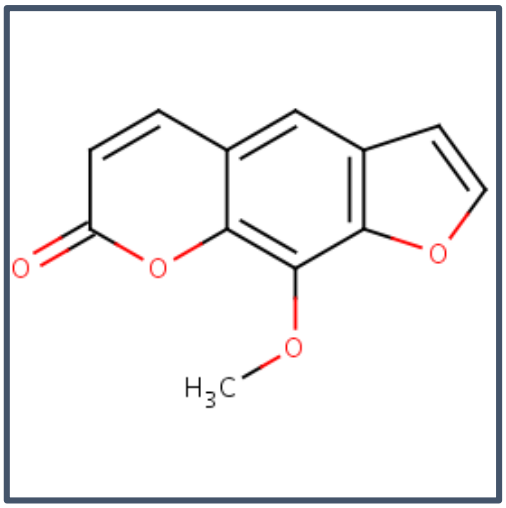
3QO



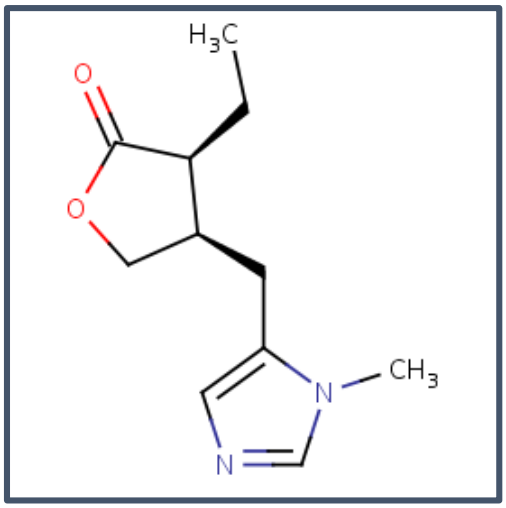
3QU



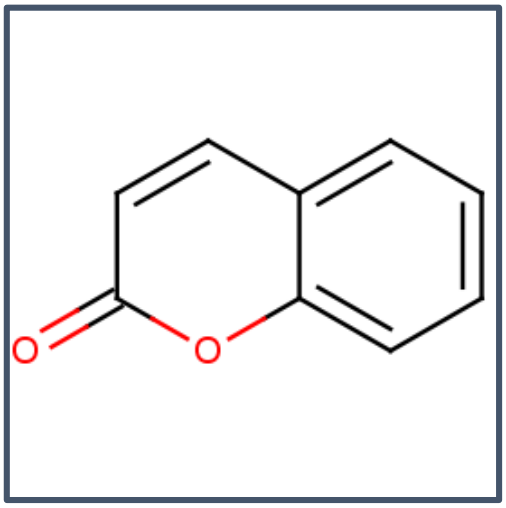
3V4



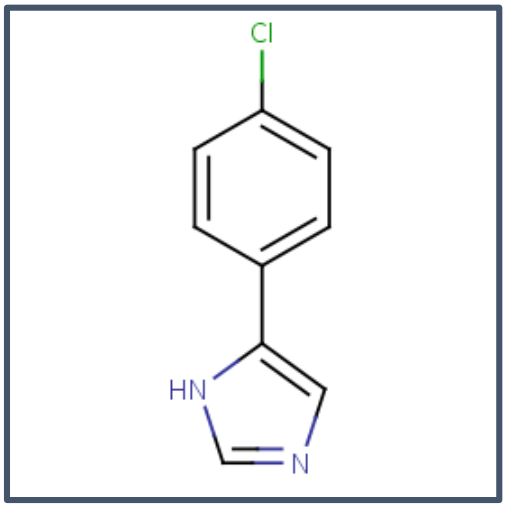
8MO



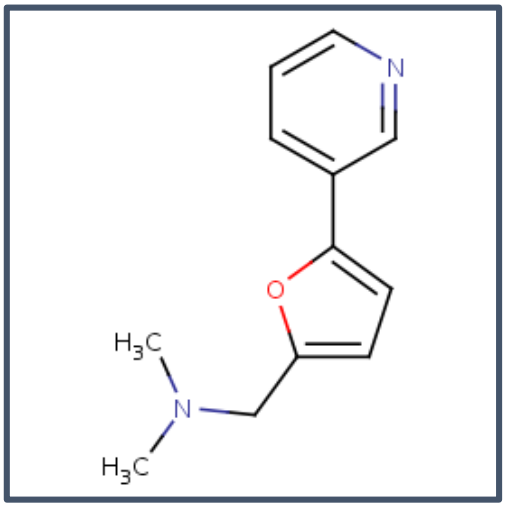
9PL



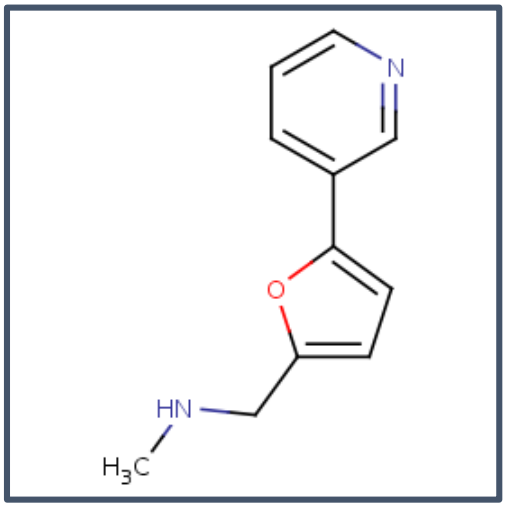
COU



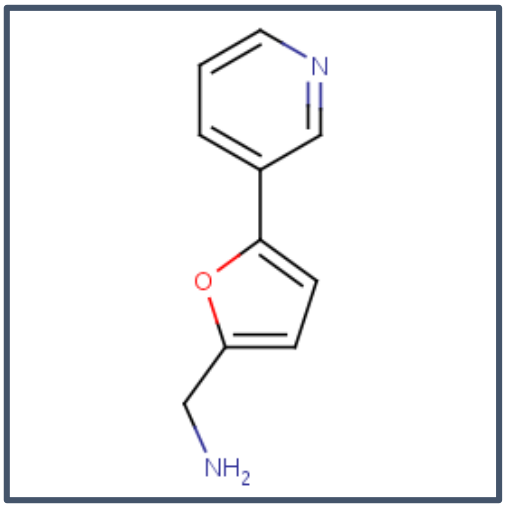
CPZ



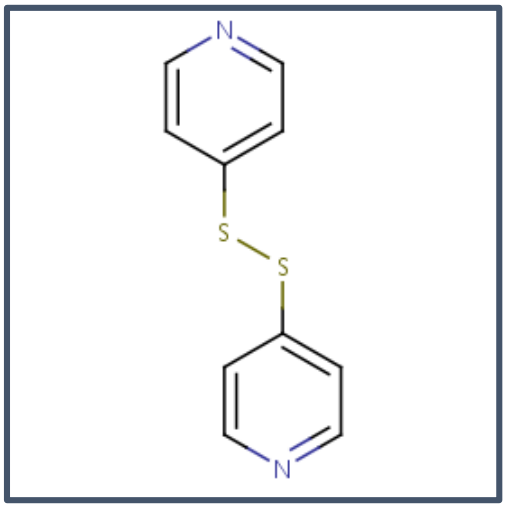
D1G or Nic2c



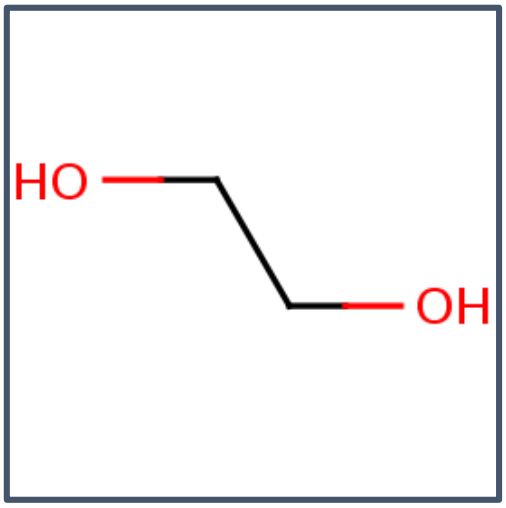
D2G or Nic2b



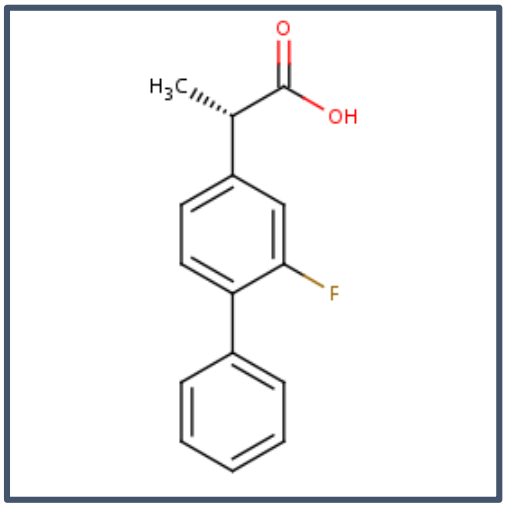
D3G or Nic2a



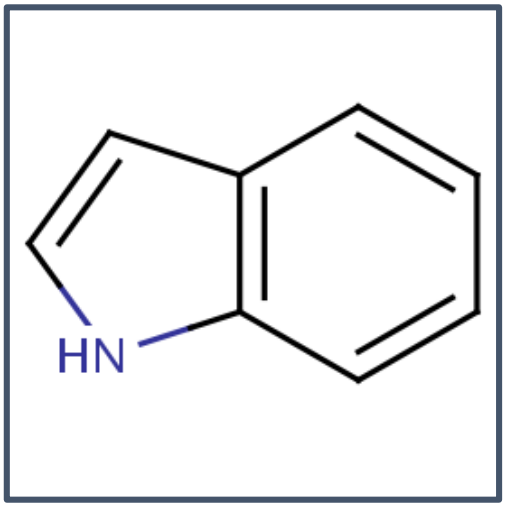
D4G



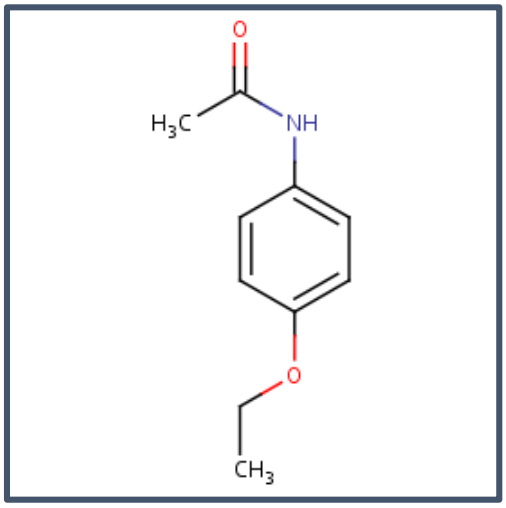
EDO



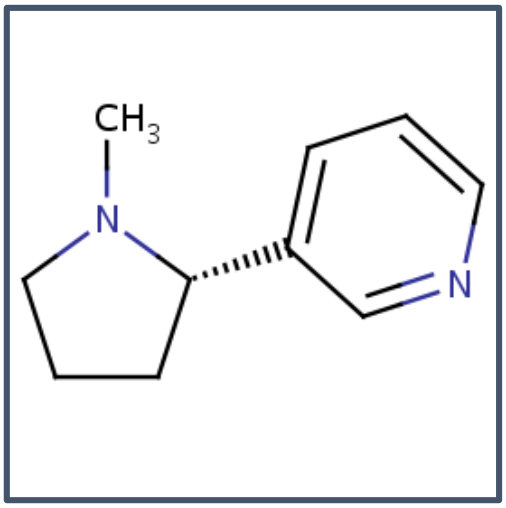
FLP



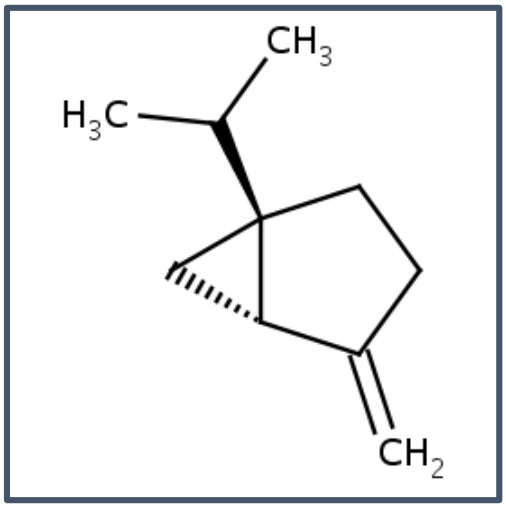
IND



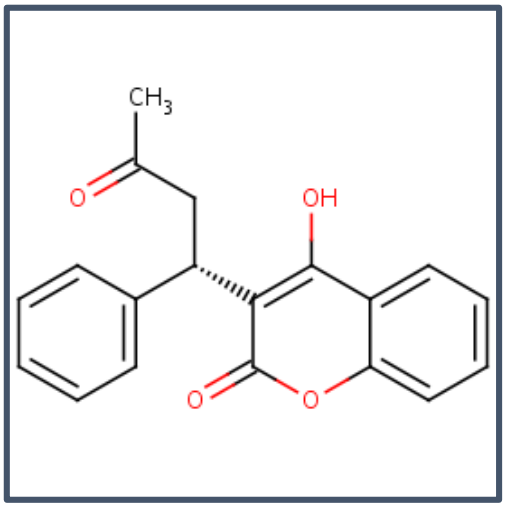
N4E



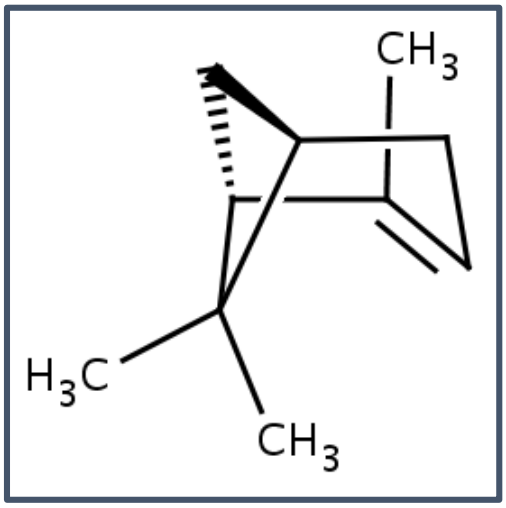
NCT



SNE



SWF



TMH

7.3 ITCs Ki values found in Literature

Type of ITC	Inhibition Mechanism	CYP Isoform	*Ki* (μM)	References
BITC	Non-Competitive Inhibition	2A6	4.1	Schlicht, K. E.; 2007
PEITC	Competitive Inhibiton	2A6	18.2 ± 2.5 [COU]	Nakajima, M. et al.; 2001
PEITC	Partial Non-Competitive Inhibition	2A6	0.37	Schlicht, K. E.; 2007
PHITC	Non-Competitive Inhibition	2A6	19.9	Schlicht, K. E.; 2007 Weymarn, L. B. von et al.; 2007
PPITC	Non-Competitive Inhibition	2A6	2.6	Schlicht, K. E.; 2007 Weymarn, L. B. von et al.; 2007
BITC	Non-Competitive Inhibition	2A13	1.26	Schlicht, K. E.; 2007
PEITC	Uncompetitive Inhibition	2A13	0.03	Schlicht, K. E.; 2007
PHITC	Non-Competitive Inhibition	2A13	1.1	Schlicht, K. E.; 2007 Weymarn, L. B. von et al.; 2007
PPITC	Non-Competitive Inhibition	2A13	0.1	Schlicht, K. E.; 2007
PPITC	Non-Competitive Inhibition	2A13	0.14	Weymarn, L. B. von et al.; 2007
PEITC	Non-competitive Inhibition	2B6	1.5±0.0	Nakajima, M. et al.; 2001
PEITC	Non-competitive Inhibition	2C9	6.5±0.9	Nakajima, M. et al.; 2001

All of these different CYPs´ Ki were determined using human CYP enzymes

7.4 Endogenous Ligands Constants found in Literature

Ligand	Ligand Type	CYP Isoform	*Ki* (a) or *Kd* (b) or *Ks* (c) (µM)	References
0QA or NNK	Substrate	2A6	(b) 1.4	DeVore, N. M.; Scott, E. E.; 2012
8MO	Non-competitive inhibitor	2A6	(a),(b) 1.9	Yano, J. K. et al.; 2005 Lu, H. et al.; 2014 Draper, A. J. et al.; 1997
8MO	Non-competitive inhibitor	2A6	(b) 1.3	DeVore, N. M. et al.; 2009 Draper, A. J. et al.; 1997
9PL	Mixed inhibitor	2A6	(a) 101 ± 17 [COU] ; 3.0 ± 0.5 [*p*-Nitrophenol]	DeVore, N. M. et al.; 2012
9PL	Competitive inhibitor	2A6* I208S/I300F/G301A/S369G	(a) 49 ± 3 [COU]	DeVore, N. M. et al.; 2012
9PL	Competitive inhibitor	2A6	(b) 3.6	DeVore, N. M. et al.; 2012
9PL	Competitive inhibitor	2A6*I208S/I300F/G301A/S369G	(b) 1.5	DeVore, N. M. et al.; 2012
COU	Substrate	2A6	(b) 3.1	DeVore, N. M. et al.; 2009
COU	Substrate	2A6*I208S/I300F/G301A	(b) 6.0	DeVore, N. M. et al.; 2009
COU	Substrate	2A6*I300F/G301A/S369G	(b) 3.1	DeVore, N. M. et al.; 2009
COU	Substrate	2A6*I308S/I300F/G301A/S369G	(b) 2.3	DeVore, N. M. et al.; 2009
D1G or Nic2c	type II spectra	2A6	(a) 14.2 ± 9	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014

D2G or Nic2b	type II spectra	2A6	(a) 0.8 ± 0.2	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014
D3G or Nic2a	type II spectra	2A6	(a) 0.1 ± 0	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014
D4G	Competitive inhibitor	2A6	(a) 0.06 [COU]	Fujita, K.; Kamataki, T.; 2001
EDO	Substrate	2A6	(b) 81.2 ± 13.7	Sansen, S. et al.; 2007
EDO	Substrate	2A6*N297Q	(b) 14.4 ± 1.1	Sansen, S. et al.; 2007
EDO	Substrate	2A6*L240C/N297Q	(b) 22.8 ± 3.6	Sansen, S. et al.; 2007
EDO	Substrate	2A6*N297Q/I300V	(b) 17.7 ± 1.8	Sansen, S. et al.; 2007
N4E	Substrate	2A6*I300F/G301A	(b) 103	DeVore, N. M. et al.; 2008
N4E	Substrate	2A6*I208S/I300G/G301A	(b) 63	DeVore, N. M. et al.; 2008
N4E	Substrate	2A6*I300F/G301A/G369S	(b) 13	DeVore, N. M. et al.; 2008
N4E	Substrate	2A6*I208S/I300F/G301A/G369S	(b) 21	DeVore, N. M. et al.; 2008
NCT	Substrate	2A6	(b) 103.0	DeVore, N. M. et al.; 2009
SNE	Substrate	2A6	(b) 4.20	Shah, M. B. et al.; 2015
TMH	Type I spectra Inhibitor	2A6	(b) 0.17	Wilderman, P. R. et al.; 2013
0QA or NNK	Substrate	2A13	(b) 4.4	DeVore, N. M.; Scott, E. E.; 2012
8MO	Non-competitive inhibitor	2A13	(a) 0.11 ± 0.03 [COU]	Weymarn, L. B. von et al.; 2005
8MO	Non-competitive inhibitor	2A13	(b) < 0.10	DeVore, N. M. et al.; 2009

				Weymarn, L. B. von et al.; 2005
9PL	Competitive inhibitor	2A13	(b) 3 ± 1	DeVore, N. M. et al.; 2012
9PL	Competitive inhibitor	2A13	(a) 48 ± 4 [COU]; 1.4 ± 0.1 [*p*-Nitrophenol]	DeVore, N. M. et al.; 2012
COU	Substrate	2A13	(b) 2.7	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*L110V	(b) 1.7	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*A117V	(b) 1.8	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*S208I	(b) 7.9	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*A213S	(b) 3.4	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*F300I	(b) 14.0	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*A301G	(b) 4.2	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*M365V	(b) 5.0	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*L366I	(b) 0.57	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*G369S	(b) 6.0	DeVore, N. M. et al.; 2009
COU	Substrate	2A13*H372R	(b) 1.5	DeVore, N. M. et al.; 2009
IND		2A13		
N4E	Substrate	2A13	(b) 34	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*L110V	(b) 18	DeVore, N. M. et al.; 2008

N4E	Substrate	2A13*A117V	(b) 20	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*S208I	(b) 108	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*A213S	(b) 65	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*M365V	(b) 135	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*L366I	(b) 2.2	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*G369S	(b) 97	DeVore, N. M. et al.; 2008
N4E	Substrate	2A13*H372R	(b) 20	DeVore, N. M. et al.; 2008
NCT	Substrate	2A13	(b) 22.0	DeVore, N. M. et al.; 2009
06X	Competitive inhibitor	2B6	(a) 1.95 ± 0.45 [BR]	Katoh, M. et al.; 2000
06X	Non-competitive inhibitor	2B6	(a) 0.68 ± 0.08 [BR]	Katoh, M. et al.; 2000
3QO	Competitive inhibitor	2B6* Y226H/K262R	(b) 0.21 ± 0.09 [7-EFC]	Shah, M. B. et al.; 2011 Korhonen, L. E. et al.; 2007
3QU	Competitive inhibitor	2B6* Y226H/K262R	(b) 0.16 ± 0.07 [7-EFC]	Shah, M. B. et al.; 2011 Korhonen, L. E. et al.; 2007
3V4	Inhibitor	2B6*Y226H, K262R	(b) 1.82	Shah, M. B. et al.; 2015 Wilderman, P. R. et al.; 2013
CPZ	Inhibitor	2B6*Y226H/K262R	(c) 0.19	Gay, S. C. et al.; 2010

D2G or Nic2b	type II spectra	2B6	(a) 179 ± 11	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014
D3G or Nic2a	type II spectra	2B6	(a) 95.5 ± 12	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014
SNE	Substrate	2B6* Y226H, K262R	(b) 0.08	Shah, M. B. et al.; 2015
TMH	Type I spectra Inhibitor	2B6* Y226H, K262R	(b) 0.22	Wilderman, P. R. et al.; 2013
TMH	Type I spectra Inhibitor	2B6* Y226H, K262R	(c) 0.38 ± 0.23	Wilderman, P. R. et al.; 2013
2QJ	type II spectra	Truncated (UNP residues 30-490) 2C9* K206E, I215V, C216Y, S220P, P221A, I222L, I223	(b) 2.0	Brändén, G. et al.; 2014 Blanc, J. et al.; 2013
D1G or Nic2c	type II spectra	2C9	(a) 111 ± 8	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014 Blanc, J. et al.; 2013
D2G or Nic2b	type II spectra	2C9	(a) 38.1 ± 5	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014 Blanc, J. et al.; 2013
D3G or Nic2a	type II spectra	2C9	(a) 5.9 ± 1.1	Yano, J. K. et al.; 2006 Lu, H. et al.; 2014 Blanc, J. et al.; 2013
FLP	Substrate	2C9*I490V	(b) 9.6	Wester, M. R. et al.; 2004
SWF	Substrate	2C9	(a) 1.5-4.5	(*)

(*) http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm093664.htm, accessed on 20 April 2015

7-benzyloxyresorufin (BR): CYP2B6 substrate – Katoh, M. et al.; 2000

7-ethoxy-4-trifluoromethylcoumarin (7-EFC): CYP2B6 substrate – Shah, M. B. et al.; 2011

All of these different CYPs´ *K*i, *K*d and *K*s were determined using human CYP enzymes.