UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di fisica

XXVI CICLO DEL

DOTTORATO DI RICERCA IN

NANOTECNOLOGIE

IN SILICO PREDICTION OF DRUG RESISTANCE: FROM CANCER TARGETED THERAPY TO CANCER TARGETED PREVENTION

Settore scientifico-disciplinare: ING-IND/24

Dottoranda
Valentina Dal Col

Coordinatore:
Chiar.ma Prof. Lucia PASQUATO

Supervisore:
Chiar.ma Prof. Sabrina PRICL

Co-supervisore:
Chiar.mo Prof. Maurizio FERMEGLIA

Anno Accademico 2012-2013

If we knew what it was we were doing,

It would not be called research, would it?

(A. Einstein)

Abstract (English)

One great innovation brought about the knowledge of the genetic processes leading to neoplastic transformation is the possibility to identify cancer-promoting molecular lesions that can be selectively targeted with rationally designed therapeutic agents.

Targeted therapy represents a revolutionary road to overcome the aspecificity of oncogenic treatment and to strike the biomolecular cause of neoplastic pathological processes.

Advances in computing are vital to cancer research. They are opening up opportunities to discover new drugs from the molecular basis of disease; this is proving to be a cost-effective method of identifying compounds that block the action of many proteins suggested by biology as important factors in different kind of cancer.

In this work we will use a wide set of molecular simulation techniques to study different problem facing during the application of chemotherapy.

The understanding of new details on mechanisms leading to cancer allowed the identification of new and more precise target protein to tackle the process without collateral damage to the neighboring healthy cells. In particular, here are reported some details about three different new targets: the σ_1 receptors, β-catenin and dendrimer development for gene silencing therapy.

However, despite the clinical success of targeted therapies, most patients that respond eventually relapse as a result of acquired resistance; the development of resistance can occur after a variable period of chronic treatment probably involving mutation that leads to a conformational rearrangement of the protein involved.

In silico molecular modeling can be used to predict the occurrence of all activating but drug-resistant secondary mutations and to develop a multi-drug targeted prevention strategy.

In this context, it will be discussed in particular mutations involved in c-Kit and Smoothened receptor, pre and after treatment.

The extensive series of examples illustrated and discussed here emphasized the role and potentiality of molecular modeling in the development of cancer targeted therapy. In silico evaluation can take into account the molecular specificity of the problem and dramatically reduced the time and cost required to formulate new compounds and therapeutic intervention.

Abstract (Italiano)

Le recenti scoperte sulla natura dei processi genetici, implicati nelle trasformazioni neoplastiche, ci hanno permesso di identificare le principali lesioni cellulari che portano alla formazione del tumore, individuando in questo modo target selettivi sulla base dei quali progettare nuovi agenti terapeutici efficaci.

La terapia *bersaglio* ha rappresentato una rivoluzione permettendo di superare l'aspecificità del trattamento antitumorale e colpendo, nello specifico, la causa biomolecolare del processo patologico neoplastico.

I recenti sviluppi della modellistica molecolare si sono resi indispensabili per la ricerca oncologica.

Questi, infatti, aprono delle opportunità per la scoperta di nuovi farmaci attraverso la comprensione delle basi molecolari delle malattie; si tratta di un metodo conveniente per identificare dei composti che siano in grado di bloccare l'azione di molte proteine, suggerite come fattori fondamentali in diversi tipi di neoplasie.

In questo lavoro di tesi verrà adoperata un'ampia gamma di tecniche di simulazione molecolare al fine di studiare differenti problemi che possono comparire durante il trattamento chemioterapico. La scoperta di nuovi dettagli nei meccanismi che conducono alla carcinogenesi ha permesso di identificare proteine target innovative e più precise per affrontare la terapia senza effetti collaterali per le cellule sane. In particolare, saranno riportati in dettaglio gli studi che coinvolgono tre nuovi differenti target: il recettore σ_1, la proteina β-catenina e lo sviluppo di dendrimeri per una terapia di silenziamento genico.

Tuttavia, nonostante i successi clinici della terapia bersaglio, molti pazienti che inizialmente hanno risposto positivamente manifestano una recidiva, come risultato di una resistenza acquisita; questo tipo di risposta può manifestarsi dopo un periodo variabile di trattamento cronico.

L'approccio delle tecniche *in silico* può essere utilizzato per predire l'avvento di mutazioni secondarie attivanti e al contempo resistenti alla terapia e quindi aiutare nello sviluppo di una strategia di prevenzione focalizzata su una terapia multi-farmaco. In questo lavoro, verranno discusse in particolare delle mutazioni che coinvolgono i recettori c-Kit e Smoothened, pre e post trattamento.

L'ampia serie di esempi illustrati e discussi qui enfatizza il ruolo e le potenzialità del modeling molecolare nello sviluppo della terapia bersaglio anti tumorale. Una valutazione *in silico* permette di prendere in considerazione la specificità molecolare del problema e ridurre drasticamente i tempi e i costi richiesti per formulare nuovi farmaci e strategie di intervento.

Index

Introduction

In the year 2013, there were estimated 1,5 million new cases of cancer and 500,000 cancer deaths worldwide. Some good news, however, has emerged: cancer mortality for both men and woman in the USA declined during the last decade of the 20th century. Thus, there has been progress, but the problem is still overwhelming. There are now dramatic improvements in therapeutic responses and 5-year survival rates with many forms of malignancy, notably the leukemias and lymphomas. A greater proportion of cancers is being cured or arrested today than ever before.[1]

Estimated Number* of New Cancer Cases and Deaths by Sex, US, 2013

	Estimated New Cases			Estimated Deaths		
	Both Sexes	Male	Female	Both Sexes	Male	Female
All Sites	1,660,290	854,790	805,500	580,350	306,920	273,430
Oral cavity & pharynx	41,380	29,620	11,760	7,890	5,500	2,390
Tongue	13,590	9,900	3,690	2,070	1,380	690
Mouth	11,400	6,730	4,670	1,800	1,080	720
Pharynx	13,930	11,200	2,730	2,400	1,790	610
Other oral cavity	2,460	1,790	670	1,640	1,260	380
Digestive system	290,200	160,750	129,450	144,570	82,700	61,870
Esophagus	17,990	14,440	3,550	15,210	12,220	2,990
Stomach	21,600	13,230	8,370	10,990	6,740	4,250
Small intestine	8,810	4,670	4,140	1,170	610	560
Colon†	102,480	50,090	52,390	50,830	26,300	24,530
Rectum	40,340	23,590	16,750			
Anus, anal canal, & anorectum	7,060	2,630	4,430	880	330	550
Liver & intrahepatic bile duct	30,640	22,720	7,920	21,670	14,890	6,780
Gallbladder & other biliary	10,310	4,740	5,570	3,230	1,260	1,970
Pancreas	45,220	22,740	22,480	38,460	19,480	18,980
Other digestive organs	5,750	1,900	3,850	2,130	870	1,260
Respiratory system	246,210	131,760	114,450	163,890	90,600	73,290
Larynx	12,260	9,680	2,580	3,630	2,860	770
Lung & bronchus	228,190	118,080	110,110	159,480	87,260	72,220
Other respiratory organs	5,760	4,000	1,760	780	480	300
Bones & joints	3,010	1,680	1,330	1,440	810	630
Soft tissue (including heart)	11,410	6,290	5,120	4,390	2,500	1,890
Skin (excluding basal & squamous)	82,770	48,660	34,110	12,650	8,560	4,090
Melanoma-skin	76,690	45,060	31,630	9,480	6,280	3,200
Other nonepithelial skin	6,080	3,600	2,480	3,170	2,280	890
Breast	234,580	2,240	232,340	40,030	410	39,620
Genital system	339,810	248,080	91,730	58,480	30,400	28,080
Uterine cervix	12,340		12,340	4,030		4,030
Uterine corpus	49,560		49,560	8,190		8,190
Ovary	22,240		22,240	14,030		14,030
Vulva	4,700		4,700	990		990
Vagina & other genital, female	2,890		2,890	840		840
Prostate	238,590	238,590		29,720	29,720	
Testis	7,920	7,920		370	370	
Penis & other genital, male	1,570	1,570		310	310	
Urinary system	140,430	96,800	43,630	29,790	20,120	9,670
Urinary bladder	72,570	54,610	17,960	15,210	10,820	4,390
Kidney & renal pelvis	65,150	40,430	24,720	13,680	8,780	4,900
Ureter & other urinary organs	2,710	1,760	950	900	520	380
Eye & orbit	2,800	1,490	1,310	320	120	200
Brain & other nervous system	23,130	12,770	10,360	14,080	7,930	6,150
Endocrine system	62,710	16,210	46,500	2,770	1,270	1,500
Thyroid	60,220	14,910	45,310	1,850	810	1,040
Other endocrine	2,490	1,300	1,190	920	460	460
Lymphoma	79,030	42,670	36,360	20,200	11,250	8,950
Hodgkin lymphoma	9,290	5,070	4,220	1,180	660	520
Non-Hodgkin lymphoma	69,740	37,600	32,140	19,020	10,590	8,430
Myeloma	22,350	12,440	9,910	10,710	6,070	4,640
Leukemia	48,610	27,880	20,730	23,720	13,660	10,060
Acute lymphocytic leukemia	6,070	3,350	2,720	1,430	820	610
Chronic lymphocytic leukemia	15,680	9,720	5,960	4,580	2,750	1,830
Acute myeloid leukemia	14,590	7,820	6,770	10,370	5,930	4,440
Chronic myeloid leukemia	5,920	3,420	2,500	610	340	270
Other leukemia‡	6,350	3,570	2,780	6,730	3,820	2,910
Other & unspecified primary sites‡	31,860	15,450	16,410	45,420	25,020	20,400

*Rounded to the nearest 10; estimated new cases exclude basal cell and squamous cell skin cancers and in situ carcinomas except urinary bladder. About 64,640 carcinoma in situ of the female breast and 61,300 melanoma in situ will be newly diagnosed in 2013. †Estimated deaths for colon and rectal cancers are combined. ‡More deaths than cases may reflect lack of specificity in recording underlying cause of death on death certificates and/or an undercount in the case estimate.

Source: Estimated new cases are based on cancer incidence rates from 49 states and the District of Columbia during 1995-2009 as reported by the North American Association of Central Cancer Registries (NAACCR), represesnting about 98% of the US population. Estimated deaths are based on US mortality data during 1995-2009, National Center for Health Statistics, Centers for Disease Control and Prevention.

Figure1.1 *Estimated number of new cancer cases and deaths by sex, US, 2013.*[1]

1.1 Molecular basis of disease

Cancer is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. If the spread is not controlled, it can result in death. Causes are both external factors (tobacco, infectious organism, chemicals and radiation) and internal factors (inherited mutations, hormones, immune conditions and mutations that occur from metabolism). These causal factors may act together or in sequence to initiate or promote the development of cancer. Ten or more years often pass between exposure to external factors and detectable cancer. Cancer is treated with surgery, radiation, chemotherapy, hormone therapy, biological therapy and targeted therapy.

Nonlethal genetic damage lies at the hearth of carcinogenesis. Such genetic damage may be acquired by the action of environmental agents, such as chemicals, radiation, or viruses, or it may be inherited in the germ line. The term “environmental”, used in this context, involves any acquired defect caused by exogenous agents or endogenous products of cell metabolism. Not all mutations, however, are “environmentally” induced. Some may be spontaneous and stochastic.[2] A tumor is formed by the clonal expansion of a single precursor cell that has incurred the genetic damage. The most commonly used method to determine tumor clonality involves the analysis of methylation patterns adjacent to the highly polymorphic locus of the human androgen receptor gene (HUMARA). The frequency of HUMARA polymorphism in the general population is more than 90%, so it is easy to establish clonality by showing that all the cells in a tumor express the same allele. For tumors with a specific translocation, such as in myeloid leukemias, the presence of the translocation can be used to assess clonality.

Four classes of normal regulatory genes (the growth-promoting proto-oncogenes, the growth-inhibiting tumor suppressor genes that regulate programmed cell death/apoptosis and genes involved in DNA repair) are the principal targets of genetic damage. Mutant alleles of proto-oncogenes are considered dominant because they transform cells despite the presence of a normal counterpart. In contrast, both normal alleles of the tumor suppressor gene must be damaged for transformation to occur, so this family of genes is sometimes referred to as recessives oncogenes. However, there are exceptions to this rule, and some tumor suppressor genes lose their suppressor activity when a single allele is lost or inactivated. This loss of functions of a recessive gene caused by damage of a single allele is called haploinsufficiency. Genes that regulate apoptosis may be dominant, as are proto-oncogenes, or they may behave as tumor suppressor genes.

DNA repair genes affect cell proliferation or survival indirectly by influencing the ability of the organism to repair nonlethal damage in other genes, including proto-oncogenes, tumor suppressor genes, and genes that regulate apoptosis. A disability in the DNA repair genes can

predispose to mutations in the genome and hence to neoplastic transformation. Such propensity to mutations is called a mutator phenotype. With some exceptions, both alleles of DNA repair genes must be inactivated to induce such genomic instability, in this sense, DNA repair genes may also be considered as tumor suppressor genes.

Carcinogenesis is a multistep process at both the phenotypic and the genetic levels. A malignant neoplasm has several phenotypic attributes, such as excessive growth, local invasiveness and the ability to form distant metastases. These characteristics are acquired in a stepwise fashion, a phenomenon called tumor progression. At the molecular level, progression results from accumulation of genetic lesions that in some instances are favored by defects in DNA repair.

1.2 The long road of cancer chemotherapy

The use of chemotherapy to treat cancer began at the start of the 20th century with attempts to narrow the universe of chemicals that might affect the disease by developing methods to screen chemicals using transplantable tumors in rodents.

Surgery and radiotherapy dominated the field of cancer therapy into the 1960s until it became clear that cure rates after ever more radical local treatments had plateaued at about 33% due to the presence of micrometastases and new data showed that combination chemotherapy could cure patients with various advanced cancers. The latter observation opened up the opportunity to apply drugs in conjunction with surgery and/or radiation treatments to deal with the issue of micrometastases, initially in breast cancer patients, and the field of adjuvant chemotherapy was born. Combined modality treatment, the tailoring of each of the three modalities so their antitumor effect could be maximized with minimal toxicity to normal tissues, then became standard clinical practice.[3] In 1990s, something else has happened to change the landscape of drug development. Information about the molecular aberrations that occur in cancer cells has become available, random screening is being replaced by screening against specific critical molecular targets. As the market for cancer drugs has grown, so has the willingness of the industry to invest in new drugs, and discovery and development are now largely in the hands of a segment of the pharmaceutical industry that did not exist before the advent of the CCNSC (Cancer Chemotherapy National Service Center). As a consequence, many new drugs and new classes of anticancer drugs have been introduced since the 1980s and are now available to clinicians.[4]

The advent of monoclonal antibodies has enhanced the effects of chemotherapy. Hybridomas were described in 1975, and monoclonal antibodies were proven clinically useful starting in the mid-1990s. Although they are not chemotherapy per se, they seem to work best when

they are used in conjunction with chemotherapy, as is the case for trastuzumab in breast cancer, cetuximab and bevacizumab in colorectal cancer, and rituximab in non–Hodgkin's lymphoma, and each are an integral part of chemotherapy regimens for these common tumors.

Chemotherapy has, in fact, transitioned to the age of "targeted therapy."

1.3 Targeted therapy

One great innovation brought about by the knowledge of the genetic processes leading to neoplastic transformation is the possibility to identify cancer-promoting molecular lesions that can be selectively targeted with rationally designed therapeutic agents.

For decades, the hallmark of medical treatment for cancer has been intravenous cytotoxic chemotherapy. These drugs target rapidly dividing cells, including cancer cells and certain normal tissue. Targeted therapies have expanded the concept of individually tailored cancer treatment because some of these drugs may be effective in patients whose cancers have a specific molecular target, but they may not be effective in the absence of such a target. This distinction may be influenced by patient ethnicity and sex, as well as by tumor histology. [5]

Traditional cytotoxic chemotherapy works primarily through the inhibition of cell division. In addition to cancer cells, other rapidly dividing cells (e.g. hair, gastrointestinal epithelium, bone marrow) are affected by these drugs. In contrast, targeted therapy blocks the proliferation of cancer cells by interfering with specific molecules required for tumor development and growth. Some of these molecules may be present in normal tissues, but they are often mutated or overexpressed in tumors.[6]

For example, most cases of chronic myeloid leukemia (CML) are caused by the formation of a gene called BCR-ABL. This gene is formed when pieces of chromosome 9 and chromosome 22 break off and trade places.

One of the changed chromosomes resulting from this switch contains part of the ABL gene from chromosome 9 fused to part of the BCR gene from chromosome 22. The protein normally produced by the ABL gene (Abl) is a signaling molecule that plays an important role in controlling cell proliferation and usually must interact with other signaling molecules to be active. However, Abl signaling is always active in the protein (Bcr-Abl) produced by the BCR-ABL fusion gene. This activity promotes the continuous proliferation of CML cells.[7]

Therefore, Bcr-Abl represents a good molecule to target.[1]

The molecular pathways most often targeted in the treatment of solid tumors are those of the epidermal growth factor receptor (EGFR)[8], vascular endothelial growth factor (VEGF)[9] and Her2. Such pathways can be inhibited at multiple levels: by binding and neutralizing ligands; by

occupying receptor binding site; by blocking receptor signaling within the cancer cell; or by interfering with downstream intracellular molecules.

The best models for this approach are proteins with kinase activity which become one of the most intensively pursued classes of druggable molecules, with approximately 30 distinct kinase-inhibiting agents undergoing clinical trials. The choice of protein kinases as drug targets is motivated by the convergence of several factors: i) every signal transduction process is regulated by kinases; ii) kinases are frequently activated by mutation/gene amplification in a variety of human cancers; iii) despite the high degree of conservation in the ATP binding site, highly selective small molecules with favorable pharmaceutical properties can be developed; iv) inhibition of kinase activity in normal cells is often tolerated. Efficient approaches, benefiting immensely from crystallography and 3D modeling, allowed the development of potent and selective inhibitors. To date the most successful clinical use of a kinase inhibitor is represented by imatinib, a drug designed specifically to target a molecular defect associated with cancer (Table 1.1). By binding the catalytic domain imatinib inactivates the BCR-ABL1 oncogene of chronic myeloid leukaemia (CML), revolutioning the management of the disease[10,11]. The discovery of imatinib as a multikinase inhibitor of c-Abl, c-KIT and PDGFRα allowed the rapid extension of the success of imatinib to gastrointestinal stromal tumours (GIST), thus providing the first example of successful kinase-targeted therapy in solid tumors.

Despite the impressive clinical success of kinase-targeted therapies, most treatment-responsive patients eventually experience relapse as a result of acquired resistance.

Among the resistance mechanisms, a Darwinian type selection appears to play a major role. In fact, under the pressure exerted by the necessary chronic treatment, there are appearance and expansion of tumour clones expressing a new mutation in the targeted kinase that is no longer drug-sensitive. It is currently debated whether the resistant mutations are pre-existing or acquired following inhibitor treatment. Both mechanisms have been recently demonstrated in CML with acquired resistance to imatinib.

The use of targeted has markedly changed outcomes for some diseases. Imatinib has had a dramatic effect on chronic myeloid leukemia, and rituximab, sunitinib and trastuzumab have revolutionized the treatment of non-Hodgkin's lymphoma, renal cell carcinoma, and breast cancer, respectively.[12-14] In addition to prolonging survival in patients with certain cancers, targeted therapies provide treatment options for some patients who may not otherwise be candidates for anticancer therapy.

Many targeted cancer therapies have been approved by the U.S. Food and Drug Administration (FDA) for the treatment of specific types of cancer. Others are being studied in clinical trials, and many more are in preclinical testing.

Drug	Target	FDA-approved indications
Imatinib (Gleevec)	BCR-ABL, c-KIT, PDGFR	Acute lymphocytic leukemia, CML, GIST, hypereosinophilic syndrome, systemic mastocytosis
Dasatinib (Sprycel)	BCR-ABL, SRC family, c-KIT, PDGFR	CML, acute lymphocytic leukemia
Erloyinib (Tarceva)	EGFR	Non-small cell lung cancer, pancreatic cancer
Gefitinib (Iressa)	EGFR	Non-small cell lung cancer
Lapatinib (Tykerb)	HER2, EGFR	Breast cancer with HER2 overexpression
Sorafenib (Nexavar)	BRAF, VEGFR, EGFR, PDGFR	Renal cell cancer, hepatocellular carcinoma
Sunitinib (Sutent)	VEGFR, PDGFR, c-KIT, FLT3	Renal cell cancer, GIST
Bosutinib (Bosulif)	BCR-ABL	CML
Vandetanib (Caprelsa)	EGFR, VEGFR	Medullary thyroid cancer

Table 1.1 *Small molecule inhibitors for cancer treatment.*

1.4 Problems during cancer therapy

Chemotherapy is one of the principal modes of treatment for cancer, but the effectiveness of chemotherapy is limited by drug resistance. Resistance to chemotherapeutics can be divided into two broad categories: intrinsic or acquired. Intrinsic resistance indicates that before receiving chemotherapy, resistance-mediating factors preexist in the bulk of tumor cells that make the therapy ineffective. Acquired drug resistance can develop during treatment of tumors that were initially sensitive and can be caused by mutations arising during treatment,[15] as well as through various other adaptive responses, such as increased expression of the therapeutic target and activation of alternative compensatory signaling pathways. Moreover, it

is increasingly recognized that tumors can contain a high degree of molecular heterogeneity, thus drug resistance can arise through therapy-induced selection of a resistant minor subpopulation of cells that was present in the original tumor.[16] As our understanding of the molecular biology of cancer has advanced, drug development has shifted towards agents that target specific molecular alterations in tumors. These "molecularly targeted therapies" have had varying degrees of success because a diverse range of resistance mechanisms have limited patients responses. Importantly, the mechanism of resistance to cytotoxic chemotherapies and targeted drugs largely overlap, thus knowledge gained from earlier research into the mechanism of resistance to cytotoxic drugs can be applied to help anticipate and elucidate mechanisms of resistance to emerging molecularly targeted agents.

Mechanisms of resistance to anticancer agents may include increased drug efflux, modifications within the target protein, activation of downstream or redundant pathways, and induction of cell survival pathways. In cancers driven by kinase oncogenes, resistance mechanisms frequently engage the underlying tumor dependency elaborated by these oncogenes. As a general rule, oncogene-addicted tumors tend to develop resistance to oncogenic pathway inhibition by reactivating that pathway rather than by engaging completely new oncogenic pathways, although there can be exceptions to this rule. In principle, then, any mechanism that reactivates an oncogenic pathway in the setting of therapeutic inhibition has the potential to engender acquired resistance.[17]

Cancers are often highly dependent on specific oncogenic mutations that occur in kinases; however, the outgrowth of cells with gatekeeper residue mutations is a common mechanism of resistance to agents that target these oncogenic kinases.

This type of target –associated resistance has also been observed, as said before, for imatinib, which is a highly effective inhibitor of the BCR-ABL oncogenic kinase that causes chronic myeloid leukaemia (CML). The first mutation identified in patients with CML that relapsed on the selected treatment was T315I, in the gatekeeper residue. This mutations hampers imatinib binding while preserving the catalytic activity that is needed for the oncogenic function of BCR-ABL1.[18] Second generation BCR-ABL1 inhibitors have been developed; nilotinib, dasatinib and bosutinib have been approved for the treatment of patients who have developed imatinib resistance, these inhibitors are active against all of the common mutations except T315 .[19,20]

Although therapeutics directed against oncogenic kinases may yield dramatic responses and improved survival in cancers driven by a dominant oncogene, such tumors invariably become resistant to these agents. Thus, elucidating the mechanism of acquired resistance (or de novo) is essential to the development of new treatment strategies that improve the clinical benefit.

In parallel, diagnostic advances that exploit increasingly powerful genomic technologies may be needed to profile individual tumors for the acquisition of specific resistance mechanisms.

New approaches to win the battle

Considering all the point explained up to here, cancer therapy is based on two different tasks: the individuation of new possible targets during the carcinogenesis processes and the study of the effects due to resistance mechanism for secondary mutations in order to predict the possible outcome of the disease exploiting, for instance, computational techniques, saving life and money.

1.5 Individuation of new more specific and accurate targets

The understanding of new details on mechanisms leading to cancer allowed the identification of new and more precise target protein to tackle in order to block the process. In particular, here are reported some details about three different new targets: the σ receptors, β-catenin in desmoid tumors therapy and dendrimers development as nanovector for gene silencing.

1.5.1 Sigma receptors as new entry in cancer therapy

σ receptors represent a class of proteins that were first discovered in 1976; initially proposed to be subtypes of opioid receptors they were subsequently confused with the PCP binding site. Finally studies revealed that σ binding sites are a distinct class of receptors that are located in the central nervous system as well as in a variety of tissues and organs.[21 22]

σ_1 receptors were first postulated based on the actions of SKF 10,047 (N-allylnormetazocine) and related benzomorphans. The name ''σ'' originated from the first letter ''S'' in SKF 10,047, which was thought to be the prototypic ligand for these binding sites. Unfortunately, SKF 10,047 is now recognized as a non-selective ligand, which contributed to the turbulent early history surrounding these enigmatic receptors. To date, two subtypes of the σ receptor have been identified, the σ_1 and σ_2 receptors, which are distinguishable by their pharmacology, function, and molecular mass.[23,24]

One distinguishing feature of the σ_1 receptor is its promiscuity in binding a wide range of different pharmacological agents although, how binding of these various compounds translates into function(s) through the σ_1 receptor is currently not clear.

Since the discovery of the σ_1 receptor, many preclinical studies have implicated the receptor in many important human diseases, from maladies of the central nervous system to cancer.

The σ_1 receptor was cloned[25]; it is a 25 kDa single polypeptide and it shares no sequence homology with any known mammalian proteins but shows approximately 30% identity with a yeast sterol C8-C7 isomerase[26]. Data about σ_2 receptor are poor: they have not yet been cloned and have an apparent molecular weight of 21.5 kDa according to photolabeling studies.[27]

The radioligand pentazocine (Figure 1.2) has a high affinity for the σ_1 receptor and a low affinity for the σ_2 receptor, whereas DTG is equipotent at both σ_1 and σ_2 receptors. Several psychoactive drugs show high to moderate affinity for σ_1 receptors such as the antipsychotic haloperidol, the psychostimulants cocaine and methamphetamine, the antidepressant drugs fluvoxamine and sertraline; furthermore phenytoin, an anticonvulsivant drug, allosterically modulates σ_1 receptors.[28] In 2009, Fontanilla et al. identified as probable σ_1 endogenous ligand N,N,dimethyltryptamine (DMT).[29]

Their localization is in the plasma membrane and in subcellular membranes, mainly in the endoplasmic reticulum, where they play a modulatory role in intracellular Ca2+ signaling.[30]

Autoradiographic studies have shown that, at the anatomical level, σ receptors are mainly distributed in different areas of the central nervous system (CNS), focused in the brain where they are concentrated in specific areas involved in memory, emotion and sensory and motor functions.[31] In the periphery σ receptors are expressed in blood vessels, genital system, gastrointestinal tract, steroid-producing organs, immune system, liver and eyes. Their large distribution in the CNS seems to explain the modulatory role in cellular, biochemical and neurotransmission systems.[32] For this reason σ_1 ligands may be involved in the treatment of diseases such as depression, anxiety, cognitive impairments, psychosis, analgesia and against the effects of drugs of abuse.[33] In addition, and also due to the modulatory role of σ_1 receptors, the combined administration of σ_1 receptor ligands and drugs with a known therapeutic effect has been shown to improve the effects of the latter, resulting in the need for lower doses to reach therapeutic concentrations.[34]

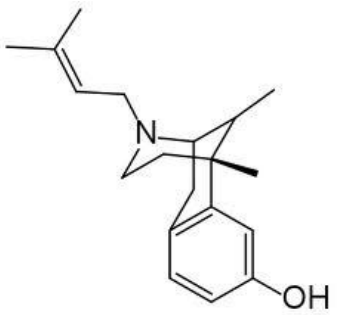


Figure 1.2 Chemical structure of *pentazocine, high selective σ_1 ligands*

σ_1 receptor structure is still unveiled. The only data known are about the secondary structure. The gene encodes a protein consisting of 223 amino acids with an intracellular N-terminal end,

two transmembrane spanning domains (10-30 and 80-100 a.a.) linked by an extracellular loop and a C-terminal end partially arranged (Figure 1.3).[35] There is also a typical arginine-arginine endoplasmic reticulum locating signal near the N terminal end. It is significant that σ_1 receptor contains two GXXXG motifs, tool which occurs with high frequency in membrane proteins that favor helix-helix interactions.[36]

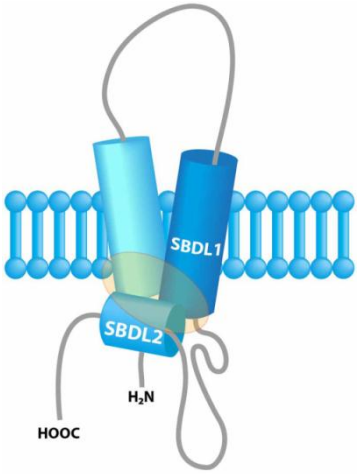


Figure 1.3 *Structure of the σ_1 receptor. The protein is constituted by two transmembrane domain and the SBDL (Steroid binding domain like-) 1, and 2.*

Involvement in cancer

Overexpression of σ receptor in cancer tissues was first shown in brain tumors, by means of homologous competition binding assays using DTG (di-o-tolylguanidine). σ receptor binding was detected in 15 of the 16 tumors analyzed and very high levels were found in a neuroblastoma and a brain metastasis from a lung adenocarcinoma. Later, elevated σ receptor expression was also observed in various cancer cell lines, including nonsmall cell lung carcinoma (NSCL), breast, lung, melanoma, leukemia, glioblastoma, neuroblastoma, and prostate tumour (Figure 1.4).[37]

Both σ_1 and σ_2 receptors exhibit a subcellular localization: plasma membrane, endoplasmic reticulum and mitochondria. These locations involve organelles that are known to play a role in both caspase-dependent and caspase-independent cell death pathways.

All these data thus suggest the conclusions that σ ligands should be evaluated as potential chemotherapeutic anticancer agents[38 39 40].

The potential usefulness of σ receptor ligands in cancer diagnosis and therapy are the overexpression in tumor cells, the ability to internalize their ligands, and the fact that several σ ligands contain an iodine or fluoride atom in their chemical structure motivated the development of σ receptor ligands for diagnostic imaging using PET[41]; different studies[42] highlighted the possibility of using σ ligands for targeting drugs to human cancer cells, in

particular a benzamide analog was incorporated into liposomes containing doxorubicin to specifically target the drug to DU-145 prostate cancer cells, which are known to overexpress σ receptors; the role of σ_1 ligands was tested for enhancing standard cancer therapy in diffusive and migrant glioblastoma.

It has been demonstrated that σ_1 receptor ligands potentiate calcium mobilization and contribute to both lipid raft formation and cytoskeletal protein modulation in normal and cancer cells. Each of the signal transduction pathways potentially involved in such actions is of interest in the context of cancer research.

They are, also, involved in sphingomyelin, glucosyl-ceramide, and cholesterol regulation [43] [44] .

It is possible to hypothesize that overexpression of σ receptors in cancer cells is useful primarily to support the high lipid production required for plasma membrane formation in highly proliferative cells. So concerning cancer therapy, it has been hypothesize that high levels of σ receptor could induce a drastic decrease of RhoGDI expression, which enables the Rho GTP ase activation required for the translocation of σ receptor/IP3R3 complexes[45] and results in the inhibition of ion flux, notably Ca2+, and a concomitant deregulation of the ceramide biosynthesis pathway. These two effects contribute to the disruption of the synthesis of major lipid membrane components, leading to mitotic arrest and/or anoikis of the cancerous cells that overexpress σ receptors. These effects might also contribute to the specific sensitization of σ_1 receptor overexpressing cancer cells to ER-stress agents.

The putative role in cancer development and the lack of σ_1 receptor's crystallographic structure have led to create a 3D model using homology modeling to find out more details about its physiology and new ligands design.

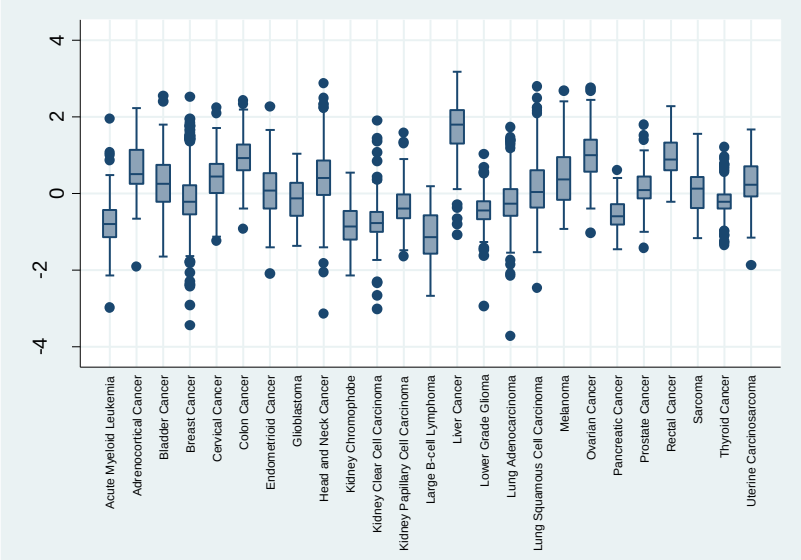


Figure 1.4 *Correlation between the expression of σ_1 receptor and their localization in normal and cancer tissues; the y axis values range from -4 to +4 where 0 is basal expression of σ receptor in human body.*

1.5.2 The role of β catenin in Desmoid tumors

Desmoid tumors are rare mesenchymal lesions with an annual incidence of five cases per million, <3% of soft tissue sarcomas and about 0.03% of all malignancies[46]. Lacking metastatic capacity, DTs are very infiltrative and possess high local recurrence risks despite adequate surgical resection with negative margins[47,48]. Most sporadic DTs (85%) harbor specific mutations in the gene encoding for β-catenin located on the short arm of chromosome 3 (*CTNNB1*)[49]; the remainder do not harbor any mutation in *CTNNB1*.

Desmoid tumors may affect all sites, including the extremities, trunk, and abdomen[50]. The age of occurrence is between 15 and 60 years, but particularly during early adolescence, and with a peak age of about 30 years. Two different types of desmoid tumors are described: sporadic and FAP-associated. The first-line therapy for patients with locally circumscribed desmoid tumors remains surgical resection. However, "wait&see" approach might be an option for a subgroup of patients. The growth pattern of these tumors is deep infiltrating, and there is no tumor capsule. Because the boundaries of the tumors are difficult to distinguish intraoperatively from scars or connective tissue, complete resection is not always possible and adjuvant radiotherapy is therefore often applied following sarcoma protocols. Desmoids, however, have a high local relapse rate after surgery and/or radiotherapy and exhibit locally aggressive growth, but can often take a multiply relapsing, multifocal course and therefore are not amenable to curative surgical treatment. In this situation, pharmacotherapy is often used to prevent disease progression[51]. The primary aim is to preserve the patient's quality of life, which is threatened by the loss of function and pain caused by proliferative disease. Therapeutic approaches to the treatment of recurrent or nonresectable desmoid tumors comprise antihormonal therapy (e.g., tamoxifen), nonsteroidal anti-inflammatory drugs (NSAIDs, e.g. cyclooxygenase [COX]-2 inhibitors), and classic chemotherapy regimens, with highly variable results[52]. Since the difficulty in standardization therapy a multidisciplinary evaluation of patients is substantial, and treatment approaches may comprise surgical intervention and/or radiotherapy and antiproliferative treatment[53].

Molecular mechanism of Wnt/β-catenin pathway

The Wnt/β-catenin pathway drives the pathogenesis of both sporadic and FAPassociated desmoid tumors. FAP-associated tumors frequently have adenomatous polyposis coli gene (APC) mutations. APC contributes to regulate the protein level of β -catenin[54].

Approximately 85% to 90% of sporadic desmoid tumors are associated with somatic mutations in the β -catenin gene.

In endothelial cells β-catenin is not only a cell-adhesion molecule, but also plays a role in nuclear transcription. Accumulation of β-catenin, by constitutive activation of the Wnt ligand pathway, loss of APC protein function, and inability to phosphorylate β-catenin or mutations in the CTNNB1 gene, allows translocation of cytoplasmic β-catenin into the nucleus and, in conjunction with other proteins, promotes abnormal proliferation (Figure 1.5).[55]

Recognizing the central role of β-catenin in desmoid tumors, evaluation of CTNNB1 could be used as a potential diagnostic test in tumors that are challenging to identify.[56] In addition, targeting the Wnt/ β-catenin pathway may be a future therapeutic option of value.

The β-catenin destruction complex sits at the heart of the canonical Wnt signaling pathway, playing a critical role in keeping the levels of β-catenin low in the absence of an exogenous Wnt signal.

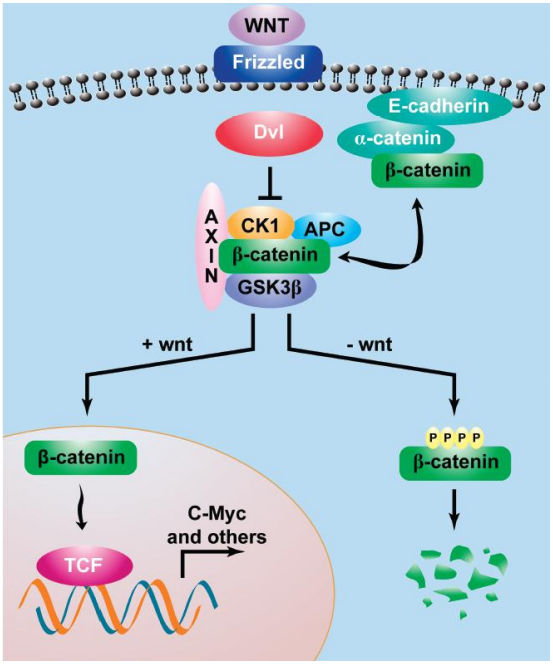


Figure 1.5 *Wnt signaling pathway*

Destruction complex

Initially, the destruction complex does not contain β-catenin. As β–catenin is synthesized in the cytosol, it joins the destruction complex by binding at a site on Axin between GSK3 and CK1-binding site. APC binds β-catenin in a C-terminal region of the armadillo repeats using one of its three 15 residues repeats. The binding of β-catenin allocates the flexible N-terminus of β-catenin in the proximity of CK1, which phosphorylates β-catenin at Ser45, priming it for GSK3 phosphorylation. The flexible N-terminal tail then encounters GSK3, which adds phosphates in a C-terminal to N-terminal direction at, successively, Thr41, Ser37 and Ser33.[57]

Phosphorilation of β-catenin results in its ubiquitination and subsequent degradation through a proteasome-dependent pathway. If catenin phosphorylation or its subsequent ubiquitination

are blocked, catenin is thus diverted from the proteasome, accumulates, and enters the nucleus where it activates expression genes, leading to the formation of tumors.

To better understand the main role in carcinogenesis processes of interaction complex β-catenin/gsk3, some details are presented below about the mechanism of phosphorylation in GSK3 (Figure 1.6).

When GSK3 is in the active state, substrate that already has a priming phosphate binds to a specific pocket, aligning them in such a way that GSK3 can phosphorylate a serine or a threonine located four residues amino terminal to the priming phosphate. After agonist stimulation, GSK3 becomes phosphorylated at a serine residue near its amino terminus, which is serine 21 in GSK3β. This transforms the amino terminus into a "pseudosubstrate" inhibitor, the phosphoserine occupying the same binding site as the priming phosphate of the substrate, and blocking access to the active site[58].

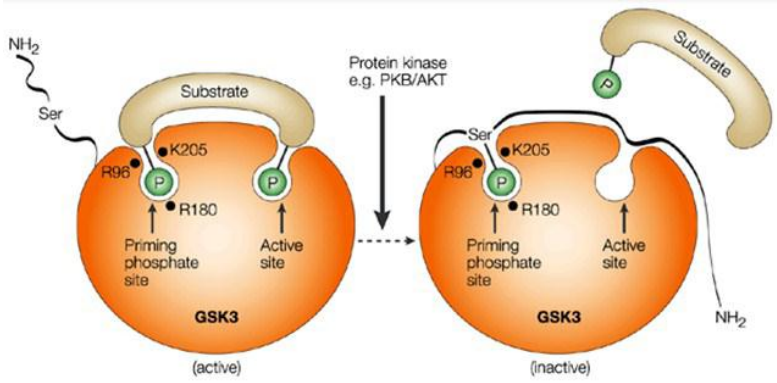


Figure 1.6 *The molecular mechanism by which phosphorylation inhibits GSK3*[58].

3D structure of β-catenin

From the structural standpoint β-catenin is composed by 781 residues and organized in several domains. The core domain is composed of 12 copies of a 42 amino acid sequence motif known as Armadillo repeats domain (Figure 1.7)[59]. The 12 repeats form a super helix of helices that feature a long, positively charged groove of the proteolyse resistant fragment. The N terminal part is about 150 aminoacids and contains the phosphorilation motif[60] while the C-ter part has the main role on the binding of co-transcriptional factors.

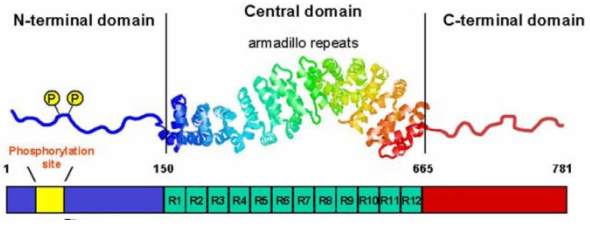


Figure 1.7 *Structure of β-catenin protein, highlighted the phosphorylation motif (in yellow) and the Armadillo repeats motif (in green).* [61]

1.5.3 Development of dendrimeric nanocarriers for gene therapy

The application of nanotechnology for targeted drug delivery is widely expected to bring breakthrough and create novel therapeutics capable of changing the landscape for pharmaceutical and biotechnology industries.[62,63] Particularly exciting are the recent efforts to apply nanotechnology to engineer nanovectors for RNA interference (RNAi) technology, giving birth to an entirely novel RNAi-therapeutics. [64]

RNAi makes use of small interfering RNA (siRNA) molecules, which are of 21-23 bp in length and break down the target mRNA in a sequence-specific manner. It is broadly, yet specifically, applicable to any target gene with known sequence. Compared with antisense oligonucleotides, siRNAs are more resistant to nuclease degradation, and show a more potent and prolonged therapeutic effect due to the catalytic mechanism. The promise that siRNA can efficiently and specifically down-regulate "undruggable" genes brings new hope to the incurable diseases or the diseases which are difficult to cure due to drug resistance and/or the absence of efficacious drug candidates, creating a new era of pharmaceutical science which have the potential to revolutionize medicine.

Nevertheless, the clinical applications of siRNA-based therapeutics are hampered by a number of undesirable physiochemical and biopharmaceutical characteristics, including enzymatic lability, lack of tissue specificity, poor cellular uptake, ecc.[65]

A breakthrough of RNAi technology in therapeutic application requires, however, safe and easy-to-handle siRNA delivery systems, which can protect siRNA during the extra- and intra-cellular delivery and bring siRNA to the site of interest. Both viral and nonviral delivery systems have been explored for siRNA delivery. Viral delivery is very effective, but the disadvantages are safety concerns and high cost. Lipids and polymers are the most common non-viral vectors which based on nanotechnology for delivery. They are able to assemble siRNA into nanoparticles, which can further protect siRNA from degradation and promote cell uptake. After internalization, siRNA will be dissociated from the vector complexes and undergo RNAi mechanism to silence the targeted gene.

Dendrimers are one of the main promising groups of nanovectors for siRNA delivery. The word "dendrimer" originated from two words, the Greek word dendron, meaning tree, and meros, meaning part.

Dendrimers are a class of well-defined nanostructured macromolecules that possess narrow mass or size polidispersivity and tree-like architecture distinguished by exponential numbers of discrete dendritic branches radiating out from a common core. These unique structural attributes confer a spherical shape to dendrimers after several layers of branching owing to

surface congestion. The precise nanoscale sizes, shapes, surface chemistries and architectures of these dendritic macromolecules distinguish these materials as model systems for understanding a wide range of crucial nano-medical issues, including polyvalent nanopharmaceuticals, DNA or siRNA delivery vectors, nanodiagnostics, MRI contrast agents and drug delivery. [66]

Dendrimers are built from a starting atom, such as nitrogen, to which carbon and other elements are added by a repeating series of chemical reactions that produce a spherical branching structure. As the process repeats, successive layers are added, and the sphere can be expanded to the size required by the investigator. The result is a spherical macromolecular structure whose size is similar to albumin and hemoglobin, but smaller than such multimers as the IgM antibody complex (Figure 1.8).

The generation number of dendrimers could be defined as the number of focal points when going from the core towards the dendrimer surface; for instance a dendrimer having five focal points when going from the centre to the periphery is denoted as the 5th generation dendrimer, usually abbreviated to simply a G5-dendrimer. The core part of the dendrimer is sometimes denoted generation "zero", in fact the core structure thus presents no focal points, as hydrogen substituents are not considered focal points. Intermediates during the dendrimer synthesis are sometimes denoted half-generations, a well-known example is the carboxylic acid-terminated PAMAM dendrimers.

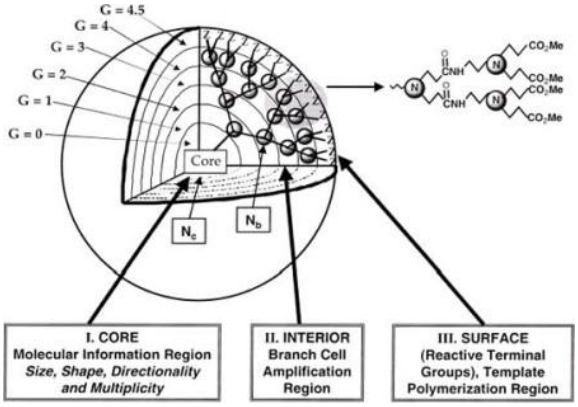


Figure 1.8 *Basic structure of dendrimer: core, interior and surface. Example of generation from G0 to G5.*

Usually, the functionalization of dendrimer was successfully obtained using Michael addition and other synthetic methodologies such as Williamson reaction, ester coupling, cross-methathesis. Unfortunately, all of these reactions showed some problem. Nevertheless, the discovery, in 2001, by Sharpless and coworkers, of "click chemistry" has overcome a lot of limitations. The catchy term "click" refers to energetically favored, specific, and versatile

chemical transformations, which lead to a single reaction product. This kind of reactions showed high yields, tolerance toward functional groups and virtually no side-reactions.[67]

Up to now, few efforts have been made for targeted siRNA delivery using dendrimers as nanovectors. One of the key issues is to critically balance the targeting ability with the delivery efficiency in order to achieve both efficient and targeted delivery. In this project, we would like to explore dendrimer conjugates for targeted delivery of siRNA, which is expected to further enhance the efficacy and safety profile, in the view to move the academic research to applications in nanomedicine.

The procedure for finding the correct vectors in preclinical stages is time, resource and cost intensive. Thus, the combination of computer-based molecular modeling and parallel assessment of biochemical and biological properties by experimental methods can open new avenues in this field of scientific research by providing powerful tools for rationale design and optimization of dendrimer-based nanovectors for RNAi therapeutics targeting different cancer type.

Application in cancer treatment

The field of oncology could soon be revolutionized by accurate strategies to identify and cure by dendrimer-based nanotherapeutics. Of course, the best anticancer drug will target only cancerous cells by individuating some parameters.

It is now possible to design specific dendrimers that can not only identify specific target on cancer cells but also deliver one or more therapeutic agents without collateral damage to the neighboring healthy cells; for achieving this objective a “Trojan horse” strategy could be employed.

In gene therapy the dendrimers seem to offer many advantages over viruses as vehicles of genes. They may be much less toxic and they may offer advantages in term of cost. For example they could be used, in couple with specific siRNA, in the treatment of prostate cancer. In Europe and the United States, prostate cancer is the most common solid neoplasm and the second leading cause of deaths due to cancer in men. The use of prostate-specific antigen (PSA) as a prostate cancer screening tool has led to a downstaging and downgrading of the disease at the time of diagnosis and a reduction in prostate cancer mortality. However now different targets have been identified and silencing through specific siRNA could be a novel strategy for therapeutic in this cancer. For example in the case of prostate cancer some proteins (e.g. Hsp27, TCTP) that seem to play a major role in the development are now under studies[68] and the gene silencing by dendrimer-siRNA approach used as innovative approach.

1.6 Study of mutations in cancer development and their interaction with drugs

The limits of cancer therapies are the development of drug resistance. For instance, recent analysis based on patterns of DNA mutations and RNA expression in 2000 specimens, revealed 10 different molecular types of breast cancer. In addition, cancer is characterized by extensive genetic and epigenetic alterations and mutations in drug targets, causing structural changes, may also be responsible for increased drug resistance. In this context, the data discussed regards in particular mutations involved c-Kit receptor in GISTs and SMO receptor in BCC.

1.6.1 Gastrointestinal Stromal Tumors (GISTs) and the c-KIT receptor role

Gastrointestinal stromal tumour (GIST) is the most common sarcoma of the intestinal tract, known to be notoriously refractory to conventional chemotherapy or radiation.

Constitutive activation of the KIT receptor tyrosine kinase is a central pathogenetic event in most GISTs and generally results from oncogenic point mutations which can involve either extracellular or cytoplasmic domains of the receptor. Oncogenic mutations enable the KIT receptor to phosphorylate various substrate proteins, leading to activation of signal transduction cascades which regulate cell proliferation, apoptosis, chemotaxis, and adhesion. KIT mutations can be broadly assigned to 2 groups, those that involve the “regulatory” regions responsible for modulating KIT enzymatic activity and those that involve the enzymatic region itself.

KIT is a member of the type III transmembrane receptor tyrosine kinase (RTK) family, which comprise five extracellular immunoglobulin domains, a single transmembrane region, an inhibitory cytoplasmic juxtamembrane domain and a split cytoplasmic kinase domain separated by a kinase insert segment[69]. Under physiological conditions, binding of the KIT ligand, stem cell factor (SCF), to the extracellular domain of the receptor results in receptor dimerization, activation of the intracellular tyrosine kinase domain through autophosphorylation of specific tyrosine residues, and receptor activation[70]. The downstream signal transduction pathways include the mitogen-activated protein kinase (MAPK), phosphatidylinositol 3-kinase (PI3K) and Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathways. The intracellular signalling through KIT plays a critical role in the development of several mammalian cell types.

In the active kinases, a characteristic DFG-motif or catalytic triad (D810, F811, G812), which is located immediately before the A-loop, adopts a conformation whereby the aspartic acid and phenylalanine are oriented toward the binding site (DFG-in). Different inactive states have been identified; one of these states is called DFG-out. In the presence of an activating mutation, the equilibrium between the DFG-in and DFG-out forms is shifted toward the active open (DFG-in) conformation of the receptor. This change allows enhanced binding of ATP and, hence, increased level of phosphorylation of the tyrosine kinase domain.[71]

The best characterized KIT regulatory region is the intracellular juxtamembrane (JXM) domain, encoded by exon 11. Contrarily to previous structural hypotheses, it has recently been shown that the JXM region of the c-Kit forms a compact hairpin loop that inserts directly into the domain interface between the kinase N- and C- lobes. Approximately 20 amino terminal residues of the JXM domain - from Y547 to G565 – comprise the inner buried half of this hairpin loop and form specific interactions that disrupt the conformation of the kinase DFG motif and A-loop.

It is well known that this β hairpin loop acts as an autoinhibitory domain: when the receptor is in the active conformation, the β hairpin loop has a peculiar shape and the DFG motif keeps the A loop in the closed conformation blocking the ATP to enter in this pocket.

To date, several types of mutations have been described and the main mutational “hot spot” is represented by c-Kit exon.

Regulatory-type mutations found in the c-Kit JXM region are associated with human gastrointestinal stromal tumors and include base substitutions and/or deletions in residues 550-560.[72]

A key feature of all these JXM region regulatory-type mutations is that the KIT oncoproteins have essentially the same active enzymatic site as that in native KIT. Therefore, kinase inhibitors such as Imatinib-mesylate (Gliveec, Novartis, Basel) are also effective inhibitors of these constitutively activated molecules.

Imatinib as first choice therapy

GIST patients harbor different oncogenic mutations in KIT and PDGFRA, which have distinct responses to imatinib. Imatinib mesylate (Gleevec) is a selective tyrosine kinase inhibitor whose targets include ABL, BCR-ABL, KIT and PDGFR (Figure 1.9). Imatinib is a2-phenyl-amino-pyrimidine derivative, which specifically binds to the inactive conformation of the ABL kinase or the inactive form of KIT. Imatinib was first applied in 1998 to refractory chronic myeloid leukaemia (CML) and subsequently to advanced GIST patients in 2000[73 74].

Imatinib achieves partial responses or stable disease in nearly 80% of GIST patients, and remarkably the 2-year survival in advanced GIST is now 75–80%. The long-term outcome of imatinib treatment for metastatic GIST has emerged from several large trials. Approximately 45% of patients with metastatic GIST have a measurable response after administration of imatinib, while about 30% will have at least stable disease.

However, responses to imatinib depend on the functional domain affected [46]. Patients with KIT exon 11 mutations have a partial response rate of 84% compared with a 0% partial response rate among patients without KIT mutations[75]. By analogy with other receptor tyrosine kinases, the juxtamembrane domain may function as a negative regulator of the KIT kinase and disruption of the conformational integrity of this domain may impair its negative regulatory function. Deletion of Y567 and V568 in the normal KIT gene have been shown to enhance agonist induced KIT signaling by altering the specificity of docking sites for cytoplasmic signaling molecules in the activated KIT receptor. By contrast, mutations in exon 9 of KIT, which encodes the fifth extracellular immunoglobulin-like loop, are less responsive, although this limitation can be partially overcome by a high-dose regimen (400 mg twice daily). Activating oncogenic mutations within the ectodomain of KIT map to the D5–D5 interface, which presumably stabilizes receptor dimers and thus activates the kinase in the absence of ligand. In spite of an activated KIT pathway, the imatinib response in GIST patients with a wild-type genotype for KIT/PDGFRA/BRAF has been disappointing. Furthermore, tumours with activation loop mutations in particular show the least response to imatinib inhibition.

Tyrosine kinase inhibitors (TKI) therapy is standard of care for patients with metastatic/advanced GIST and, although the optimal duration of treatment is not known, it is recommended to continue therapy indefinitely or as long as they are experiencing a clinical benefit. Imatinib interruption in patients with advanced GIST controlled with the drug is often associated with a high risk of disease progression within 1 year. In most patients, however, the disease is controlled with imatinib re-challenge, with no statistically significant differences in imatinib-refractory progression-free survival between interruption and continuation [76].

No apparent difference was seen in sunitinib activity between the intermittent and continuous daily dosing schedules.

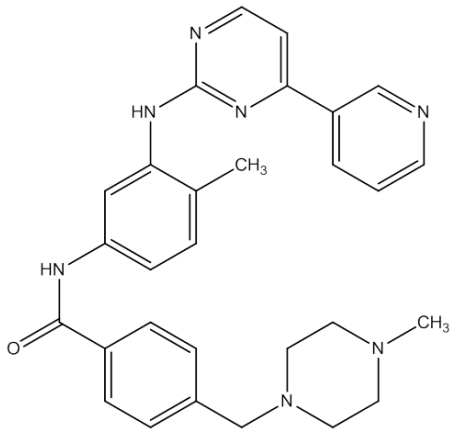


Figure 1.9 *Chemical structure of Imatinib (Gleevec).*

1.6.2 Basal cell carcinoma and the activation of Hedgehog pathway

Basal cell carcinoma (BCC) is the most frequent cutaneous cancer accounting for approximately, 70% of all malignant diseases of the skin. The incidence has increased dramatically over the years, and the number of women under 40 years diagnosed with BCC has more than doubled in the last 30 years.[77]

The incidence of BCC is strongly associated with exposure to UV radiation; tumors develop primarily on the sun-exposed skin of elderly individuals with fair skin phototypes, are rarely found on palmoplantar surfaces or in children, and never appear on the mucosa. Additional established risk factors include ionizing radiation (IR), arsenic, and immune suppression[78].

However, a variety of hereditary syndromes can result in an increased risk of developing BCC tumors, including nevoid BCC syndrome, Bazex–Dupré–Christol syndrome, Rombo syndrome, Oley syndrome, and xeroderma pigmentosum.

Treatment for BCC mainly consists in local approaches, with surgery (including different techniques) as the most effective strategy, followed by radiotherapy, photodynamic therapy, imiquimod as alternative treatments[79,80]. Radiation therapy is employed in case of surgical residual not amenable to resection or in patients not surgical candidates, if feasible.

However, the occurrence of locally aggressive and invasive tumors, a bare prognosis upon metastatic spread, and a significant rate of recurrence often associated with increased aggressiveness, as well as the multitude of tumors appearing in high-risk populations such as BCNS patients, provide strong reasons to search for new preventive and therapeutic strategies.

The Hedgehog pathway plays a critical role in embryonic development and is not active in most adult tissues, with the exception of stem cells, hair follicles, and skin cells in which the pathway is important for cell maintenance.[81] Hedgehog pathway signaling starts with the Hedgehog (Hh) ligand binding to a 12-pass transmembrane receptor known as PTCH (Figure

1.10). In 1993, three mammalian homologs of the Hh ligand were described, including Sonic Hedgehog, and also Indian Hedgehog and Desert Hedgehog. The "Hedgehog" name was coined based on the description of Drosophila melanogaster larvae, which took on the appearance of the spiky hedgehog when the gene was mutated.[82]

In the absence of the Hh ligand, the PTCH receptor acts as a tumor suppressor by inhibiting the next protein in the pathway known as Smoothened (SMO), which is a G-protein-coupled receptor. When the Hh ligand binds to PTCH, the inhibitory effects on SMO are released allowing the signal to propagate[83]. Although the mechanisms following SMO inhibition release have not been completely elucidated, SMO activation ultimately results in the release of inhibition of glioma-associated protein (Gli) through the suppressor of fused molecule (Sufu). The Gli family of proteins (Gli-1–3) are zinc finger transcription factors that are capable of activating a number of target genes, which can result in an oncogenic effect on the cell. Among the genes that are upregulated through Gli transcriptional activation are PTCH1 (provides negative feedback of pathway), Gli-1 (positive pathway feedback), and other gene pathways that aid in the survival of the cell, such as angiogenesis, cell cycle regulation, and antiapoptosis pathways. The Hedgehog pathway also conducts significant crosstalk with other molecules and pathways including p53, Wnt, PI3 K/aKT/mTOR, and retinoic acid. These interactions create a complex network, which may promote a variety of resistance mechanisms for drug targeting of this pathway. A variety of diseases have been linked to abnormal Hedgehog pathway signaling besides BCC, including medulloblastoma, hematologic malignancies, and other solid tumors.

In particular, the inappropriate activation of Hh pathway is involved in the pathogenesis of BCC: loss of function of the transmembrane receptor Patched (PTCH1), or gain of function of the Smoothened (SMO), Sonic Hh (SHH), and GLI family transcription factors can be ligand-independent oncogenic drivers of this disease[84].

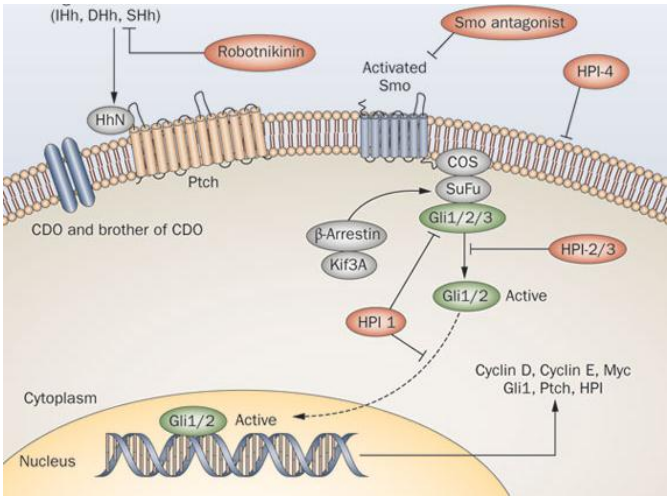


Figure 1.10 *Hedgehog signaling pathway inhibition, arranged*[85].

New therapeutics

The first evidence that the Hh signaling pathway is sensitive to inhibition by small molecules stemmed from the observation of cyclopia in lambs, induced by the maternal ingestion of corn lilies (Veratrum californicum).

Cyclopamine was, then, isolated from the Veratrum californicum plant. This discovery has led to a variety of more potent and selective SMO antagonists, which have been developed and incorporated into clinical research for a variety of cancer types, including patients with advanced BCC.

New derivatives of cyclopamine with improved pharmaceutical properties are now in clinical trials. Excellent results were obtained with the orally administered SMO inhibitor GDC-0449 (vismodegib) in a phase I trial of patients with locally advanced or metastatic BCC[86,87]. Vismodegib is a first-in-class, orally bioavailable inhibitor of SMO (Figure 1.11). Based on the efficacy and tolerability results of recent clinical studies, vismodegib (150 mg once daily) received US Food and Drug Administration (FDA) approval on January 2012 for the treatment of adults with BCC that has recurred following surgery or patients unsuitable for surgery or radiation. It represents a breakthrough in the management of advanced BCC, especially, for patients 'with high surgical morbidity and inoperable tumors. The median duration of treatment in trials has been 10.2 months and is continued until disease progression or unacceptable toxicity.

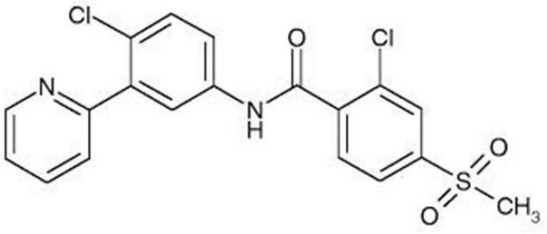


Figure 1.11 *Chemical structure of GDC-0449 (Vismodegib)*

Case of primary resistance in medulloblastoma

Although treatment of BCC with vismodegib and similar agents has resulted in dramatic responses, resistance to SMO inhibition occurs resulting in new tumor development or growth of previously responding tumors. A recent analysis into the mechanism of resistance to vismodegib has recently shed light into the complex nature of this process. In a patient with medulloblastoma with a known PTCH mutation, the initial response to vismodegib was seen with subsequent progression of disease[88]. Comparison of before treatment tumor samples and samples obtained from a vismodegib-resistant tumor demonstrated a novel finding. The pre-existing PTCH mutation found in the pre-treatment tumor was still present in the resistant

tumor; however, a new mutation in SMO (D473) was seen. Similar to alterations in BCR–ABL, which confer resistance of CML cells to imatinib, this SMO-D473 mutation was found to confer resistance to vismodegib. Further evaluation implicated that this particular mutation affected vismodegib binding to Smo.

1.7 Molecular simulation in cancer therapy

The complexity of cancer and the vast amount of experimental data available have made computer-aided approaches necessary.

Many details on a given protein's functions and how it is altered by different mutations, become clear once the protein 'structure is known. Proteins, however, are not static and it is their dynamic landscape that determines their roles. Computer-aided studies can complement molecular studies and yield details that are not available to the experiment. Molecular modeling approaches therefore become increasingly useful in many clinically-oriented studies.

Today, "molecular modeling" refers to the application of computer-generated models in molecular studies ranging from a few atoms to multitude of biomolecules. These models are used to simulate processes that may be as fast as 10^{-5} s or as slow as a few seconds. Clearly, the accuracy and level of detail depend on the size and timescale of the system. In general the more accurate the method at hand, the more time and computational resources will be needed to get a meaningful result. The choice of the modeling method is therefore made based upon the problem at hand.[89]

A fast and noninvasive global approach based on an in silico analysis represents a useful tool that, coupled with traditional biochemical molecular evidence, may help to overcome the problem of resistance in cancer treatment.

In silico analysis, which uses the three-dimensional structure of the target protein, has multiple uses. The method can predict structural changes introduced by mutations, measure the strength of the interactions between the protein model and the drug, determine whether the drug can be effective or not by calculating the binding affinity, and evaluate to what extent decreasing efficacy could still be counteracted by increasing the drug dose or by using a different compound.

The individuation of new target, that unfortunately could not biologically evaluated cause of lack of structural information, could be helped by the application of new techniques that exploiting homology modeling and advanced simulation assay give details for further strategies in order to inhibit the process of cancer development.

1.8 Future perspectives

The Darwinian selection of mutated variants of target proteins could be harnessed with the intelligent design of drugs directed against the single new variants. This innovative multi-drug targeted prevention, which can potentially change the natural history of a disease, has been proposed in the setting of chronic myeloid leukemia.

By predicting the mutations, and classifying them according to their degree of probability, it could be possible to administer a drug as soon as patients display such mutations, so that disease recurrence can be prevented, rather than waiting for the mutation to emerge to select the right compound.

Bibliography

1 Society, A. C. *http://www.cancer.org/research/cancerfactsstatistics/index*.

2 Kumar, V. A., A. Fausto, N. *Robbins & Cotran Pathologic basis of disease*. 7 edn, (Elsevier, 2004).

3 Papac, R. J. Origins of cancer therapy. *The Yale journal of biology and medicine* **74**, 391-398 (2001).

4 DeVita, V. T. & Chu, E. A History of Cancer Chemotherapy. *Cancer research* **68**, 8643-8653, doi:10.1158/0008-5472.can-07-6611 (2008).

5 Calvo, E. & Baselga, J. Ethnic differences in response to epidermal growth factor receptor tyrosine kinase inhibitors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 2158-2163, doi:10.1200/JCO.2006.06.5961 (2006).

6 Gerber, D. E. Targeted therapies: a new generation of cancer treatments. *American family physician* **77**, 311-319 (2008).

7 Daley, G. Q., Van Etten, R. A. & Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* **247**, 824-830 (1990).

8 Mendelsohn, J. & Baselga, J. Epidermal growth factor receptor targeting in cancer. *Seminars in oncology* **33**, 369-385, doi:10.1053/j.seminoncol.2006.04.003 (2006).

9 Cardones, A. R. & Banez, L. L. VEGF inhibitors in cancer therapy. *Current pharmaceutical design* **12**, 387-394 (2006).

10 Kantarjian, H. *et al.* Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *The New England journal of medicine* **346**, 645-652, doi:10.1056/NEJMoa011573 (2002).

11 Kantarjian, H. M. *et al.* Imatinib mesylate therapy in newly diagnosed patients with Philadelphia chromosome-positive chronic myelogenous leukemia: high incidence of early complete and major cytogenetic responses. *Blood* **101**, 97-100, doi:10.1182/blood-2002-02-0545 (2003).

12 Romond, E. H. *et al.* Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *The New England journal of medicine* **353**, 1673-1684, doi:10.1056/NEJMoa052122 (2005).

13 Moore, M. J. *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **25**, 1960-1966, doi:10.1200/JCO.2006.07.9525 (2007).

14 Willett, C. G. *et al.* Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer. *Nature medicine* **10**, 145-147, doi:10.1038/nm988 (2004).

15 Longley, D. B. & Johnston, P. G. Molecular mechanisms of drug resistance. *The Journal of pathology* **205**, 275-292, doi:10.1002/path.1706 (2005).

16 Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer research* **72**, 4875-4882, doi:10.1158/0008-5472.CAN-12-2217 (2012).

17 Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature reviews. Cancer* **13**, 714-726, doi:10.1038/nrc3599 (2013).

18 Gorre, M. E. *et al.* Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**, 876-880, doi:10.1126/science.1062538 (2001).

19 Shah, N. P. *et al.* Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* **305**, 399-401, doi:10.1126/science.1099480 (2004).

20 Golas, J. M. *et al.* SKI-606, a 4-anilino-3-quinolinecarbonitrile dual inhibitor of Src and Abl kinases, is a potent antiproliferative agent against chronic myelogenous leukemia cells in culture and causes regression of K562 xenografts in nude mice. *Cancer research* **63**, 375-381 (2003).

21 Matsumoto, R. R., Liu, Y., Lerner, M., Howard, E. W. & Brackett, D. J. Sigma receptors: potential medications development target for anti-cocaine agents. *European journal of pharmacology* **469**, 1-12 (2003).

22 Martin, W. R., Eades, C. G., Thompson, J. A., Huppler, R. E. & Gilbert, P. E. The effects of morphine- and nalorphine- like drugs in the nondependent and morphine-dependent chronic spinal dog. *Journal of Pharmacology and Experimental Therapeutics* **197**, 517-532 (1976).

23 Quirion, R. *et al.* A proposal for the classification of sigma binding sites. *Trends in pharmacological sciences* **13**, 85-86 (1992).

24 Hellewell, S. B. & Bowen, W. D. A sigma-like binding site in rat pheochromocytoma (PC12) cells: decreased affinity for (+)-benzomorphans and lower molecular weight suggest a different sigma receptor form from that of guinea pig brain. *Brain research* **527**, 244-253 (1990).

25 Hanner, M. *et al.* Purification, molecular cloning, and expression of the mammalian sigma1-binding site. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8072-8077 (1996).

26 Moebius, F. F., Striessnig, J. & Glossmann, H. The mysteries of sigma receptors: new family members reveal a role in cholesterol synthesis. *Trends in pharmacological sciences* **18**, 67-70 (1997).

27 Hellewell, S. B. *et al.* Rat liver and kidney contain high densities of sigma 1 and sigma 2 receptors: characterization by ligand binding and photoaffinity labeling. *European journal of pharmacology* **268**, 9-18 (1994).

28 Cobos, E. J., del Pozo, E. & Baeyens, J. M. Irreversible blockade of sigma-1 receptors by haloperidol and its metabolites in guinea pig brain and SH-SY5Y human neuroblastoma cells. *Journal of neurochemistry* **102**, 812-825, doi:10.1111/j.1471-4159.2007.04533.x (2007).

29 Fontanilla, D. *et al.* The hallucinogen N,N-dimethyltryptamine (DMT) is an endogenous sigma-1 receptor regulator. *Science* **323**, 934-937, doi:10.1126/science.1166127 (2009).

30 Hayashi, T., Maurice, T. & Su, T. P. Ca(2+) signaling via sigma(1)-receptors: novel regulatory mechanism affecting intracellular Ca(2+) concentration. *The Journal of pharmacology and experimental therapeutics* **293**, 788-798 (2000).

31 Alonso, G. *et al.* Immunocytochemical localization of the sigma(1) receptor in the adult rat central nervous system. *Neuroscience* **97**, 155-170 (2000).

32 Bermack, J. E. & Debonnel, G. The role of sigma receptors in depression. *Journal of pharmacological sciences* **97**, 317-336 (2005).

33 Maurice, T. & Su, T. P. The pharmacology of sigma-1 receptors. *Pharmacology & therapeutics* **124**, 195-206, doi:10.1016/j.pharmthera.2009.07.001 (2009).

34 Cobos, E. J., Entrena, J. M., Nieto, F. R., Cendan, C. M. & Del Pozo, E. Pharmacology and therapeutic potential of sigma(1) receptor ligands. *Current neuropharmacology* **6**, 344-366, doi:10.2174/157015908787386113 (2008).

35 Aydar, E., Palmer, C. P., Klyachko, V. A. & Jackson, M. B. The sigma receptor as a ligand-regulated auxiliary potassium channel subunit. *Neuron* **34**, 399-410 (2002).

36 Polgar, O. *et al.* Mutational analysis of ABCG2: role of the GXXXG motif. *Biochemistry* **43**, 9448-9456, doi:10.1021/bi0497953 (2004).

37 Megalizzi, V., Le Mercier, M. & Decaestecker, C. Sigma receptors and their ligands in cancer biology: overview and new perspectives for cancer therapy. *Medicinal research reviews* **32**, 410-427, doi:10.1002/med.20218 (2012).

38 Zeng, C. *et al.* Subcellular localization of sigma-2 receptors in breast cancer cells using two-photon and confocal microscopy. *Cancer research* **67**, 6708-6716, doi:10.1158/0008-5472.CAN-06-3803 (2007).

39 Megalizzi, V. *et al.* Screening of anti-glioma effects induced by sigma-1 receptor ligands: potential new use for old anti-psychiatric medicines. *Eur J Cancer* **45**, 2893-2905, doi:10.1016/j.ejca.2009.07.011 (2009).

40 Megalizzi, V. *et al.* 4-IBP, a sigma1 receptor agonist, decreases the migration of human cancer cells, including glioblastoma cells, in vitro and sensitizes them in vitro and in vivo to cytotoxic insults of proapoptotic and proautophagic drugs. *Neoplasia* **9**, 358-369 (2007).

41 Caveliers, V. *et al.* Evaluation of 99mTc-RP128 as a potential inflammation imaging agent: human dosimetry and first clinical results. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **42**, 154-161 (2001).

42 John, C. S. *et al.* A malignant melanoma imaging agent: synthesis, characterization, in vitro binding and biodistribution of iodine-125-(2-piperidinylaminoethyl)4-iodobenzamide. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **34**, 2169-2175 (1993).

43 Palmer, C. P., Mahen, R., Schnell, E., Djamgoz, M. B. & Aydar, E. Sigma-1 receptors bind cholesterol and remodel lipid rafts in breast cancer cell lines. *Cancer research* **67**, 11166-11175, doi:10.1158/0008-5472.CAN-07-1771 (2007).

44 Crawford, K. W., Coop, A. & Bowen, W. D. sigma(2) Receptors regulate changes in sphingolipid levels in breast tumor cells. *European journal of pharmacology* **443**, 207-209 (2002).

45 Su, T. P., Hayashi, T. & Vaupel, D. B. When the endogenous hallucinogenic trace amine N,N-dimethyltryptamine meets the sigma-1 receptor. *Science signaling* **2**, pe12, doi:10.1126/scisignal.261pe12 (2009).

46 Papagelopoulos, P. J. *et al.* Current trends in the management of extra-abdominal desmoid tumours. *World journal of surgical oncology* **4**, 21, doi:10.1186/1477-7819-4-21 (2006).

47 Gronchi, A. *et al.* Quality of surgery and outcome in extra-abdominal aggressive fibromatosis: a series of patients surgically treated at a single institution. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 1390-1397 (2003).

48 Lev, D. *et al.* Optimizing treatment of desmoid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **25**, 1785-1791, doi:10.1200/JCO.2006.10.5015 (2007).

49 Alman, B. A., Li, C., Pajerski, M. E., Diaz-Cano, S. & Wolfe, H. J. Increased beta-catenin protein and somatic APC mutations in sporadic aggressive fibromatoses (desmoid tumors). *The American journal of pathology* **151**, 329-334 (1997).

50 Reitamo, J. J., Hayry, P., Nykyri, E. & Saxen, E. The desmoid tumor. I. Incidence, sex-, age- and anatomical distribution in the Finnish population. *American journal of clinical pathology* **77**, 665-673 (1982).

51 Bertario, L. *et al.* Genotype and phenotype factors as determinants of desmoid tumors in patients with familial adenomatous polyposis. *International journal of cancer. Journal international du cancer* **95**, 102-107 (2001).

52 Caspari, R. *et al.* Familial adenomatous polyposis: desmoid tumours and lack of ophthalmic lesions (CHRPE) associated with APC mutations beyond codon 1444. *Human molecular genetics* **4**, 337-340 (1995).

53 Lefevre, J. H. *et al.* Risk factors for development of desmoid tumours in familial adenomatous polyposis. *The British journal of surgery* **95**, 1136-1139, doi:10.1002/bjs.6241 (2008).

54 Xing, Y., Clements, W. K., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/axin complex suggests a mechanism for the beta-catenin destruction complex. *Genes & development* **17**, 2753-2764, doi:10.1101/gad.1142603 (2003).
55 MacDonald, B. T., Tamai, K. & He, X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell* **17**, 9-26, doi:10.1016/j.devcel.2009.06.016 (2009).
56 Colombo, C. *et al.* CTNNB1 45F mutation is a molecular prognosticator of increased postoperative primary desmoid tumor recurrence: an independent, multicenter validation study. *Cancer* **119**, 3696-3702, doi:10.1002/cncr.28271 (2013).
57 Kimelman, D. & Xu, W. beta-catenin destruction complex: insights and questions from a structural perspective. *Oncogene* **25**, 7482-7491, doi:10.1038/sj.onc.1210055 (2006).
58 Cohen, P. & Frame, S. The renaissance of GSK3. *Nature reviews. Molecular cell biology* **2**, 769-776, doi:10.1038/35096075 (2001).
59 Huber, A. H., Nelson, W. J. & Weis, W. I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90**, 871-882 (1997).
60 Graham, T. A., Weaver, C., Mao, F., Kimelman, D. & Xu, W. Crystal structure of a beta-catenin/Tcf complex. *Cell* **103**, 885-896 (2000).
61 Megy, S. *et al.* Solution structure of a peptide derived from the oncogenic protein beta-Catenin in its phosphorylated and nonphosphorylated states. *Peptides* **26**, 227-241, doi:10.1016/j.peptides.2004.09.021 (2005).
62 Ferrari, M. Cancer nanotechnology: opportunities and challenges. *Nature reviews. Cancer* **5**, 161-171, doi:10.1038/nrc1566 (2005).
63 Jain, K. K. Challenges of drug discovery for personalized medicine. *Current opinion in molecular therapeutics* **8**, 487-492 (2006).
64 McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature reviews. Genetics* **3**, 737-747, doi:10.1038/nrg908 (2002).
65 Shukla, G. C. *et al.* A boost for the emerging field of RNA nanotechnology. *ACS nano* **5**, 3405-3418, doi:10.1021/nn200989r (2011).
66 Menjoge, A. R., Kannan, R. M. & Tomalia, D. A. Dendrimer-based drug and imaging conjugates: design considerations for nanomedical applications. *Drug discovery today* **15**, 171-185, doi:10.1016/j.drudis.2010.01.009 (2010).
67 Lallana, E., Fernandez-Trillo, F., Sousa-Herves, A., Riguera, R. & Fernandez-Megia, E. Click chemistry with polymers, dendrimers, and hydrogels for drug delivery. *Pharmaceutical research* **29**, 902-921, doi:10.1007/s11095-012-0683-y (2012).
68 Cornford, P. A. *et al.* Heat shock protein expression independently predicts clinical outcome in prostate cancer. *Cancer research* **60**, 7099-7105 (2000).
69 Yarden, Y. *et al.* Human proto-oncogene c-kit: a new cell surface receptor tyrosine kinase for an unidentified ligand. *The EMBO journal* **6**, 3341-3351 (1987).
70 Lev, S., Yarden, Y. & Givol, D. Dimerization and activation of the kit receptor by monovalent and bivalent binding of the stem cell factor. *The Journal of biological chemistry* **267**, 15970-15977 (1992).
71 Antonescu, C. R. The GIST paradigm: lessons for other kinase-driven cancers. *The Journal of pathology* **223**, 251-261, doi:10.1002/path.2798 (2011).
72 Pierotti, M. A., Tamborini, E., Negri, T., Pricl, S. & Pilotti, S. Targeted therapy in GIST: in silico modeling for prediction of resistance. *Nature reviews. Clinical oncology* **8**, 161-170, doi:10.1038/nrclinonc.2011.3 (2011).
73 Druker, B. J. *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature medicine* **2**, 561-566 (1996).
74 Joensuu, H. *et al.* Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *The New England journal of medicine* **344**, 1052-1056, doi:10.1056/NEJM200104053441404 (2001).

75 Heinrich, M. C. *et al.* Primary and secondary kinase genotypes correlate with the biological and clinical activity of sunitinib in imatinib-resistant gastrointestinal stromal tumor. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 5352-5359, doi:10.1200/JCO.2007.15.7461 (2008).
76 Duffaud, F. & Le Cesne, A. Imatinib in the treatment of solid tumours. *Targeted oncology* **4**, 45-56, doi:10.1007/s11523-008-0101-x (2009).
77 Kasper, M., Jaks, V., Hohl, D. & Toftgard, R. Basal cell carcinoma - molecular biology and potential new therapies. *The Journal of clinical investigation* **122**, 455-463, doi:10.1172/JCI58779 (2012).
78 Epstein, E. H. Basal cell carcinomas: attack of the hedgehog. *Nature reviews. Cancer* **8**, 743-754, doi:10.1038/nrc2503 (2008).
79 Rubin, A. I., Chen, E. H. & Ratner, D. Basal-cell carcinoma. *The New England journal of medicine* **353**, 2262-2269, doi:10.1056/NEJMra044151 (2005).
80 Avril, M. F. *et al.* Basal cell carcinoma of the face: surgery or radiotherapy? Results of a randomized study. *British journal of cancer* **76**, 100-106 (1997).
81 Gupta, S., Takebe, N. & Lorusso, P. Targeting the Hedgehog pathway in cancer. *Therapeutic advances in medical oncology* **2**, 237-250, doi:10.1177/1758834010366430 (2010).
82 Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes & development* **15**, 3059-3087, doi:10.1101/gad.938601 (2001).
83 Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892-897, doi:10.1038/nature00989 (2002).
84 Tilli, C. M., Van Steensel, M. A., Krekels, G. A., Neumann, H. A. & Ramaekers, F. C. Molecular aetiology and pathogenesis of basal cell carcinoma. *The British journal of dermatology* **152**, 1108-1124, doi:10.1111/j.1365-2133.2005.06587.x (2005).
85 Takebe, N., Harris, P. J., Warren, R. Q. & Ivy, S. P. Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nature reviews. Clinical oncology* **8**, 97-106 (2011).
86 Von Hoff, D. D. *et al.* Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. *The New England journal of medicine* **361**, 1164-1172, doi:10.1056/NEJMoa0905360 (2009).
87 LoRusso, P. M. *et al.* Phase I trial of hedgehog pathway inhibitor vismodegib (GDC-0449) in patients with refractory, locally advanced or metastatic solid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 2502-2511, doi:10.1158/1078-0432.CCR-10-2745 (2011).
88 Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572-574, doi:10.1126/science.1179386 (2009).
89 Friedman, R., Boye, K. & Flatmark, K. Molecular modelling and simulations in cancer research. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1836**, 1-14, doi:http://dx.doi.org/10.1016/j.bbcan.2013.02.001 (2013).

Materials & Methods

In recent years, the use of computer simulations as a tool for bridging between microscopic length and time scales and the macroscopic world of the laboratory has been increasing exponentially. This success is attributed, by using computational chemistry and physics, to the guess for the interactions between molecules can be provided, and `exact' predictions of bulk properties can be obtained. The predictions are `exact' in the sense that they can be made as accurate as we like, subject to the limitations imposed by the available computer budget.

The actual computational modeling of biological macromolecules, mainly based on molecular dynamics (MD) simulations, commonly revolves around structure representations in atomic or near-atomic detail, with a classical description of physical interactions. In a typical MD simulation, the atomic trajectories of a system of N (e.g., 223) particles are generated by numerical integration of Newton's equation of motion, for a specific interatomic potential, with certain initial and boundary conditions. Such models have been quite successful in complementing experimental data with structural, dynamic, and energetic information, but involve substantial computational resources for larger systems, or when long time scales have to be considered.

2.1 The atomistic scale of simulation

The molecular mechanic (MM) is the branch of the molecular simulation that studies the equilibrium conditions of a certain molecular system. The goal is the determination of the reliable configuration that a certain molecule assumes in given conditions (equilibrium in physiological condition). The instrument used to reach this result is the minimization of the total energy of the molecular system. It is worth noting that all the measured energetic and structural quantities are statistic and thus the result is a statistical equilibrium.

There are some minimum information that are necessary to characterize the whole systems and analyze the equilibrium conditions: i) the *position of atoms* that compose the molecule in space; ii) the description of *how these atoms are bonded* together and how "they feel" each other.

For an accurate description of the energetic of a system, it is necessary to determine all the forces that are applied on the atoms that constitute the molecule (Figure 2.1). Thus, the determination of the total energy of the system is determined as the sum of all the forces that are acting on each atom of the system itself.

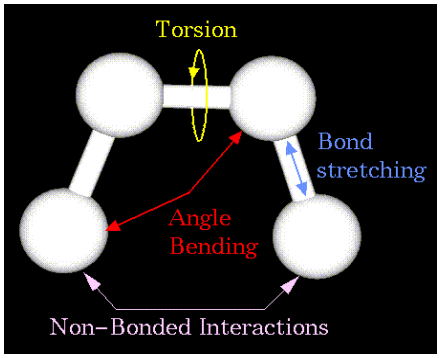


Figure 2.1 *Schematic representation of energies.*

Fundamentally, proteins derive their properties from the interaction between their constituent atoms. These basic interactions make the atoms assemble in a particular structure. The same interactions also define how the atoms prefer to arrange themselves on the surface or around a vacancy. Therefore, to understand the behavior of a protein, it is necessary and sufficient to study the collective behavior of atoms that constitutes the complex. This handout discusses the basic features of the interactions between atoms and different models that describes them.

When put close together, atoms interact by exerting forces on each other. Depending on the atomic species, some interatomic interactions are relatively easy to describe, while others can be very complicated. This variability stems from the quantum mechanical motion and interaction of electrons. Henceforth, rigorous treatment of interatomic interactions should be based on a solution of the Schrodinger's equation for interacting electrons, which is usually referred to as the first principles or ab initio theory.

The atomistic scale of molecular simulation has a pure classical expression and no quantum term is taken into account, since the electrons of each atom are not explicitly considered. To represent the atoms that constitute a certain molecule it could be useful to think as balls, identified by coordinates and connected each other to obtain a certain geometry. While the starting position is described as cartesian x,y,z coordinates in a configuration file, the connection between atoms is a rather more complex issue. The starting point is that molecules follow classical laws in MM and MD, the atoms are simply connected by springs with a certain force constant. These constants depend on the atoms coupled and on the bond type they are able to form.

The force field could be used to represent these interactions. The definition of force field by Otto Ermer said "the analytical expression that describes the potential energy of the studied molecular complex and all the parameters needed to describe". The force field is a file that

contains a data bank of bond and non-bond parameters for different atoms. There are different force fields dedicated to different branches of applications of molecular modeling, e.g. life science. When some parameters are missing in the force filed – i.e. when not all the atoms in the system are present in the force field, like a Fluorine atom in a life science dedicated force filed, the software is not able to recognize what this atom is and how it is supposed to bond to each other. The missing parameters could be derived by experimental results or, if it is not possible, by ab initio techniques.

Therefore, when the system can describe completely, the reaching of the equilibrium is described by equation.

$$\frac{\partial E_{pot}}{\partial r_1} = 0$$

where the potential energy E_{pot} is function of the coordinates r_i of the atoms of the system. Even if molecular mechanic is useful to reduce bad energetic configurations that typically occur in the early steps of the simulation, it is considered nothing more than a preliminary step for complex systems, since it does not take care of the temperature, that in MD appears only in the kinetic energy term and, thus, where the time evolution is considered.

$$E_{kin}=\frac{3}{2}k_bT$$

Molecular dynamic (MD) is simply an iteration in time of molecular mechanic – at each time step of the simulation the system moves toward the equilibrium. The concept of this technique is conceptually rather simple: given a starting configuration for the molecular system of interest, corresponding to the position of atoms, the forces that are acting on each atom are calculated, and thus trajectories can be computed as the evolution at each time step of the system toward the equilibrium. Differently form the MM, in MD an initial set of random velocities are calculated from temperature– a starting condition that is defined by temperature. This creates atomic movements – the system moves to reach the equilibrium – and new positions and velocities for all the atom of the system are calculated at each dynamic step. Obviously, the system fluctuates toward the equilibrium maintaining the total energy (E_{tot}) constant.

$$E_{tot}= E_{pot} + E_c = \text{constant}$$

It means that, if the new velocities are too low or too high to maintain the constancy, at each step they are rescaled in order to be consistent with the temperature of the system and with the E_{tot} with the so-called thermostats and velocity scaling algorithms.

As starting point we could consider the Schrödinger equation, that could describe, the motion of molecules, time indipendent:

$$H\Psi(R,r)=E\Psi(R,r)$$

Where:

- H is the Hamiltonian operator
- Ψ is the wave function depending on r (electronic coordinates) and R (nuclear coordinates)
- E is the energy of the state

During MD the electrons are not taken in account in the calculation. Thus, the Schrödinger equation assumes the following form, according to the Born-Oppenheimer approximation that allows the wave function of a molecule to be broken into its electronic and nuclear components.

This equation loses the explicit description of electrons - the system becomes dependent only on the nucleus position and velocity of each atom. All the quantum effects are lost, and can be expressed in classical form as follows:

$$-\frac{dE_{pot}(R)}{dR} = m\frac{d^2R}{dt^2}$$

In fact, to solve a MD problem, means to solve across time an equation like:

$$a_i(t) = \frac{F_i(t)}{m_i}$$

It is thus necessary to calculate the force F(t), dependent on the position of each atom in space, from the potential V(R). This will be done for all the atoms that constitute the system:

$$F_i = -\nabla V_i(R_1,\ldots\ldots,R_n)$$

The potential V(R) is, as said, defined by the force field. Once F(t) is calculated – and thus the acceleration a(t) is known – by integration it is possible to calculate the new coordinates for all the atoms of the system with the so-called Verlet algorithm that is represented by the following equations:

$$R_i(t+\Delta t) = R_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2$$

$$R_i(t-\Delta t) = R_i(t) - v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2$$

that can be summed to give:

$$R_i(t+\Delta t) = 2R_i(t) - R_i(t-\Delta t) + a_i(t)\Delta t^2$$

With this algorithm it is possible to compute the new positions of each atom of the system and thus to calculate the new potential energy, as sum of the potential energetic contribution

given by each atom, E_{pot}. Moreover, it is possible to calculate the new velocities of all the atoms:

$$v_i(t) = \frac{[R_i\,(t+\Delta t) - R_i\,(t-\Delta t)]}{2\Delta t}$$

in order to calculate the new values of the kinetic energy as:

$$E_{kin,\,i} = \frac{1}{2} m_i\,[y_i\,(t)^2]$$

and the constancy of the total energy can be expressed as in described equation. This procedure is iteratively repeated at each time step for the whole length of molecular dynamic simulation.

2.2 Force field and software

Many of the problems that involve the definition of a system in molecular modeling are linked to the parameterization.

The application of force field methods that could calculate the energy of a system taking into account only the nuclear position is a possible solution.

The reliability of the model depends on the choice and usage of the proper force field, dedicated to the simulation of a certain class of molecules. For instance, a force field that is dedicated to general material science could give completely wrong results for proteins simulations, or in most cases could be unable even to start the calculation.

Transferability is an indispensable property to choose a force field, indeed it allows to use a set of parameters tested on a small number of cases and apply it in very different situations.

The studies presented in this work were all carried out exploiting AMBER (version 11)[1] is a set of molecular mechanics force fields for molecular dynamics of biomolecules originally developed by the Peter Kollman's group at the University of California (San Francisco) and, at the same time, refers to a simulation package that implements these force fields.

The form of the AMBER force field is:

$$V(R^N) = \sum_{bonds} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angles} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] +$$

$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{Rij} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{Rij} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} \right\}$$

Where:

$$\sum_{bonds} \frac{1}{2} k_b (l - l_0)^2$$

It represents the harmonic force (ideal spring) linked to the energy between covalently bonded atoms. This is a good approximation near the equilibrium bond length, but becomes increasingly poor as atoms separate. K_b is a constant that determined the strength of the kind of atoms involved in the bond;

$$\sum_{angles} k_a(\theta - \theta_0)^2$$

It Is associated to the variations in the angle term θ respect to the ideal value θ_0. It represents the energy due to the geometry of electron orbitals involved in covalent bonding;

$$\sum_{torsions} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)]$$

The third term is the potential function associated with the torsional angles. It represents the energy for twisting a bond depending on bond order (e.g. double bonds) and neighboring bonds or lone pairs of electrons;

$$\sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{Rij} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{Rij} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} \right\}$$

The last term involves the non-bonded energy between all atom pairs as decomposed by van der Waals (first term) and electrostatic (second term of summation) energies.

All the constant (k_b, k_a, etc.) express in the main equation are unknown. In order to use the AMBER force field, it is necessary to have values for the parameters of the force field (e.g. force constants, equilibrium bond lengths and angles, charges). The AMBER software contains a fairly large number of these parameter sets. Each parameter set has a name, and provides parameters dedicated to certain types of molecules.

- "ff" followed by the year of creation (i.e. 95): peptide, protein and nucleic acid parameters.
- GAFF (Generalized AMBER force field): parameters for small organic molecules to facilitate simulations of drugs and small molecule ligands in complex with biomolecules (e.g. protein).
- GLYCAM: parameters for simulating carbohydrates.

In all the calculations presented in this thesis, the ff03[2] was used, being the most widely used in the study of proteins. All the amino acids forming proteins are parameterized with the constant values present in ff03.

However, ff03 provides a very exhaustive description of proteins and nucleic acids, but parameters for molecules different for the conventional organic ones are missing. GAFF is

completely compatible with AMBER ff, for this region is used to integrate missing parameters. The full list of modules present in the AMBER package are the following:

- leap is the tool that is used for preparing files necessary to start the simulation.antechamber calculates the missing parameters for small organic molecules using GAFF forcefield[3].
- sander (Simulated Annealing with NMR-Derived Energy Restraints) is the central simulation program and provides facilities for energy minimization and molecular dynamics with a wide variety of options.
- Pmemd (Particle mesh Ewald molecular dynamics) is a new version of Sander by Bob Duke. It is optimized to improve the performance of the simulations.
- Nmode (Normal Mode analysis) calculates the entropy of the system.
- ptraj is an analysis program for simulation results. AMBER does not include visualization capabilities.

In these studies the most used visualization tools was UCSF Chimera, that is a program to visualize and analyze molecular structures and their data deriving from Molecular Dynamics simulation. It is possible to analyze trajectories, sequence alignments, docking results, conformational changes and obtain images.

By using ab initio (or semiempirical) techniques, we can calculate parameters for small molecules that are not originally present in the force field (i.e. small drug molecules). In all the studies described in this thesis, a semiempirical Austin model (AM1) calculation method with the antechamber module within AMBER 11 was use for missing GAFF parameters.

- Antechamber is used during MM studies to recognize atom type, bond type, generate residue topology file and find missing force field parameters and supplying reasonable and similar substitute.

This Austin model is implemented with BCC method (ab initio term bond) that is one of the more precise charge calculation method and can guarantee results with precision comparable to the one of pure quantum-mechanic calculations reducing the computational time.

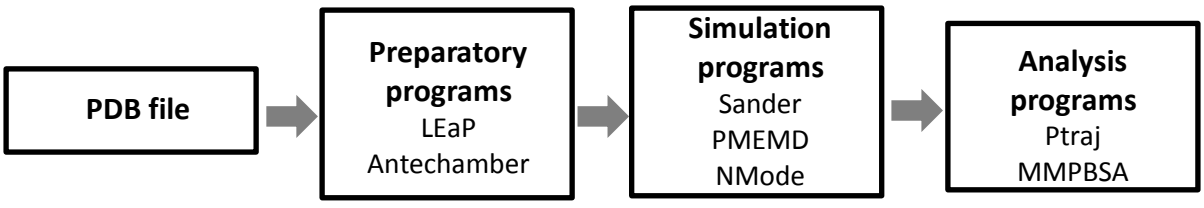


Figure 2.2 *Steps of the molecular simulation process.*

2.3 Creation of the “wild type” complexes

The molecular simulation allowed us to explore different properties for the system analyzed, however to start a simulation first of all we need to proceed building the system of interest.

In this work we evaluated different kind of systems: protein structure, protein/drug complex and protein/protein complex. In addition, some of those were studied in different conformational states, starting from wild type (WT) to the mutated complex.

Here we present the procedure that was followed for the creation of the WT molecular complexes, subject of the simulation studies.

It is worth noting that, since we want to simulate a system as close as possible to the reality, it is fundamental to mimic its real conditions. In these studies, the main characters are proteins and drugs that are naturally inserted in physiological condition of the human body.

To obtain an accurate simulation we will evaluate the complex in the correct salt concentration corresponding to the human ionic strength (NaCl 150 mM) and surrounded by a water environment.

The starting point is the building of the initial file with coordinates information.

As default topology files, AMBER 11 is able to read a series of different formats (.pdb, .xyz, etc.). Typically, for protein structure, the format that is most widely used is the .pdb file. The extension .pdb means Protein Data Bank.

This database (http://www.rcsb.org/pdb/home/home.do) is a free online tool containing the .pdb files for X-ray/NMR studied protein and organic structures. The file .pdb is a simple coordinates file and contains the atom numbers and names, the name of the residues to whom these atoms belong and the coordinates of each atom within the structure. In our case the residues belonging to the class of aminoacids since the systems studied involve protein (Figure 2.3).

The first column refers to ATOM meaning known residue. There could be atoms within the structure with the statement HETATM also – this is used to identify which atoms belong to a residue that common force fields are able to recognize or not.

Typically, HETATM is used to identify a small inhibitor drug inside the .pdb file, that is not included in any force field and require a separate parameterization (Figure 2.4).

```
ATOM      1  N   ILE   542      36.745  36.126-105.381  1.00 27.48           N
ATOM      2  CA  ILE   542      35.942  37.308-105.798  1.00 26.74           C
ATOM      3  C   ILE   542      34.454  37.056-105.678  1.00 25.60           C
ATOM      4  O   ILE   542      34.044  35.981-105.244  1.00 26.17           O
ATOM      5  CB  ILE   542      36.315  37.610-107.271  1.00 20.00           C
ATOM      6  CG1 ILE   542      37.832  37.861-107.399  1.00 20.00           C
ATOM      7  CG2 ILE   542      35.550  38.863-107.743  1.00 20.00           C
ATOM      8  CD1 ILE   542      38.195  38.150-108.867  1.00 20.00           C
ATOM      9  N   LEU   543      33.616  38.033-106.054  1.00 24.84           N
ATOM     10  CA  LEU   543      32.160  37.864-105.997  1.00 23.89           C
ATOM     11  C   LEU   543      31.699  36.734-106.911  1.00 24.16           C
ATOM     12  O   LEU   543      32.187  36.594-108.033  1.00 24.62           O
ATOM     13  CB  LEU   543      31.619  39.221-106.452  1.00 24.32           C
ATOM     14  CG  LEU   543      30.729  39.872-105.370  1.00 20.00           C
ATOM     15  CD1 LEU   543      30.269  41.254-105.863  1.00 20.00           C
ATOM     16  CD2 LEU   543      29.490  39.001-105.080  1.00 20.00           C
ATOM     17  N   THR   544      30.763  35.927-106.424  1.00 23.03           N
ATOM     18  CA  THR   544      30.199  34.847-107.223  1.00 23.16           C
ATOM     19  CB  THR   544      30.603  33.464-106.664  1.00 24.62           C
ATOM     20  OG1 THR   544      30.291  33.399-105.267  1.00 27.94           O
ATOM     21  CG2 THR   544      32.094  33.225-106.863  1.00 25.92           C
ATOM     22  C   THR   544      28.681  34.942-107.268  1.00 21.75           C
ATOM     23  O   THR   544      28.068  35.655-106.469  1.00 18.67           O
ATOM     24  N   TYR   545      28.081  34.232-108.217  1.00 20.94           N
ATOM     25  CA  TYR   545      26.634  34.249-108.384  1.00 21.85           C
ATOM     26  CB  TYR   545      26.282  34.529-109.847  1.00 20.58           C
ATOM     27  CG  TYR   545      26.707  35.908-110.318  1.00 18.14           C
ATOM     28  CD1 TYR   545      25.942  37.029-110.018  1.00 16.67           C
ATOM     29  CD2 TYR   545      27.868  36.087-111.060  1.00 19.07           C
ATOM     30  CE1 TYR   545      26.322  38.292-110.443  1.00 15.88           C
ATOM     31  CE2 TYR   545      28.255  37.344-111.492  1.00 15.52           C
ATOM     32  CZ  TYR   545      27.477  38.443-111.180  1.00 16.71           C
ATOM     33  OH  TYR   545      27.845  39.697-111.616  1.00 16.14           O
ATOM     34  C   TYR   545      26.003  32.936-107.922  1.00 24.59           C
ATOM     35  O   TYR   545      24.785  32.765-107.980  1.00 25.14           O
```

Figure 2.3 *Example of a .pdb file. Columns represent (starting from the second left colum) the atom number, name, the residue and chain of belonging. The first three floating point numbers are x, y and z coordinates of each atom expressed in angstroms (Å). The next three columns are the occupancy, temperature factor, and the element name, respectively.*

```
HETATM    1  C6  STI   932      55.052  12.887 -91.790  1.00  0.00           C
HETATM    2  H6  STI   932      54.465  12.861 -92.708  1.00  0.00           H
HETATM    3  C1  STI   932      56.297  12.335 -91.771  1.00  0.00           C
HETATM    4  H1  STI   932      56.693  11.856 -92.667  1.00  0.00           H
HETATM    5  C2  STI   932      57.053  12.392 -90.593  1.00  0.00           C
HETATM    6  H2  STI   932      58.065  11.990 -90.627  1.00  0.00           H
HETATM    7  N3  STI   932      56.587  12.911 -89.462  1.00  0.00           N
HETATM    8  C4  STI   932      55.360  13.471 -89.428  1.00  0.00           C
HETATM    9  H4  STI   932      54.979  13.850 -88.478  1.00  0.00           H
HETATM   10  C5  STI   932      54.555  13.496 -90.561  1.00  0.00           C
HETATM   11  C7  STI   932      53.239  14.038 -90.471  1.00  0.00           C
HETATM   12  N8  STI   932      52.835  14.466 -89.253  1.00  0.00           N
HETATM   13  C12 STI   932      52.365  14.077 -91.601  1.00  0.00           C
HETATM   14 2H1  STI   932      52.679  13.803 -92.608  1.00  0.00           H
HETATM   15  C11 STI   932      51.098  14.484 -91.344  1.00  0.00           C
HETATM   16 1H1  STI   932      50.409  14.460 -92.188  1.00  0.00           H
HETATM   17  N10 STI   932      50.672  14.895 -90.152  1.00  0.00           N
HETATM   18  C9  STI   932      51.548  14.871 -89.122  1.00  0.00           C
HETATM   19  N13 STI   932      51.081  15.307 -87.888  1.00  0.00           N
HETATM   20 3H1  STI   932      51.683  15.107 -87.076  1.00  0.00           H
HETATM   21  C14 STI   932      49.889  16.065 -87.710  1.00  0.00           C
HETATM   22  C15 STI   932      48.743  15.389 -87.277  1.00  0.00           C
HETATM   23  H15 STI   932      48.773  14.313 -87.106  1.00  0.00           H
HETATM   24  C19 STI   932      49.877  17.470 -87.876  1.00  0.00           C
HETATM   25  C20 STI   932      51.065  18.221 -88.376  1.00  0.00           C
HETATM   26 1H20 STI   932      51.082  18.190 -89.465  1.00  0.00           H
HETATM   27 2H20 STI   932      51.007  19.257 -88.043  1.00  0.00           H
HETATM   28 3H20 STI   932      51.975  17.764 -87.987  1.00  0.00           H
HETATM   29  C18 STI   932      48.681  18.153 -87.657  1.00  0.00           C
HETATM   30  H18 STI   932      48.661  19.234 -87.797  1.00  0.00           H
HETATM   31  C17 STI   932      47.530  17.491 -87.268  1.00  0.00           C
HETATM   32  H17 STI   932      46.604  18.047 -87.122  1.00  0.00           H
HETATM   33  C16 STI   932      47.536  16.090 -87.074  1.00  0.00           C
HETATM   34  N21 STI   932      46.325  15.482 -86.663  1.00  0.00           N
HETATM   35 1H2  STI   932      45.458  16.016 -86.812  1.00  0.00           H
```

Figure 2.4 *Coordinates .pdb file for the imatinib drug (residue name: STI). The meaning of other columns is the same as above.*

As it appears evident, .pdb files do not include the Hydrogen atoms. In fact, AMBER software, automatically, adds H atoms once the .pdb file is loaded.

In this thesis work, different proteins and drugs were studied corresponding to a wide set of .pdb files.

Below we will proceed with a general procedure that could be used in all the cases studied; the single exception of protein complexes will be further described in their relative case study. Conceptually, to run a simulation it is necessary to obtain from the coordinate .pdb file two different files, whose extensions are .prmtop and .inpcrd. From a .pdb file, in fact, we can obtain only the positions of atoms and their name while no information about how they are connected and how they feel each other are present. The software reads the .pdb file and interprets it depending on the force field that was chosen for the correct system description. Thus, .prmtop and .inpcrd files are coupled and descript plenty the molecular system.

- inpcrd: coordinates file, the position of atoms that compose the molecule in space.
- .prmtop: parameters file, the values of the constants taken from the force field that was chosen for calculations (bond constants, torsions, atomic partial charges, etc.).

In the case of a protein/drug complex, a preliminary step was required before to import the .pdb complex inside the AMBER11 suite of programs and to obtain the respective .prmtop and .inpcrd files. This consists with the creation of missing parameters for the drug (imatinib, vismodegib, etc.). While importing the .pdb complex file, about these small molecules AMBER11 recognises only positions, but no information about the bond and non-bond constants are present. In fact, in importing a .pdb for both protein and drug molecules, it is obvious to obtain an error due to the impossibility to recognize the drug by the leap module – the module dedicated to the creation of complexes.

In order to obtain the missing parameters for drug molecules, it is necessary to resort to calculation methods that are able to obtain all the parameters for a molecule from its coordinates only.

There are the quantum calculations methods (ab initio) that, taking into account in the calculation the electrons, are able to give these parameters but these calculation methods are based on the explicit solving of the Schrödinger equation and are extremely heavy from the point of view of computational time.

The antechamber module of AMBER 11 allows different methodology for calculation of missing parameters (HF, semiempirical methods) and it presents a different request of CPU time for the resolution. In the studies presented in this thesis we used, as said, the AM1-BCC calculation method, able to produce for such small molecules data of similar precision of the

most advanced quantum mechanics methods with a notable reduction in the computational time.

The antechamber module accept as input file the .pdb file of the molecule analyzed and generates for this an input file .prepin. The .prepin file lists all the calculated parameters, readable by GAFF, for the drug. In the case antechamber fails to complete the calculations for all the constants, the parmchk utility available in AMBER 11 substitutes the missing bond and angle values with the one that it find to be more "similar" in the data base of the selected force field. In the case of the small molecules the .pdb file needs to contain the Hydrogen atoms.

Ones this step is accomplished, it is possible to load into leap the .pdb file of the whole complexes. As said, the ".ff" AMBER force fields are not "closed", but they accept the parameters that are calculated with antechamber. While the AMBER force fields use caps lock names (i.e. C3) for force field types of atoms, the "GAFF" force field type calculated with antechamber are expressed with small capitals (i.e. c3). The leap module is able to accept both force field types without making any confusion and to write a .prmtop file in which the atoms of the complex are identified by both indexes.

```
This is a remark line
molecule.res
STI    INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.522   111.1        .0      .00000
   4  C6    ca    M    3   2   1     1.540   111.208   180.000  -0.07182
   5  H6    ha    E    4   3   2     1.090    57.414    88.378   0.14336
   6  C1    ca    M    4   3   2     1.362   156.076    -1.971  -0.25375
   7  H1    ha    E    6   4   3     1.090   119.902    74.738   0.14586
   8  C2    ca    M    6   4   3     1.401   119.244  -105.255   0.40611
   9  H2    h4    E    8   6   4     1.089   117.390  -176.727   0.02388
  10  N3    nb    M    8   6   4     1.329   122.845     3.148  -0.67013
  11  C4    ca    M   10   8   6     1.349   120.166    -3.171   0.43686
  12  H4    h4    E   11  10   8     1.091   118.916   177.225   0.03877
  13  C5    cp    M   11  10   8     1.390   120.903     1.101  -0.27287
  14  C7    cp    M   13  11  10     1.426   119.327   176.975   0.53602
  15  C12   ca    B   14  13  11     1.429   121.657  -179.405  -0.41133
  16  H12   ha    E   15  14  13     1.090   123.199    -6.889   0.15422
  17  C11   ca    B   15  14  13     1.355   115.466   173.190   0.46179
  18  H11   h4    E   17  15  14     1.090   115.948  -176.487   0.02834
  19  N10   nb    E   17  15  14     1.331   124.179     3.437  -0.72149
  20  N8    nb    M   14  13  11     1.353   116.915    -1.957  -0.74896
  21  C9    ca    M   20  14  13     1.356   117.716  -174.142   0.87259
  22  N13   nh    M   21  20  14     1.390   119.910   179.366  -0.70389
  23  H13   hn    E   22  21  20     1.030   116.267   -13.062   0.43911
  24  C14   ca    M   22  21  20     1.424   124.017   162.993   0.14589
  25  C19   ca    B   24  22  21     1.415   121.408   -83.401  -0.12949
  26  C20   c3    3   25  24  22     1.492   122.175     6.242  -0.05485
  27  H201  hc    E   26  25  24     1.090   109.457    83.634   0.04586
  28  H202  hc    E   26  25  24     1.090   109.499  -156.361   0.04641
  29  H203  hc    E   26  25  24     1.090   109.521   -36.268   0.04303
  30  C18   ca    B   25  24  22     1.395   118.372  -178.593  -0.09610
  31  H18   ha    E   30  25  24     1.090   118.763  -179.134   0.13339
  32  C17   ca    S   30  25  24     1.384   121.551     0.914  -0.18384
  33  H17   ha    E   32  30  25     1.090   120.024  -178.841   0.12967
  34  C15   ca    M   24  22  21     1.399   117.852   100.784  -0.11418
  35  H15   ha    E   34  24  22     1.090   120.197    -1.246   0.18747
```

Figure 2.5 *Part of the .prepin file for the imatinib drug (residue name: STI). The third column report the force field types for each atom.*

```
remark goes here
MASS

BOND

ANGLE
nb-ca-cp    69.200      122.630    same as ca-ca-nb
h4-ca-cp    48.200      121.090    same as ca-ca-h4
ca-c3-n3    66.152      111.475    Calculated with empirical approach

DIHE
cp-cp-nb-ca    1    4.800       180.000          2.000      same as X -ca-nb-X
ca-cp-nb-ca    1    4.800       180.000          2.000      same as X -ca-nb-X

IMPROPER
ca-cp-ca-ha          1.1          180.0         2.0          General improper torsional angle (2 general atom types)
ca-ca-ca-ha          1.1          180.0         2.0          General improper torsional angle (2 general atom types)
ca-h4-ca-nb          1.1          180.0         2.0          Using default value
cp-h4-ca-nb          1.1          180.0         2.0          Using default value
ca-ca-cp-cp          1.1          180.0         2.0          Using default value
ca-cp-cp-nb          1.1          180.0         2.0          Using default value
nb-nb-ca-nh          1.1          180.0         2.0          Using default value
ca-ca-nh-hn          1.1          180.0         2.0          Using default value
ca-ca-ca-nh          1.1          180.0         2.0          Using default value
ca-ca-ca-n           1.1          180.0         2.0          Using default value
c -ca-n -hn          1.1          180.0         2.0          General improper torsional angle (2 general atom types)
ca-n -c -o          10.5          180.0         2.0          General improper torsional angle (2 general atom types)
c -ca-ca-ca          1.1          180.0         2.0          Using default value

NONBON
```

Figure 2.6 *Part of the .frcmod file for the imatinib drug (residue name: STI). Unknown parameters in .prepin files are here substituted with the closer available inside the force field.*

Ones .prepin and .frcmod files are loaded and accepted by the leap module, the residue STI (used here as an example – the procedure is the same for all the drug considered) will be listed between all the available inside the force field, as it would has ever been comprised in the known residues.

After this step, it was possible to load the .pdb file of the protein-inhibitor correctly. The leap screenshot report the total charge of the protein-inhibitor complex.

The total charge of the complex is -1. Since the imatinib does not carry any charge, this is due to the protein and, more precisely, to some of the residues that constitutes the proteic chain. In fact, as we already known, protein is composed by 20 different residues (Figure 2.7). These are aminoacids, organic compounds composed by an amine and a carboxylic acid functional group and characterized by a side-chain specific to each amino acid. The sidechain influence the total charge of the residue at physiological pH (7.4), in particular: Aspartic and Glutamic acids are characterized by a negative charge while Lysine and Arginine by a positive charge.

During the building of the complex it is important to consider the total charge of our protein to reach a neutral final complex.

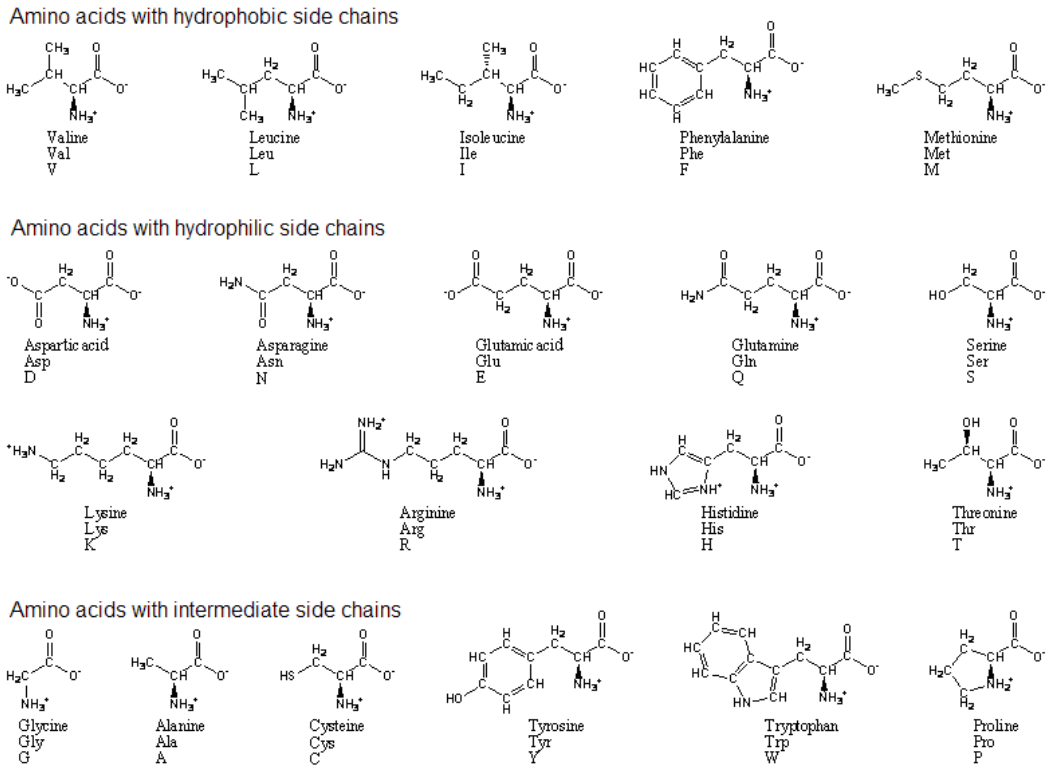


Figure 2.7 *Chemical representation of all the amino acids.*

As the complex is loaded in leap without giving errors, and before proceeding to arrange the environment with the same conditions of the human body, three couples of .prmtop and .inpcrd files were saved, for the protein and the drug taken separately and for the complex drug/protein. This operation is requested for the further energetical analysis.

The next two steps for the creation of the molecular system that will be simulate are:

- Solvation of the system: the complex solute is immersed in a water box of variable dimensions.
- Addition of counterions: a suitable number of Na+ and Cl- ions is added in the solution in order to guarantee the system neutrality and to reproduce the correct physiological ionic strength (150 mM NaCl).

About these two steps, few comments must be done. The simulations that will be presented in the next sections will be carried out under periodic boundary conditions (PBC) – the geometry of the water box have to be consistent with this condition that will be expressed as the replication in the three dimensions of the same box.

A realistic model of a solution requires a very large number of solvent molecules to be included along with the solute. Simply placing the solute in a box of solvent is not sufficient,

however, since while some solvent molecules will be at the boundary between solute and solvent and others will be within the bulk of the solvent a large number will be at the edge of the solvent and the surrounding vacuum. This is obviously not a realistic picture of a bulk fluid. In order to prevent the outer solvent molecules from "boiling" off into space, and to allow a relatively small number of solvent molecules to reproduce the properties of the bulk, periodic boundary conditions are employed. In this method the particles being simulated are enclosed in a box which is then replicated in all three dimensions to give a periodic array, a two-dimensional representation of which is shown below

During the simulation only one of the particles is represented, but the effects are reproduced over all the image particles with each particle not only interacting with the other particles but also with their images in neighboring boxes. Particles that leave one side of the box re-enter from the opposite side as their image. In this way the total number of particles in the central box remains constant.

Upon initial inspection such a method would appear to be very computationally intensive requiring the evaluation of an infinite number of interacting pairs. However, by employing a technique known as the Ewald sum, or its more modern equivalent the Particle Mesh Ewald (PME)[4] method it is possible to obtain the infinite electrostatics.

This involves dividing the calculation up between a real space component and a reciprocal space component. The VDW interactions are still needed though which means we cannot make the cut-off too small. For production calculations the ideal range is between 8 and 10 Å.

This from the calculation point of view is carried out with the condition that all the atoms and molecules that get off the original box from one side, come in again from the opposite. In this way it is guaranteed that the content of the periodic box is always the same.

Basically, there are two geometries for the water box that are used for this purpose: the tetrahedric and octahedric geometries. The dimensions can be chosen explicitly, or only the minimum distance between the solute atoms and the side of the box can be defined. In this case, the water buffer must be grater that the cutoff radius that is used for calculations.

In general, the geometry of the water box is chosen dependently on the geometry of the solute, and in order to reduce as much as possible the presence of water molecules in the system. In fact, if the solute presents a spherical symmetry, to choose an octahedric water box allows a consistent saving of "useless" atoms in the system. On the other hand, if the solute is strongly un-spherical and present a geometry that is not isotropic (i.e. DNA strands), to choose an octahedric box (very similar to a soccer ball) will create a final system full of water molecule

that are practically useless for the description of the system. In this case, the tetrahedic geometry is more indicated and guarantees more flexibility.

There are different molecular models to describe the water inside AMBER. We used the TIP3P water model.[5]

This simple water models use between three and five interaction sites and a rigid water geometry. The TIP3P model uses a total of three sites for the electrostatic interactions; the partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge located on the oxygen atom. The Van der Waals interaction between two water molecules is computed using a Lennard-Jones function, no Van der Waals interactions involving the hydrogen atoms are calculated.

This model for water has the peculiarity that the angle between the two hydrogen atoms is maintained rigid. Thus the water molecules are something "rigid" in the calculation – this is a simplification, but decreases notably the CPU time required to simulate the system (the greatest part of atoms in the system belong to water molecules of the solution).

Once we "immersed" the complex into the water, it is necessary to add the counterions in the system. In fact, ions can affect consistently the ability to move of the solute, influencing the final dynamic trajectories. The correct amount of ions to add to the system depends on the volume of the water box, in order to respect the 150 mM [NaCl] condition. The Na+ and Cl- ions can be added in the system with the Leap module – they are inserted in the system following the electrostatic potential surface of the solute. If during the addition of counterions positional conflict (superposition) with water molecules, constituting the box, are verified, the entire water molecule is substitute with the ion that is placed in the water molecule centre of mass. Since the number of water molecules in the box is rather high, usually almost all the counterions substitute a water molecule. This create a non-uniformity in the box, since one single atom (Na+ or Cl-) is replacing three atoms (H-O-H) – there are single points of lower density. That is why, for instance, during the first period of NPT dynamic (which allows the volume of the water box to change) the box reduces its dimensions in order to adjust the density. In conclusion, in the calculation of how many ions should be added to the system, the fact that each ion will replace an entire water molecule and that the final equilibrated volume will be smaller than the one of the initial box constructed by leap must be taken into account.

Once the system is completed it is possible to save the .prmtop and .inpcrd files for the whole system – these are files that characterize the system that is going to be simulated.

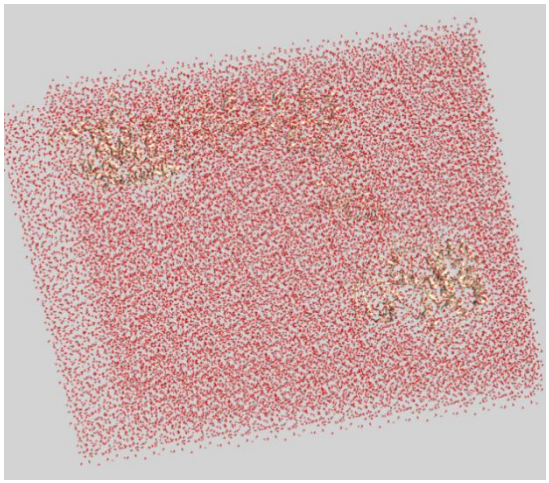


Figure 2.8 *The β catenin/GSK3 receptor complex in theTIP3P box water.*

2.4 Creation of mutated complexes

The section above described the wild type complex, so in absence of mutations. However, during the carcinogenesis process, it is very frequent to find mutations in important proteins that could be targeted for the therapy or could lead to drug resistance, usually called “secondary” mutations. Changes in the DNA sequence can occur at the level of the chromosome, in which large segments of chromosomes are altered. In this case, fragments of chromosomes can be deleted, duplicated, inverted, translocated to different chromosomes, or otherwise rearranged, resulting in changes such as modification of gene dosage, the complete absence of genes, or the alteration of gene sequence. The type of variation that occurs when entire areas of chromosomes are duplicated or lost, called copy number variation (CNV), has especially important implications for human disease and evolution.

While a mutation is defined as any alteration in the DNA sequence, biologists use the term "single nucleotide polymorphism" (SNP) to refer to a single base pair alteration that is common in the population. Specifically, a polymorphism is any genetic location at which at least two different sequences are found, with each sequence present in at least 1% of the population. Note that the term "polymorphism" is generally used to refer to a normal variation, or one that does not directly cause disease. Moreover, the cutoff of at least 1% prevalence for a variation to be classified as a polymorphism is somewhat arbitrary; if the frequency is lower than this, the allele is typically regarded as a mutation.

SNPs are important as markers, or signposts, for scientists to use when they look at populations of organisms in an attempt to find genetic changes that predispose individuals to certain traits, including disease. On average, SNPs are found every 1,000–2,000 nucleotides in the human genome, and scientists participating in the International HapMap Consortium have

mapped millions of these alterations (International Human Genome Sequencing Consortium, 2001).

1.Point mutation:

- Substitution: one base is incorrectly added during replication and replaces the pair in the corresponding position on the complementary strand.
- Insertion: one or more extra nucleotides are inserted into replicating DNA, often resulting in a frameshift. Deletion: one or more nucleotides is “skipped” during replication or otherwise excised, often resulting in a frameshift.

2. Chromosomal mutation:

- Inversion: one region of a chromosome is flipped and reinserted.
- Deletion: a region of a chromosome is lost, resulting in the absence of all the genes in that area.
- Duplication: a region of a chromosome is repeated, resulting in an increase in dosage from the genes in that region.
- Translocation: a region from one chromosome is aberrantly attached to another chromosome.

In this section, we proceed building the “mutated” complex in order to compare these results between wild type and mutates complex and to analyze the effects in the protein/protein or protein/drug interaction occurred following the mutation. We will briefly describe how we introduced point mutations or amino acid deletions inside the protein chain.

All the computational mutagenesis procedure has been carried out exploited the Rotamers module of Chimera (UCSF Chimera package, University of California, San Francisco).

Each mutated complex was then equilibrated in explicit water solvent to allow for relaxation of the structure following the stress from mutation. Each system was subjected to a minimization of 1000 steps, and then the solvent molecules were equilibrated for 100 ps at 310K. Finally, the entire system is equilibrated for 2 ns at 310K. Then the complex is subjected to MD simulation.

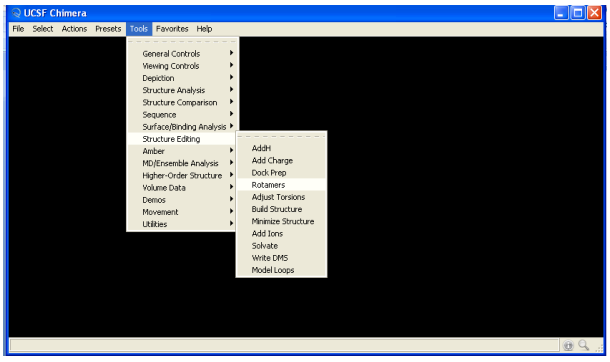


Figure 2.9 *Panel of Chimera visualization program for generation of rotamers.*

For the aminoacids that have several possible rotamers (for instance acid aspartic and glutamic), each rotamer structure is generated and minimize separately. We proceed with a systematic conformational search for 100 different conformations for the mutated residue. For each conformation obtained, a check of the steric bump is planned to avoid sterical clash between all the residue in the protein and the mutant. In particular, the conformation of the mutated residue is not taken in account if the new residue is closer than 1 Å to any aminoacid in the protein. Each surviving complex is minimized using the Sander module of Amber11 with steepest descent algorithm for the first 200 steps, and then subsequently with a conjugate gradient approach. A cutoff of 10 Å and a distance-dependent dielectric are applied. Then we perform a minimization consisting on different steps. At first, the mutated residue is minimized for maximum 1000 steps and all the system is fixed; then, the hydrogen atom are minimized for maximum 1000 steps and at the end all the atoms are minimized for 2000 steps.
The structure containing the rotamer characterized by the lowest final energy is selected to proceed with MD assays.

Looking at the final structure, while a point mutation provokes only adjustments in the secondary structure of the protein, a consistent deletion can provoke drastic distortions and final configurations that are considerably distant from the "wild type" protein. Thus, it is perfectly clear that the huger the deletion, the more consistent the simplification and the risk that the final mutated protein conformation in the model could be completely different from the real one.

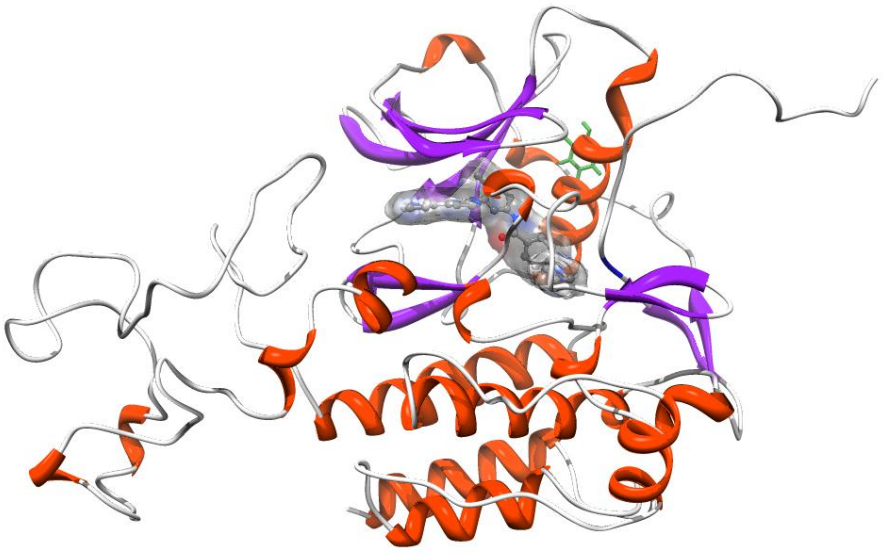


Figure 2.10 *Examples of point mutation* in the complex between c-KIT receptor and Imatinib. Highlighted in blue the Trp557 mutated into Gly, in green Tyr578.

Once the mutation has been inserted, the same steps of solvation and addition of counterions described in the previous section followed – a .prmtop and an .inpcrd file were obtained also for the mutated case, ready to be simulated.
With the two files containing parameters (.prmtop) and starting configurations (.inpcrd) we had everything is needed as a starting point to run a molecular simulation.

2.5 Molecular simulation

Once we have our input files we can progress to the next step of the simulation process.
Since the default geometry that we obtained, building our structure, might not correspond to the actual minima in the selected force field and could result in conflict and overlap with atom in other residues, it is always a good idea to minimize the structure before starting molecular dynamics.
Sander.MPI (Simulated Annealing with NMRDerived Energy Restraints) and pmemd modules of AMBER were used to conduct a short minimization run. Both modules are parallel calculation tools, but while Sander.MPI allows many facilities and tools for energy minimization, molecular dynamic, restraints, etc., pmemd presents only few options. In fact, with pmemd, only periodic boundary condition simulations without restraints applied on the molecule are allowed, but a better scaling during the parallelization is possible.
The system is in the configuration described by the .inpcrd file.
All the cases studied in this thesis followed this procedure:

- Minimization: a preliminary phase, only necessary to remove bad contacts between atoms after the solvation, obtaining for each step .out file that contains energetical information and .rst file to give the input for the next step.
- Constrained dynamic: a phase in which the solute is constrained in its starting position, during this phase the simulation temperature is reached, and start the solution density optimization.
- Equilibrium dynamic: without any restraint – it is the phase in which the system reach the equilibrium (NPT, temperature and pressure constant).
- Production dynamic: without restraints in NVT condition (temperature and volume constant), the trajectories of this phase will be further processed and analyzed.

In detail, the molecular systems were relaxed for 500 steps using Sander.MPI applying a restraint on all the atoms of the solute with elastic constant of 2 N Å-1 in order to let move all

the atoms belonging to the solution by first. This was followed by other 500 steps of conjugate-gradient minimization identical to the first, but in absence of restraints on the solute was run to remove bad contacts between atoms of the complex. After the minimization part, the system was gradually heated to a temperature of 300 K in intervals of 50 ps of NVT dynamic, with a time step of 1 femtosecond (fs). Both these two first dynamic steps were conducted with the solute maintained restrained on the starting position applying an elastic force with a 2 N Å-1 elastic constant. The density of the system was subsequently equilibrated via MD runs in the isothermal- isobaric ensemble (NPT), with isotropic position scaling and a pressure relaxation time of 1.0 ps, for 50 ps with a time step of 1 fs. All restraints on the protein atoms were then removed, and each system was further equilibrated using NPT MD runs at 300K, with a pressure relaxation time of 2 ps.

More precisely, this phase can take from few to tens of nanoseconds, depending on the conformation of the solute. We preferred to use in the previous steps 1 fs time step in order to avoid too large deviations due to the rising of temperature and to the high fluctuations of the molecules belonging to the solvent.

This Particle Mesh Ewald (PME) method was applied to treat long-range electrostatic interactions. Briefly, since the electrostatic interaction can be divided in short- and long- range effects.

The SHAKE[6] method was also used to constrain all of the covalently bound hydrogen bonds, while long-range nonbonded Van der Waals interactions were truncated by using dual cutoffs of 6 and 12 Å.

The equilibrium phase, thus, has not a pre-defined length, it depends on how long the reaching of the equilibrium takes to the complex. During the simulation, some values could be checked to verify the status of the equilibrium of the system. The quantities that usually are plotted to the time to check the status of the system are temperature (T), pression (P), Volume (V), Epot, RMSD (root mean square deviation), in particular:

- The potential energy (Epot): strictly correlated to the total energy of the system Etot. The energy of the system usually converges almost immediately (i.e. until the first picoseconds of the equilibrium phase).

- The root mean square deviation (RMSD) of the atoms of the backbone

While the first parameter tends reach an equilibrium immediately (in correspondence of the stable reaching of the 300K temperature), the second takes a longer.

If the equilibrium is reached, both the curves of potential energy and RMSD should be horizontal. On the other hand, if the RMSD has not reached the equilibrium (it is growing) despite a constancy of the energetic of the system, it means that the complex, in order to maintain the constancy of the energy, is vibrating stronger and stronger at each dynamic step - thus, while the energetic equilibrium is reached, the vibrational one has not.

As reported by Figure 2.11, the energies converge sooner than RMSD – the real equilibrium is reached after ~6 ns in this case, thus this is the starting point from which the data collection can be considered as reliable. In conclusion, the equilibrium phase has not a pre-defined length, but it depends on the time required to the solute to reach the equilibrium. It is also worth noting that a good oscillation for RMSD is between 1 and 3 Å around the equilibrium position, as shown from the graph in the terminal part of the RMSD curve (last 2-4 ns) the fluctuation of atoms around their equilibrium position is really below these values.

The equilibration phase was followed by a data production run consisting of 4 ns of MD simulation in the canonical (NVT) ensemble. Only the last 2 ns of each equilibrated MD trajectory were considered for statistical data collection. In all these simulations step a cutoff was used. The role of the cutoff is to exclude from an explicit non-bond calculation all the atoms that are not inside a sphere defined by the cutoff radius (R cutoff). This is done for shake of CPU time saving. Since the cutoff value can vary depending on the single systems, the individual parameters and details changed for each case studied.

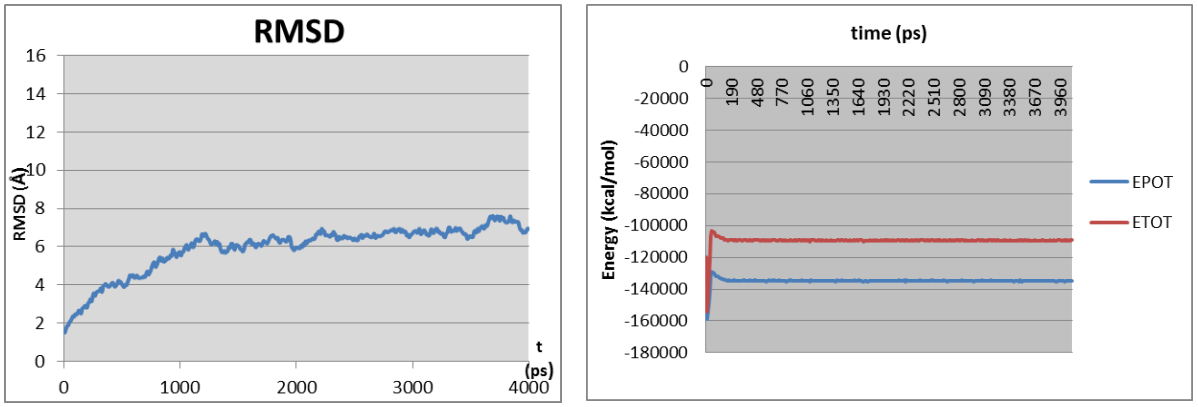


Figure 2.11 *Examples of Epot, Etot and RMSD profiles of a protein-drug complex (KIT "wild type"-imatinib). The time scale is expressed in nanoseconds, the energies in kcal mol^{-1} and the RMSD in angstroms (Å). The last 2 ns of these profiles belong to the production phase of the dynamic.*

However, since all the atoms do not enter in the calculation explicitly, a corrective term (an external constant) must be inserted, in order to correct the error given by the exclusion of part of the atoms from calculation.

In conclusion the total length of each simulation is around 10 ns (depending on each system, but the complexes presented in this thesis are rather similar).

The single simulation details (number of atoms of the system, number of water molecules, cutoff used, etc.) concerning each single case studied will be reported in the each paragraph dedicated to each single system simulated.

Once the simulation is finished we can proceed analyzing the results; in particular we can look at the movie, through a visualization program. The files .mdcrd (trajectory) and .prmtop are used for this purpose. Have a look at the trajectory could help to understand conformational changes occur during the simulation and particular behavior of our system. However it is also interesting to compute the binding free energy for the system studied, applying MMPBSA methodology. Let see more in details, the MMPBSA approach in the next section.

2.6 Affinity energy analysis (MM-PBSA)

Once the simulation, described in the previous section, is finished, the output files could be processed by the application of the MM/PBSA approach[7] to estimate the binding free energies. At the end of the simulations the output files obtained are below:

- .out: the output file. It reports all detail about the energetic of the system at each step of the simulation. It also list all the parameters listed in the input files.
- .rst: file restart. It is equivalent to the .inpcrd – it is a coordinate file that is written after a given time interval. In dynamic regime, differently from the .inpcrd file, it reports also the velocities of the atoms.
- .mdcrd: the trajectory file. These files have huge dimensions and report the “history” of the simulation for each atom of the system.

These last trajectory files are used to perform a MM-PBSA analysis on the complex in order to evaluate, at each step of the simulation, the free energy for each molecular species, complex, ligand and receptor.

The MM-PBSA methodology (Molecular Mechanic – Poisson Boltzmann Solvent Accessible area) is a tool, implemented in AMBER, that allow to calculate the affinity energy for the association of two molecules (protein/protein and protein/drug) in solution.

The main aim of the MM-PBSA method is to calculate the free energy difference between two states which most often represent the bound and unbound state of two solvated molecules or to compare the free energy of two different solvated conformations of the same molecule.

Ideally, the free energy of binding for the association of a ligand (L: drug) and a receptor (R: protein) can be expressed by the following representation:

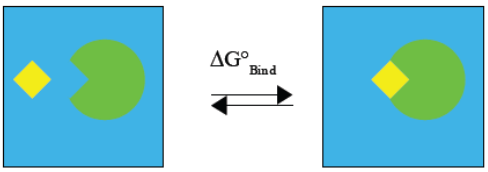


Figure 2.12 *The two states "bound" and "unbound" of ligand (yellow) and receptor (green) in solution (blue).Adapted*[8]

The ability to accurately calculate G, the average free energy for a given macromolecular system in various different conformations structures, adds a very important methodology to the available tools. This has been possible before with free energy perturbation, but only for small systems and very limited conformational or topological changes. By using a continuum model, we are implicitly integrating out all the solvent coordinates and simplifying the problem. Also, by calculating the absolute free energy directly between the two "end points" instead of calculating the relative free energy along a mapping coordinate, we are avoiding computations on less interesting intermediate states. Nonetheless, it is clear that applications of the equation will have intrinsically much larger errors than free energy perturbation/thermodynamic integration calculations. However, despite the larger uncertainties, we can most of the times, calculate ΔG in respectable agreement with experiment. It could be used to study: relative free energies of macromolecules, protein-ligand interactions and its deconvolution per residues energies, protein-protein interactions.

In the simulation of these solvated states the majority of the energy contributions would come from solvent-solvent interactions and the fluctuations in total energy would be an order of magnitude larger than binding energy of the two molecules of interest. Thus the calculation would take an enormous amount of CPU time to converge. In order to simplify, the MM-PBSA method is based on the following thermodynamic cycle:

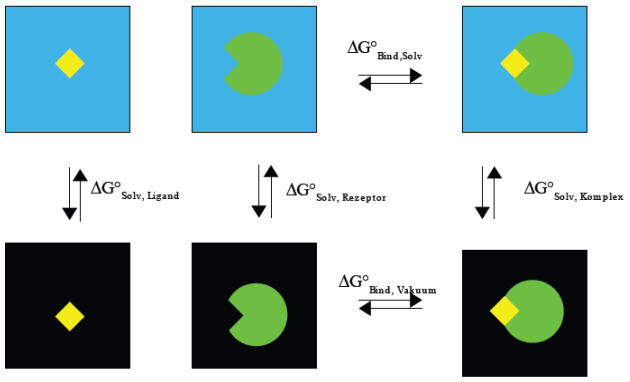


Figure 2.13 *Thermodynamic cycle to exploit $\Delta G_{bind,\ vacuum}$ to solve $\Delta G_{bind,\ solv}$. Adapted*[8]

In the MM-PBSA approach the different contributions to the binding free energy above are calculated in various ways: solvation free energies are calculated by either solving the linearized Poisson Boltzman equation[9] for each of the three states (this gives the electrostatic contribution to the solvation free energy) and adding an empirical term for hydrophobic contributions.

From the scheme we derived:

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta G_{solv,complex-}\left(\Delta G_{solv,ligand} + \Delta G_{solv,receptor}\right)$$

That could simplify in:

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta\, G_{solv}$$

Whereas ΔG_{solv} is coming from the solvatation of the complex:

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{NP}$$

ΔG_{PB} is estimated using DelPhi[10] which solves the Poisson-Boltzmann equations numerically.

The non polar solvation contribution was determined using the following relationship[9]:

$$\Delta G_{NP} = \gamma(SASA) + \beta$$

In which γ = 0.00542 kcal/(mol Å^2), β = 0.92 kcal/mol and SA is the molecular surface area estimated by means of the MSMS software[11].

While $\Delta G_{bind,vacuum}$ is obtained by calculating the average interaction energy between receptor and ligand and taking the entropy change upon binding into account:

$$\Delta G_{bind,vacuum} = \Delta E_{MM} - T\Delta S_{bind}$$

Where $T\Delta S_{bind}$ is the entropy contribution and it is estimated by performing a normal mode analysis with the Nmode of AMBER11. The normal mode analysis calculations are computationally expensive and are the one that introduce in the results the larger margin of error and the higher uncertainty in the result.

The molecular mechanics energy ΔEMM was calculated as the sum of the Van der Waals and electrostatic interactions:

$$\Delta E_{MM} = E_{complex} - (E_{receptor} + E_{ligand})$$

ΔEMM is a measure for the tendency of the ligand and complex to bind each other in a stable complex, rather than keeping two isolate conformations. The more negative the ΔE, the more

stable the complex, in other words in certain condition the complex has lower energy than the two of protein and drug taken separately and, thus, its formation is favored.

So it is possible to rewritten the binding energy as:

$$\Delta G_{bind,solv} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S$$

Or

$$\Delta G_{bind,solv} = \Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

The methodology that is adopted in this thesis for the calculation of each energetic term is the "single trajectory" technique. In fact, starting from the assumption that the conformational change of protein and drug will not be heavily affected due to the binding, the single conformations of protein, drug and complex for which the energies will be calculated at each steps will be taken from the single trajectory of the system we simulated. Practically, from the single dynamic trajectory of all the periodic box, we define "what" is complex and "what" is ligand and receptor respectively – this is what is needed for the calculation of each energetic component for the calculation of the free energy of binding ΔGbind.

However, in such simplified approach, the presence of water and ions is no more taken into account into the energetic calculations but only in the effect on the trajectory of these molecules along the dynamic.

In practice, to minimize the effects due to different conformations adopted by individual snapshots, and due to the high computational demand of this approach, we averaged the estimation of entropy over MD 40 snapshots for each molecular complex that were evenly extracted from the last 2 ns of each corresponding trajectory.

So in summary, this is a post-processing method in which representative snapshots from an ensemble of conformations are used to calculate the free energy change between two states (typically a bound and free state of a receptor and ligand). Free energy differences are calculated by combining the so-called gas phase energy contributions that are independent of the chosen solvent model as well as solvation free energy components (both polar and non-polar) calculated from an implicit solvent model for each species. Entropy contributions to the total free energy may be added as a further refinement. The entropy calculations can be done successively.

MM/PBSA typically employs the approximation that the configurational space explored by the systems are very similar between the bound and unbound states, so every snapshot for each species is extracted from the same trajectory file, although MMPBSA.py will accept separate trajectory files for each species. Furthermore, explicit solvent and ions are stripped from the

trajectory file to hasten convergence by preventing solvent-solvent interactions from dominating the energy terms.

With AMBER software package there are two different approaches to calculate the MMPBSA (both applied in this thesis work). The first uses the mm_pbsa.pl script of AMBER for the complex, and the protein and drug unbound. After this, the same mm_pbsa.pl script was used in order to compute the free energy of binding of each complex of interest. The second one applied the MMPBSA.py and it doesn't need to extract snapshot with the .pl script.

The effectiveness of a certain inhibitor drug is defined by the half maximal inhibitory concentration (IC50) that is defined as the amount of drug to put into the solution in order to inhibit the activity of the specific biological target of the 50%.

The correspondence between IC50 and the free energy of binding is defined as follows:

$$\Delta G = -RT \ln (IC50)$$

Thus, until this point, we were able to produce results consistent with experimental measurement, indicating that the procedures and the molecular models used for the cases faced in this thesis are reliable.

2.7 Per residue decomposition analysis

The interactions between the drug and each separate aminoacids may be evaluated applying a further energetical analysis, in previous section we explained the methodology used for the calculation of the free energy of binding ΔG_{bind}.

In particular this methodology allows us to identify the role of a point mutation in the total protein systems, evaluating energetical contribution of the surrounding residue.

The concept is rather simple; it is possible to split the total interaction energy that exists between ligand and receptor on all the residues constituting the protein:

$$\Delta E_{MM,j} = \Delta E_{vdW,i} + \Delta E_{ele,i}$$

By calculating each ΔE_j it is possible to evaluate the role of each residue in the stabilization of a protein-drug complex, the more negative the ΔE_j, the stronger is the favorable action of the residue in stabilizing the complex. In details, the gas-phase in vacuo energies (E_{gas}) for each residue is composed of electrostatic and van der Waals interaction contributions (E_{ele} and E_{vdW}, respectively).

The result is an enthalpy division between all the residues constituting the protein, and gives precious indications in which are the residues that are crucial in binding – the residue that most interact favorably with the inhibitor molecule.

The single energetical contribute per residue could help to understand the particular “domino effect” that rearrange the all structure or the identification of particular residue for the binding sites.

The $\Delta\Delta E_{i,tot}$ value represents the effect of the mutation on the affinity between a precise residue and the drug. In fact, if $\Delta\Delta E_{i,tot} < 0$ it indicates a negative effect of the mutation on the attraction ability of the single residue toward the drug (the $\Delta E_{i,tot}$) mutated is less negative (favorable) than ΔEi,tot-“wild type”, meaning that this precise residue tends to stabilize more the binding in the naive conformation of the protein than in presence of a given mutation along the amino acidic chain. It is worth noting that of course the residues that will report the most evident getting worse due to the presence of the mutation will be the ones in close proximity to the mutation itself.

This energetic analysis can be rather useful because it is able to provide unique details in the binding between protein and inhibitor drugs. Such details are impossible to obtain with other experimental techniques, and evidence this approach and the modelling in general as an extremely powerful instrument for the exploration and the analysis of the potentiality of different small inhibitor molecules.

The new method separates nonpolar contribution into two terms: the attractive (dispersion) and repulsive (cavity) interactions. Doing so significantly improves the correlation between the cavity free energies and solvent accessible surface areas for branched and cyclic organic molecules.

This is in contrast to the commonly used strategy that correlates total nonpolar solvation energies with solvent accessible surface areas, which only correlates well for linear aliphatic molecules. In the new method, the attractive interaction is computed by a numerical integration over the solvent accessible surface area that accounts for solute solvent attractive interactions with an infinite cutoff.

This analysis was carried out using the MM/GBSA approach,[12,13] and was based on the same snapshots used in the binding free energy calculation.

The general procedure is to edit the decomp.in file, you just need to prepare prmtop files for receptor, ligand, and complex using LEaP; if you are just doing "stability" calculations, only one of the prmtop files is required.

The output files are labeled ".out", and the most useful summaries are in the "statistics.out" files.

2.8 Pharmacophore modeling

The pharmacophore is the group of properties that form a vital part of a drug. A pharmacophore may be defined as the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger its biological response. Pharmacophore mapping is one of the major elements of drug design in the absence of structural data of the target receptor. The tool initially applied to discovery of lead molecules now extends to lead optimization. Pharmacophores can be used as queries for retrieving potential leads from structural databases (lead discovery), for designing molecules with specific desired attributes (lead optimization), and for assessing similarity and diversity of molecules using pharmacophore fingerprints. It can also be used to align molecules based on the 3D arrangement of chemical features or to develop predictive 3D QSAR models.

The model structures of all compounds are built using the 2D-3D sketcher of Discovery Studio Catalyst (DS v. 2.5, Accelrys, San Diego, CA, USA). High quality conformational models are crucial for the development of predictive pharmacophore models. Accordingly, in this thesis we employed an ad hoc procedure to derive molecular conformations, instead of using those automatically generated by DS Catalyst, for a better quality in covering the low-energy conformational space. Each molecular structure is subjected to energy minimization using the generalized CHARMM force field until the gradient dropped below 0.05. The minimized structures are used as the starting point for subsequent conformational searches. A 10000-step Monte Carlo torsional sampling conformational search is conducted for each compound. Unique low energy conformations within 20 kcal/mol of the corresponding global energy minimum are collected for each molecule. A conformation is considered unique only when the maximum displacement of at least one heavy atom is greater than 0.5 Å. A maximum of 250 unique conformations are recovered for each compound. The classical conformational search is also carried out using the Poling algorithm and the CHARMM force field as implemented in DS Catalyst program for comparison. The “best quality” generation option is adopted to select representative conformers over a 0 – 20 kcal/mol interval above the computed global energy minimum in the conformational space, and again the number of conformers generated for each compound is limited to a maximum of 250. The main drawback of this technique, however, is that it takes considerable longer to generate the relevant conformational models. Nonetheless, as the spirit of the work is the generation of a predictive 3D pharmacophore model for these classes of compounds, we considered it worthwhile to use more accurate conformational models.

Based on the conformations for each compound, the DS Catalyst Hypothesis module was used to generate three-dimensional pharmacophore models. During hypotheses generation, the software attempts to minimize a cost function containing two main terms: the first penalizes the deviation between the estimated affinities of the training set molecules and their experimental values, whilst the second penalizes the complexity of the hypothesis. The uncertainty factor for each compound represents the ratio range of uncertainty in the affinity value based on the expected statistical irregularity of biological data collection. Uncertainty influences the first step – also called the constructive phase – of the hypothesis generating process. An uncertainty value of 3.0 is the default factor, but sometimes if the experimental affinities of compounds barely span the required four orders of magnitude an uncertainty of 1.1 is preferred.

Different functional groups characterizing compounds allow to individuate defined features. For example in the case studied in this thesis work for σ_1 receptor the identified features correspond to hydrophobic aromatic (HYAr), aliphatic (HYAl) and generic hydrophobic (HY) features, hydrogen bond acceptors (HBA), and positive ionizable (PI).

The overall costs of a model consist of three cost components, namely, the weight cost, the error cost and the configuration cost. The weight component is a value that increases in a Gaussian form as this function weights in a model deviate from the ideal value of two. The error cost represents the difference between estimated and measured activities of the training set. The configuration cost quantifies the entropy of the hypothesis space.

In addition the following three cost values are calculated during the generation of pharmacophore models: the fixed cost, the total cost and the null cost. The fixed cost is the lowest possible cost representing a hypothetical simplest model that fits all data perfectly. Another cost parameter, the null cost, represents the maximum cost of a pharmacophore with no features and estimates activity to be the average of activity data of training set molecules. The null cost value is equal to the maximum occurring error cost. For every pharmacophore generation ten total cost values and each of fixed cost and null cost values are calculated by the pharmacophore generation protocol in the unit of bits. For a meaningful pharmacophore model, the fixed cost should be lower and the null cost should be higher and the total cost value should be closer to the fixed cost and away from the null cost value.

Three validation procedures are, usually, used to determine the statistical relevance and the validity of the proposed 3D pharmacophore models: the test set prediction method, the CatScramble method, and the leave-one-out procedure. A larger difference between the fixed and null costs than that between the fixed and total costs signifies the quality of a pharmacophore model. In this work, the first procedure consisted of the collection of further,

different compounds into a test set, and in performing a regression analysis by mapping the test set molecules onto the best pharmacophore hypothesis. The high correlation coefficients obtained using the test set compounds revealed the good correlation between the actual and estimated affinities and, hence, the predictive validity of the corresponding 3D hypothesis. The CatScramble validation procedure is based on Fisher's randomization test. The goal of this type of validation is to check whether there is a strong correlation between the chemical structures and the biological activity. This is done by randomizing the affinity data associated with the training set compounds, generating pharmacophore hypotheses using the same features and parameters employed to develop the original pharmacophore model.

The statistical significance is calculated according to the following formula:

$$\text{significance} = 100 \times [1 - (1 + x / y)]$$

where x is the total number of hypotheses having a total cost lower than the original (best) hypothesis, and y is the total number of DS Catalyst Hypothesis runs (initial + random runs). Thus, 49 random spreadsheets (i.e., 49 DS Catalyst Hypothesis) have to be generated to obtain a 98% confidence level. Should any randomized data set result in the generation of a 3D pharmacophore with similar or even better cost values, root-mean-square deviations, and correlation coefficients, then it is likely that the original hypothesis does reflect a chance correlation.

Finally, the leave-one-out test checks if the correlation between experimental and computed affinities is heavily dependent on one particular molecule of the training set by re-computing the pharmacophore model with the exclusion of one molecule at a time. For each run, the hypothesis characterized by the lowest total cost was employed to predict the affinity of the excluded compound and to estimate the new correlation coefficient.

The data used in the particular case analyzed in this thesis work will be discussed in the appropriate section of Results & Discussion.

2.9 Homology modeling technique

One of the goals of protein modeling is to predict a structure from its sequence with an accuracy that is comparable to the best results achieved experimentally. This could help in different contexts when experimental techniques fail (i.g. the protein is too large for NMR analysis): structure-based drug design, analysis of protein function, interactions, antigenic

behavior, and rational design of proteins with increased stability or novel functions. In addition, protein modeling is the only way to obtain structural information if experimental techniques fail.

The 3D protein structure is entirely determined by its aminoacidic sequence; however when proteins are in optimal temperature and solvent conditions, as into the cell, they spontaneously fold to reach the native state structure. In some cases it is necessary the help of some “chaperonine” to fold in the correct way, but they just act to catalyse the folding and not interfere with the direction.

Since the sequence contains the sufficient data to specify the 3D structure of protein, it would be possible to develop an alghorithm to predict the structure of the protein using the only aminoacidic sequence. Really the sequence could only determine 60-70% of the total secondary structure and you need a lot of other information to complete the secondary and tertiary structures.

During evolution, the structure is more stable and changes much slower than the associated sequence, so that similar sequences adopt practically identical structures, and distantly related sequences still fold into similar structures. Thanks to the exponential growth of the Protein Data Bank (PDB), Rost (1999) could derive a precise limit for this rule. As long as the length of two sequences and the percentage of identical residues fall in the region marked as “safe,” the two sequences are practically guaranteed to adopt a similar structure.

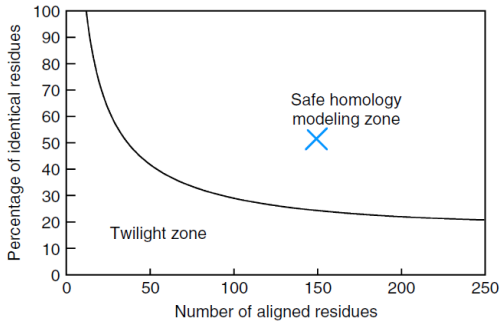


Figure 2.14 *Different zones of sequence alignments (twilight and safe). Sequence are guaranteed to fold into the same structure if their sequence identity fall into the region marked as safe (for example gray cross).*

It is possible to distinguish in three different methods to predict the 3D structure: homology modeling, fold recognition, and ab initio prediction.

The first approach can be used when there is a high degree of similarity (more than 30%) between the protein of interest and a known structure template. In this case, we can use the template structure as reference to build the model of study.

In the case of no homology with other known protein, an alternative method consists on Threading, based on fold recognition. The aim is identified the most compatible fold with the aminoacidic sequence of the target protein. It is a necessary requirement that the fold of the unknown protein already exists in the structure protein data bank.

In practice, homology model is a multistep process as summarized below:

1. Template recognition and initial alignment
2. Alignment correction
3. Backbone generation
4. Loop modeling
5. Side chain modeling
6. Model optimization via Molecular Dynamics (MD) simulation
7. Model validation

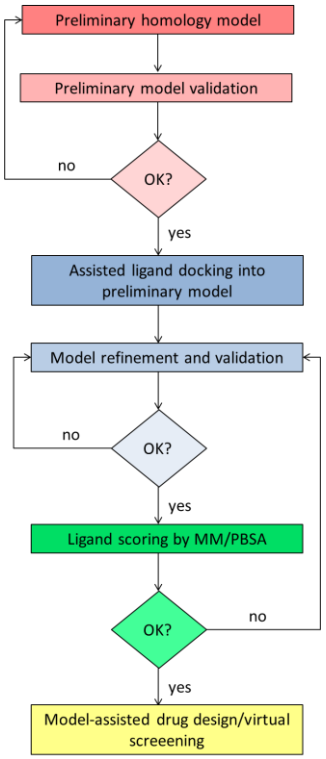


Figure 2.15 *Schematic representation of homology modeling procedure.*

The steps to homology model

The first step consists on the downloading of the primary sequence of the target protein from UniProt database. Then we resort to the application of a comparative PSI-Blast search through different database: MODWEB (http://modbase.compbio.ucsf.edu/scgi/modweb.cgi) , ExPASy Proteomic Server (http://expasy.org) and NCBI website (www.ncbi.nlm.nih.gov).

To identify hits, the programs work comparing the query sequence to all the sequence belonging to the known structures in the protein data bank using two different matrices: a residue exchange matrix and an alignment one.

Having identify one or more possible modeling templates, it is now necessary to evaluate the possibility of improve our first alignment applying a multiple sequence alignment. This could implicitly contains information about the hardest possibility to have an exchange on the hydrophobic core than on the surface. If there is an exchange between hydrophobics aminoacids the residue is probably buried. Considering these aspects during the alignment you could derive position specific scoring matrices, the profiles.

Other information could be very useful to proceed toward a realistic model. In particular, for the identificatiuon of transmembrane zone or the overall prediction of the secondary structure, it is possible to, comparatively, employ TMPRED, DISOPRED, PredictProtein, HMMTOP and Secondary Structure Prediction tool of DS (Discovery Studio).

At this point the predictions and all the alignments are ready, the actual model building can start.

Experimentally determined proteins are sources of errors, for instance poor electron density in the X-ray diffraction map. So the best option is to combine the good parts of both templates in one model.

The crystal structures of the template proteins identified are used as template for building the 3D structure. Each initial template-based structure is obtained using MODWEB, checked by Build and Homology Models of DS and compared to the corresponding template through Structure Comparison-Match Maker feature of Chimera.

Loop modelling

In the majority of cases, the alignment between model and templates sequence contains gap.

Two different approaches could be applied: a knowledge based (involving search in PDB for known loops with similar endpoints) and an energy based (ab initio fold prediction, judging quality by energy function).

At least for short loops, these two methods have a reasonable chance of predicting a loop conformation that superimposes well on the true structure.

Some extended portions of the protein, usually the intrinsically disordered as N-ter and C-ter parts, could sometimes show no homology with any other protein in databases; in these particular cases the procedure employs a de novo building through Build and Edit Protein modulus of DS.

The structure is, finally, adjusted taking into account the results of secondary structure predictions and optimized for side chain conformation. All successful approaches to side chain placement are at least partly knowledge based. They use libraries of common rotamers extracted from high resolution X-ray structures. The various rotamers are tried successively and scored with a variety of energy functions.

Model optimization

Once we checked all the alignment with previsions and rearranged loop conformation, the most straightforward approach to model optimization is simply to run a molecular dynamics simulation of the model that contains implicitly entropic effects that are otherwise difficult to treat.

Model validation

The stereochemistry quality of the obtained model is checked using different online program: PROCHECK, WHATIF, PROSA, VERIFY 3D.

PROCHECK generates a Ramachandran plot of the main torsional angles of the model distinguishing between most favored region, additionally allowed regions, disallowed region. Other stereochemical parameters to consider are peptide planarity, bad nonbonded interactions, main-chain hydrogen bonding energy and standard deviation of xi angle.

The same program could also give the G-factor value that for reasonable structures fall between 0 and –0.5.

WHATIF gives an idea of the reliability of the model by the normality index (z-score) that needs to be < -5.0.

Concerning PROSA we will obtain a quality of the model by corresponding PROSA normalized z-score, <0.70 is predictive for a good structure. This parameter is derived from the nature of residue-residue interaction energies, directly linked to backbone contacts.

VERIFY 3D has a compatibility score depending on the side chain environments, a good value means an overall self-consistency in terms of sequence-structure compatibility.

A further optimization could be done by another cycle of MD simulation and a comparison between the parameters, previously, obtained and the second ones.

In this thesis, we exploit the homology modeling to build different proteins having different degree of complexity, and we wil interpret further the results obtained.

2.10 Steered Molecular Dynamics (SMD)

In the discovery and study of safe and effective drugs the aim is to achieve inhibitors that bind tightly to active sites, in other words that possess a negative free energy of binding; this allow the use at low doses. As explained in the section regarding the calculation of binding free energy, the most used approach uses MD trajectories by MMPBSA technique. However a direct calculation of the change in free energy in some complexes is difficult for the natural change in conformation. An alternative approach is to use Steered molecule dynamics (SMD) to compute the force that is required to extract inhibitors from complexes with protein.[14,15] The simulation runs and all atoms adjust to the forced change in structure so that conformations may be sampled along a particular pathway. In theory the change could happen spontaneously during the simulation, but we would have to wait a very long time to see it. Thus, we accelerate a process in order to understand how the systems energy changes. As the force is executed and motion occurs along a coordinate, the potential energy of the system is calculated. This potential of mean force (PMF) is related to the free energy change for the process. SMD is therefore one of the methods used to calculate free energy changes.

The potential of mean force (PMF) is an averaged quantity. Formally, it is obtained from theory using the Jarzynski equality:

$$PMF = -RT \ln (\exp(-W/RT))$$

where W is the work, R is the gas constant and T is the temperature. The bracket means that a configurational average is to be carried out. In practice we will run 10 simulations and average them to obtain the effect of the bracket. The work, W results from the resistance met by the pulling force. When the system meets with resistance it does not “keep up” with the pulling trajectory. The force needed to keep it moving is calculated as it moves and output to the out file. These forces will then be used to calculate the work W, and ultimately the potential of mean force.

Since the average factor is related to the equilibrium constant, the PMF is related to the free energy change for the process.

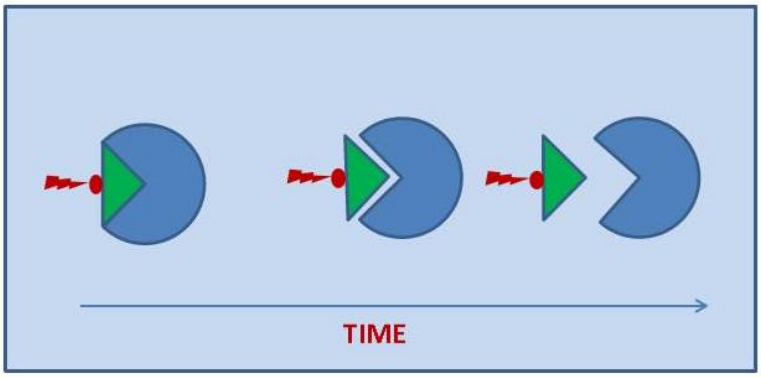


Figure 2.16 *System in which a small spring is connected to the ligand in the protein-ligand complex and then pulled at constant velocity into the surrounding water.*

Bibliography

1 AMBER11 (University of California, San Francisco, 2010).

2 Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999-2012, doi:10.1002/jcc.10349 (2003).

3 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).

4 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092, doi:doi:http://dx.doi.org/10.1063/1.464397 (1993).

5 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:doi:http://dx.doi.org/10.1063/1.445869 (1983).

6 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341, doi:http://dx.doi.org/10.1016/0021-9991(77)90098-5 (1977).

7 Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *Journal of the American Chemical Society* **120**, 9401-9409, doi:10.1021/ja981844+ (1998).

8 http://ambermd.org/tutorials/advanced/tutorial3/.

9 Sitkoff, D., Sharp, K. A. & Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *The Journal of Physical Chemistry* **98**, 1978-1988, doi:10.1021/j100058a043 (1994).

10 Gilson, M. K., Sharp, K. A. & Honig, B. H. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry* **9**, 327-335, doi:10.1002/jcc.540090407 (1988).

11 Sanner, M. F., Olson, A. J. & Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305-320, doi:10.1002/(sici)1097-0282(199603)38:3<305::aid-bip4>3.0.co;2-y (1996).

12 Onufriev, A., Bashford, D. & Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B* **104**, 3712-3720, doi:10.1021/jp994072s (2000).

13 Feig, M. *et al.* Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry* **25**, 265-284, doi:10.1002/jcc.10378 (2004).

14 Jorgensen, W. L. Drug discovery: Pulled from a protein's embrace. *Nature* **466**, 42-43 (2010).

15 Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M. & Cavalli, A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *Journal of the American Chemical Society* **132**, 7361-7371, doi:10.1021/ja100259r (2010).

Results & Discussion

As anticipated in the introduction in this thesis different case studies will be presented.

The individuation of new targets for cancer therapy could involve the interest for some protein that are characterized by partially unknown structure or mechanism or simply are in an indirect pathway that leads to carcinogenesis paragraph 3.1, 3.2, 3.3).

In the meantime some kind of cancer could present particular mutations, primary or acquired as response to the therapy; an accurate evaluation of the conformational and energetical effects due to these variations help clinicians in a tailored therapy (see more on paragraphs 3.4, 3.5).

3.1 σ_1 RECEPTOR AS NEW TARGET IN CANCER THERAPY: STRUCTURAL CHARACTERIZATION AND DRUG DESIGN

3.1.1 BUILDING OF 3D HOMOLOGY MODEL OF σ_1 RECEPTOR[1]

In this work, we present for the first time a 3D model for the σ_1 receptor based on a molecular modeling and simulation ansatz involving the following sequential steps: i) generation and optimization of the complete 3D model of the receptor by homology modeling techniques;[2] ii) further 3D homology model refinement exploiting information derived from ligand docking,[3] a 3D pharmacophore model[4] and mutagenesis experiments[5] as spatial restraints; iii) docking of a series of bioactive ligands into the putative binding site, estimation of their binding affinity via the MM/PBSA methodology,[6] and comparison with the available experimental activities; iv) design of a set of new ligands with high affinity based on the derived receptor model. All detailed information about each single step mentioned above can be found in the next Experimental Section.

The initial stage of step i) consisted in the modeling of the transmembrane (TM) regions (three α-helices spanning residues 10-30, 80-100, and 180-200, respectively), and the prediction of the overall secondary structure of the protein, which included a few β-strands in the C-terminal half (e.g., residues 155-158 and 181-185), and some loops. The search for a suitable template to build the 3D receptor structure yielded 4 different protein sequences having both a considerable identity (≥ 30%) with specific portions of the σ_1 sequence and an X-ray structure available in the Protein Data Bank (PDB entries 3CIA,1I24, 2Z2Z, and 2Q8I).

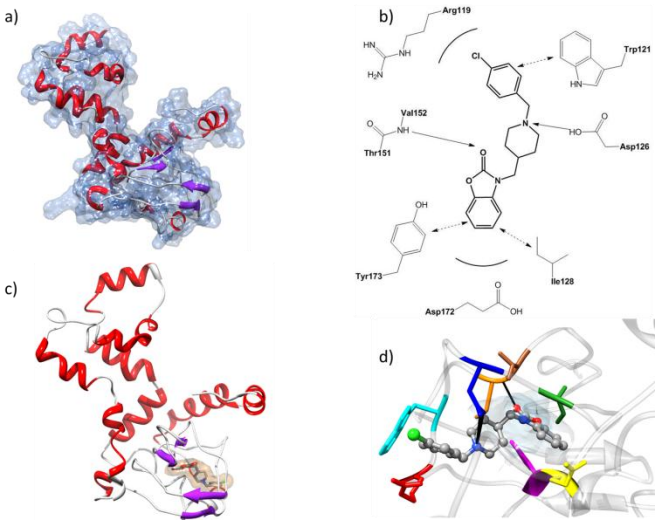


Figure 3.1.1 *The σ_1 receptor and its ligand 1c. a) View of the σ_1 receptor MD-optimized homology model in complex with 1c. b) 2D schematic representation of postulated interactions between the σ_1 receptor and 1c, established by analyzing published sequence-structure relationships, mutagenesis data, 3D pharmacophore model requirements, and direct affinity measurements. The lines/arrows indicate proposed key interaction between the receptor and its ligand. The interactions essential for the binding of ligand 1c are 1) an hydrogen bond (HB) between the 1c piperidine –NH donor and the protein Asp126 –COO- acceptor (1st 3D pharmacophore requirement); 2) the HB between the backbone –NH donor of the peptidic bond linking Thr151 and Val 152 on the receptor side and the acceptor oxygen atom of the benzoxazolone moiety of 1c (2nd 3D pharmacophore requirement); 3) T-stacking π-π- interaction between the protein residue Trp121 and the heteroaromatic condensed rings of 1c (3rd 3D pharmacophore requirement); 4) parallel-stacking π-π- interaction between the protein residue Tyr173 and the benzoxazolone phenyl ring of 1c (4th 3D pharmacophore requirement); 5) highly stabilizing van der Waals and electrostatic interactions between Arg119 and Glu172 with the p-Cl-phenyl and the benzoxazolonic heterocycle of 1c, respectively; 6) general hydrophobic favorable interactions between the side chains of Ile128, Thr151, and Val152 with both the aliphatic and aromatic portions of the ligand. c) View of the final σ_1 receptor homology model in complex with 1c. d) Modeled complex of the σ_1 receptor with 1c showing the key interactions proposed in the topographical interaction model depicted in part b. The main protein residues involved in these interactions are Arg119 (red), Trp121 (cyan), Asp126 (blue), Ile128 (forest green), Thr151 (sienna), Val152 (orange), Glu172 (yellow), and Tyr173 (magenta).*

Accordingly, each corresponding receptor/template sequence was aligned, and the Cartesian coordinates for σ_1 residues in structurally conserved regions were obtained from the corresponding sequence in the template PDB file. The first part of the N-terminal domain of the protein (residues 1-16) showed no homology with any other proteins in all queried

databases and thus was built de novo. Finally, the overall receptor 3D structure was built linking the different template-based homology models and creating/optimizing the missing loop portions via several refinement processes. The structure was further adjusted manually to fully match the results of secondary structure predictions, and optimized for backbone and side chain conformation. This initial 3D model was then subjected to extensive molecular dynamics (MD) simulations in explicit solvent to reach an energetically reasonable conformation. The MD-refined homology model (Figure 3.1.1a), characterized by a very high stereochemical quality (more than 98% of residues in the favored or additional allowed regions of Ramachandran plot and only two residues in the disallowed region), was employed in the successive step.

Step ii) required the initial identification of a putative binding site in the receptor model, based on preliminary information on sequence-structure relationships, mutagenesis studies,[5]and some ligand-binding pharmacophore requirements.[4] In detail, a protein isoform missing residues 119-149 was found devoid of ligand binding capacity, and the conversion of residues D126 and E172 to glycine led to a several-fold reduction in ligand-binding function for the σ_1 receptor.[5] Moreover, a hydrophobicity analysis performed in the present study identified, aside for the TM regions, a third hydrophobic region matching the SBDLII region and centered on Asp188, a residue specifically photolabeled by [125I]IACoc (3-iodo-4-azido cocaine).[7] Having localized this protein region as a possible zone for ligand binding, a thorough search for a sequence satisfying the chemical features imposed our recently developed 3D model was performed and successfully retrieved. Thus, compound 1c was into the σ_1 putative binding pocket. In the set of docked ligand conformations a solution was found that best reproduced the key interactions depicted in the derived topographical model (Figure 3.1.1b). Subsequently, new homology models were generated explicitly considering bound compound 1c. The best-scored amino acids were merged in a combinatorial fashion, the resulting molecular complex was relaxed by energy minimization and MD simulations, and finally a model without unfavorable van der Waals interactions was obtained (Figure 3.1.1c). Figure 3.1.1d shows the ligand-binding site of the resulting σ_1/1c complex, which convincingly reproduces the proposed interactions in the 2D topological model.

In step iii), the refined σ_1 homology model was used for the docking of a series of bioactive ligands (including 1c) into the putative binding site, estimation of their binding affinity via the MM/PBSA methodology,[6] and comparison with the available experimental activities. Table 3.1.1 shows the results of the MM/PBSA ranking, from which the excellent agreement between computed and experimental affinities of these ligand series is apparent.

Table 3.1.1. *Binding free energies ΔGbind (kcal/mol) for a set of 9 bioactive ligands and the σ_1 receptor homology model. Errors are given in parenthesis as standard errors of the mean. The experimental and calculated Ki (nM), as estimated from the corresponding ΔGbind values, are also reported for comparison.* [a] *Ki was calculated from ΔGbind according to the following relationship: ΔGbind = -RT lnKi.*

Cpd	1c	1a	1j	2a	2d	3a
ΔG_{bind}	-12.05 (0.37)	-10.93 (0.44)	-8.38 (0.46)	-8.87 (0.48)	-8.22 (0.42)	**-10.53 (0.39)**
K_i(calc.)[a]	1.48	9.80	730	315	945	**19.1**
K_i(exp.)	0.098	358	1017	223	1147	**2.60**
Cpd	**3d**	**7a**	**7b**	**FEN**	**PTZ**	**HALO**
ΔG_{bind}	-10.67 (0.45)	-10.24 (0.39)	-10.99 (0.43)	-12.44 (0.43)	-10.02 (0.45)	**-10.61 (0.42)**
K_i(calc.)[a]	15.3	31.4	8.70	0.77	45.5	**16.8**
K_i(exp.)	**7.1**	**22.5**	**12.9**	**0.01**	**15.0**	**5.7**

The quality of the model was further validated by probing its ability to accommodate other known σ_1 ligands belonging to structurally diverse compound classes. Thus, the strong σ_1 agonist pentazocine (PTZ),[8] and the two high-affinity σ_1 ligands fenpropimorf (FEN)[9] and haloperidol (HALO) were docked into the σ_1 binding site and scored according to the same protocol employed for 1c and related compounds. Importantly, the adopted binding modes of these three compounds (Figure 3.1.2) were also in agreement with the proposed 2D interaction model (Figure 3.1.1b), and the MM/PBSA was able to rank their affinity not only in the correct order but also with an outstanding agreement between computed and experimental Ki values.[10]

Lastly, in step iv) all information presented above and the new 3D homology model derived in this work were exploited for the design of a new series of σ_1 ligands starting from the structure of the lead compound 1c. Subsequently, step iii) of the multistep procedure outline above was applied to predict their affinity towards the protein. According to preliminary tests performed in vitro, several of them showed affinity in the nanomolar range. In particular, compound EL-1 exhibited a very promising Ki value of 1.87 nM, in utter agreement with the MM/PBSA predicted value of 8.40 nM. As shown in Figure 3.1.3, compound EL-1 adopted a binding mode within the receptor putative binding pocket which satisfies all proposed 2D/3D interaction model.

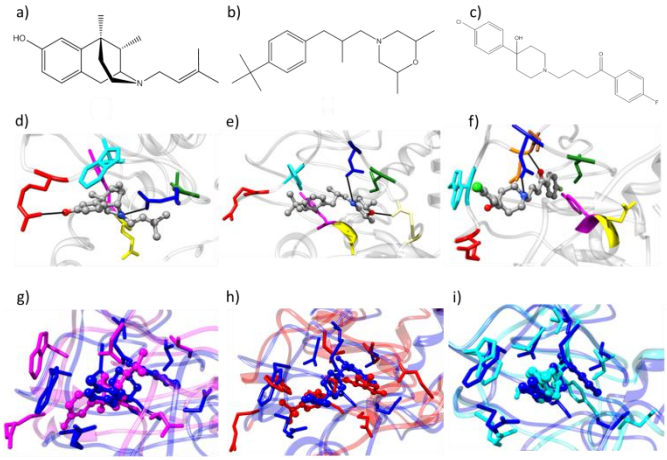


Figure 3.1.2. *Structure of a) pentazocine (PTZ), b) fenprofimorf (FEN), and c) (haloperidol (HALO). Modeled complex of the σ_1 receptor with d) PTZ, e) FEN, and f) HALO, showing the key interactions proposed in the topographical interaction model depicted in Figure 1b. The main protein residues involved in these interactions are Arg119 (red), Trp121 (cyan), Asp126 (blue), Ile128 (forest green), Thr151 (sienna), Val152 (orange), Glu172 (yellow), and Tyr173 (magenta). c) Superposition of the σ_1 receptor binding site in complex with 1c (blue) and g) PTZ (magenta), h) FEN (red), and i) HALO (cyan).*

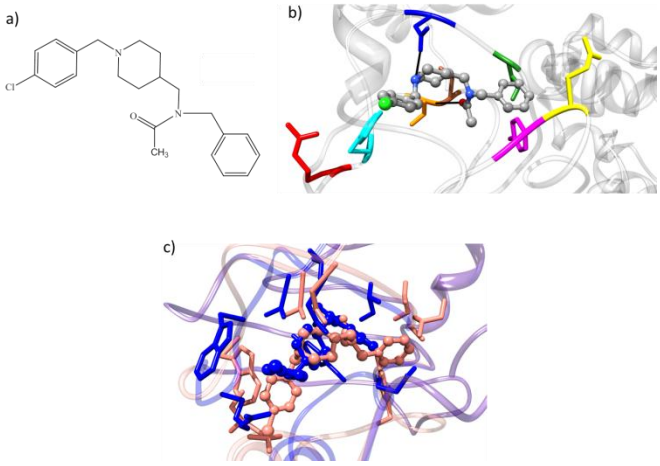


Figure 3.1.3. *Structure of compound EL-1 designed exploiting the 2D/3D interaction model and the 3D homology model of the σ_1 receptor developed in this work. b) Modeled complex of the σ_1 receptor with EL-1 showing the key interactions proposed in the topographical interaction model depicted in Figure 3.1.1b. The main protein residues involved in these interactions are Arg119 (red), Trp121 (cyan), Asp126 (blue), Ile128 (forest green), Thr151 (sienna), Val152 (orange), Glu172 (yellow), and Tyr173 (magenta). c) Superposition of the σ_1 receptor binding site in complex with 1c (blue) and EL-1 (rosy brown).*

In conclusion, the multistep homology modeling procedure adopted in this work produced a reliable 3D structure of the σ_1 receptor that can successfully be employed as a platform for structure-based drug design. Overall, this works illustrates that explicit ligand information can be successfully employed to finely shape, optimize, and the binding site model. This approach

can be especially helpful when modeling proteins with highly flexible binding sites, or targets that are difficult to characterize experimentally, and/or pockets with low homology to available template structures, as in the present case of σ_1 receptors. A thorough model validation was performed by a docking/MMPBSA-based small-scale virtual screening, and by the receptor model-based design of a new series of potent σ_1 ligands featuring activity in the nM range. Although medicinal chemistry is still eager to obtain new x-ray structures of protein targets, this study with σ_1 receptor demonstrates that its derived 3D homology model is already an excellent substrate. Given the intense ongoing efforts in designing new compounds as potent σ_1 ligands, this work is rather an encouraging account.

Experimental section

For the identification of the transmembrane (TM) helical domains and the prediction of the overall secondary structure of the σ_1 receptor, we employed the programs TMPRED, DISOPRED, Predict Protein, HMMTOP, I-TASSER, and the Secondary Structure Prediction tool of Discovery Studio (DS) (v. 2.1, Accelrys, San Diego, USA) in a comparative fashion.

All programs identified the same three TM domains, spanning the protein residues 10-30, 80-100, and 180-200, respectively. A prediction of the secondary structure of the entire protein was also obtained by running different programs and comparing the corresponding output for quality, reliability, and matching. Again, very similar results were obtained, according to which the σ_1 secondary structure should feature, aside the already described TM domains, a few β-strands in the C-terminal half (residues 111-116, 133-135, 144-146, and 158-164), and some loops, as shown in Figure 3.1.4.

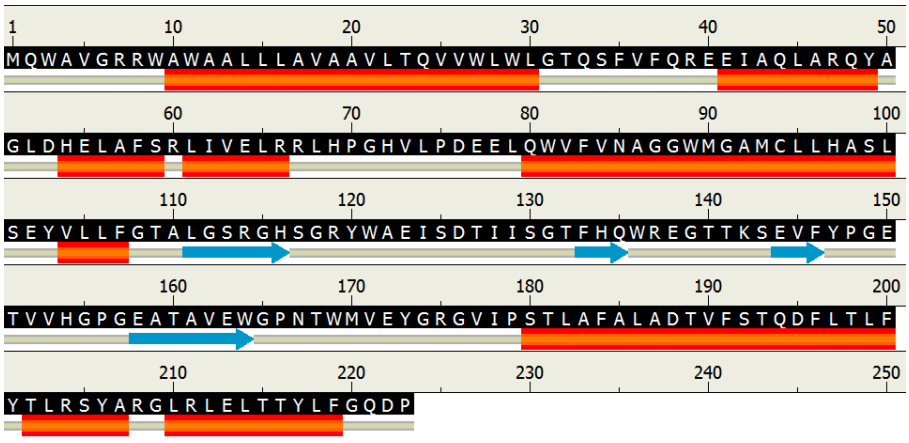


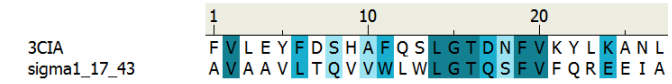
Figure 3.1.4 *General secondary structure prediction of the σ_1 receptor. The main protein structural motifs (α-helices, β-sheets, and loop/coils) are highlighted below the primary sequence as follows: orange/red large tubes: α-helices; light blue arrows: β-sheets; gray thin tubes: loops/coils.*

Template searching, initial σ_1 receptor 3D model building, and preliminary model validation

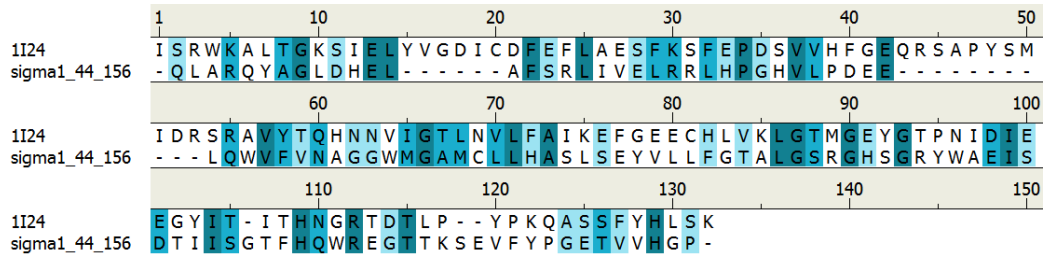
The primary sequence of the σ_1 receptor was obtained from SWISS-PROT database (accession number Q99720)[11]. A comparative PSI-Blast search against the non-redundant protein database via the NCBI website (http://www.ncbi.nlm.nih.gov), the ExPASy Proteomic Server (http://expasy.org), and MODWEB (http://modbase.compbio.ucsf.edu/scgi/modweb.cgi) did not identify one single protein with high homology for the target receptor sequence, as expected. However, all these searches retrieved 4 different protein sequences having both a considerable identity (≥ 30%) with specific portions of the σ_1 sequence and an X-ray structure available. The sequence alignment between the receptor target sequence and the templates was conducted automatically by applying different scoring matrices, gap penalties, and gap length penalty parameters in order to achieve the highest sequence homology. In detail: i) cold-aminopeptidase (PDB code 3CIA) gave 22% sequence identity (SI) and 37% sequence similarity (SS) between residues 432-458 and σ_1 residues 17-43; ii) Udp-Sulfoquinovose Synthase (PDB code 1I24) featured 17% SI and 31% SS between residues 58-186 and σ_1 residues 44-156; iii) for Tk-subtilisin (PDB code 2Z2Z), 366-396 residues shared 39% SI and 55% SS with σ_1 residues 157-187; and iv) pyruvate dehydrogenase kinase isoform 3 (PDB code 2Q8I) residues 140-178 yielded 42% SI and 55% SS with residues 188-223 of the σ_1 receptor (see Figure 3.1.5).

Accordingly, the crystal structures of these 4 proteins were used as templates for building the different sections of the σ_1 homology structure. Each initial template-based structure was obtained using MODWEB, was checked using the Build Homology Model of DS, and compared to the corresponding template through the Structure Comparison -> Match Maker feature of Chimera.

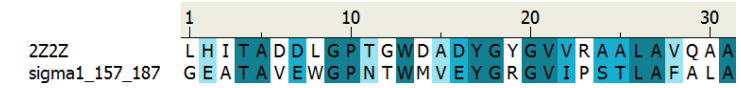
The first part of the N-terminal domain of the protein (residues 1-16) showed no homology with any other protein in all queried databases; therefore, it was built de novo using the Build and Edit Protein modulus of DS. Finally, the overall receptor 3D structure was built linking the different template-based homology models and creating/optimizing the missing loop portions via several refinement processes. The structure was further adjusted manually to fully match the results of secondary structure predictions and optimized for side chain conformation. The stereochemistry quality of the resulting model was checked using PROCHECK[12], WHATIF[13], and PROSA[14].



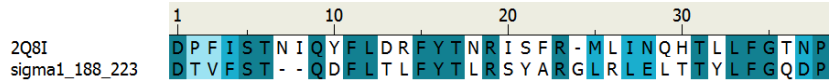
(A)



(B)



(C)



(D)

Figure 3.1.5 *Sequence alignment of different portions of the human σ_1 receptor with : (A) cold-aminopeptidase (PDB code: 3CIA); (B) Udp-Sulfoquinovose Synthase (PDB code 1I24); (C) Tk-subtilisin (PDB code 2Z2Z); and (D) pyruvate dehydrogenase kinase isoform 3 (PDB code 2Q8I).*

The analysis of the Ramachandran plot produced by PROCHECK of the main torsional angles of the preliminary model revealed that a total of 95.8% of the protein residues were in the most favored regions of the plot, with 2.7% in additionally allowed regions, giving a total of 98.5%. Bond lengths and angles were all found within their normal range, and no bumps (corresponding to high van der Waals energies) were detected. Other stereochemical parameters, such as peptide planarity, bad nonbonded interactions, main-chain hydrogen bonding energy, and standard deviations of σ_1 angle (i.e., the first torsion angle of the side chains) were also examined and checked for validity. Accordingly, the total quality G-factor was – 0.27, revealing the good quality of the initial σ_1 model structure since reasonable values for the G-factor in PROCHECK fall between 0 and – 0.5, with the best models displaying values closest to 0. The packing quality of the 3D model was assessed for the absence of steric clashes between any pair of atoms. The results of the WHATIF quality report showed a normality index (z-score) of -1.89, a value which falls in the acceptable range for a valid structure (z-score < -5.0). The analysis of the PROSA results revealed that, for most regions in the initial σ_1 model, the residue-residue interaction energies were all attractive, indicating no bad backbone contacts for the model in these regions. Lastly, the good overall quality of the model was also

reflected in the corresponding value of the PROSA normalized z-score = 0.92, being z-score > 0.70 indicative of a good structure.

Model refinement via molecular dynamics (MD) simulations and final model validation

The preliminary 3D homology model of the σ_1 receptor was further refined and its stability verified via a cycle of MD simulations in water environment. All MD simulations were carried out using the Sander and Pmemd modules of the AMBER 11 suite of programs[15]. The protein was solvated by a TIP3P[16] box of water extending at least 10.0 Å in each direction from the solute; then, 79 Cl- and 73 Na+ counterions were added to ensure the overall charge neutrality of the system and to mimic a physiological ionic strength of 0.15 M. The energy of the resulting system was relaxed via a multi-step minimization procedure using the AMBER ff03 force field[17], as follows. First, the water molecules and counterions were minimized for 5000 cycles of steepest-descent minimization. Next, the side chains of the σ_1 receptor and the solvent molecules were further relaxed using a 1000- cycle energy minimization to relieve possible unexpected side-chain clashes. Lastly, the whole system was energy relaxed for further 1000-cycles of conjugate-gradient minimization. The system was then gradually heated to T = 300 K in three MD intervals , allowing a 20 ps interval per each 100 K, followed by NPT MD simulations at 300 K for other 60 ps with the backbone of the protein restrained with an harmonic potential, while all water molecules, counterions, and the protein side chains were allowed to move. Then, the whole system was further equilibrated in the NVT ensemble for 15 ns at 300 K, to attain a well relaxed receptor structure and to verify the stability of the corresponding MD trajectory. The time step used in the MD simulation was 2 fs. The periodic boundary condition with Berendsen temperature coupling and P =1 atm, with isotropic molecule-based scaling. The SHAKE algorithm[18] was used to fix all covalent bonds containing hydrogen atoms. Long-range nonbonded van der Waals interactions were truncated by using a dual cut off of 6 and 12 Å, respectively, where energies and forces due to interactions between 6 and 12 Å were updated every 20 time steps. The particle mesh Ewald method[19] was used to treat the long- range electrostatics. A residue-based cut off of 8 Å was used for the non-covalent interactions. During the NVT MD simulations, the coordinates of the protein were saved every 1 ps.

The σ_1 model structure remained stable for all 15 ns of the MD trajectory, as indicated by the small fluctuations of the root-mean-square (RMS) deviation of the simulated position of the backbone atoms with respect to those of the initial structure and the corresponding total potential energy of the system E shown in Figure 3.1.6.

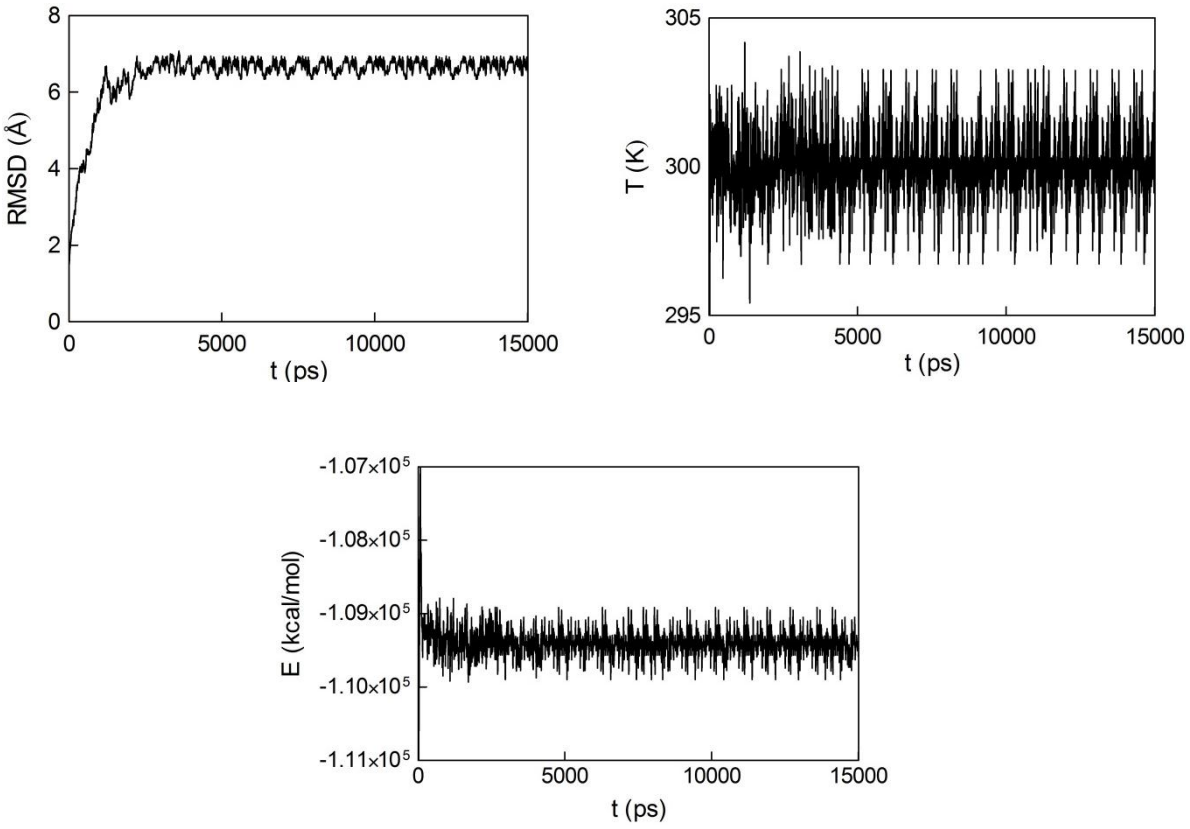


Figure 3.1.6 *(A) RMDS of the coordinates of the backbone atoms of the initial homology model of the σ_1 receptor along the 15 ns MD simulation compared with those of the initial structure. (B) time behavior of the total potential energy and (C) temperature of the σ_1 receptor 3D model in water during the same MD simulation.*

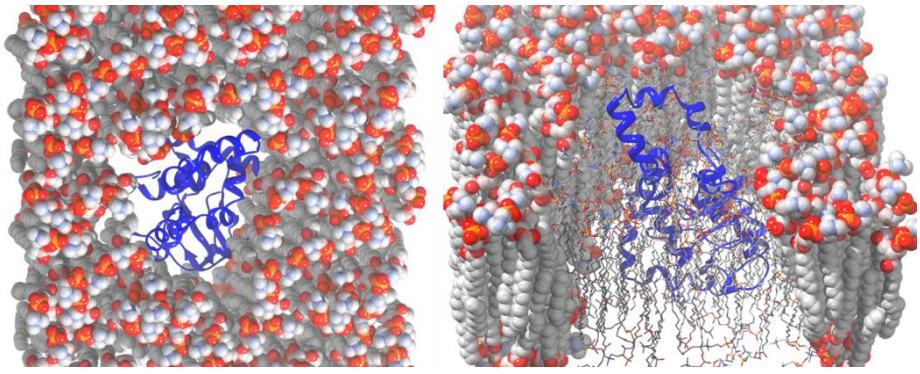


Figure 3.1.7 *Top (left) and side view (right) of the last frame extracted from the equilibrated MD trajectory of the σ_1 receptor 3D homology model in a solvated membrane environment. The protein secondary structure is depicted as a blue ribbon. POPC molecules are shown as atom-colored CPK spheres. In the right panel, a mixed representation of CPK spheres and sticks was chosen to represent the lipid molecules for graphical purposes. Atom color code: O, red; C, light gray; N, light blue; H, white; P, orange. Water molecules and ions are not shown for clarity.*

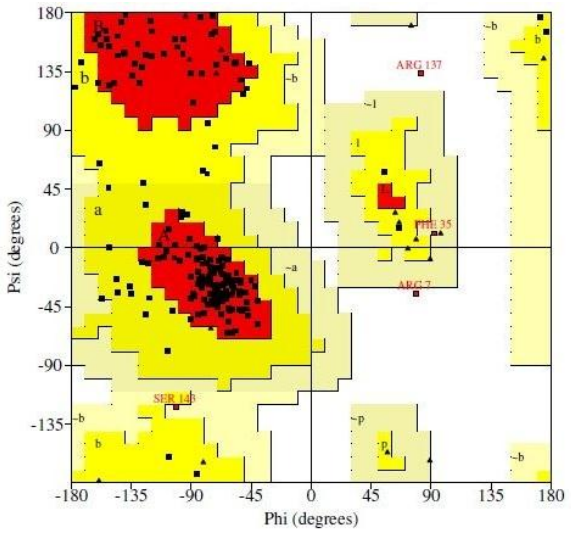


Figure 3.1.8 *Ramachandran plot generated from the MD refined 3D homology model of the σ_1 receptor.*

From the above described 15 ns equilibrated MD trajectory, the protein coordinates were extracted from last frame of the membrane complex (see Figure 3.1.7), and the corresponding 3D σ_1 receptor model was evaluated for quality and reliability.

By comparison with the pre-refined structure, the G-factor increased from -0.27 to -0.21, indicating that the MD simulations relieved some bad contacts between nonbonded atoms without increasing the number of bad dihedral angles of the structure. The result of the π-π combination displayed by the Ramachandran plot of the MD σ_1 relaxed model in Figure 3.1.8 is consistent with this change in G-factor, with a higher percentage of the residues in the most favored regions (97.9%), and a smaller number in generously allowed regions (2.1%), giving a total percentage of 100%. Other stereochemical parameters such as dihedral angles, covalent geometry, and planarity were also examined. The respective G-factors were all close to zero, and all values were well within the acceptable limits. Utterly analogous results were obtained by running MOLPROBITY[20]. The packing quality of the MD relaxed model was checked by the atomic contact analysis of WHATIF, yielding a z-score of -0.89 which testifies the improvement of the model after MD relaxation. The packing quality of each protein residue was also evaluated by the VERIFY-3D method[21]: the compatibility score above zero in the VERIFY-3D graph shown in Figure 3.1.9 is a song of side-chain environments and suggests that the model is characterized by an overall self-consistency in terms of sequence-structure compatibility. The enhancement of the overall 3D homology model quality was also reflected in the value of the corresponding PROSA normalized z-score of 0.98. Such high values of the PROSA z-score approach those typical of high resolution crystal structures, further supporting the fact that the proposed model is of very good quality in backbone conformation.

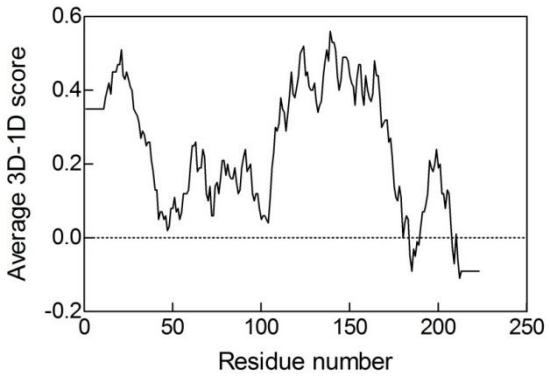


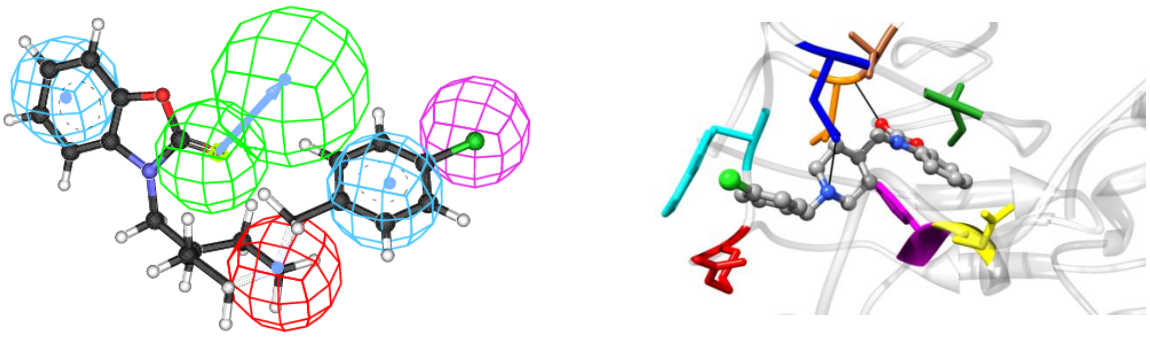
Figure 3.1.9 *VERIFT-3D graph generated for the MD refined 3D homology model of the σ_1 receptor.*

The overall accuracy of a comparative model is related to the percentage sequence identity on which it is based, correlating with the relationship between the structural and sequence similarity of the target and template proteins. High-accuracy comparative models are based on more than 50% sequence identity to their templates. Medium-accuracy comparative models are based on 30 to 50% SI, while low-accuracy comparative models rely on less than 30% of SI. Nonetheless, other factors such as template selection and alignment accuracy usually have a large impact on the resulting final model quality, especially for models based on less than 40% sequence identity to the templates. We are well aware of the fact that many methods often fail to correctly align protein pairs with 20-30% pairwise sequence identity, and that in the present work a consistent portion of the σ_1 receptor was modeled in this "twilight zone". However, since i) it is also often possible to correctly predict features of the target protein that do not occur in the template structure, ii) errors in functional important regions in comparative models are many times relatively low because the functional regions (e.g., binding sites) tend to be more conserved in evolution than the rest of the fold, and iii) all σ_1 ligands considered were ranked for their affinity towards the σ_1 receptor 3D model developed in this work consistently with the corresponding experimental values, we are confident that the proposed homology model of this important protein is characterized by an overall correct folding

Assisted-ligand docking into the σ_1 receptor 3D homology model

To retrieve a putative binding site in the σ_1 receptor model structure we exploited the currently available preliminary information on sequence-structure relationships and mutagenesis studies[5], and some ligand-binding pharmacophore requirements.[4] In detail, some experiments revealed that an alternative splice variant of the σ_1 receptor encoding gene lacking exon 3 expresses a protein that has no ability to bind ligands.[5] The region deleted in the splice variant consists of the protein residues 119-149. Thus, the putative membrane spanning domain should lie upstream the region coded by exon 3 and, since the transcript of this splice

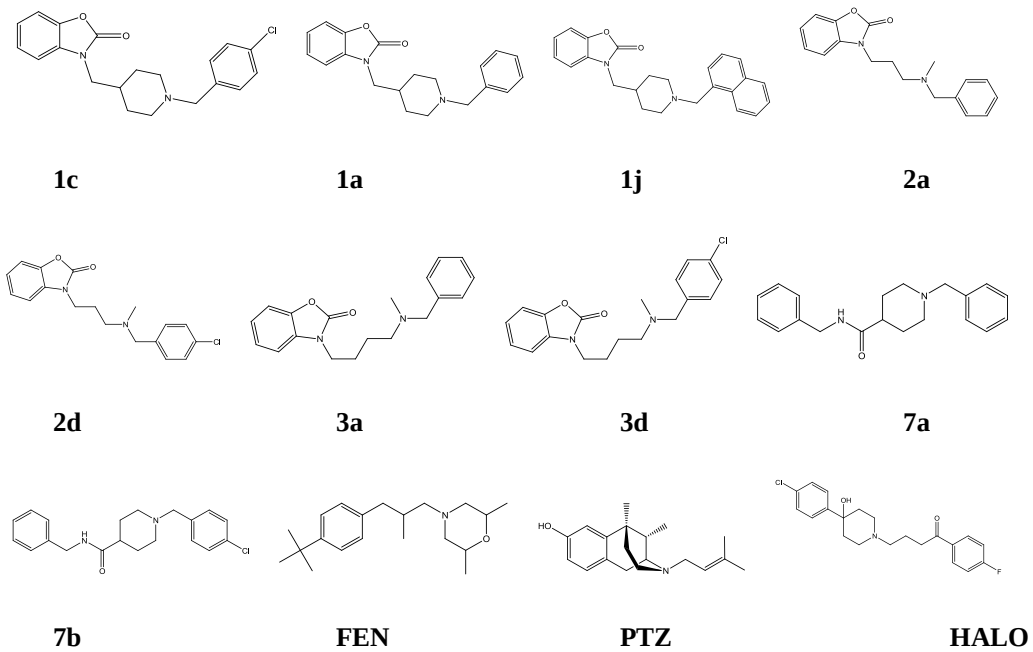
variant is devoid of ligand binding function, we may hypothesize that the ligand binding site may sensible reside in the C- terminal half of the protein. This region contains 12 anionic residues, among which D126 and E172 were found to be directly involved in ligand binding by mutagenesis studies. Moreover, a further hydrophobicity analysis performed in the present study using PONDR[7] identified, aside the TM regions, a third hydrophobic region matching the SBDLII region and centered on Asp188, a residue specifically photolabeled by [125I]IACoc (3-iodo-4-azido cocaine)[22]. Having localized this protein region as a possible zone for ligand binding, a thorough search for a sequence satisfying the chemical features imposed our recently developed 3D pharmacophore model was performed (see Figure 3.1.10(A)), and successfully retrieved. Thus, compound 1c was initially selected for docking into the putative σ_1 binding site. AUTODOCK 4[23] with AUTODOCK TOOLS 1.4.6 on a win64 platform was used for all docking experiments. The general docking methodology was based on a consolidated procedure[24,25]; accordingly, it will be described here only briefly. The size of the binding site was defined via DISCOVERY STUDIO (DS) Studio (v.2.1, Accelrys Inc., San Diego, CA, USA), using an opening site of 10 Å and a grid size of 0.7 Å.



(A) (B)

Figure 3.1.10 *(A) σ_1 receptor 3D pharmacophore mapping of compound 1c. The hypothesis features are portrayed as mashed spheres, color-coded as follows: red, positive ionizable; light blue, hydrophobic aromatic; purple, hydropbobic aliphatic, light green, hydrogen bond acceptor. The last feature is actually represented as a pair of spheres (the smaller sphere represents the location of the hydrogen bond acceptor atom on the ligand and the larger one the location of an hydrogen bond donor on the receptor). (B) Compound 1c docked into the putative binding pocket of the σ_1 3D homology model. Black lines denote polar interactions. The protein residues involved in these interactions are: Arg119 (red), Trp121 (cyan; π–π interaction, hydrophobic aromatic from pharmacophore model), Asp126 (blue; salt bridge NH-piperidine/OD2Asp, positive ionizable), Ile128 (forest green), Thr151 (in sienna), Val152 (in orange) (H-bond O-benzoxazolone/NH-backbone-Val; hydrogen bond acceptor), Glu172 (yellow), and Tyr173 (in magenta; π–π interaction, hydrophobic aromatic from pharmacophore model).*

The AUTODOCK grid box dimensions, based on the identified cavity from DS, were large enough to cover all possible rotations of the ligand. AMBER12 and Lennard-Jones parameters were used in modeling van der Waals interactions and hydrogen bonding (N-H, O-H and S-H), respectively. In the generation of the electrostatic grid maps, the distance dependent relative permittivity of Mehler and Solmajer was applied[26]. 300 hundred Monte Carlo/Simulated Annealing (MC/SA) runs were performed, with 100 constant temperature cycles for simulated annealing. For these calculations, the GB/SA implicit water mode[27] was used to mimic the solvated environment. The rotation of the angles ϕ and φ, and the angles of side chains were set free during the calculations. All other parameters of the MC/SA algorithm were kept as default. Following the docking procedure, the structure of all compounds were subjected to cluster analysis with a tolerance of 1 Å for an all-atom root-mean-square (RMS) deviation from a lower-energy structure representing each cluster family. The resulting docked conformations were clustered and visualized. Interestingly, the most populated cluster was not only the one characterized by the lowest (i.e., more favorable) AUTODOCK energy but also the only one in which the 3D pharmacophore required interactions were satisfied by compound 1c, as shown in Figure 3.1.10(B).



Scheme 3.1.1 *Chemical structures of the σ_1 ligands considered for docking, scoring, and model assisted drug design/virtual screening.*

The same docking procedure was then applied to the entire list of compounds shown in Scheme 3.1.1 (see Table 3.1.2 for IUPAC nomenclature) with a twofold purpose: i) checking the validity of the entire docking procedure on an extensive set of molecules (compounds 1a-7b), and ii) probing the ability of the 3D σ_1 homology model to accommodate other known σ_1 ligands belonging to structurally diverse compound classes (PTZ, FEN, and HALO).

Table 3.1.2 *Chemical formula and IUPAC names of the compounds*

Compound	Formula	IUPAC name
1c	$C_{20}H_{21}ClN_2O_2$	3-[[1-(4-Chlorobenzyl)piperidin-4-yl]methyl]benzo[d]oxazol-2(3H)-one
1a	$C_{20}H_{22}N_2O_2$	3-((1-benzylpiperidin-4-yl)methyl)benzo[d]oxazol-2(3H)-one
1j	$C_{24}H_{24}N_2O_2$	3-((1-(naphthalen-1-ylmethyl)piperidin-4-yl)methyl)benzo[d]oxazol-2(3H)-one
2a	$C_{18}H_{20}N_2O_2$	3-(3-(benzyl(methyl)amino)propyl)benzo[d]oxazol-2(3H)-one
2d	$C_{18}H_{19}ClN_2O_2$	3-(3-((4-chlorobenzyl)(methyl)amino)propyl)benzo[d]oxazol-2(3H)-one
3a	$C_{19}H_{22}N_2O_2$	3-(4-(benzyl(methyl)amino)butyl)benzo[d]oxazol-2(3H)-one
3d	$C_{19}H_{21}ClN_2O_2$	3-(4-((4-chlorobenzyl)(methyl)amino)butyl)benzo[d]oxazol-2(3H)-one
7a	$C_{20}H_{24}N_{2O}$	N,1-dibenzylpiperidine-4-carboxamide
7b	$C_{20}H_{23}ClN_{2O}$	N-benzyl-1-(4-chlorobenzyl)piperidine-4-carboxamide
FEN	$C_{20}H_{33}NO$	4-(3-(4-tert-butylphenyl)-2-methylpropyl)-2,6-dimethylmorpholine
PTZ	$C_{19}H_{27}NO$	2-dimethylallyl-5,9-dimethyl-2'-hydroxybenzomorphan
HALO	$C_{21}H_{23}ClFNO_2$	4-(4-(4-chlorophenyl)-4-hydroxypiperidin-1-yl)-1-(4-fluorophenyl)butan-1-one

Ligand scoring by MM/PBSA

The resulting σ_1/1c complex shown in Figure 3.1.10(B), and all other best receptor/ligand complexes resulting from the automated docking procedure were further refined in AMBER 9 using the quenched molecular dynamics (QMD) method.[28] In this case, 1 ns molecular dynamics (MD) simulation at 300 K were employed to sample the conformational space of each ligand-receptor complex in the GB/SA continuum solvation environment. The integration step was equal to 1 fs. After each ps, each system was cooled to 0 K, the structure was extensively minimized, and stored. To prevent global conformational changes of the protein, the backbone atoms of the protein binding site were constrained by an harmonic force constant of 100 kcal/Å, whereas the amino acid side chains and the ligands were allowed moving without any constraint. The best energy configuration of each complex resulting from the previous step was allowed to relax in an 80Å × 80Å × 80Å box of TIP3P water molecules. The resulting systems were minimized with a gradual decrease in the position restraints of the protein atoms. Finally, to achieve electroneutrality, a suitable number of counterions neutralizing ions were added; further, the solution ionic strength was adjusted to the physiological value of 0.15 M by adding the required amounts of Na+ and Cl- ions. After energy minimization of the added ions for 1500 steps, keeping the protein, the ligand, and the pre-

existing waters rigid, followed by an MD equilibration of the entire water/ion box with fixed solute for 5 ns, further unfavorable interactions within the structures were relieved by progressively smaller positional restraints on the solute (from 25 to 0 kcal/(mol Å^2) for a total of 10 ns. Each hydrated complex system was gradually heated to 300 K in three intervals, allowing a 2 ns interval per each 100 K, and then equilibrated for 5 ns at 300 K, followed by 20 ns of data collection runs, necessary for the estimation of the free energy of binding.

For the calculation of the binding free energy between σ_1 and each ligand in water, a total of 20000 snapshots were saved during the MD data collection period described above.

The binding free energy ΔGbind of each σ_1 receptor/ligand complex in water was calculated according to a validated computational recipe28, based on the so-called Molecular Mechanic/Poisson-Boltzmann Surface Area (MM/PBSA) ansatz, and originally proposed by Srinivasan et al.[6] For any details see the chapter Materials&methods.

As discussed in the main text, by applying the MM/PBSA procedure described above we were able to correctly rank not only the series of compounds 1c-7b, but also the three structurally diverse σ_1 ligands (FEN, PTZ, and HALO); for all these molecules, an excellent agreement (R^2 = 0.93) between computed and experimental affinities of these ligand series was indeed obtained, as can be inferred from the scatter plot shown in Figure 3.1.11 (see also Table 3.1.1).

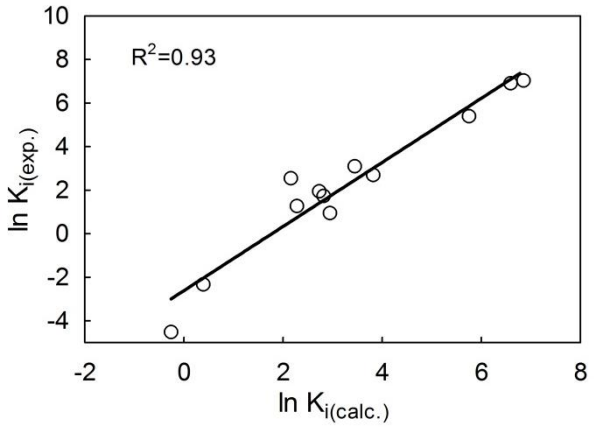
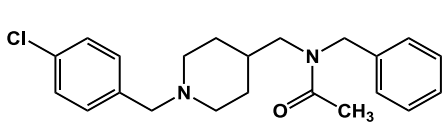


Figure 3.1.11 *Linear correlation obtained between the calculated σ_1 receptor-ligand Ki values and the corresponding experimental Ki values (correlation coefficient R^2 = 0.93) for 12 complexes.*

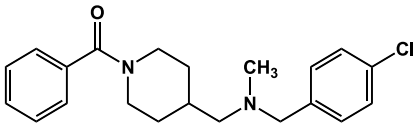
Model-assisted drug design/virtual screening

The new 3D homology model of the σ_1 receptor was then exploited for the design of three new σ_1 ligands EL-1, EL-2, and EL-3 (Scheme 3.1.2), starting from the structure of the lead compound 1c. In details, the chemical structures of these molecules were designed according to the following rationale: compound EL-1 was conceived as the non-cyclic analogous of 1c; indeed, the benzaoxazolone moiety was replaced by a N-benzylacetamide group featuring an oxygen atom capable to fit the hydrogen bond acceptor feature (see Figure 3.1.3). The N-4-

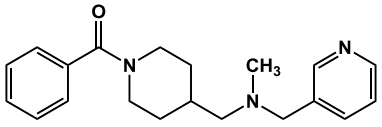
chlorobenzylpiperidine portion, essential for protein binding, was retained. Therefore, this compound was predicted to be a high affinity σ_1 ligand. On the other hand, compound EL-2 was designed for a moderate affinity toward the receptor. In fact, despite the structure possesses all the necessary chemical requisites to satisfy the features of our models, its piperidin-1-yl-phenylmethanone scaffold yields an element of structural rigidity which, in turn, lowers the ability of the derivative to adopt the best conformational fitting in the binding site. In order to obtain a low affinity ligand, we further synthetized compound EL-3. The low conformational freedom of this molecule (as for EL-2), which hampers a complete fitting of the chemical features onto the pharmacophoric maps, and the replacement of the 4-chlorophenyl group with a piperidine ring, resulting in its inappropriate placement in a highly hydrophobic region of the receptor, concur to the quite modest affinity of this compound towards the protein. The steps outline above were then repeated to predict the affinity of this small new molecular set towards the protein.



EL-1



EL-2



EL-3

Scheme 3.1.2 *Chemical structures of the new σ_1 ligands design to test the validity of the docking, scoring, and model assisted drug design/virtual screening.*

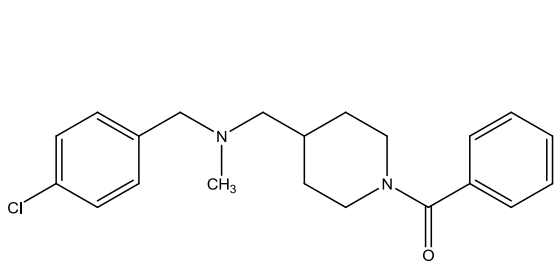
3.1.2 EXTENSIVE VALIDATION OF THE MODEL FOR σ_1 RECEPTOR[29]

In this work we present a further, extensive validation of the predictive capability of our σ_1 3D model in computer-assisted drug design. In brief, exploiting our 3D pharmacophore model, previously described, for σ_1 ligands we designed 33 new σ_1 ligands, with affinity for the receptor spanning 5 orders of magnitude. All these compounds were then docked – using a pharmacophore-guided procedure – into the putative binding site of our σ_1 3D receptor model [1]and ranked for affinity by extensive molecular dynamics simulation based free energy calculations. The binding modes of all compounds and the key receptor residues involved in

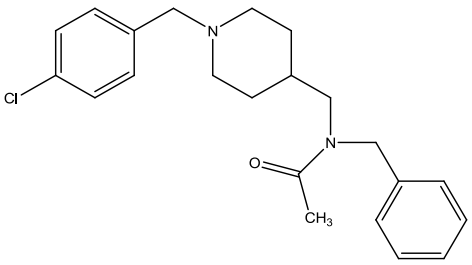
each ligand binding were further assessed by applying two further computational techniques: a per residue free energy deconvolution and in silico alanine scanning. Subsequently, all 33 compounds were synthesized in our laboratory and then tested for σ_1 binding activity in vitro.

3D pharmacophore-based design of new σ_1 ligands.

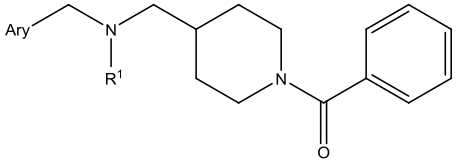
The initial step in the present computer-assisted drug design recipe consisted in exploiting our previously developed 3D pharmacophore model for the in silico design of new σ_1 ligands. Recalling that the two major purposes of the present work were i) the ultimate validation of the 3D homology model of the σ_1 protein and its reliability in assisting σ_1 ligands computer-based design, and ii) the formulation of a molecular-base rationale for the receptor affinity of the ligands, we took advantage of the validated 3D pharmacophore model and, starting from compounds 1e and 2c[29] (Scheme Figure 3.1.3, top panels), we designed two new molecular series 1(a-r) and 2(a-l) (Scheme 3.1.3, bottom panels) with σ_1 receptor affinity σ_1Ki ranging from sub-nanomolar to micromolar values, as shown in Table 3.1.3.



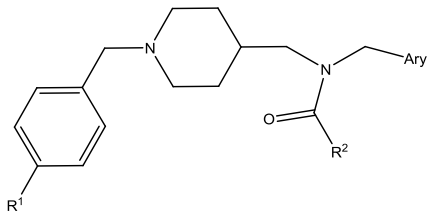
1e



2c



1(a-r)



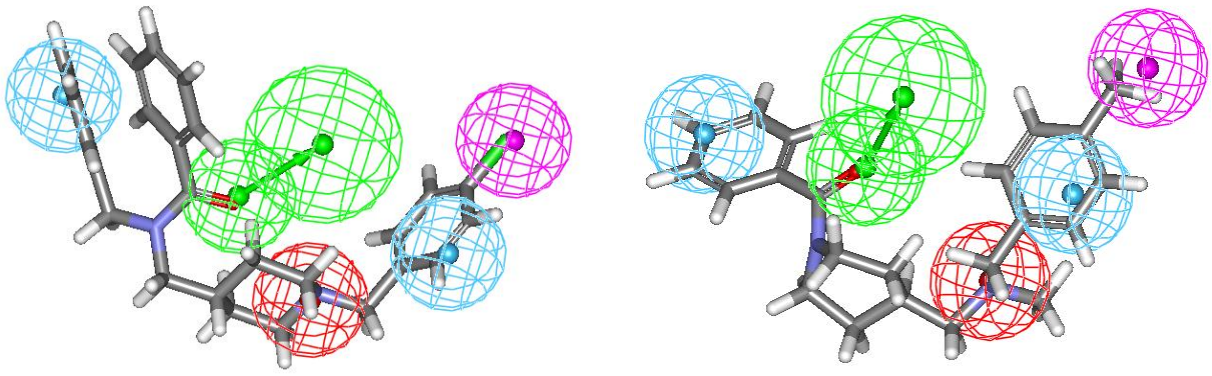
2(a-l)

Scheme 3.1.3 *(Top) Chemical structure of compounds 1e and 2c. (Bottom) Molecular structures of compounds of series 1 and 2.*

Table 3.1.3 *3D Pharmacophore Predicted σ_1 Receptor Affinities σ_1Ki,(3DPh) of Compounds 1a-r and 2a-l.*

cmpd	Ary	R^1	$\sigma_1K_{i,3DPh}$ (nM)	cmpd	Ary	R^1	R^2	$\sigma_1K_{i,3DPh}$ (nM)
1a	phenyl	H	228	**1p**	pyridin-2-il	Bz	-	810
1b	4-chlorophenyl	H	32.5	**1q**	pyridin-3-il	Bz	-	1961
1c	4-methylphenyl	H	116	**1r**	pyridin-4-il	Bz	-	785
1d	phenyl	CH_3	27.0	**2a**	phenyl	H	CH_3	24.1
1e	4-chlorophenyl	CH_3	66.7	**2b**	phenyl	H	phenyl	75.2
1f	4-methylphenyl	CH_3	43.7	**2c**	phenyl	Cl	CH_3	1.04
1g	phenyl	Bz	32	**2d**	phenyl	Cl	phenyl	9.7
1h	4-chlorophenyl	Bz	83	**2e**	4-chlorophenyl	H	CH_3	0.26
1i	4-methylphenyl	Bz	462	**2f**	4-chlorophenyl	H	phenyl	37.1
1j	pyridin-2-il	H	520	**2g**	4-chlorophenyl	Cl	CH_3	20.4
1k	pyridin-3-il	H	4433	**2h**	4-chlorophenyl	Cl	phenyl	25.7
1l	pyridin-4-il	H	1148	**2i**	4-methylphenyl	H	CH_3	6.8
1m	pyridin-2-il	CH_3	1186	**2j**	4-methylphenyl	H	phenyl	319
1n	pyridin-3-il	CH_3	2683	**2k**	4-methylphenyl	Cl	CH_3	86.5
1o	pyridin-4-il	CH_3	3731	**2l**	4-methylphenyl	Cl	phenyl	86.6

As can be seen from this Table, both series of compounds are indeed characterized by a broad range (5 decades) of predicted σ_1Ki values; also, quite interestingly, compounds belonging to series 1 are generally endowed with a lower affinity towards the σ_1 receptor than the series 2-based counterparts. This aspect can be easily explained by comparing, for instance, the mapping of two of the most active compounds of both series – i.e., 1f (σ_1Ki,3DPh = 43.7 nM) and 2d (σ_1Ki,3DPh = 09.7 nM), respectively – on the corresponding 3D pharmacophore features, as shown in Figure 3.1.12.



A) B)

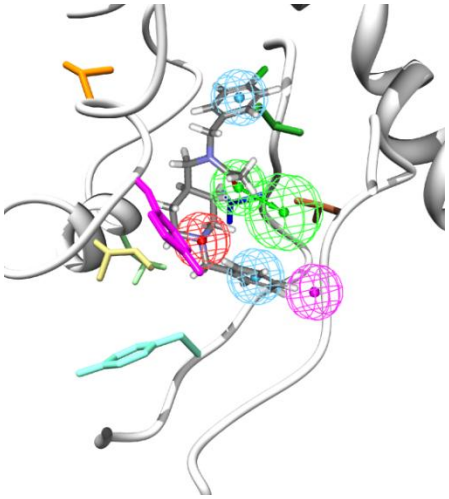
Figure 3.1.12 *Pharmacophore mapping of 2d (A) and 1f (B). The 3D pharmacophore hypothesis features are portrayed as meshed spheres, color-coded as follows: red, PI; light blue, HYAr; pink, HY; light green, HBA.*

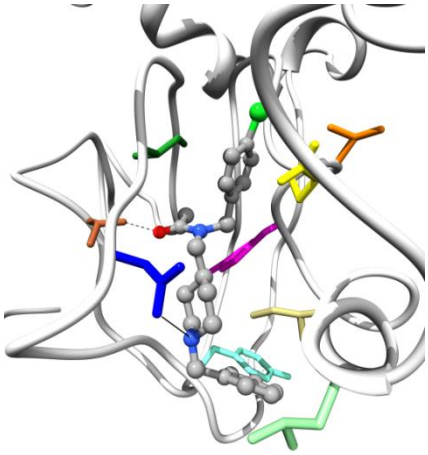
It might be apt to recall here that the best 3D pharmacophore hypothesis for our σ_1 ligands was characterized by the following five chemical features: one positive ionizable (PI) site, one

hydrogen bond acceptor (HBA) group, two hydrophobic aromatic (HYAr) moieties, and one further hydrophobic (HY) site. Thus, when superposed on the pharmacophore model, compound 2d maps all the chemical features with perfect overlap (Figure 3.1.12(A)): the phenyl ring match the aromatic hydrophobic function, the chlorine atom fills the hydrophobic group of the model, the basic nitrogen atom has the function of proton acceptor, while the carbonyl group matches the hydrogen bond feature. On the other hand, the chemical groups present on compound 1f, although suitable to fulfill all chemical requirements of the 3D pharmacophore, cannot be perfectly superposed to the corresponding pharmacophore features due to a different molecular conformation and steric rigidity characterizing this molecule (Figure 3.1.12(B)).

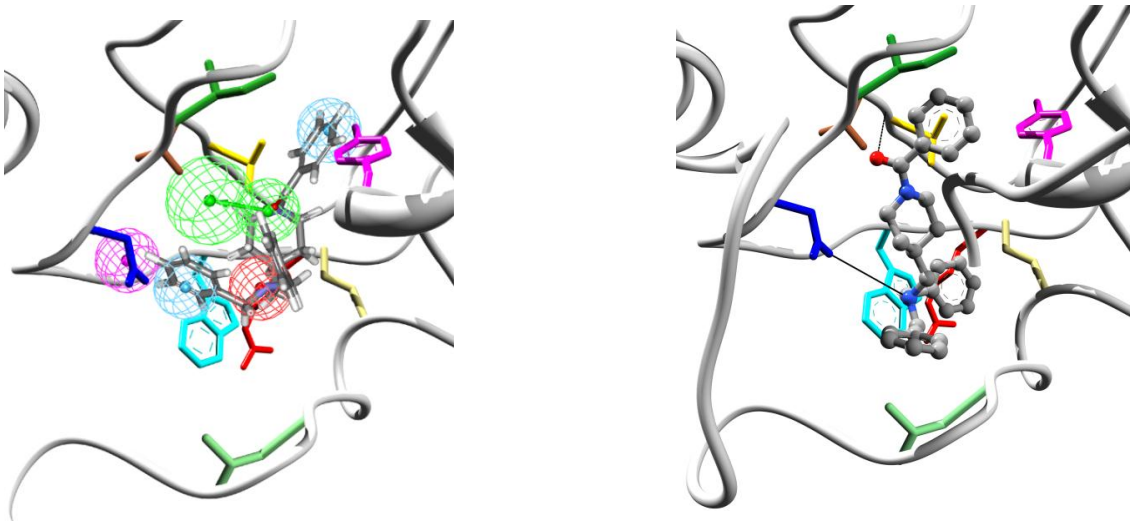
Assisted-ligand docking into the σ_1 receptor 3d homology model

Following our pioneer work, the putative binding site and binding modes of all compounds from both series 1 and 2 in the σ_1 receptor 3D homology model structure were retrieved exploiting the currently available preliminary information on sequence-structure relationships and mutagenesis studies, the ligand-binding pharmacophore requirements, and the docking poses and receptor affinity ranking of compounds 1e and 2c.[5] Thus, all compounds 1a-r and 2a-l were docked into the putative binding site of receptor 3D model. In the set of docked ligand conformations, for each compound a solution was found that best reproduced the key 3D pharmacophore requirements (see Figure 3.1.13). Each resulting receptor/drug complex was then relaxed by energy minimization and molecular dynamics (MD) simulations and, finally, the relevant values of the free energy of binding - ΔGbind - between all compounds of series 1 and 2 and the σ_1 receptor were evaluated by applying the well-known Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) computational ansatz[7,30] as listed in Table 3.1.4.

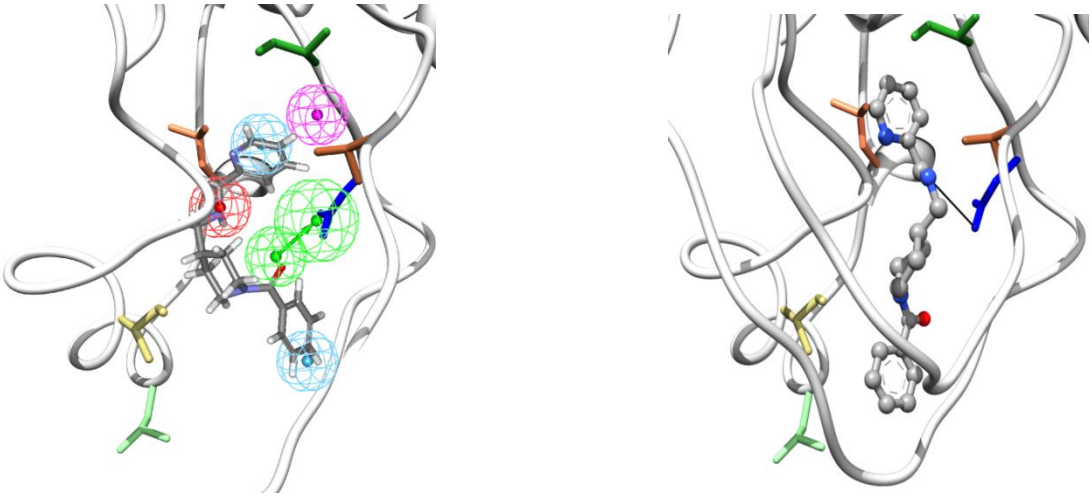




(A) (B)



(C) (D)



(E) (F)

Figure 3.1.13 *(A, C, and E) Zoomed view of the 3D pharmacophore-assisted identification of the binding modes of 2e, 1g, and 1j in the putative binding site of the σ_1 receptor. The main protein residues involved in ligand/receptor interactions are Arg119 (red), Tyr120 (aquamarine), Trp121 (cyan), Asp126 (blue), Ile128 (forest green), Thr151 (sienna), Val152 (gold), Val 171 (orange), Glu172 (yellow), Tyr173 (magenta), Ile178 (khaki), Leu182 (light green), and Leu186 (coral). (B, D, and F) Details of the key interactions detected in the equilibrated MD snapshots of 2e, 1g, and 1j in complex with the σ_1 receptor 3D homology model. In the left panels, compounds 2e, 1g and 1j are shown in atom-colored sticks: C, gray; O, red; N, blue, Cl, green, H, white. The 3D pharmacophore features are shown as in Figure 2. In the right panels, the same compounds are portrayed as atom-colored sticks-and-balls; in this case, H atoms are not shown but H-bonds and salt bridges are indicated as black dotted and continuous lines, respectively. In all images, water molecules, ions and counterions are not shown for clarity.*

Table 3.1.4 *Enthalpy (ΔH), entropy (-TΔS), free energy of binding (ΔGbind), and corresponding σ_1Ki, ΔGbind values for compounds 1a-r and 2a-l and the σ_1 receptor homology model, as estimated using the MM/PBSA approach. The values of two σ_1 ligand reference compounds, (+)pentazocine (PTZ) and haloperidol (HAL) are also reported for comparison. Errors are given in parenthesis as standard errors of the mean (SEM).*

cmpd	ΔH (kcal/mol)	-TΔS (kcal/mol K)	ΔG_{bind} (kcal/mol)	$\sigma_1 K_{i\Delta G_{bind}}$ (nM)	cmpd	ΔH (kcal/mol)	-TΔS (kcal/molK)	ΔG_{bind} (kcal/mol)	$\sigma_1 K_{i\Delta G_{bind}}$ (nM)
1a	-36.29 (0.19)	-26.36 (0.34)	-9.93 (0.39)	53.0	**1p**	-31.89 (0.16)	-23.03 (0.37)	-8.86 (0.40)	322
1b	-34.39 (0.21)	-23.87 (0.32)	-10.52 (0.38)	19.6	**1q**	-30.92 (0.18)	-23.67 (0.34)	-7.25 (0.38)	4870
1c	-36.18 (0.17)	-25.75 (0.31)	-10.43 (0.35)	22.8	**1r**	-29.97 (0.20)	-22.11 (0.33)	-7.86 (0.39)	1740
1d	-35.02 (0.18)	-24.62 (0.35)	-10.40 (0.39)	24.0	**2a**	-38.31 (0.20)	-25.81 (0.35)	-12.50 (0.40)	0.69
1e	-37.90 (0.18)	-26.71 (0.30)	-10.02 (0.35)	6.32	**2b**	-39.01 (0.15)	-26.19 (0.36)	-12.82 (0.39)	0.40
1f	-36.21 (0.20)	-25.09 (0.31)	-11.12 (0.37)	7.12	**2c**	-38.81 (0.19)	-25.54 (0.33)	-11.31 (0.38)	0.19
1g	-33.19 (0.16)	-23.24 (0.35)	-9.95 (0.38)	51.2	**2d**	-38.39 (0.19)	-24.91 (0.33)	-13.48 (0.38)	0.13
1h	-32.60 (0.20)	-23.39 (0.36)	-9.21 (0.41)	178	**2e**	-40.83 (0.21)	-25.71 (0.30)	-15.12 (0.37)	0.01
1i	-33.38 (0.19)	-24.37 (0.33)	-9.01 (0.38)	250	**2f**	-37.76 (0.17)	-25.97 (0.35)	-11.79 (0.39)	2.30
1j	-30.33 (0.18)	-24.49 (0.35)	-5.84 (0.39)	52600	**2g**	-38.33 (0.19)	-26.38 (0.34)	-11.95 (0.39)	1.75
1k	-29.89 (0.21)	-23.02 (0.31)	-6.87 (0.37)	9250	**2h**	-37.04 (0.18)	-25.11 (0.36)	-11.93 (0.40)	1.81
1l	-30.69 (0.17)	-22.43 (0.36)	-8.26 (0.40)	880	**2i**	-38.61 (0.19)	-26.09 (0.35)	-12.52 (0.40)	0.67
1m	-29.88 (0.21)	-23.35 (0.31)	-6.53 (0.37)	16400	**2j**	-33.95 (0.18)	-24.35 (0.32)	-9.60 (0.37)	92.5
1n	-29.24 (0.18)	-22.65 (0.32)	-7.75 (0.37)	14800	**2k**	-34.61 (0.19)	-25.03 (0.31)	-9.58 (0.36)	95.6
1o	-28.87 (0.22)	-22.27 (0.32)	-6.60 (0.39)	14600	**2l**	-34.49 (0.21)	-24.62 (0.35)	-9.87 (0.41)	58.6
PTZ	-27.56 (0.22)	-17.54 (0.39)	-10.02 (0.45)	45.5	**HAL**	-35.17 (0.19	-24.56 (0.38)	-10.61 (0.42)	16.8

Figure 3.1.13 shows three MD snapshots extracted from the corresponding equilibrated trajectories of 2e, 1g, and 1j in complex with the σ_1 receptor, as an example. In particular, the left-hand panels in Figure 3.1.13 clearly indicate that the predicted binding mode for each of these compounds satisfies the 3D pharmacophore requirements. Among both series, 2e is the molecule characterized by the highest MM/PBSA predicted affinity toward the σ_1 receptor, σ_1Ki, ΔGbind = 0.01 nM (see Table 3.1.4), in agreement with the corresponding estimation from the 3D pharmacophore model (σ_1Ki,3DPh = 0.26 nM, Table 3.1.3). From the right panel of

Figure 3.1.13 it is instructive to observe how, for 2e, an hydrogen bond (HB) is established between the carbonyl oxygen of the ligand and the side-chain –OH group of Thr151 on the receptor. This HB, which persists throughout the entire MD simulation period, is characterized by an average dynamic length (ADL) of 1.82 ± 0.21 Å, yield a substantial contribution to binding. A permanent salt bridge (SB, ADL = 3.81 ± 0.10 Å) is detected between the –COO- group of Asp126 on the receptor side and the piperidine –NH+ moiety of the 2e ligand. The 4-Cl-phenyl group of the compound is encased in a hydrophobic pocket, mainly lined by the side chains of Val171 and Ile128. An outer hydrophobic region accommodates the second phenyl ring of 2e, with the contribution of residues Leu182, Ile178 and Tyr120. Also, this phenyl ring is further engaged in a parallel $\pi-\pi$ stacking interaction with Tyr120. The synergistic effect of all these stabilizing interactions ultimately reflects in the highly negative (i.e., favorable) value of ΔGbind (-15.12 kcal/mol, Table 3.1.4) between 2e and the σ_1 protein and, hence, of the predicted sub-nanomolar value of the corresponding σ_1Ki, ΔGbind.

For compound 1g, MM/PBSA calculations yield an intermediate affinity for the σ_1 receptor - σ_1Ki, ΔGbind = 51.2 nM (see Table 3.1.4) – a value quite close to that predicted by the 3D pharmacophore model (σ_1Ki3DPh = 32 nM, Table 3.1.3). Panels (C) and (D) in Figure 3.1.13 illustrate how the stable hydrogen bond and the salt bridge outlined for 2e are still present, although the hydrogen donor on the receptor side in this case is the –NH group of the backbone peptide bond between Thr151 and Val152. Both these interactions, however, are somewhat weaker than for 2e, as the corresponding ADL are 2.09 ± 0.12 Å (HB) and 4.53 ± 0.18 Å (SB), respectively. The hydrophobic pockets again enwrap two of the aromatic rings, residue Trp121 playing a role in determining a T-stacked $\pi-\pi$ interaction with one of the two aryl groups. However, the conformation of the molecule is such that last phenyl ring is not mapped by any suitable pharmacophore feature of 1g (Figure 3.1.13(C)); correspondingly, this interaction is lost within the receptor binding site (Figure 3.1.13(D)). In line with this analysis, the presence of these less effective, albeit still favorable interactions, reflect in the higher (less negative) value of the estimated free energy of binding (ΔGbind = -9.95 kcal/mol, Table 3.1.4).

According to both 3D pharmacophore modeling and MM/PBSA simulations, compound 1j is endowed with the lowest affinity towards the σ_1 receptor, with Ki value of σ_1Ki3DPh = 520 nM and σ_1Ki, ΔGbind = 52600 nM, respectively. In this case, the role exerted by Tyr120 and Thr151 in binding 1j is negligible, while the two hydrophobic pockets described above are still able to accommodate the two aromatic moieties of this compound. More importantly, the carbonyl group is positioned far too distant from any possible proton donor group on the receptor to set up any stabilizing HB bond. In agreement with these considerations, a very low activity is predicted for this compound, with a corresponding ΔGbind value of -5.84 kcal/mol.

Generally speaking, according to the MM/PBSA analysis the binding affinity of the two series of compounds 1 and 2 towards the σ_1 receptor is an enthalpy-driven process: panel (A) in Figure 3.1.14 indeed shows that for both molecular sets the entropic contribution $T\Delta S$ to ΔGbind – unfavorable to ligand binding – is overwhelmed by the favorable enthalpic component ΔH, resulting in an overall negative value of the free energy of binding ΔGbind. It is also interesting to note that the higher MM/PBSA average σ_1 affinity value predicted for all compounds of series 2 originates solely from the difference in the enthalpic contributions between the two molecular sets ($\Delta\Delta H = \Delta H(2) - \Delta H(1) = -4.91$ kcal/mol), as the corresponding difference in the entropic terms is significantly smaller ($\Delta T\Delta S = T\Delta S(2) - T\Delta S(1) = 1.56$ kcal/mol).

The calculated binding free energy values are encouraging also in the light of the balance of the energy terms contributing to them. In more detail, as observed in many other drug/receptor complex simulations including our own studies,[31] the favorable contribution of the electrostatic interactions between the σ_1 protein and both series of compounds 1 and 2 (ΔEele), is more than compensated by the electrostatic desolvation free energy upon complexation (ΔGPB), so that the total electrostatic term contributes unfavorably to the binding (see Figure 3.1.14(B)). Interestingly, several recent papers have demonstrated that (natural) receptor-ligand pairs often show suboptimal electrostatic interactions that may be optimized, leading to increased affinity.[32-35] On the contrary, van der Waals interactions (ΔEvdW) contribute favorably to the binding affinity of the receptor, as does the nonpolar part of the solvation free energy, again in agreement with other studies mentioned above[30,31,36] (Figure 3.1.14(B)). Therefore, for both 1 and 2 molecular series, the favorable binding free energy for receptor/ligand complex formation stems predominantly from the nonpolar terms (ΔEvdW + ΔGnp), while the polar interactions provide most of all the directional constraint for the complexation, that is, the relative positions of the molecules.

The analysis of the entire MD simulation trajectories for all 33 compounds in complex with the σ_1 receptor 3D model further reveals that the overall conformation of the protein backbone undergoes only minima global conformational changes upon complex formation with the different compounds, while a rearrangement of the side chains of several residues lining the receptor binding site is required for ligand binding. This is revealed by the superposition of the average structures of all trajectories with their respective starting structures.

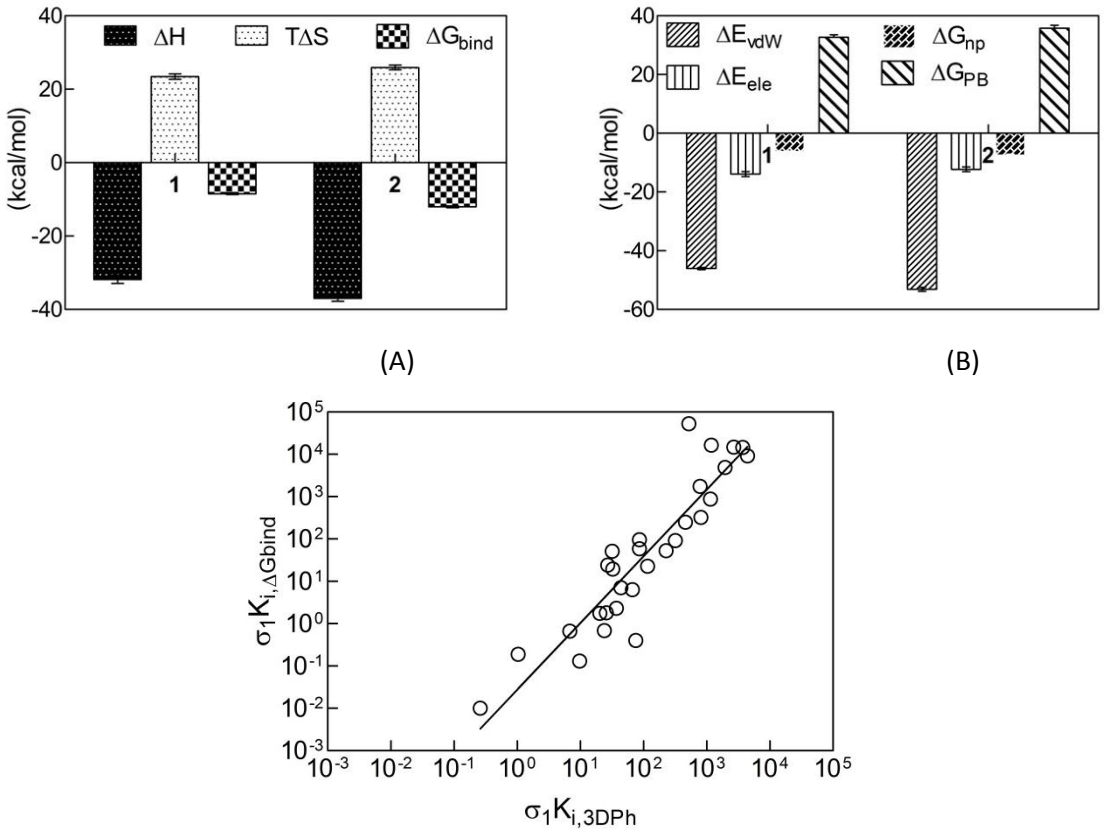


Figure 3.1.14. *(A) Average MM/PBSA values of the enthalpy (ΔH), entropy (TΔS), and free energy of binding (ΔGbind) for compound series 1 and 2 in complex with the σ_1 receptor. (B) Average values of the enthalpic components of ΔGbind (van der Waals term ΔEvdw, electrostatic term ΔEele, polar solvation term ΔGPB, and nonpolar solvation term ΔGnp) for compound series 1 and 2 in complex with the σ_1 receptor. (C) Correlation between the σ_1 affinity constant Ki values for compounds 1a-r and 2a-l estimated using the 3D pharmacophore model (σ_1Ki,3DPh) and the MM/PBSA methodology (σ_1Ki, ΔGbind).*

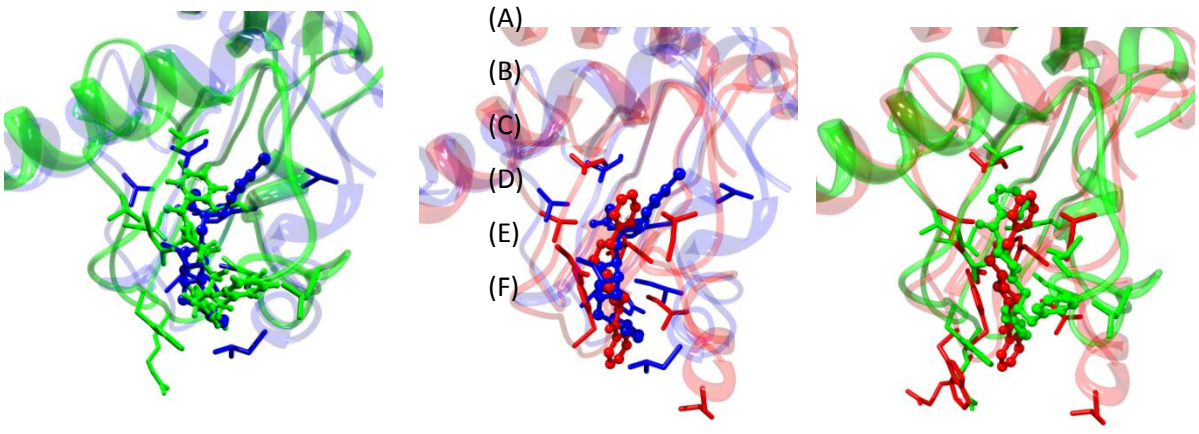
In its uncomplexed form, the σ_1 model structure remained stable for the entire 10 ns MD trajectory, as revealed by the small fluctuations of the root-mean-square (RMS) deviation of the simulated position of the backbone atoms with respect to those of the initial structure and the corresponding total potential energy of the system. The same parameters showed very low fluctuations during both MD equilibration and data harvesting steps also in the cases of all receptor/ligand complexes, indicating that the presence of a bound ligand does not result in large protein structural deviations. The first part of Table 3.1.5 reports the average RMS deviations of the binding site region of the σ_1 receptor (i.e., from residue 100 to residue 200) determined between average structures of the proteins in the unbound and bound states for 2e, 1g, and 1j as an example: this region deviates by very small amounts, confirming that the

σ_1 binding pocket does not experience a significantly larger-than-average conformational change upon complex formations with ligands 1 and 2.

Table 3.1.5 *Average root-mean-square deviations (RMSDs) of the binding site of the σ_1 protein in complex with 2e, 1g, and 1j with respect to the unbound protein and to each alternative complex. All values are reported in Å.*

	2e	**1g**	**1j**
σ_1(bound/unbound)	0.9	1.6	2.1
	2e/1g	**2e/1j**	**1g/1j**
	2.5	4.2	3.5

This can be further visually inspected and quantified considering the superposition of equilibrated snapshots extracted from the MD trajectory of the receptor in complex with compounds 2e, 1g, and 1j reported in Figure 3.1.15 and the relevant RMSDs values listed in the second part of Table 3.1.5. These evidences allows us to conclude that the present binding site of the σ_1 receptor is able to accommodate all ligands of series 1 and 2 with no major conformational readjustments, the difference in affinity towards the different compounds being ascribable to a better/worse re-arrangement of the binding pocket residue side chains.



(A) (B) (C)

Figure 3.1.15 *Superposition of equilibrated MD snapshots of the σ_1 receptor in complex with (A) 2e (blue) and 1g (green), (B) 2e (blue) and 1j (red), and (C) 1g (green) and 1j (red). The images are zoomed views of the receptor binding site. The ligands are portrayed in sticks-and-ball colored according to the protein in the corresponding complex. Water, ions and counterions are not shown for clarity.*

In concluding this section, it is interesting to observe that the affinity values of all 33 compounds towards the σ_1 receptor as ranked by the 3D pharmacophore model are in good

agreement with the corresponding values obtained from the MM/PBSA scoring, with a correlation coefficient of R^2 = 0.84, as shown in Figure 3.1.14(C). The great quality of the overall linear correlation between these two Ki data sets undoubtedly constitutes the first step toward the definitive validation of the σ_1 3D homology model and the location of its putative binding site.

Residue-based description of ligand binding to σ_1: per residue binding free energy decomposition and computational alanine scanning

Per residue binding free energy decomposition

Insight into the origin of binding of σ_1 to compounds 1 and 2 at an atomistic level may be obtained by decomposing the total free energy of binding ΔGbind in terms of contributions from structural subunits of both binding partners. A non-perturbing alternative approach related to MM-PBSA, where the electrostatic contribution to the solvation free energy is determined using a generalized Born model and, hence, known as Molecular Mechanics/Generalized Born Surface Area (MM/GBSA),[36] is an attractive variant to MM/PBSA in that the GB model is much faster than the PB approach and allows the decomposition of the electrostatic solvation free energy into atomic contributions in a straightforward manner. This, in turn, allows an easy and rapid per residue binding free energy decomposition (PRBFED), yielding the residue-based ΔHGB values required for the detailed study of the ligand/protein interactions at each single amino acid level, including the backbone atoms. Therefore, we proceeded in our analysis of the binding modes of compound series 1 and 2 to the σ_1 receptor by applying PRBFED to the analysis of some residues that , as qualitatively discussed above (Figure 3.1.13), are predicted to be important for the binding of these ligand to the protein.

Figure 3.1.16 illustrates the results of the PRBFED analysis obtained for compounds 2e, 1g, and 1j, again taken as proof-of-concept as they constitute an example of high affinity (2e), intermediate affinity (1g), and very low affinity (1j) receptor ligands. As can be seen from this Figure, in all cases three clusters of residues (I, II, and III) are found, centered around Glu123, Thr151, and Val177. In the case of 2e (Figure 3.1.16(A)), according to analysis of the corresponding MD trajectory (Figure 3.1.13(A) and (B)) the side chain of Asp126 is engaged in a fundamental salt bridge with the piperidine –NH+ moiety of the compound; indeed, this residue is responsible for the -2.54 kcal/mol favorable contribution to binding, mostly provided by stabilizing electrostatic interactions. The stabile hydrogen bond detected between the –CO group of 2e and the side chain –OH group of Thr151 is responsible for the favorable -1.31 kcal/mol contribution provided by the electrostatic interactions.

The PRBFED approach then confirms that residues belonging to the two major clusters yield the required van der Waals and hydrophobic interactions to favorably encase the aromatic portions of the ligand: e.g., Arg119 (-1.58 kcal/mol), Tyr120 (-2.16 (kcal/mol), Trp121 (-1.76 kcal/mol), and Ile128 (-3.15 kcal/mol) of cluster I, and Glu172 (-1.73 kcal/mol)), Tyr 173 (-1.91kcal/mol), and Leu182 (-1.04 kcal/mol) of cluster II, respectively.

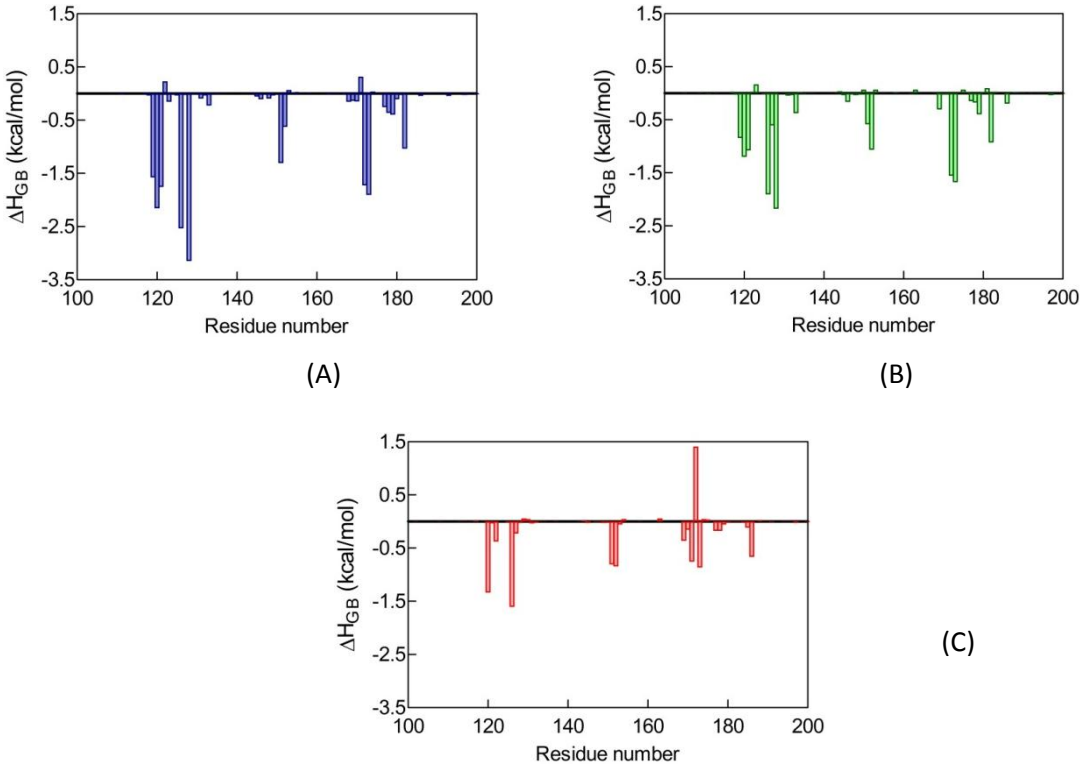


Figure 3.1.16 *Per residue binding free energy decomposition for σ_1 in complex with 2e (A), 1g (B), and 1j (C). Only σ_1 amino acids from position 100 to 200 are shown, as for all the remaining protein residues the contribution to ligand binding is irrelevant.*

Panel (B) in Figure 6 illustrates the PRBFED results for compound 1g. The presence of weaker (i.e., longer) salt bridge and hydrogen bond interactions detected along the 1g/σ_1 MD trajectory (Figure 3(C) and (D)) is confirmed by the lower ΔHGB values of the corresponding residues: -1.91 kcal/mol (SB) for Asp126 and -1.07 kcal/mol (HB) for Val152, respectively. The minor entity of the contributions to binding from residues belonging to clusters I and II (120-128 and 171-182) support the evidence that part of the hydrophobic and van der Waals interactions are lost in this ligand/protein complex. Of note, the proposed T-stacked π–π interaction of Trp121 with one of the two aryl groups of 1g is captured by the PRBFED analysis, according to which the contribution afforded by this residue to binding is equal to -1.08 kcal/mol.

According to our predictions, compound 1j is the ligand characterized by the lowest affinity for σ_1 in both series 1 and 2. The PRBFED results shown in panel (C) of Figure 3.1.16 clearly justify these calculations. While the presence of a salt bridge between the –NH+ moiety of 1j and the –COO- of Asp126 is preserved, which provides a favorable contribution of -1.61 kcal/mol, the

absence of any H-bond made by the –CO group of 1j with possible acceptor groups on the receptor is responsible for the small contributions of residues from cluster II (e.g., Thr151, -0.81 kcal/mol). Some residues of clusters I and II are still able to provide some favorable contribution to the binding, although the reorientation of Glu172 upon binding 1j is in part responsible for the +1.41 kcal/mol unfavorable dispersion/electrostatic contribution to the binding made by this side chain.

To gain additional insights into the contributions to the binding free energy change, Figure 3.1.17 depicts the decomposition of ΔHGB values on a per residue basis into contributions from the nonpolar terms (i.e., the van der Waals energy ΔEvdW and the nonpolar term of the solvation free energy ΔGnp) and the sum of the Coulombic interaction and the polar solvation free energy (ΔEele + ΔGGB) for residues with |ΔHGB| ≥ 1 kcal/mol for the three σ_1 complexes discussed above. The sum of electrostatic interactions in the gas phase plus the change of the polar part of the solvation free energy is shown instead of the separate contributions, since, in most of the cases, the numbers are strongly anti-correlated.

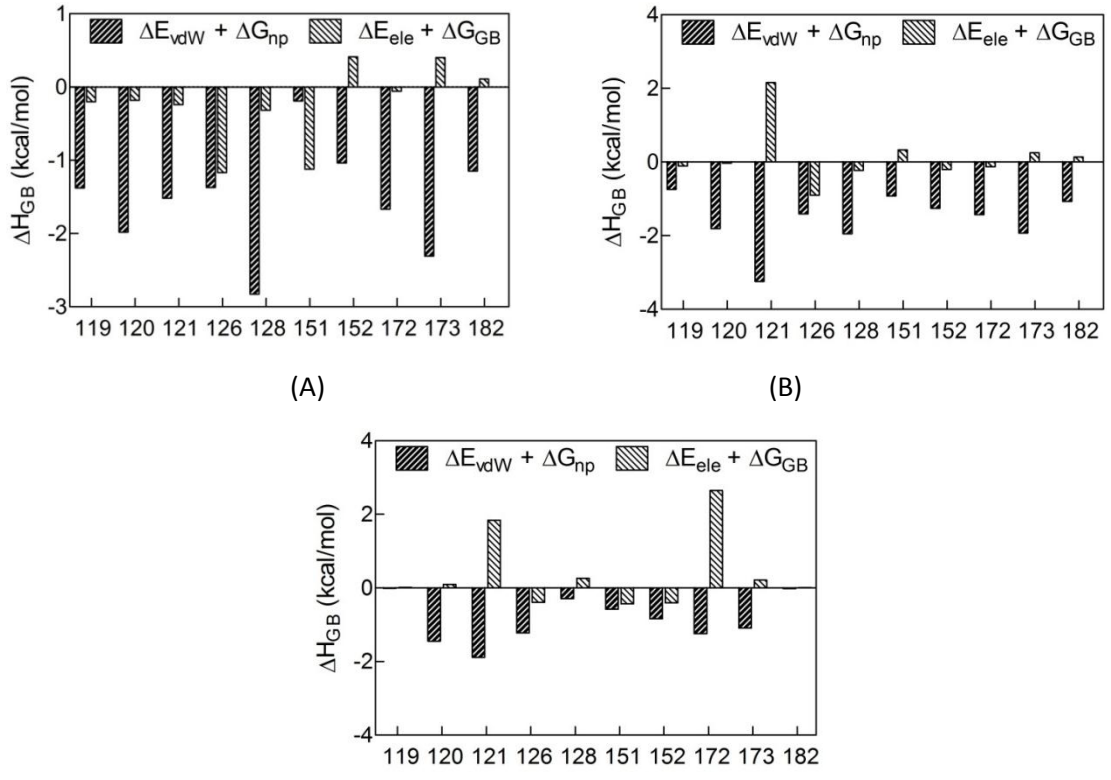


Figure 3.1.17 *Decomposition of ΔGGB on a per residue basis into contribution of the nonpolar (ΔEvdW + ΔGnp) and polar (ΔEele + ΔGnp) terms for residues of σ_1 in complex with 2e (A), 1g (B), and 1j (C) for which |ΔHGB| ≥ 1 kcal/mol.*

Qualitatively, major differences are obvious between residues located in the binding site in the case of high, intermediate, and low affinity ligands, respectively. For compound 2e (σ_1Ki, ΔGbind = 0.01 nM), stabilizing van der Waals and nonpolar interactions, reinforced by favorable overall electrostatic/desolvation terms, prevail for almost all residues lining the σ_1 binding pocket (Figure 3.1.17(A)). Interestingly, the overall stabilization of the salt bridge of 2e with Asp126 (ΔGGB = -2.54 kcal/mol, see PRBFED analysis) is almost equivalently contributed by the dispersive (ΔEvdW + ΔGnp = -1.37 kcal/mol) and electrostatic (ΔEele + ΔGGB = -1.17 kcal/mol) components, while the pharmacophoric H-bond involving the side chain of Thr151 (ΔGGB = -1.31 kcal/mol) mainly gains from a favorable electrostatic interaction (ΔEele + ΔGGB = -1.12 kcal/mol while ΔEvdW + ΔGnp = -0.19 kcal/mol). Figure 3.1.17(A) also confirm the dominant role played by the hydrophobic (i.e., overall nonpolar) interactions in binding of 2e to σ_1, well exemplified by the values of ΔEvdW + ΔGnp for residues Arg119 (-1.38 kcal/mol, ΔHGB = -1.58 kcal/mol), Tyr120 (-1.98 kcal/mol, ΔHGB = -2.16 kcal/mol), Trp121 (-1.52 kcal/mol, ΔHGB = -1.76 kcal/mol), Ile128 (-2.83 kcal/mol, ΔHGB = -3.15 kcal/mol), Glu172 (-1.67 kcal/mol, ΔHGB = -1.73 kcal/mol), Tyr 173 (-2.31 kcal/mol, ΔHGB = -1.91 kcal/mol), and Leu182 (-1.15 kcal/mol, ΔHGB = -1.04 kcal/mol).

In the case of compound 1g (σ_1Ki, ΔGbind = 51.2 nM), the dispersive forces benefit by a lower synergistic interaction with the polar terms of ΔHGB with respect to the case of 2e (Figure 3.1.17(B)). Notably, while the salt bridge between 1g and Asp126, although increased in length and, thus, decreased in strength (ΔHGB = -1.91 kcal/mol) shows a relative contribution from the different ΔHGB components which parallels that discussed for 2e, (i.e., ΔEvdW + ΔGnp amounts to approximately 74% of the total ΔHGB), the other, distinctive pharmacophoric element (i.e., H-bond with Val152) is not only weaker but features contributions from the two main ΔHGB components in reverse trend with respect to 2e (ΔEvdW + ΔGnp = -1.26 kcal/mol and ΔEele + ΔGGB = -0.20 kcal/mol, respectively). Of importance is the quantification of the $\pi-\pi$ stacking interaction between one aromatic moiety of 1g and the side chain of Trp121 (ΔHGB = -1.08 kcal/mol), for which the dispersive and the electrostatic interactions show strong contributions of opposite sign (ΔEvdW + ΔGnp = -3.24 kcal/mol and ΔEele + ΔGGB = +2.16 kcal/mol, respectively).

Lastly, for compound 1j (σ_1Ki, ΔGbind = 52600 nM) not only the number of useful contacts with the amino acids belonging to the putative σ_1 binding site, but also the overall intensities of the interactions between these residues and the ligand are highly decreased with respect to the two compounds discussed above. Also, at some specific positions already highlighted

during PRFEBD analysis (e.g., Trp121 and Glu172), a gain in favorable van der Waals interactions and the nonpolar part of solvation free energy is overcompensated by unfavorable contributions from the ΔEele + ΔGGB components of ΔHGB.

Computational alanine scanning

The MD simulations performed in the MM/PBSA framework of theory can be further exploited to perform the so-called computational alanine scanning (CAS)[37,38], in which the absolute binding free energy is calculated for the wild type protein, as well as for several mutants in which one residue has been replaced by an alanine. Aside yielding complementary information to those obtained from a PRBFED analysis, the difference in the binding free energy of the wild type and of the mutants estimated by CAS may be compared directly to the results of an experimental alanine scanning (ASM) mutagenesis. Undoubtedly, in the CAS approach it is questionable whether simply modifying a given side chain to alanine in the corresponding MD simulation trajectory of the wild type system can lead to a good representation of the conformational space of the mutant, since no eventual conformation induced by the mutation is investigated. However, it is also questionable how the binding free energy contribution of a given side chain in the wild type complexation may be always representative of the change in the binding free energy upon mutation, since the conformational modifications induced by the mutations are not included in the model either. Also, on the contrary to the total free energy, the free energy components are not state functions, and the values of these contributions are thus dependent on the decomposition scheme adopted. Obviously, both CAS and the PRBFED methods cannot be expected to provide results exactly comparable to experimental values obtained from an experimental ASM; nonetheless, the application of both methodologies can give a good, preliminary indication of which protein residues play a key role in ligand binding, ultimately enabling the biochemist to avoid trial-and-error tests and perform targeted ASM experiments with the obvious advantages of cost and time saving.

The CAS was applied to the entire set of molecules belonging to both classes 1 and 2; for the sake of brevity and in keeping with the previous discussion, Table 3.1.6 gives the results for the CAS for compounds 2e, 1g, and 1j. Note that, according to the definition adopted in this work, a negative value of ΔΔGbind value corresponds to a residue for which the wild type (wt) side chain is more favorable to the binding than an alanine side chain. As a first, important observation, from the values listed in Table 3.1.6 the pivotal role exerted by Asp 126 in ligand binding is clearly highlighted by the highly unfavorable free energy of binding of the Ala126 mutant with respect to the wt protein. Also, Tyr173 is confirmed to play a substantial role in the complex stabilization for each compound. Interestingly, residues Ile128 and Leu182 afford

a significant contribution to the stabilization of the protein/ligand complex for those ligands with high or intermediate affinity (2e and 1g) but seem to be less critical for compounds endowed by a poor affinity for the receptor.

Table 3.1.6 *Computational alanine scanning (CAS) mutagenesis results for the σ_1 receptor in complex with ligands 2e, 1g, and 1j. Errors are given in parenthesis as standard errors of the mean (SEM).*

		$\Delta\Delta G_{bind} = \Delta G_{bind,wt} - \Delta G_{bind,mut}$						
cmpd	$\Delta G_{bind,wt}$	**D126A**	**I128A**	**T151A**	**V152A**	**E172A**	**Y173A**	**L182A**
2e	-15.12 (0.37)	-3.69 (0.42)	-2.27 (0.43)	-0.67 (0.38)	-0.56 (0.40)	-2.04 (0.37)	-1.79 (0.38)	-1.11 (0.37)
1g	-9.95 (0.38)	-3.25 (0.39)	-1.35 (0.40)	-0.71 (0.39)	-0.45 (0.39)	-1.19 (0.40)	-2.03 (0.41)	-0.79 (0.43)
1j	-5.84 (0.39)	-2.82 (0.43)	-0.12 (0.42)	-0.57 (0.38)	-0.39 (0.43)	-0.21 (0.39)	-1.27 (0.38)	0.02 (0.44)

The importance of these hot spot residues detected by CAS can be verified by experimental binding assays of various σ_1 mutants. For instance, in their seminal work Seth et al. clearly proved, via in vitro binding assays of the Asp126Gly and Glu172Gly σ_1 mutants to radiolabeled haloperidol, the obligatory nature of the fully conserved Asp126 and Glu172 for the ligand binding function of the σ_1 receptor[39]. Also, other mutational studies identified Tyr173 as a residue belonging to the σ_1 binding pocket critical for the cholesterol binding activity of the protein.[40]

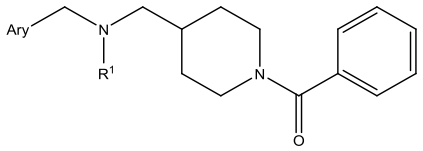
To summarize all in silico work discussed above, we used extensive MM/PBSA-based analysis to rank all 33 ligands belonging to series 1 and 2 according to their affinity towards our recently proposed 3D model of the σ_1 receptor. The σ_1Ki values derived from the MM/PBSA simulations are in excellent agreement (R^2 = 0.84) with those obtained using a 3D pharmacophore model, previously shown to be extremely reliable in reproducing and/or predicting the affinity of similar compounds to the same receptor. Lastly, the combined application of a per residue free energy deconvolution and computational alanine scanning experiments allowed to dissect the contribution of each single residue belonging to the putative σ_1 receptor binding pocket to ligand binding, yielding vital information for further drug desing and development. Furthermore, and perhaps more importantly in the perspective of the present manuscript, those σ_1 residues experimentally found to be vital for the ligand binding activity of the receptor were also found critical in our 3D model, according to our PRFEBD and CAS simulations.

Synthesis and activity of new σ_1 ligands, and comparison with in silico predictions

On the basis of the estimated σ_1 receptor affinity predicted by our complex computer-assisted drug design recipe presented above, the two series of phenylmethanone (1a-r) and amide (2a-l) derivatives were then synthesized, (see details about synthesis in paper[29]).

Compounds were subsequently subjected to in vitro binding assays, in order to assess their experimental affinity toward the σ_1 receptor (see Tables 3.1.7 and 3.1.8). σ_1Ki values for all 33 compounds were determined using a protocol based on the competitive displacement of [3H](+)-pentazocine in a rat liver homogenate preparation. ,

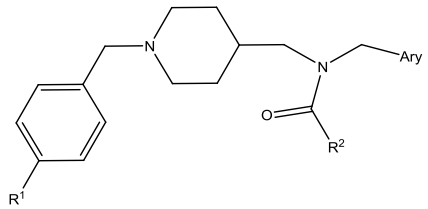
Table 3.1.7 *Experimental σ_1 receptor affinities (σ_1Ki) and corresponding free energy of binding values (ΔGbind,exp) of compounds 1a-r. The σ_1Ki/ΔGbind,exp values for (+)-pentazocine and haloperidol as reference compounds are also reported, for comparison. Errors are given in parenthesis as standard errors of the mean (SEM).*



1a-r

compd	aryl	R^1	σ_1K_i (nM)± SEM[a] (nM)	$\Delta G_{bind\ (sptml)}$[b] (Kcal/mol)
1a	phenyl	H	114 ± 7	**-9.47**
1b	4-chlorophenyl	H	42.3 ± 5.3	**-10.05**
1c	4-methylphenyl	H	139 ± 26	**-9.35**
1d	phenyl	CH_3	75.5 ± 9.7	**-9.71**
1e	4-chlorophenyl	CH_3	30.3 ± 3.9	**-10.25**
1f	4-methylphenyl	CH_3	36.4 ± 3.6	**-10.14**
1g	phenyl	Bz	146 ± 27	**-9.32**
1h	4-chlorophenyl	Bz	124 ± 24	**-9.42**
1i	4-methylphenyl	Bz	201 ± 36	**-9.13**
1j	pyridin-2-il	H	1664 ± 147	**-7.88**
1k	pyridin-3-il	H	1031 ± 95	**-8.16**
1l	pyridin-4-il	H	638 ± 24	**-8.45**
1m	pyridin-2-il	CH_3	1305 ± 105	**-8.02**
1n	pyridin-3-il	CH_3	1578 ± 115	**-7.91**
1o	pyridin-4-il	CH_3	1066 ± 130	**-8.14**
1p	pyridin-2-il	Bz	506 ± 92	**-8.58**
1q	pyridin-3-il	Bz	1401 ± 184	**-7.98**
1r	pyridin-4-il	Bz	1020 ± 162	**-8.17**
(+)-pentazocine	-	-	15 ± 3	**-10.67**
haloperidol	-	-	**5.7 ± 1**	**-11.24**

Table 3.1.8 *σ_1 receptor affinities of compounds 2a-l. The Ki values for (+)-pentazocine and haloperidol as reference compounds are also reported, for comparison.*



2a-l

compd	aryl	R^1	R^2	σ_1K_i (nM± SEM[a] (nM)	$\Delta G_{bind\ (sptml)}$[b] (Kcal/mol)
2a	phenyl	H	CH_3	9.62 ± 1.81	**-10.93**
2b	phenyl	H	phenyl	18.8 ± 2.2	**-10.53**
2c	phenyl	Cl	CH_3	1.87 ± 0.29	**-11.90**
2d	phenyl	Cl	phenyl	24.3 ± 3.6	**-10.38**
2e	4-chlorophenyl	H	CH_3	0.09 ± 0.03	**-13.70**
2f	4-chlorophenyl	H	phenyl	23.2 ± 3.2	**-10.41**
2g	4-chlorophenyl	Cl	CH_3	14.6 ± 1.4	**-10.68**
2h	4-chlorophenyl	Cl	phenyl	10.7 ± 2.5	**-10.87**
2i	4-methylphenyl	H	CH_3	14.2 ± 2.5	**-10.70**
2j	4-methylphenyl	H	phenyl	118 ± 28	**-9.45**
2k	4-methylphenyl	Cl	CH_3	61.8 ± 7.6	**-9.83**
2l	4-methylphenyl	Cl	phenyl	66.6 ± 12.5	**-9.78**
(+)-pentazocine	-	-	-	15 ± 3	**-10.67**
haloperidol	-	-	-	**5.7 ± 1**	**-11.24**

With respect to compounds 1a-r, the results confirm the importance of the presence of a basic nitrogen atom substituted with a small group (H, CH3) to endow the compound with σ_1 affinity.

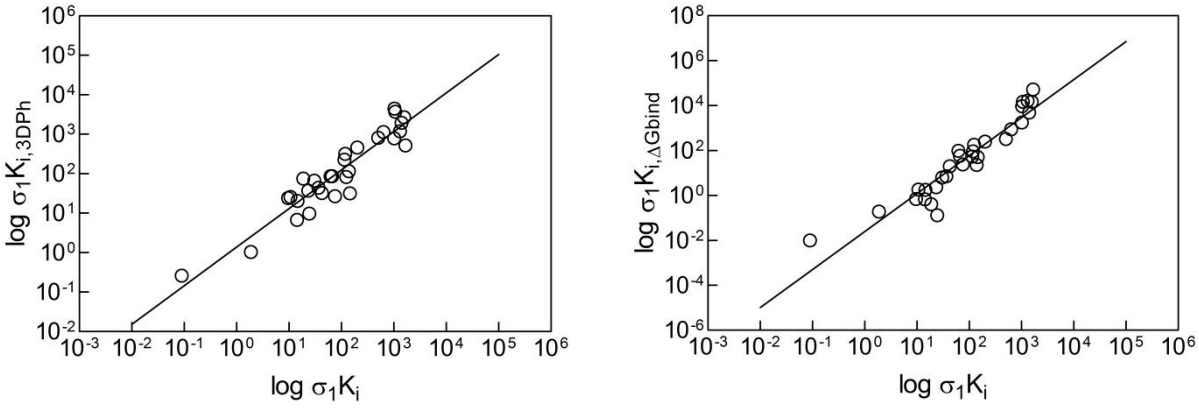
Actually, the N-substitution with a benzyl group decreases the σ_1 affinity of the derivatives 1g,i with respect to the corresponding derivatives 1a-f; contrarily, the presence of a -CH3 group linked to the basic nitrogen atom in compounds 1e (σ_1Ki = 30.3 nM) and 1f (σ_1Ki = 36.4 nM) improves the σ_1 affinity of this compound. However, the simultaneous absence of a small substituent, as chlorine or methyl group, on the para position of the benzyl moiety decreases the σ_1-receptor affinity of compound 1d (σ_1Ki = 75.5 nM) with respect to the corresponding para-substituted compounds 1e and 1f.

Compound 1e, characterized by the presence of a chlorine atom on the para position of the benzyl residue and of a methyl group linked to the basic nitrogen atom, explicates the highest σ_1 affinity. The para substitution with chlorine in compound 1b (σ_1Ki = 42.3 nM) maintain some level of σ_1 affinity, inferior to that of the corresponding N-methyl derivative 1e, but superior to that of the analogues 1a (σ_1Ki = 114 nM) and 1c (σ_1Ki = 139 nM). The replacement of the phenyl or substituted phenyl residues in compounds 1j-r with the pyridine-2-yl, pyridine-3-yl or pyridine-4yl moieties abolishes the σ_1 affinity of the corresponding compounds. In the derivative series 2a-l, the amide nitrogen atom is linked to variously substituted benzyl residues and to a 4-methylpiperidin-1-yl spacer, substituted on the benzene ring. The basic nitrogen atom that allows the ionic bond with a receptor acid site belongs to the piperidine cycle. The experimental σ_1 affinity of compound 2c is rather high (σ_1Ki = 1.87 nM). Compound 2c is substituted with chlorine on the para position of the benzyl group linked to piperidine nitrogen atom. The most potent compound of the series, however, is the acetamide derivative 2e, characterized by a Ki value as low as 0.09 nM. Interestingly, the corresponding benzamide derivative 2f shows a much lower σ_1-receptorial affinity (σ_1Ki = 23.2 nM). Actually, acetamide derivatives are endowed with σ_1-affinity higher than that of corresponding benzamide derivatives, except for compound 2g. The superior affinity of acetamide derivatives may be attributed to the electron donating effect of the acetamide methyl group that may increase the electronegative character of the carbonyl oxygen and further contribute to the σ_1 binding affinity as hydrogen bond acceptor.

As declared in the beginning, one of the main purposes of the entire work was an extensive validation of our originally proposed 3D model of the σ_1 receptor. Therefore, the most important point of the entire discussion is undoubtedly the direct comparison of the results stemming from the experimental ligand binding assays and the corresponding values predicted by the application of the entire computational ansatz. Figure 3.1.18 thus illustrates the results of this direct comparison. In detail, Figure 3.1.18(A) shows the excellent agreement between the affinities for σ_1 of compounds 1 and 2 predicted by the 3D pharmacophore model and the corresponding experimental σ_1Ki values, quantified by the corresponding value of the correlation coefficient of 0.89. If this could be a somewhat expected result, the astounding correlation between the experimental Ki values and those derived from the application of the MM/PBSA ranking is even more gratifying: indeed, Figure 3.1.18(B) shows how well the calculated σ_1KiΔGbind reproduce the experimental ones (R^2 = 0.89), with a more than satisfactory ranking order. To further confirm the capability of the entire MM/PBSA computational ansatz in correctly ranking the affinities of all 33 compounds towards the σ_1 receptor, if we compare the values of the MM/PBSA calculated ΔGbind values (Table 3.1.4)

with those derived from the biological assays (Tables 3.1.7 and 3.1.8), we can observe that the average unsigned error between these two data sets is 0.93 kcal/mol, and the corresponding root-mean-square deviation is 1.19 kcal/mol.

Thus, the remarkable quality of these correlations couple with the wide range of the σ_1 affinity values properly ranked constitute a further, decisive validation of the putative σ_1 binding site and, overall, of the entire 3D homology model of this intriguingly enigmatic receptor.



(A) (B)

Figure 3.1.18 *(A) Plot of the experimental vs. 3D pharmacophore predicted σ_1Ki values for the 33 compounds of series 1 and 2 (R^2 = 0.89). (B) Plot of the experimental vs. MM/PBSA predicted σ_1Ki values for the 33 compounds of series 1 and 2 (R^2 = 0.89).*

Conclusions

Notwithstanding many pharmacologic responses have been linked to the sigma receptors, the function of the σ_1 protein is still a subject of intense study and current debate. Importantly, until very recent times relatively little information regarding the structure of the σ_1 receptor or its ligand binding site was available to the scientific community. Cloning of the σ_1 receptor revealed that the rat brain receptor σ_1 protein consists of 223 amino acids, which results in a molecular weight of 23 kDa. Although human and animal σ_1 receptors show a similarity of more than 95%, unfortunately there is no resemblance of this receptor to other known mammalian proteins.

Lately our group published for the first time a 3D model of the σ_1 receptor protein as obtained from a complex multistep computational recipe based on homology modeling techniques. [5]The reliability of the proposed σ_1 model and the validity of its putative ligand binding site was assessed by a docking/MMPBSA-based small-scale virtual screening of a series of available σ_1 ligands, and by the receptor model-based design of three new σ_1 ligands, featuring activity in a wide range (from 1.87 to 1578 nM). To ultimately confirm the validity of our σ_1 3D model and its complete reliability as a platform for σ_1-ligand structure-based drug design, in the present work we expanded our study by designing 33 new σ_1 ligands, with affinity for the receptor

spanning 5 orders of magnitude. All these compounds were then ranked for receptor affinity by extensive molecular dynamics simulation-based free energy calculations, and the main interactions/receptor residues involved in ligand binding were thoroughly analyzed by applying per residue free energy deconvolution and in silico alanine scanning experiments. All compounds were subsequently synthesized in our laboratory and then tested for σ_1 binding activity in vitro.

Remarkably, the experimental affinity ranking for all 33 compounds towards the σ_1 receptor was found to be fully consistent with the corresponding predictions obtained from our in silico procedure. Therefore, we are convinced that the computational recipe adopted here, based on the σ_1 receptor 3D structure we recently developed, cane be generally employed to estimate the activity of new σ_1 ligands prior to their synthesis, with an obvious optimization of time and resources. Furthemore, if we reconsider all experimental affinity data listed in Tables 3.1.7 and 3.1.8, we can see that both sets of compounds 1 and 2 can be classified on the basis of their experimental activity as highly affine (Ki ≤ 40.0 nM, +++), moderately affine (40.0 < Ki < 600 nM, ++) and poorly affine (Ki ≥ 600 nM, +). According to this classification, and looking the in silico ranking shown it Table 3.1.4, we can also conclude that all compounds classified as highly affine (+++) are characterized by ΔGbind values ≤ -11.00 kcal/mol, those with a moderate affinity (++) have -11.00 < ΔGbind < -8.50 kcal/mol, and finally the less affine ones have ΔGbind ≥ -8.50 kcal/mol. Taking into account that the entire computational recipe are based on a protein model obtained via homology modeling techniques, these ranking capacity and the related results ultimately confirm the reliability of the σ_1 3D model in structure-based ligand design.

3.1.3 STUDY OF PARTICULAR BINDING MODE OF σ_1 SPYROCICLIC COMPOUNDS[41]

Our interest has been focused on the development of novel compounds with high σ_1 receptor affinity and selectivity over related receptor systems in the central nervous system. Recently we have shown that additional lipophilic substituents on spirocyclic σ_1 ligands favor the interactions with the σ_1 receptor protein. (Scheme 3.1.4) Replacement of the methyl moiety of 1a (Ki = 21 nM) with a phenyl group (1b, Ki = 1.5 nM) led to a 15-fold increase of σ_1 receptor binding while maintaining high σ_1/ σ_2 selectivity.[42,43] In the series of thiophene annulated spirocyclic σ_1 ligands 2,3 an additional aryl moiety in α-position of the S-atom is well tolerated by the σ_1 receptor protein[44]. These observations are in good accordance with established

pharmacophore models indicating the acceptance of large lipophilic substituents by the $\sigma_1$receptor protein.[45]

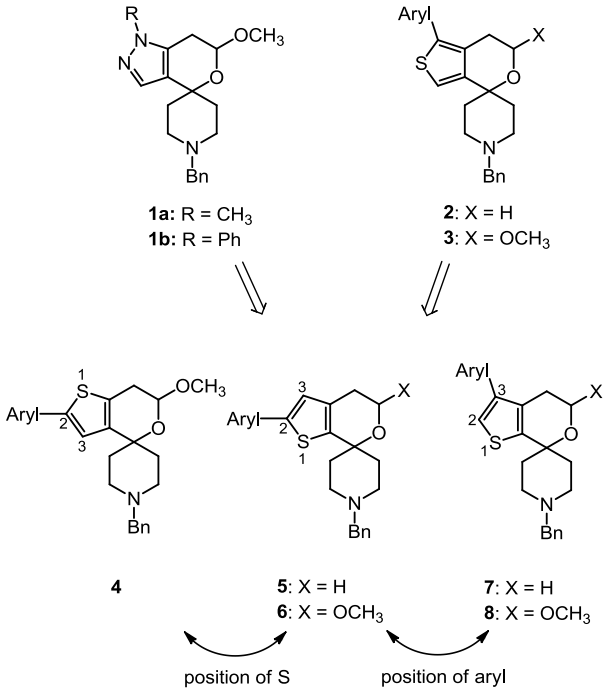


Figure 3.1.4 *Design of arylated spirocyclic thiophenes.*

Herein, we wish to report on the introduction of various aryl moieties in α- (4-6) and β-position (7,8) of regioisomeric thiophenes by direct C-H-bond arylation using recently developed Pd-catalysts. Whereas the position of the additional aryl moiety of 7 and 8 correlates with the position in the lead compounds 1b, 2 and 3, the aryl moiety of the spirocyclic ligands 4-6 is shifted to the "left" position of the thiophene ring. On the other hand compounds 4 and 5/6 represent regioisomers with respect of the position of the S-atom in the thiophene ring. The σ_1 affinity of the arylated thiophenes 4-8 are investigated in receptor binding studies with radioligands. The resulting data together with the affinity data of arylated spirocyclic σ_1 ligands 2 and 3 are used for the establishment and validation of a σ_1 pharmacophore model. Moreover, docking of the ligands into the σ_1 receptor 3D model and affinity scoring will result in the analysis of the exact detailed ligand receptor interactions.

Details about the chemical synthesis could be found in the paper[41].

Receptor Affinity

The σ_1 and σ_2 receptor affinities of the arylated spirocyclic thiophenes 4-8, the synthetic precursors 9-11, as well as the lead and reference compounds were determined in competition experiments with radioligands. The potent and selective tritium labeled radioligand [3H]-(+)-pentazocine was used as competitor for the test compounds in the σ_1 assay. Membrane

preparations of guinea pig brains served as receptor material. In the σ_2 assay, homogenates of rat liver were the source for σ_2 receptors and [3H]-di-o-tolylguanidine was employed as radioligand. Since di-o-tolylguanidine also binds to σ_1 receptors, an excess of non-radiolabeled, σ_1 selective (+)-pentazocine was added to mask σ_1 receptors.[46,47]

Table 3.1.9 *σ_1 and σ_2 receptor affinities of the spirocyclic thiophenes and reference compounds.*

			K_i ± SEM [nM] (n = 3)		selectivity
compd.	X	Aryl	σ_1	σ_2	σ_1/σ_2
1a[a)]	OCH_3	CH_3	21 ± 2.3	> 1 μM	> 47
1b[a)]	OCH_3	C_6H_5	1.5 ± 0.08	> 1 μM	> 660
2a[b)]	H	H	0.35 ± 0.06	230	657
2b[b)]	H	C_6H_5	0.22 ± 0.06	806	3664
3a[b)]	OCH_3	H	4.5 ± 2.9	> 1 μM	> 222
3b[b)]	OCH_3	C_6H_5	1.0 ± 0.4	> 1 μM	> 1000
4a	OCH_3	C_6H_5	5.5 ± 1.5	530	96
4b	OCH_3	*p*-$MeOC_6H_4$	111 ± 27	> 1 μM	> 9
4c	OCH_3	*p*-CNC_6H_4	89 ± 72	> 1 μM	> 11
4d	OCH_3	*p*-C_6H_4-C_6H_5	220 ± 99	> 1 μM	> 4
5a	H	C_6H_5	2.4 ± 0.69	283	118
6a	OCH_3	C_6H_5	16 ± 5.8	> 1 μM	> 63
6b	OCH_3	*p*-$MeOC_6H_4$	56 ± 23	> 1 μM	> 18
6c	OCH_3	*p*-MeC_6H_4	29 ± 1.0	> 1 μM	> 34
6d	OCH_3	*p*-$NO_2C_6H_4$	16 ± 0.9	> 1 μM	> 63
6e	OCH_3	*p*-AcC_6H_4	59 ± 15	> 1 μM	> 1
6f	OCH_3	*p*-CNC_6H_4	20 ± 6.7	> 1 μM	> 50
6g	OCH_3	*p*-$CF_3C_6H_4$	43 ± 18	> 1 μM	> 23
6h	OCH_3	1-naphthyl	12 ± 1.8	142	12
6i	OCH_3	*p*-C_6H_4-C_6H_5	211 ± 74	> 1 μM	> 5
6j	OCH_3	3-pyridyl	2.5 ± 0.46	> 1 μM	> 400
7a	H	C_6H_5	1.6 ± 0.7	445	278
8a	OCH_3	C_6H_5	5.4 ± 0.97	> 1 μM	> 185
8b	OCH_3	*p*-$MeOC_6H_4$	7.1 ± 2.4	> 1 μM	> 141
8c	OCH_3	*p*-MeC_6H_4	2.7 ± 1.6	> 1 μM	> 370
9[c)]	OCH_3	H	0.32 ± 0.10	> 1 μM	> 3000
10	H	H	1.0 ± 0.3	147	147
11	OCH_3	H	1.9 ± 0.44	> 1 μM	> 500
15	OH	H	16 ± 2.0	849	53
16	C^4=C^5	H	3.4 ± 1.0	236	69
haloperidol			3.9 ± 1.5	78 ± 2.0	20
di-o-tolylguanidine			61 ± 8	42 ± 15	0.7

Table 3.1.9 clearly shows that introduction of aryl moieties in position 2 of the spirocyclic compound 9 bearing the S-atom in the "top" position 1 led to significant decrease of σ_1 receptor affinity. The Ki-values of the α-arylated compounds 4a-4d are 17-500-fold higher than the Ki-value of the parent thiophene 9. The very low σ_1 affinity of the biphenylyl derivative 4d indicates that the biphenylyl residue is sterically too demanding for the binding pocket of the σ_1 receptor.

α–arylation of the regioisomeric thiophene derivatives 10 and 11 led to slightly reduced σ_1 affinities compared with the σ_1 affinities of their parent compounds 10 and 11. However, some of the arylated compounds show Ki values below 20 nM (e.g. 6a: Ki = 16 nM; 6d: Ki = 16 nM; 6h: Ki = 12 nM), indicating considerable σ_1 affinities. High σ_1 affinities are achieved with either an unsubstituted aryl residue (e.g. 6a (phenyl): Ki = 16 nM; 6h (naphthyl): Ki = 12 nM) or an electron deficient aryl residue in 2-position (e.g. 6d (nitrophenyl): Ki = 16 nM; 6j (3-pyridyl): Ki = 2.5 nM). As seen for the series 4 spirocyclic thiophenes with an electron rich aryl moiety in 2-position gave reduced σ_1 affinities (e.g. 6b (methoxyphenyl): Ki = 56 nM). The most promising σ_1 ligands of this series are 5a without acetalic methoxy group (Ki = 2.4 nM) and the 3-pyridiyl derivative 6j (Ki = 2.5 nM) showing similar σ_1 affinities as the parent compounds 10 and 11.

The highest σ_1 affinities were found for the α-arylated spirocyclic thiophenes 7 and 8. In this compound class even compounds with donor substituents (e.g. 8b (methoxyphenyl): Ki = 7.1 nM, compare the Ki values of 4b and 6b) interact in the low nanomolar range with σ_1 receptors. Unfortunately, α-arylation of 10 and 11 with acceptor substituted aryl iodides failed to give the arylated products.

Generally the σ_1 affinities decreased in the order 7/8 > 5/6 > 4. This tendency is depending on the orientation of the aryl moiety. In compounds 7/8, the phenyl ring adopts a position at the "top" of the molecule, which is favorable for receptor interaction. Due to the increased C-S bond length and the larger size of the S-atom, the spirocyclic thiophenes 5/6 with the S-atom in the "bottom" position direct the aryl moiety in a position similar to its position in 7/8, whereas in the regioisomers 4 with the S-atom in the "top" position the additional aryl moiety is directed in an unfavorable position.

With exception of the naphthyl derivative 6h and the unsubstituted compound 10, the σ_2 affinities of the (arylated) spirocyclic thiophenes 4-11 are very low indicating high selectivity for the σ_1 subtype over the σ_2 subtype. Despite the moderate σ_2 affinity of 10 (Ki = 147 nM), its σ_1/ σ_2 selectivity is still high (147) due to the very high σ_1 affinity. On the contrary the naphthyl derivative 6h shows the lowest σ_1/ σ_2 selectivity (12) of this series of compounds.

Molecular modeling of ligand/protein interactions

To investigate the role played by the position of the aromatic moiety on the thiophene ring in determining the affinity of our compounds towards the σ_1 receptor, we applied a sequential molecular modeling procedure, based on the following steps: i) generation of a 3D pharmacophore model for the present series of compounds; ii) 3D pharmacophore-guided docking of all ligands into the putative binding site of our 3D σ_1 receptor model, iii) estimation of their binding affinity via the MM-PBSA methodology, and iv) comparison with the available experimental activities.

Pharmacophore model generation

These pharmacophore models can be used in different ways in drug design programs, e.g., 1) as 3D query tools in virtual screening to identify potential new compounds from 3D databases of drug-like molecules with patentable structures different from those already discovered; 2) to predict the activities of a set of new compounds yet to be synthesized or, as in the present case, 3) to understand the possible mode/mechanism of drug/receptor interaction.

3D pharmacophore hypothesis generation by the Catalyst software is one of the popular approach that has been successfully used in drug discovery and toxicology research so far. The most critical aspect of Catalyst pharmacophore hypothesis generation is the selection of the molecular training set; however, some basic guidelines could be find in Material & Methods chapter.

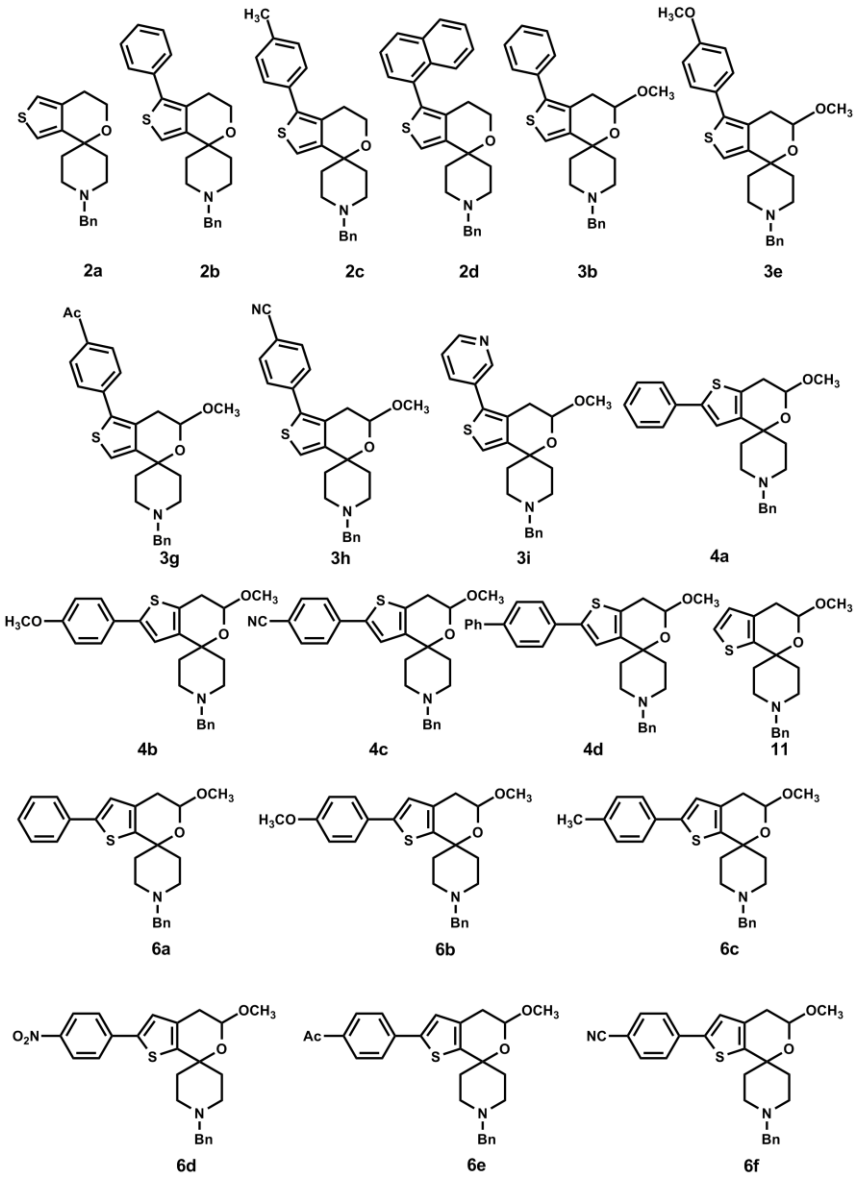
On the basis of the above criteria, a set of 10 pharmacophore hypothesis was generated with Catalyst using 27 training set compounds listed in Scheme 3.1.5.

The results of the hypotheses, which include different cost values calculated during hypothesis generation along with the corresponding root-mean-square deviation (rmsd), correlation coefficient (ρ), and pharmacophore features, are listed in Table 3.1.10.

The total hypothesis cost of these ten best models varies between 122.9 for the best ranked model (Hypo1) to 178.9 for the lowest ranked one (Hypo10). Such a confined difference (56 bits) reflects both the homogeneity of the generated hypotheses and the adequacy of the molecular training set. The difference between the null and the fixed cost, which should be higher than 70 to guarantee a robust correlation, is 89.6 in our case. This corresponds to a chance of true correlation in the data greater than 90%. Furthermore, in all the generated hypotheses the total costs are much closer to the fixed cost (115.7) than to the null cost (205.3), indicating that meaningful models are obtained. Finally, the root mean square deviations (RMSD) and the correlation coefficients (ρ) between estimated and experimental

affinities range from 0.922 to 1.874, and from 0.905 to 0.695, respectively. As all the generated pharmacophores map the molecules of the training set in a similar way, the first model (Hypo1), characterized by the highest cost difference, the lowest RMDS, and the best ρ values, was selected for further analysis.

In analogy with other pharmacophore models formerly developed for σ_1 ligands, [1,10,39,48] the main features of the present Hypo1 are a positive ionizable group (PI), an hydrogen bond acceptor group (HBA), two hydrophobic moieties (HY), and one hydrophobic aromatic point (HYAr) (see Figure 3.1.19).

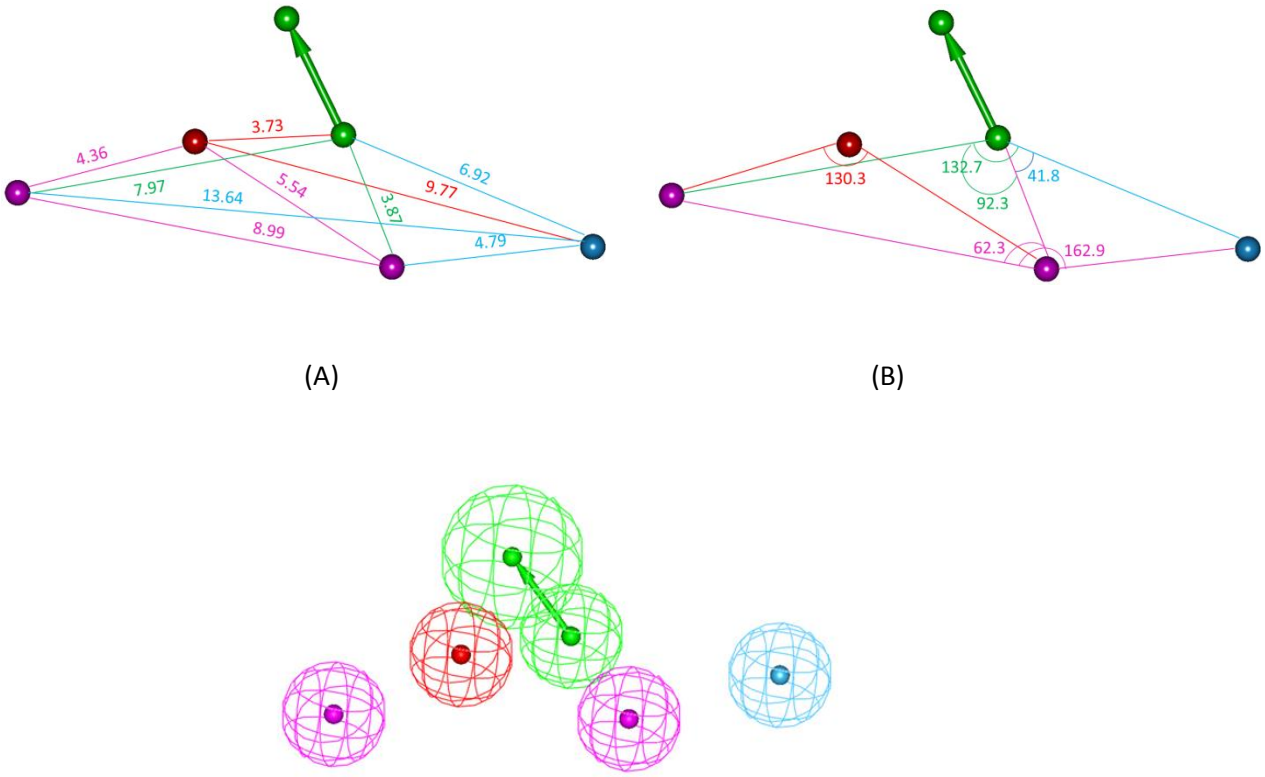


Scheme 3.1.5 *Chemical structures of the 27 training set compounds.*

Table 3.1.10 *Composition (features), costs (bits), and statistical parameters (rmsd and ρ) associated with the 10 best hypotheses (pharmacophore models) generated with Catalyst using the training set molecules (Scheme 3.1.5).*

Hypothesis	total cost	Δcost[a]	rmsd	ρ	features[b]
1	122.9	82.6	0.922	0.905	HBA, HYAr, HY, HY, PI
2	131.5	73.8	1.305	0.839	HBA, HY, HY, HY, PI
3	136.4	68.9	1.319	0.832	HBA, HYAr, HY, HY, PI
4	137.8	67.5	1.377	0.813	HBA, HYAr, HYAr, HY, PI
5	139.6	65.7	1.418	0.792	HBA, HY, HY, HY, PI
6	142.8	62.5	1.490	0.775	HBA, HYAr, HY, PI
7	148.5	56.8	1.558	0.760	HBA, HY, HY, PI
8	153.4	51.9	1.704	0.739	HBA, HYAr, HYAr, PI
9	173.6	31.7	1.792	0.701	HBA, HYAr, HY, PI
10	178.9	26.4	1.874	0.695	HBA, HYAr, HY, PI

The affinities of the 27 compounds for the σ_1 receptor estimated using Hypo1 are reported in Table 3.1.11, along with the experimental values and the relevant errors (expressed as the ratio between estimated and experimental values). This Table clearly shows that 20 out of 27 molecules in the training set have errors equal or less than 2.2, whilst the remaining 7 have errors less than 5.0.



(A) (B)

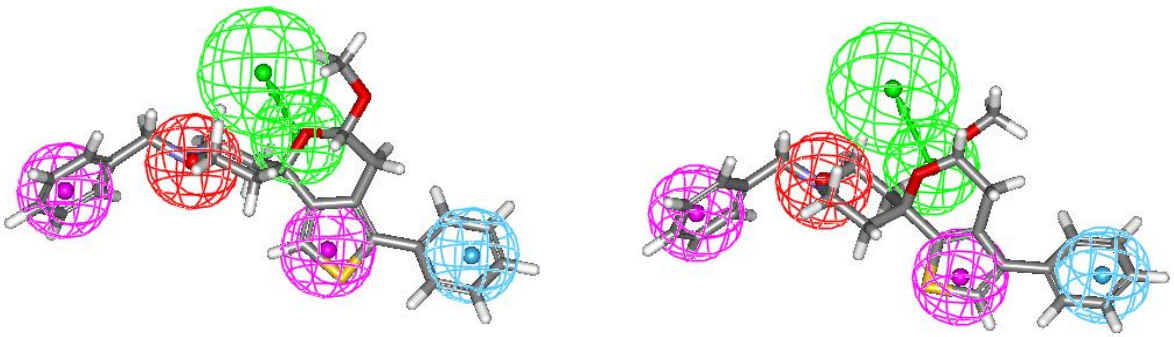
(C)

Figure 3.1.19 *Geometrical relationships (A,B) among the features of the top-scoring pharmacophore Hypo1 (C). The hypothesis features are portrayed as meshed spheres, color-coded as follows: red, PI; light blue, HYAr; pink, HY; light green, HBA. HBA is actually represented as a pair of spheres (the smaller sphere represents the location of the HBA atom on the ligand and the larger one the location of an HB donor on the receptor).*

Pharmacophore mapping of compounds belonging to series 3 and 8 – all bearing an aromatic ring in the top position with respect to the tetrahydropyran ring - indicates that these molecules can map all 5 pharmacophore features quite nicely (see Figure 3.1.20). Indeed, the presence of the N-benzylic group provides one of the aromatic function (HY) to the molecules, the other HY being provided by the thiophene ring. The piperidinic nitrogen atom is aptly positioned over the positive ionizable feature PI, the pyran ring oxygen fulfills the HBA requirement and, lastly, the phenyl substituent overlays the hydrophobic aromatic feature (HYAr) independently of its α- (3b) or β-position (8a) with respect to the heterocyclic sulfur atom. Referring to compounds 3b and 8a of Figure 4 as an example, the in silico estimated Ki (σ_1) values (Ki = 1.2 nM for 3b and Ki = 3.4 nM for 8a, respectively) are in excellent agreement with the corresponding experimentally determined Ki values (1.0 nM for 3b and 5.4 nM for 8a, respectively, see Table 3.1.11).

Table 3.1.11 *Experimentally determined and estimated Ki values of the training set compounds calculated on the basis of the pharmacophore hypothesis 1 (Hypo1).*

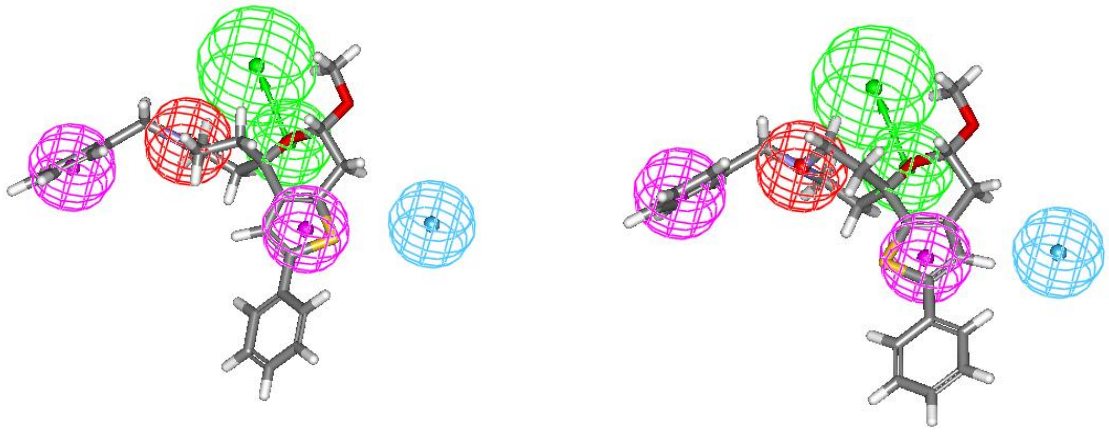
compd. this article	compd. of ref.[22]	Aryl	X	σ_1 affinity: K_i (nM)		Error[a]
				estimated	experimental	
-	**2a**	-	H	1.8	0.35	5.0
-	**2b**	C_6H_5	H	2.7	4.5	-1.6
-	**2c**	p-MeC_6H_5	H	2.8	3.6	-1.3
-	**2d**	1-napthyl	H	6.9	4.0	1.7
-	**3b**	C_6H_5	$-OCH_3$	1.2	1.0	1.2
-	**3e**	*p*-$MeOC_6H_5$	$-OCH_3$	3.1	2.2	1.4
-	**3g**	*p*-AcC_6H_5	$-OCH_3$	3.5	1.6	2.2
-	**3h**	*p*-CNC_6H_5	$-OCH_3$	0.51	0.25	2.0
-	**3i**	3-pyridyl	$-OCH_3$	1.9	2.2	-1.1
4a	**-**	C_6H_5	$-OCH_3$	13	5.5	2.2
4b	**-**	*p*-$MeOC_6H_5$	$-OCH_3$	50	111	-2.2
4c	**-**	*p*-CNC_6H_5	$-OCH_3$	60	89	-1.5
4d	**-**	*p*-biphenyl	$-OCH_3$	51	220	-4.3
11	**-**	-	$-OCH_3$	5.7	1.9	3.0
6a	**-**	C_6H_5	$-OCH_3$	23	16	1.4
6b	**-**	*p*-$MeOC_6H_5$	$-OCH_3$	67	56	1.2
6c	**-**	*p*-MeC_6H_5	$-OCH_3$	27	29	-1.1
6d	**-**	*p*-$NO_2C_6H_5$	$-OCH_3$	51	16	3.2
6e	**-**	*p*-AcC_6H_5	$-OCH_3$	34	59	-1.7
6f	**-**	*p*-CNC_6H_5	$-OCH_3$	37	20	1.8
6g	**-**	*p*-$CF_3C_6H_5$	$-OCH_3$	21	43	-2.0
6h	**-**	1-naphtyl	$-OCH_3$	22	12	1.8
6i	**-**	*p*-biphenyl	$-OCH_3$	45	211	-4.7
7a	**-**	C_6H_5	H	1.8	1.6	1.1
8a	**-**	C_6H_5	$-OCH_3$	3.4	5.4	-1.6
8b	**-**	*p*-$MeOC_6H_5$	$-OCH_3$	2.2	7.1	-3.3
8c	**-**	*p*-MeC_6H_5	$-OCH_3$	3.5	2.7	1.3



3b 8a

Figure 3.1.20 *Mapping of compounds 3b and 8a onto the developed σ_1 3D pharmacophore model (Figure 3.1.19). Compounds are portrayed as atom-colored sticks (red, O; yellow, S; gray, C; blue, N; white, H).*

On the contrary, compounds belonging to series 4 and 6 featuring the aryl substituent in the left position are all endowed with a lower affinity towards the σ_1 receptor since all explored molecular conformations failed to access the HYAr feature, as exemplified for compounds 4a and 6a in Figure 3.1.21. The HYAr pharmacophore interaction point is, however, the only feature unmapped by compounds of series 4 and 6. Thus, given the excellent overlay between the molecular chemical moieties and the remaining 3D pharmacophore requirements shown in Figure 3.1.21, only a small decrease in σ_1 affinity for these two molecular series is predicted for, e.g., compounds 4a and 6a (Ki = 13 nM for 4a and Ki = 23 nM for 6a) in agreement with the experimental binding constants (Ki = 5.5 nM for 4a and Ki = 16 nM for 6a, see Table 3.1.11).



4a 6a

Figure 3.1.21 *Mapping of compounds 4a and 6a onto the developed σ_1 3D pharmacophore model (Figure 3.1.19). Compounds are portrayed as atom-colored sticks (red, O; yellow, S; gray, C; blue, N; white, H).*

Importantly, however, the activity of compounds 4 and 6 is strongly affected by the presence of a substituent at the para position of the thiophene-linked phenyl ring. On the contrary, all derivatives 3 and 8 bearing the same substituted aromatic moieties are endowed with comparable (if not higher) σ_1 affinities with respect to the unsubstituted counterparts.

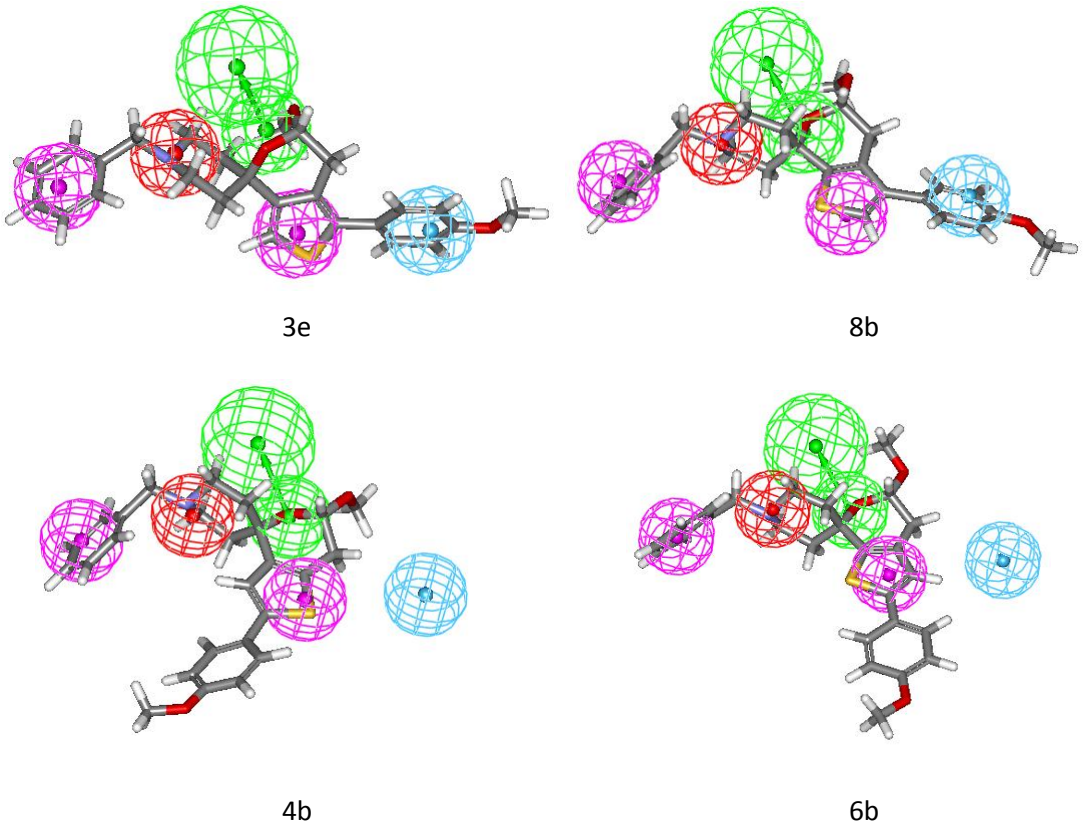


Figure 3.1.22 *Mapping of compounds 3e, 8b, 4b, and 6b onto the developed σ_1 3D pharmacophore model (Figure 3.1.19). Compounds are portrayed as atom-colored sticks (red, O; yellow,S; gray, C; blue, N; white, H).*

In this respect, the analysis of the top panels of Figure 3.1.22 indeed reveals how, for compounds 3e and 8b bearing the p-methoxyphenyl substituent in the top position, all five 3D pharmacophore features are mapped very well (as in the case of their unsubstituted analogs, see Figure 3.1.20), resulting in high predicted/experimental σ_1 affinity values (Ki(estimated) = 3.1 nM/Ki(experimental) = 2.2 nM for 3e and Ki(estimated) = 2.2 nM/Ki(experimental) = 7.1 nM for 8b, respectively, see Table 3.1.11).

On the other hand, all derivatives characterized by the presence of the aromatic ring in the left position not only miss the HYAr feature as discussed above but the presence of the p-OCH3 substituent on the phenyl ring induces a further conformational change so that all the remaining four pharmacophoric features are mapped less precisely with respect to the

unsubstititedunsubstituted compounds (compare the lower panels in Figure 3.1.22 and Figure 3.1.21). Accordingly, for this set of compounds the model predicts lower σ_1 affinities for 4b (Ki = 50 nM) and 6b (Ki = 67 nM) in line with the experimental σ_1 affinities (Ki = 111 nM for 4b and Ki = 56 nM for 6b, see Table 3.1.11).

Pharmacophore assessment and cross-validation study

The quality of the developed pharmacophore was assessed using the CatScramble technique available in Catalyst. The purpose of this technique is to randomize the affinity data among the training set compounds and to generate pharmacophore hypotheses using the same features and parameters used to develop the original pharmacophore hypothesis. If the randomized sets generates pharmacophores with similar or better cost values, rmsd, and correlation, then the original pharmacophore can be considered as generated by chance. The results of the CatScramble runs are listed in Table 3.1.12. Since a 98 % confidence level was selected for this test, 49 random hypothesis runs were performed. The results clearly indicate that randomization produced hypotheses with no predictive values similar or close to the corresponding Hypo1. Indeed, none of the outcome hypotheses had a lower cost score, better correlation or smaller root-mean-square deviation than the initial hypothesis. Table 3.1.12 lists the first 10 lowest total score values of the resulting 49 hypotheses for our training set molecules. In conclusion, there is a 98 % chance for the best hypothesis to represent a true correlation in the training set affinity data for the present classes of compounds.

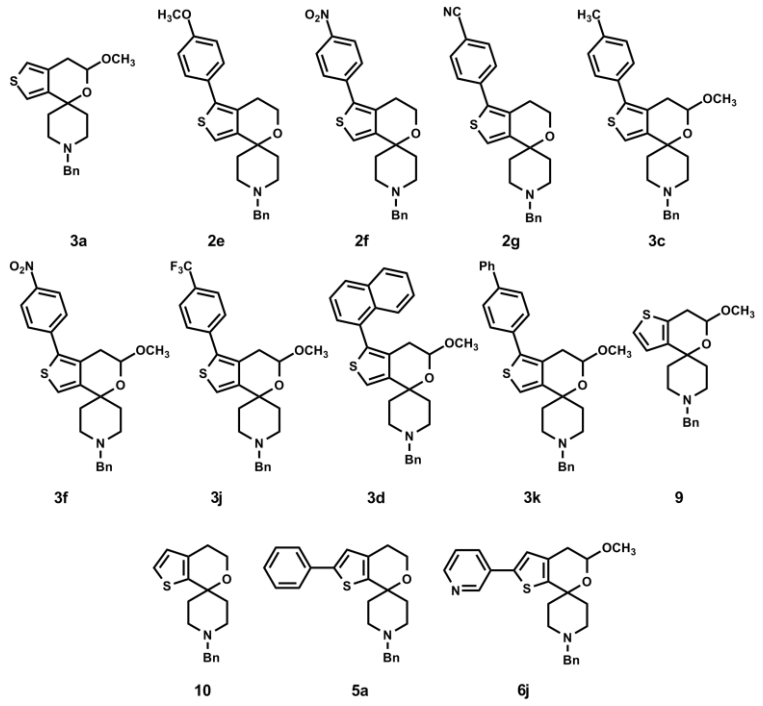
Table 3.1.12 *Output parameters of the 10 lowest cost hypotheses resulting from the statistical evaluation according to the CatScramble validation procedure.*

Hypothesis	ρ	rmsd	total cost
1	0.812	1.154	142.9
2	0.717	1.389	151.9
3	0.700	1.397	154.0
4	0.724	1.453	155.1
5	0.622	1.496	157.1
6	0.603	1.601	160.6
7	0.576	1.657	163.8
8	0.558	1.708	165.7
9	0.515	1.755	167.8
10	0.507	1.823	168.9
Hypo1	0.905	0.922	122.9

Finally, a further statistical test, the leave-one-out method, which consists of re-computing the hypothesis by excluding from the training set one molecule at a time, was carried out. Basically, this test is performed to verify whether or not the correlation is strongly dependent on one particular compound in the training set. The test is positive if the affinity of each excluded molecule is correctly predicted by the corresponding one-missing hypothesis. The value of ρ, the feature composition of the pharmacophore, and the quality of the predicted affinity of the excluded molecule were used as measures for the assessment of the statistical test. For each of the 27 new hypotheses generated according to this method we did not obtain meaningful differences between Hypo1 and each hypothesis resulting from the exclusion of one compound at a time.

Pharmacophore validation using test set compounds

A critical step in automated pharmacophore generation is model validation, especially in those cases where the model has been generated for the purpose of predicting the activity either of external sets of compounds or to estimate the activity of newly conceived molecular entities prior to their synthesis. Thus, to check the robustness of our correlation we used the pharmacophore model (Hypo1) to predict the affinity of a further set of molecules, also called the test set, composed of 13 additional molecules from our series (see Scheme 3.1.6).



Scheme 3.1.6 *Chemical structures of the 13 test set compounds.*

Interestingly, a good correlation coefficient 0.80 was observed when a regression analysis was performed by mapping the test set onto the features of the best pharmacophore hypothesis Hypo1. The predicted and the experimental Ki(σ_1) values for the test set along with the respective errors are shown in Table 3.1.13. The average error in predicting the affinity of the test set molecules is 3.0. Given the inherent simplicity of the pharmacophoric approach, and considering the intrinsic variability of the biological responses, we can conclude that the ability of the present 3D pharmacophore model to predict the affinity of these series of σ_1 receptor ligands is quite satisfactory.

Table 3.1.13 *Experimentally determined and estimated Ki values of the test set compounds calculated on the basis of the pharmacophore hypothesis 1 (Hypo1).*

				σ_1 affinity: K_i (nM)		
compd. this article	compd. of ref.[22]	Aryl	X	estimated	Experimental	Error
-	**3a**	-	- OCH_3	1.8	0.22	7.7
-	**2e**	p-$MeOC_6H_5$	H	1.6	1.5	1.1
-	**2f**	p-$NO_2C_6H_5$	H	3.4	1.7	2.0
-	**2g**	p-CNC_6H_5	H	2.7	3.4	-1.3
-	**3c**	p-MeC_6H_5	- OCH_3	1.1	2.0	-1.9
-	**3f**	p-$NO_2C_6H_5$	- OCH_3	3.6	1.0	3.6
-	**3j**	p-$CF_3C_6H_5$	- OCH_3	2.1	5.7	-2.7
-	**3d**	1-naphtyl	- OCH_3	5.8	5.0	1.2
-	**3k**	p-biphenyl	- OCH_3	18	30	-1.7
9	**-**	-	- OCH_3	2.0	0.32	6.3
10	**-**	-	H	3.2	1.0	3.2
5a	**-**	C_6H_5	H	11	2.4	4.6
6j	**-**	3-pyridyl	- OCH_3	5.4	2.5	2.2

Docking of ligands into the σ_1 receptor 3D model and affinity scoring via MM-PBSA calculations

Following our pioneer work,[1] the putative binding site and binding modes of all compounds considered in this work in the σ_1 receptor 3D homology model structure were retrieved exploiting the currently available preliminary information on sequence-structure relationships and mutagenesis studies, and the ligand-binding pharmacophore requirements presented above. Having localized the protein binding site as described in paragraph 3.1.2, a thorough search for a sequence satisfying the chemical features imposed by Hypo1 was performed, and successfully retrieved. Thus, all compounds were docked into the putative binding site of receptor 3D model. In the set of docked ligand conformations, for each compound a solution was found that best reproduced the key 3D pharmacophore requirements.

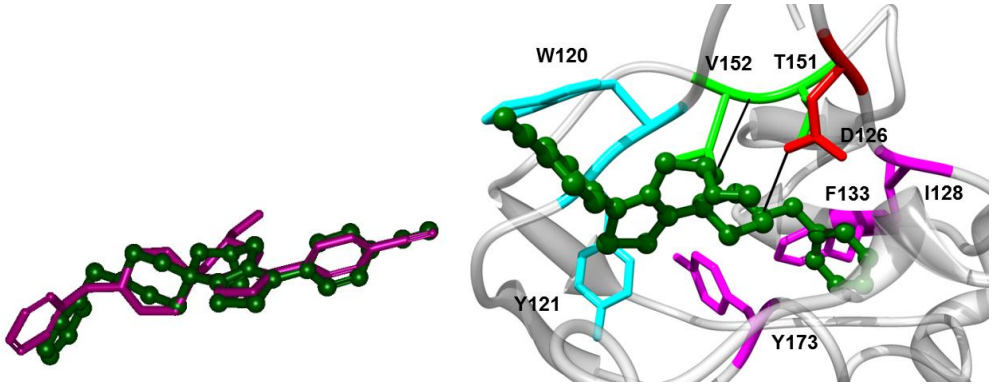


Figure 3.1.23 *(Left) Comparison between the best conformation assumed by compound 3h upon mapping the 3D pharmacophore model (purple sticks) and after docking in the putative σ_1 binding site (green sticks-and-balls). (Right) Equilibrated MD snapshots of the σ_1 receptor in complex with 3h. The image is a zoomed view of the receptor binding site. The ligand is portrayed in green sticks-and-balls, while the protein residues mainly involved in the interaction with 3h are highlighted as colored sticks and labeled. Salt bridge and H-bonds interactions are shown as black lines. Water, ions and counterions are not shown for clarity.*

Taking compound 3h as a prototypical example of the class of compounds featuring the phenyl ring in the top position, from the top panel of Figure 3.1.23, we can observe the striking similarity between the molecular conformation adopted by the compound in mapping the 3D pharmacophore model and that assumed by the same compound in the σ_1 pocket upon binding. Most importantly, however, the best binding mode detected for 3h after docking satisfies all ligand/protein intermolecular interactions required by pharmacophore hypothesis Hypo1. In detail: i) the piperidinic nitrogen is engaged in a salt bridge with Asp 126 (PI feature); ii) the pyran oxygen atom H-bonds with the–NH group of the backbone peptidic linkage

between Thr151 and Val152 (HBA feature); iii) the aryl group linked to the thiophene ring is involved in stabilizing interactions with the aromatic components of Trp121 and Tyr 120 (HYAr feature); and iv) a network of hydrophobic interactions stabilizes the position of the thiophene and the N-benzylic rings among the residues Tyr173, Ile128, and Phe133, respectively (2 HY features).

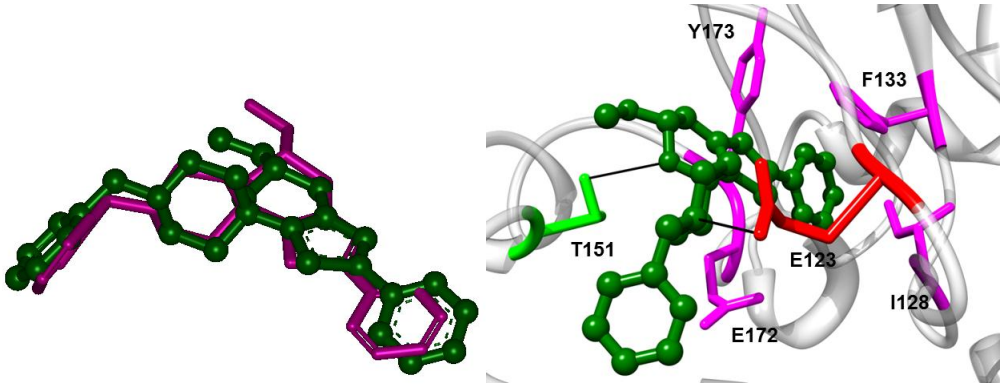


Figure 3.1.24 *(Left) Comparison between the best conformation assumed by compound 6a upon mapping the 3D pharmacophore model (purple sticks) and after docking in the putative σ_1 binding site (green sticks-and-balls). (Right) Equilibrated MD snapshots of the σ_1 receptor in complex with 6a. The image is a zoomed view of the receptor binding site. The ligand is portrayed in green sticks-and-balls, while the protein residues mainly involved in the interaction with 6a are highlighted as colored sticks and labeled. Salt bridge and H-bonds interactions are shown as black lines. Water, ions and counterions are not shown for clarity.*

Considering compound 6a as a representative example of the derivatives featuring the phenyl ring in the left position, from the top panel of Figure 3.1.24 we can appreciate that, also in this case, the best molecular conformation mapping the 3D pharmacophore Hypo1 and the best binding mode of 6a into the σ_1 binding pocket are superimposable. Quite surprisingly, however, the detailed inspection of the binding pose of 6a reveals that, in order to comply with the pharmacophoric requirements upon protein binding, this molecule must adopt a "reverse" orientation with respect to that assumed by, e.g., 3h, as shown in the bottom panel of Figure 3.1.24. Accordingly, the basic nitrogen of 6a creates a salt bridge with Glu123 (PI feature,), the oxygen atom of the pyran ring is engaged in a H-bond with the side-chain –OH group of Thr131 (HBA feature,), and the thiophene ring is stabilized by hydrophobic interactions with the side-chains of both Tyr173 and E172 (HY feature). Notably, at variance with 3h, the remaining hydrophobic pharmacophore requirement is fulfilled by the thiophene-

linked phenyl ring, which aptly fits a small hydrophobic cavity lined by the side chains of Ile128 and Phe133.

All 40 receptor/drug complexes obtained from the pharmacophore-based docking procedure were then relaxed by energy minimization and molecular dynamics (MD) simulations and, finally, the relevant values of the free energy of binding - ΔGbind - between all compounds and the σ_1 receptor were evaluated by applying the well-known Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) computational ansatz[6,30,49], as listed in Table 3.1.14.

Table 3.1.14 *Experimentally determined (ΔGbind.(exp)) and calculated (ΔGbind) free energy of ligand - σ_1 receptor complexes*[a]

compd. this article	compd. of ref.[22]	Aryl	X	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG_{bind} (kcal/mol)	$\Delta G_{bind,(exp)}$[a] (kcal/mol)
-	**2a**	-	H	-36.66 (0.18)	23.56 (0.37)	-13.10 (0.41)	-12.91
-	**3a**	-	$-OCH_3$	-36.90 (0.19)	23.67 (0.39)	-13.23 (0.43)	-13.18
-	**2b**	C_6H_5	H	-37.40 (0.18)	25.76 (0.42)	-11.64 (0.45)	-11.39
-	**2e**	p-$MeOC_6H_5$	H	-38.36 (0.21)	25.83 (0.39)	-12.53 (0.44)	-12.04
-	**2c**	p-MeC_6H_5	H	-38.50 (0.19)	25.69 (0.38)	-12.81 (0.42)	-11.52
-	**2f**	p-$NO_2C_6H_5$	H	-38.70 (0.20)	25.91 (0.38)	-12.79 (0.43)	-11.97
-	**2g**	p-CNC_6H_5	H	-38.63 (0.20)	26.02 (0.40)	-12.61 (0.45)	-11.56
-	**2d**	1-napthyl	H	-39.06 (0.17)	26.18 (0.41)	-12.88 (0.44)	-11.46
-	**3b**	C_6H_5	$-OCH_3$	-38.52 (0.22)	25.81 (0.37)	-12.71 (0.43)	-12.28
-	**3e**	p-$MeOC_6H_5$	$-OCH_3$	-38.71 (0.18)	25.77 (0.39)	-12.94 (0.43)	-11.82
-	**3c**	p-MeC_6H_5	$-OCH_3$	-38.51 (0.20)	25.79 (0.39)	-12.72 (0.44)	-11.87
-	**3f**	p-$NO_2C_6H_5$	$-OCH_3$	-38.58 (0.18)	25.99 (0.40)	-12.59 (0.45)	-12.28
-	**3g**	p-AcC_6H_5	$-OCH_3$	-38.43 (0.17)	25.56 (0.40)	-12.87 (0.43)	-12.00
-	**3h**	p-CNC_6H_5	$-OCH_3$	-39.11 (0.16)	25.82 (0.40)	-13.29 (0.42)	-13.10
-	**3j**	p-$CF_3C_6H_5$	$-OCH_3$	-38.13 (0.19)	25.64 (0.41)	-12.49 (0.43)	-11.25
-	**3d**	1-naphtyl	$-OCH_3$	-38.38 (0.20)	26.03 (0.37)	-12.35 (0.42)	-11.33
-	**3i**	3-pyridyl	$-OCH_3$	-38.67 (0.21)	25.98 (0.36)	-12.69 (0.41)	-11.82
-	**3k**	p-biphenyl	$-OCH_3$	-38.08 (0.20)	26.30 (0.38)	-11.78 (0.43)	-10.27
9	-	-	$-OCH_3$	-36.82 (0.20)	23.61 (0.37)	-13.21 (0.42)	-12.96
4a	-	C_6H_5	$-OCH_3$	-36.31 (0.18)	24.48 (0.39)	-11.83 (0.43)	-11.27
4b	-	p-$MeOC_6H_5$	$-OCH_3$	-34.83 (0.19)	24.82 (0.41)	-10.01 (0.43)	-9.49
4c	-	p-CNC_6H_5	$-OCH_3$	-34.78 (0.18)	24.79 (0.42)	-9.99 (0.44)	-9.62

4d	-	p-biphenyl	$-OCH_3$	-34.71 (0.22)	24.95 (0.36)	-9.76 (0.42)	-9.09
10	-	-	H	-36.61 (0.21)	23.60 (0.36)	-13.01 (0.41)	-12.28
11	-	-	$-OCH_3$	-36.78 (0.18)	23.71 (0.41)	-13.07 (0.44)	-11.90
5a	-	C_6H_5	H	-36.07 (0.19)	24.53 (0.39)	-11.54 (0.43)	-11.76
6a	-	C_6H_5	$-OCH_3$	-35.80 (0.19)	24.74 (0.39)	-11.06 (0.43)	-10.64
6b	-	p-$MeOC_6H_5$	$-OCH_3$	-35.62 (0.20)	24.68 (0.40)	-10.94 (0.45)	-9.90
6c	-	p-MeC_6H_5	$-OCH_3$	-35.78 (0.17)	24.61 (0.38)	-11.17 (0.41)	-10.29
6d	-	p-$NO_2C_6H_5$	$-OCH_3$	-35.61 (0.17)	24.59 (0.37)	-11.02 (0.40)	-10.64
6e	-	p-AcC_6H_5	$-OCH_3$	-35.50 (0.17)	24.67 (0.41)	-10.83 (0.42)	-9.87
6f	-	p-CNC_6H_5	$-OCH_3$	-35.67 (0.21)	24.71 (0.38)	-10.96 (0.43)	-10.51
6g	-	p-$CF_3C_6H_5$	$-OCH_3$	-35.49 (0.19)	24.70 (0.39)	-10.79 (0.43)	-10.05
6h	-	1-naphtyl	$-OCH_3$	-35.72 (0.21)	24.81 (0.40)	-10.91 (0.45)	-10.81
6i	-	p-biphenyl	$-OCH_3$	-35.13 (0.20)	24.98 (0.40)	-10.15 (0.45)	-9.11
6j	-	3-pyridyl	$-OCH_3$	-36.08 (0.22)	24.82 (0.36)	-11.26 (0.42)	-11.74
7a	-	C_6H_5	H	-38.46 (0.18)	25.73 (0.38)	-12.73 (0.42)	-12.00
8a	-	C_6H_5	$-OCH_3$	-38.49 (0.18)	25.81 (0.38)	-12.68 (0.42)	-11.28
8b	-	p-$MeOC_6H_5$	$-OCH_3$	-38.51 (0.21)	25.91 (0.38)	-12.60 (0.43)	-11.12
8c	-	p-MeC_6H_5	$-OCH_3$	-38.69 (0.20)	25.88 (0.41)	-12.81 (0.46)	-11.69

A plot of the ΔGbind values calculated from simulations vs. those obtained experimentally using the corresponding Ki(σ_1) (ΔGbind,(exp), last column in Table 3.1.14) shows a very good correlation coefficient (R^2 = 0.81, see Figure 3.1.25).

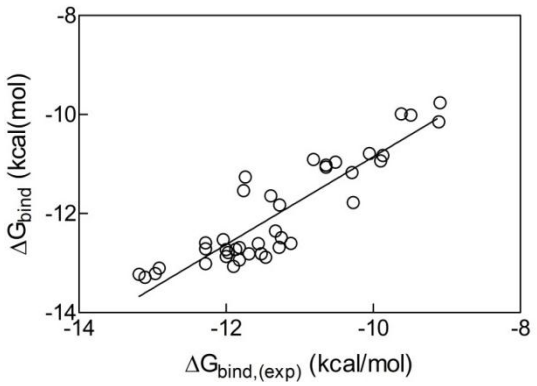
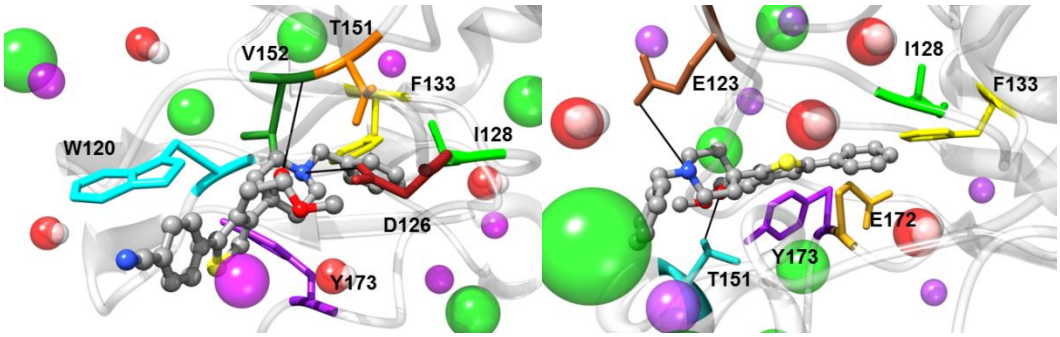


Figure 3.1.25 *Correlation displaying the experimental vs. estimated free energy of binding for all 40 ligand/ $\sigma_1$receptor complexes considered in this work.*

Although this might be a somewhat expected result, given the structural similarity of the compounds considered, it is interesting to observe how MM-PBSA is able not only to rank all molecules in the correct order with respect to their affinity towards the σ_1 receptor but also to discriminate between high affinity (eg., ΔGbind values ranging from approximately -12 kcal/mol to > -13 kcal/mol), intermediate affinity (-10.5 kcal/mol < ΔGbind < -12 kcal/mol), and low affinity compounds (ΔGbind < -10.5 kcal/mol). The most notable results, however, is the fact that these calculations do support the "reverse" binding mode of the molecules with the phenyl substituent in the left position with respect to those featuring the same group in the top position. In fact, if we consider the MD snapshot of the ligand 3h/σ_1 receptor complex shown in Figure 3.1.26(A), we can observe how the main interactions detected after docking are totally preserved along the MD trajectory: a permanent salt bridge (average dynamic distance (ADL) = 4.03 $\pm$ 0.12 Å) is established between the pyperidine NH+ group and the –COO- moiety of Asp126; a persistent H-bond, characterized by an ADL of 1.93 $\pm$ 0.18 Å, is detected between the backbone NH group of Thr151 and Val152 and the O atom of the pyran ring of 3h; a number of stabilizing π–π interactions takes place, involving the side chains of Trp121 and Tyr173 and the ligand benzonitrile and thiophene rings, respectively; and, lastly, the N-benzyl group of 3h remains nicely encased within the hydrophobic pocket generated by residues Ile128 and Phe133. These plethora of favorable, stabilizing interactions results in a calculated value of ΔGbind of -13.29 kcal/mol, in excellent agreement with the corresponding experimental value (-13.10 kcal/mol, see Table 3.1.14), thereby confirming the ranking of 3h as one of the most potent σ_1 ligands of the entire series.

In the case of compound 6a, template molecule representative of all derivatives bearing the phenyl ring in the left position, the protein/ligand interactions depicted at the end of the pharmacophore-assisted docking phase are again detected and preserved during the entire MD simulation. In detail, a stable H-bond (ADL = 1.88 $\pm$ 0.15 Å) is found between the side chain –OH group of Thr181 and the pyran ring oxygen of 6a while a salt bridge (ADL = 3.98 $\pm$ 0.22 Å) between the CO2- group of E123 Glu123 and the pyperidine NH+ of 6a is also evidence for the whole simulation period. Favorable hydrophobic interactions are generated between the phenyl substituent and the side chains of Ile128 and Phe133 and, lastly, the thiophene moiety of 6a is engaged in a positive π–π-interaction with the aromatic side chain of Tyr173, and is further stabilized by the influence of the vicinal alkylic chain of E172Glu172.



(A) (B)

Figure 3.1.26 *Equilibrated MD snapshots of 3h (A) and 6a (B) in complex with the σ_1 receptor. The images are zoomed views of the receptor binding site. The protein structure is depicted as a transparent gray ribbon, while both ligands are shown in atom-colored sticks-and-balls (C, gray; O, red; S, yellow; N, blue. H atoms are omitted). The protein residues mainly involved in the interaction with the ligands are highlighted as colored sticks and labeled. Salt bridge and H-bond interactions are shown as black lines. Some water molecules, ions and counterions are displayed as atom-colored spheres (O, red; H, white; Na+, purple; Cl-, green).*

Conclusion

The introduction of aryl moieties in α- and β-position of complex spirocyclic σ_1 ligands was performed by regioselective direct C-H bond arylation using the catalyst systems PdCl2/2,2'-bipyridyl/Ag2CO3 and PdCl2/P[OCH(CF3)2]3/Ag2CO3, respectively. This late stage functionalization allows the synthesis of a large library of diversely substituted spirocyclic σ_1 ligands, which represent the basis for the development of a pharmacophore model. The main features of the pharmacophore are a positive ionizable group, an hydrogen bond acceptor group, two hydrophobic moieties and one hydrophobic aromatic point. Validation using test set compounds confirms the high quality of the pharmacophore model. The interactions of the spirocyclic σ_1 ligands with the receptor protein were analyzed by docking into a 3D homology model of the σ_1 receptor. These docking experiments led to a very good correlation between the experimentally determined free energy of the ligand binding (calculated from the Ki value) and the estimated free energy form the model. The most striking result of these calculations is the reverse binding mode of ligands with the phenyl moiety in the "top" position compared with ligands bearing the aryl moiety in the "left" position. Moreover, the contribution of single amino acid residues to the overall binding of both types of ligands was analyzed using the docking experiments.

3.1.4 BINDING AFFINITY OF S1RA, A σ_1 RECEPTOR, LACKING A PHARMACOPHORE REQUIREMENTS[49]

In the field of σ_1 drug discovery and development, during the past years several pharmacophore models for σ_1 ligands have been published. All these models share the common features of a basic amino group and at least two hydrophobic substituents at the N-atom.[48,50] Considering the chemical structure of S1RA, however, in a recent viewpoint article,[51] Wünsch observed that, whilst the 2-[1-(2-naphthyl)pyrazol-3-yloxy]-ethyl moiety of S1RA could satisfy the requirement of the first hydrophobic N-substituent, the second hydrophobic substituent was represented only by the morpholine ring. According to the reported pharmacophoric models the high affinity of S1RA is then somewhat unexpected, since this heterocyclic moiety alone is likely not sufficient to fill the second hydrophobic pocket of the binding site of the σ_1 receptor. Here we report a computational study to investigate the origin of this interesting molecular behavior.

A B

Figure 3.1.27 *Chemical structure of S1RA (A) and the benzoxazolone derivative 1c (B).*

Results and discussion

In order to find a molecular-based rationale for the low Ki σ_1 value of S1RA (17 nM),9b) we first mapped the optimized molecular model of S1RA onto our σ_1 three-dimensional (3D) pharmacophore model, as shown in Figure 3.1.28. The corresponding mapping of our best σ_1 ligand, the benzoxazolone derivative 1 (Figure 3.1.27, bottom, Ki σ_1 = 0.098 nM), is also shown for comparison.

As we can see, the oxygen atom of the S1RA ether group serves as an hydrogen bond acceptor, and the basic nitrogen of morpholine fulfills the positive ionizable pharmacophoric requirement. However, as predicted by Wünsch,[51] whilst the 2-napthyl moiety of S1RA maps the first hydrophobic aromatic feature of the pharmacophore, the other hydrophobic pharmacophoric features remain unmapped. Accordingly, the pharmacophoric predicted Ki σ_1 is 148 nM. On the contrary, all suitable chemical groups of 1 nicely overlay the corresponding

pharmacophoric features, resulting in a predicted lower Ki σ_1 value (0.27 nM), in line with experiment.

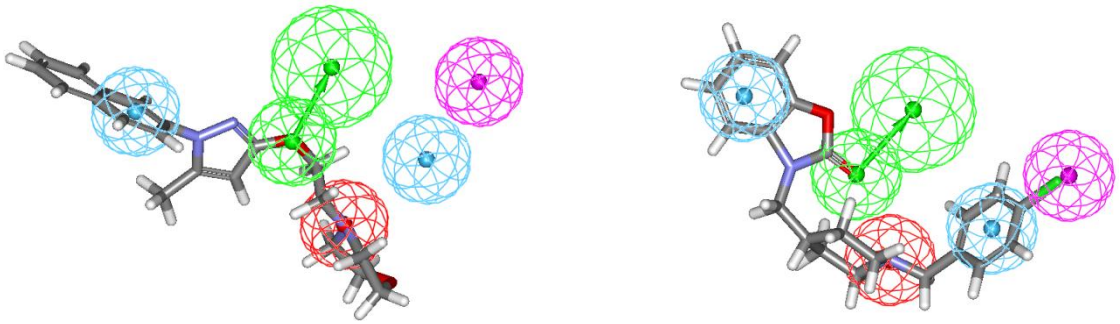


Figure 3.1.28 *Mapping of S1RA (left) and 1 (right) on the σ_1 3D pharmacophore model. The hypothesis features are portrayed as mashed spheres, color-coded as follows: red, positive ionizable (PI); light blue, hydrophobic aromatic (HYAr); pink, generic hy-drophobic (HY), light green, hydrogen bond acceptor (HBA). HBA is actually represented as a pair of spheres (the smaller sphere represents the location of the HBA atom on the ligand and the larger one the location of an HB donor on the receptor).*

To progress further with the analysis, the optimized struc-ture of S1RA was docked into the putative binding pocket on the 3D σ_1 homology model by applying a consolidated procedure,[25,29] and the drug/protein free energy of binding (ΔGbind) was estimated via MM/PBSA calculations[6,52] (see also Materiasl&Methods chapter). As shown in Figure 3.1.29 A, a typical snapshot of the MD-simulated σ_1/S1RA complex reveals the presence of a stable and persistent salt bridge between the –NH+ moiety of the S1RA morpholine ring and the COO- group of Asp126 on σ_1, characterized by an average dynamic length (ADL) of 3.90 ± 0.09 Å. Also, the pyrazole ring and the 2-napthyl moiety of the ligand are engaged in stabilizing π–π interactions with the aromatic side chains of Tyr173 and Phe133, respectively, while a permanent H-bond is detected between the oxygen atom of the 2-aminoethoxy group of S1RA and the –NH group of the σ_1 backbone peptide bond between Thr151 and Val152 (ADL = 2.05 ± 0.06 Å). Finally, several hydrophobic interactions via the side chains of the σ_1 residues Tyr120, Ile128, and Glu172 further contribute to drug/receptor binding stability (Figure 3.1.29 A). With respect to 1c, however, some crucial stabilizing contacts are clearly missing (Figure 3.1.29 B), such as van der Waals and electrostatic interactions between Arg119 and Glu172 with the p-Cl-phenyl and the benzoxazolonic heterocycle of 1, a T-stacking π–π interaction between the protein residue Trp121 and the heteroaro-matic condensed rings of 1, and a parallel-stacking π–π interaction between the protein residue Tyr173 and the benzoxazolone phenyl ring of 1c.
In agreement with the experimental findings, S1RA was endowed with a good affinity for σ_1, as the calculated ΔGbind = -10.41 kcal/mol. In accordance with the components of the binding free energy in Figure 3.1.30 A, for both complexes van der Waals (ΔEVDW) and electrostatic

(ΔEELE) interactions in the gas phase provide the major favorable contribution to ligand binding. Nonpolar solvation energies (ΔGNP), resulting from the burial of S1RA solvent accessible surface area, also afford positive contributions to the binding affinity. Conversely, polar solvation energies (ΔGPB) and entropy components (-TΔS) oppose binding by originating the unfavorable contribution to ΔGbind.

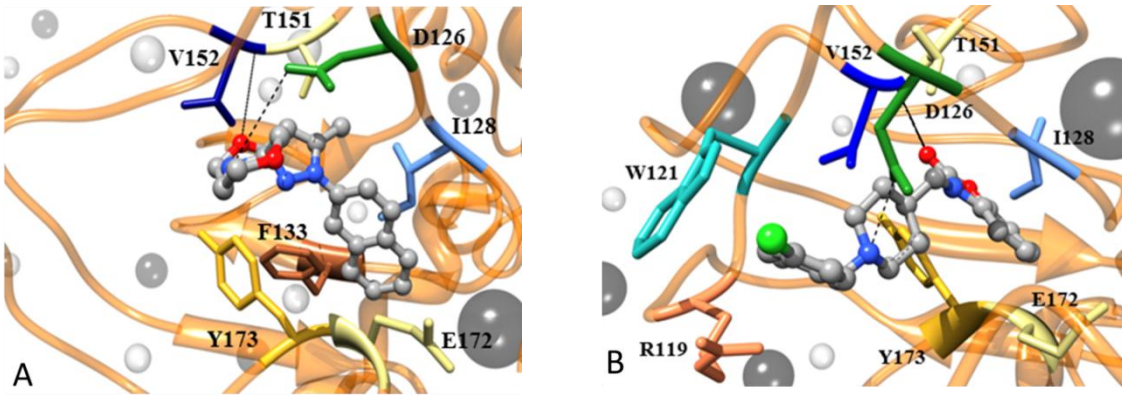


Figure 3.1.29 *Equilibrated MD snapshot of S1RA (A) and 1 (B) in complex with the σ_1 receptor. The ligands are shown in atom-colored sticks-and-balls whilst the secondary structure of the receptor is portrayed as an orange ribbon. The main protein residues involved in ligand/receptor interactions are evidenced as colored sticks and labeled. Salt bridges and hydrogen bonds are highlighted as dotted and continuous lines, respectively. Some Na+ and Cl- counterions are shown as light and dark gray spheres, respectively. Hydrogen atoms and water molecules are omitted for clarity.*

To analyze in details similarities and differences in the binding mode of S1RA and 1 to the σ_1 receptor, the binding free energy was further decomposed into ligand-residue pairs to create the receptor/residue interaction spectrum shown in Figure 3.1.30 B. As we can see, the interaction spectra of the two complexes share some common features, e.g., the major stabilizing contributions stemming from a few groups centered around residues Asp126, Ile128, and Val151.

However, for those interacting residues less shared by the two compounds upon binding to the σ_1 receptor (e.g., Phe133 and Tyr173), a sort of compensatory effect is detected. Where the interaction spectra differ substantially is on the left-hand side: as mentioned above, 1 is strongly stabilized by hydrophobic and $\pi-\pi$ interactions between its p-Cl-phenyl ring and the side chains of Arg119 and Trp121, respectively. Such interactions are obviously absent in the case of S1RA, and this loss of stabilization is only partly compensated by a $\pi-\pi$ interaction between the 2-naphtyl moiety of S1RA and the lateral chain of Phe133. Interestingly, this interaction is forbidden to 1 as, due to its conformation, it cannot penetrate so deeply into the corresponding hydrophobic pocket.

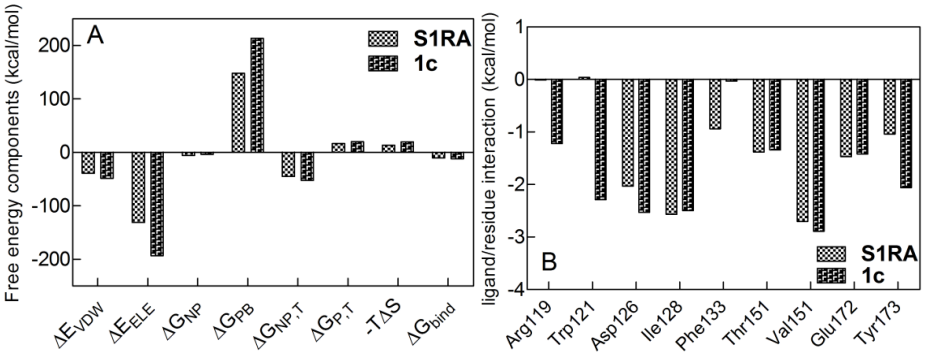


Figure 3.1.30 *(A) Energy components for the binding of S1RA and 1 to the σ_1 receptor. ΔEVDW, van der Waals energy; ΔEELE, electro-static energy; ΔGNP, nonpolar solvation energy; ΔGPB, polar solvation energy; ΔGNP,T = ΔEVDW + ΔGNP; ΔGP,T = ΔEELE + ΔGPB; -TΔS, total entropy contribution; ΔGbind = (ΔGNP,T + ΔGP,T) –TΔS). (B) Per-residue free energy contribution to the binding of S1RA and 1 to the σ_1 receptor.*

Table 3.1.15 *Binding free energy (ΔGbind) and its components for S1RA and 1 in complex with the σ_1 receptor. All energy values are in kcal/mol. The experimental and calculated Ki values (nM) are also reported for comparison.*

S1RA		1	
ΔE_{VDW}	-39.11 ± 0.07	ΔE_{VDW}	-48.68 ± 0.05
ΔE_{ELE}	-131.24 ± 0.10	ΔE_{ELE}	-192.98 ± 0.09
ΔE_{MM}	-170.35 ± 0.12	ΔE_{MM}	-241.66 ± 0.10
ΔG_{PB}	147.96 ± 0.13	ΔG_{PB}	213.39 ± 0.15
ΔG_{NP}	-5.72 ± 0.02	ΔG_{NP}	-3.64 ± 0.03
ΔG_{SOL}	142.24 ± 0.13	ΔG_{SOL}	209.75 ± 0.15
ΔH_{bind}	-28.11 ± 0.18	ΔH_{bind}	-31.90 ± 0.18
$-T\Delta S_{bind}$	17.70 ± 0.28	$-T\Delta S_{bind}$	19.85 ± 0.32
ΔG_{bind}	-10.41 ± 0.33	ΔG_{bind}	-12.05 ± 0.37
$K_{i(calcd)}$	**24 nM**	**$K_{i(calcd)}$**	**1.48 nM**
$K_{i(exp)}$	**17 ± 7 nM**	**$K_{i(exp)}$**	**0.098 ± 0.03 nM**

Perhaps the most important finding is that, according to the predicted binding mode, S1RA leaves a moiety with polar characteristic such as the morpholine ring partially exposed to the solvent. In fact, as seen in the MD trajectory, the morpholine oxygen atom is always surrounded by water molecules; this in turn, results in a lower desolvation energy upon binding of S1RA (ΔGPB = +147.96 kcal/mol, Table 3.1.15) with respect to 1, for which the

deeply burying of the p-Cl-phenyl group into the binding pocket requires a substantial desolvation penalty (ΔGPB = +213.39 kcal/mol, Table 3.1.15).

Conclusions

The present in silico study was spurred by the inspiring observation that S1RA - a potent σ_1 antagonist that could possibly become the first σ_1 selective drug in clinical use – notwithstanding its molecular structure does not fulfill all well-established pharmacophoric requirements of σ_1 ligands, is still endowed with a high affinity for the receptor. The results of the investigation led to the discovery that, aside a substantial number of stabilizing interactions between S1RA and σ_1, a significant favorable contribution to binding is afforded by a low molecular desolvation penalty paid upon binding, due to the exposition of a polar moiety to the solvent.

This underlying binding mode of S1RA (never described so far for any other σ_1 ligand), if aptly exploited in association with all other fundamental σ_1 ligand pharmacophoric requirements might lead to the design and synthesis of new molecules endowed with ideal affinity towards the receptor. And it is just under these exciting perspectives that our jointed laboratories are now spending their present efforts in computer-assisted design and synthesis of new, potent σ_1 ligands.

3.1.5 STUDY OF ENANTIOMERIC COMPOUND RC-33 AS σ_1 RECEPTOR AGONIST[53]

In this study we present the synthesis, and biological investigation of new σ_1 receptor ligands based on arylalkenylamine and arylalkylamine scaffolds (Figure 3.1.31).[54]

Among our compound library, the most promising molecule is 1-[3-(1,1'-biphen)-4-yl]butylpiperidine hydrochloride (RC- 33·HCl, Figure 3.1.31), showing excellent σ_1 receptor affinity (Ki= 0.700.3 nm) along with high selectivity over the σ_2 subtype, m- and k-opioid receptors, and the PCP binding site of NMDA receptors.[55] RC-33·HCl also turned out to be a potent σ_1 receptor agonist in our validated PC12 cell model of neuronal differentiation.

Indeed, this molecule is able to potentiate both nerve growth factor (NGF)-induced neurite outgrowth and elongation in the absence of toxic effects and at lower doses than the well-characterized σ_1 receptor agonist PRE-084.[55] Finally, RC-33·HCl was found to be metabolically stable in several biological matrices such as mouse and rat blood, and rat, dog and human plasma. Significant degradation was only observed in rat and human liver S9 fractions in the presence of NADPH.[56] According to these results, we identified RC-33·HCl as an optimal

candidate for the hit-to-lead process. Because the enantiomers of a chiral biologically active molecule may behave differently under physiological conditions (i.e., they can exhibit different pharmacological and/or toxicological profiles), the investigation of enantiopure forms is a key step in the drug discovery and development processes.[57] Accordingly, in the present work we directed our efforts toward investigating the role of chirality in the biological activity and metabolic processing of RC-33. To this end, we isolated the pure enantiomers of RC-33 by chiral chromatography and carried out the assignment of their absolute configuration.

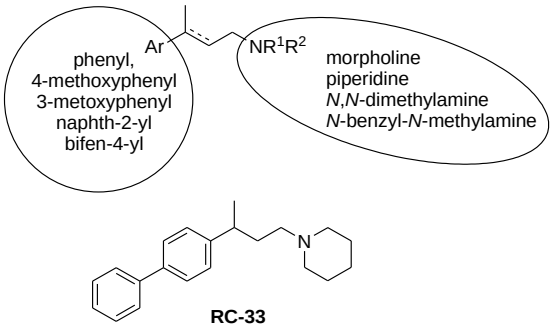


Figure 3.1.31 *Structure of arylalkenyl- and arylalkylamines and of the most promising compound RC-33.*

The binding of enantiopure RC-33 at the σ_1 receptor was then investigated both in silico, by molecular dynamics (MD) simulations using our recently developed homology model for the σ_1 receptor,[1] and in vitro, by competition experiments with radioligand. The in vitro binding affinities of RC-33 enantiomers for opioid receptors and the PCP binding site of the NMDA receptor were also investigated, together with their capacity to promote NGF-induced neurite outgrowth in PC12 cells. As the last step in the present hit-to-lead process, both RC-33 enantiomers underwent toxicity and in vitro metabolic studies.

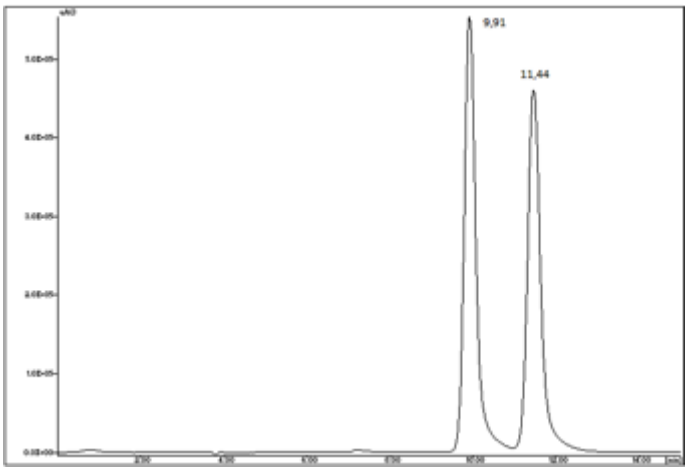


Figure 3.1.32 *Analytical separation of rac-1 on a Chiralcel OJ-H (0.46 cm diameter x 15 cm length, 5 µm) column. Mobile phase, methanol/diethylamine (100/0.1, v/v), flow rate, 0.5 ml/min; detection at 250 nm.*

Receptor binding studies

The affinities of the pure RC- 33·HCl enantiomers for σ_1 and σ_2 receptors were determined in competition experiments with radioligands. The highly σ_1-selective radioligand [3H](+)-pentazocine and homogenates of guinea pig brain cortex were used in the σ_1 assay. Nonspecific binding was recorded in the presence of non-radiolabeled (+)-pentazocine in large excess. In the σ_2 assay, membrane preparations of rat liver served as the source for σ_2 receptors. The nonselective radioligand [^{3}H]DTG was employed in the σ_2 assay because a σ_2-selective radioligand is not yet commercially available. To mask the σ_1 receptors, an excess of non-tritiated (+)-pentazocine was added to the assay solution, while a high concentration of non-tritiated DTG was used to determine nonspecific binding. To investigate receptor selectivity, the binding properties of pure RC-33·HCl enantiomers were also investigated for m-, k-, d-opioid receptors and the PCP binding site of NMDA receptors using guinea pig brain (μ and κ), rat brain (δ) and pig brain cortex (PCP binding site of NMDA) as sources of receptor material. [^{3}H]DAMGO (μ), [^{3}H]U- 69,593 (κ), [^{3}H]DPDPE (δ), and [^{3}H]MK-801 (NMDA) were used as specific radioligands. Nonspecific binding was determined with unlabeled naloxone (μ), unlabeled U-69,593 (κ), unlabeled morphine (μ) and unlabeled (+)-MK-801 (PCP binding site of NMDA). The residual binding of the radioligand is given at a test compound concentration of 1 mm (opioid receptor binding assays) or 10 mm (NMDA receptor binding assays). Binding results are summarized in Table 3.1.16, together with data for RC-33·HCl, reported for comparison purposes.

The enantiomeric arylalkylamines (S)-RC-33·HCl and (R)-RC-33·HCl herein presented showed interesting binding properties towards the σ_1 receptor subtype (Ki values < 2 nM, Table 3.1.16) and a good selectivity over the σ_2, μ-, κ-, δ- opioid, and the PCP binding site of NMDA receptors. As concerns the role of chirality in the σ_1 receptor binding, the two enantiomers and RC-33·HCl are endowed with a comparable affinity for the σ_1 receptor. Importantly, this finding supports the notion that the interaction of these compounds with this biological target is non-stereoselective.

Table 3.1.16 Affinities of compounds **RC-33**·HCl towards σ_1, σ_2, NMDA, μ-, κ- and δ-opioid receptors.

Compd[a]	K_i ± SEM (nM)[b]		Displacement (%) of specific radioligands			
	σ_1	σ_2	μ-opioid	κ-opioid	δ-opioid	NMDA
RC-33	0.9±0.3	29±5	0	36	0	25
(*S*)-**RC-33**	1.9±0.2	34±8	12	0	0	11
(*R*)-**RC-33**	1.8±0.1	45±16	14	0	5	8

[a] Compounds were tested as hydrochlorides. [b] Values are means ± SEM of three experiments.

MD simulations

For a molecular-level explanation of the non-stereoselective binding of RC-33 enantiomers to the σ_1 receptor, the putative binding modes of (R)- and (S)-RC-33 on our σ_1 receptor 3D homology model were retrieved by exploiting the currently available preliminary information on sequence–structure relationships and mutagenesis studies.[1,58]

The two enantiomers of RC-33 were docked into the putative binding site of our 3D receptor model, and their affinity toward the receptor was scored by MM/PBSA analysis.[6]

In a typical structure of the MD-simulated σ_1–ligand complexes, both (R)- and (S)-RC-33 are oriented horizontally inside the receptor binding pocket and adopt similar binding poses, as illustrated in Figure 3.1.33 For both enantiomers, the -NH+ moiety of the ligand piperidine ring is anchored around the negatively charged side chain of Asp126 from the σ_1 protein, interacting with each other through a permanent salt bridge.

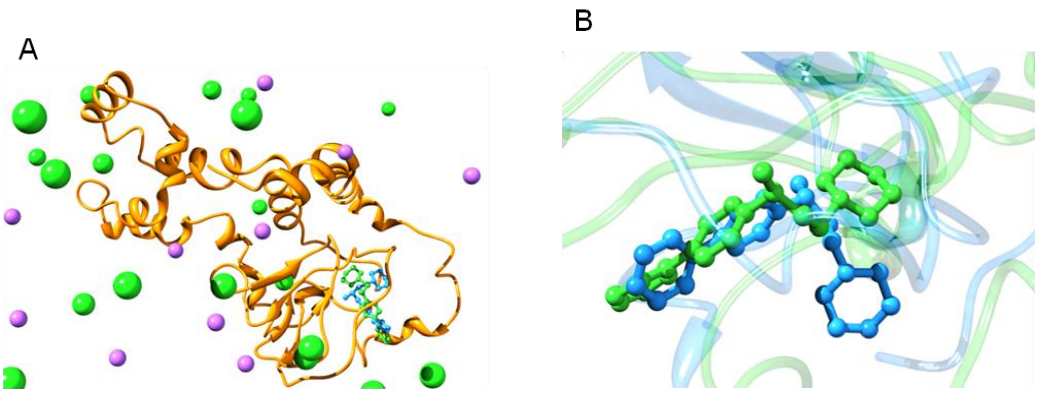


Figure 3.1.33 *A) Overlay of representative structures for (R)-RC-33 (light blue balls-and-sticks) and (S)-RC-33 (spring green balls-and-sticks) in complex with the σ_1 receptor (orange ribbon), taken from equilibrated MD snapshots. Some Na+ and Cl- ions are also shown as purple and light green spheres, respectively. Hydrogen atoms and water molecules are not shown for clarity. B) Zoomed view of the binding modes of (R)-RC-33 and (S)-RC-33 in the σ_1 binding pocket. In this image, the protein is portrayed as transparent ribbon, colored according to the respective ligand.*

As tracked by MD simulations, the average distance for the salt bridge through the proton at the cationic moiety of RC-33 and the COO- group of σ_1 Asp126 is 2.9 ± 0.1 Å for the R enantiomer and 2.7 ± 0.1 Å for the S enantiomer (Figure 3.1.34). The two aromatic rings of RC-33 are packed parallel to Tyr120 and perpendicularly with respect to Tyr173, resulting in strongly stabilizing π–π interactions. Lastly, as can be appreciated from Figure3.1.3 4, in both cases the entire RC-33 molecule is nicely lined by the side chains of several other σ_1 residues,

and the relevant hydrophobic interactions which originate from these intermolecular contacts all contribute to stabilize receptor– ligand binding.

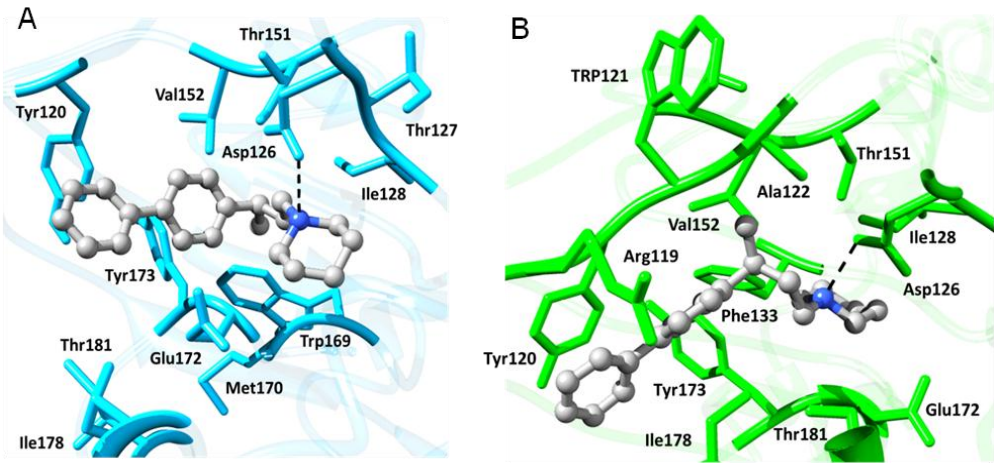


Figure 3.1.34 Equilibrated MD snapshots of A) (R)-RC-33 and B) (S)-RC-33 in complex with the σ_1 receptor. The images are zoomed views of the receptor binding site. The protein structure is depicted as a transparent light blue and light green ribbon, respectively, while both enantiomers are shown in atom-colored sticks-and-balls (C, gray; N, blue; H atoms are omitted). The protein residues mainly involved in the interaction with the ligands are highlighted as colored sticks and labeled. Salt bridge interactions are shown as dotted black lines. Water molecules and counterions are not shown for clarity.

The similar binding mode and the equivalent type and number of interactions predicted for (R)- and (S)-RC-33 in complex with the σ_1 receptor constitute the molecular rationale, according to which our MM/PBSA estimated binding affinities (ΔGbind) for these two enantiomers for the protein are nearly identical, that is, ΔGbind = -11.42 ± 0.29 kcal mol^{-1} for (R)-RC-33 and ΔGbind = -11.15 ± 0.31 kcal mol^{-1} for (S)-RC-33. In accordance with the components of the binding free energy in Figure 3.1.35 a, for both σ_1–RC-33 enantiomer complexes, van der Waals (ΔEVDW) and electrostatic (ΔEELE) interactions in the gas phase provide the major favorable contributions to ligand binding. Nonpolar solvation energies (ΔGNP), resulting from the burial of RC-33 solvent-accessible surface area, also afford positive contributions to the binding affinity. Conversely, polar sol-vation energies (ΔGPB) and entropy components (-TΔS) oppose binding by making unfavorable contributions to ΔGbind.

To analyze in detail the similarities and differences in the binding modes of the two RC-33 enantiomers with the σ_1 receptor, the binding free energy was further decomposed into ligand–residue pairs to create the receptor–residue interaction spectrum shown in Figure 3.1.35b. As can be observed, the interaction spectra of the two complexes are similar to each other, with the major contributions stemming from a few groups centered around residues Asp126, Tyr120, and Glu172 in both cases, as expected. Furthermore, for interacting residues

that are not shared by the two enantiomers of RC-33 upon binding to the s1 receptor (i.e., Thr127, Trp169, and Met170 for the R enantiomer, and Arg119, Trp121, and Phe133 in the case of the S enantiomer), a sort of compensatory effect is present, ultimately resulting in the non-stereoselective binding of the σ_1 receptor to RC-33.

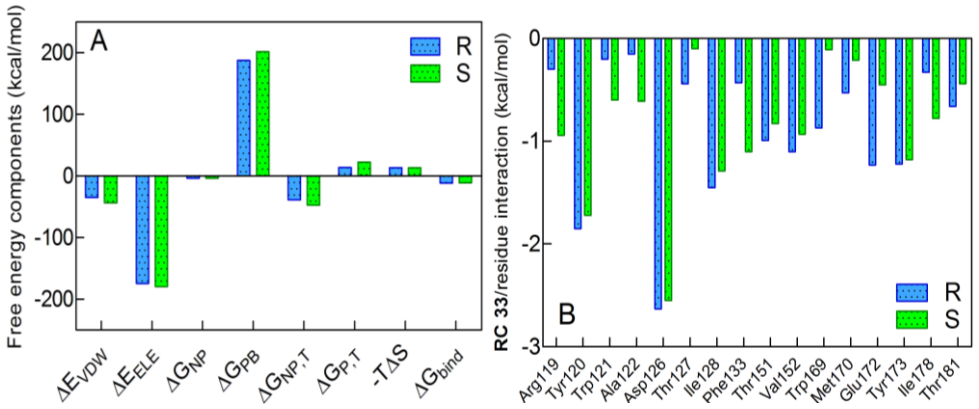


Figure 3.1.35 *A) Energy components for the binding of (R)-RC-33 and (S)-RC-33 to the σ_1 receptor. ΔEVDW, van der Waals energy; ΔEELE, electrostatic energy; ΔGNP, nonpolar solvation energy; ΔGPB, polar solvation energy; ΔGNP,T = ΔEVDW + ΔGNP; ΔGP,T = ΔEELE + ΔGPB; -TΔS, total entropy contribution; ΔGbind = ΔGNP,T + ΔGP,T –TΔS. B) Per-residue free energy contribution to the binding of (R)-RC-33 and (S)-RC-33 to the σ_1 receptor.*

NGF-induced neurite outgrowth in PC12 cells and cytotoxicity

Pure enantiomers of RC-33·HCl were tested in our validated PC12 cell model of neuronal differentiation to analyze their effect on NGF-induced neurite outgrowth. In previous experiments on PC12 cells rac-RC-33·HCl displayed an agonist profile, consistently and significantly potentiating NGF-induced neurite outgrowth at concentrations as low as 0.25 mm.[55] Therefore, we chose this concentration to test (S)- and (R)-RC-33·HCl on PC12 cells. rac-RC-33·HCl and its enantiomers appear to be almost equally effective. Indeed, both enantiomers were able to increase the percentage of cells with neurite outgrowth with respect to NGF alone (152.8±7.6 %, $p<0.01$ and 143.1 3.8% $p<0.01$ for (S)- and (R)-RC-33·HCl, respectively), showing an effect very similar to that of rac-RC-33·HCl (157.7±11.0 %, $p<0.001$) (Figure 3.1.36). Additionally, their effect was totally blocked by co-administration of the selective σ_1 antagonist NE- 100. Taken together, these results confirm that (S)- and (R)-RC-33·HCl similarly modulate NGF-induced neurite outgrowth, specifically acting as σ_1 receptor agonists.

Finally, an MTT-based cytotoxicity assay, performed after treating HaCaT cells with (S)- or (R)-RC-33·HCl and with rac-RC- 33·HCl for 48 h (Figure 3.1.37) or 72 h confirmed that, similarly to

the racemic mixture, both enantiomers are nontoxic at the concentration used in our NGF-induced neurite outgrowth experiments.

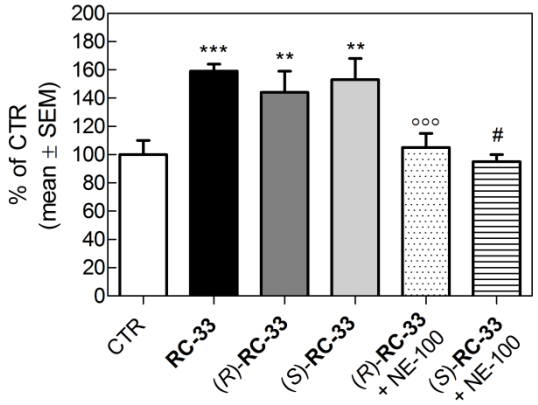


Figure 3.1.36 *Potentiating effect of σ1 receptor ligands RC-33·HCl and enantiomers (at 0.25 μM) on NGF (2.5 ng mL^{-1})-induced neurite outgrowth. NE-100 (3 μM) co-administration totally blocked the potentiating effect. Histograms represent the mean ± SEM of at least three different experiments. Data are expressed as percentage of control (CTR) (NGF alone). ANOVA: p<0.0005; Tukey's post-hoc test: ***=p<0.001; **=p<0.01 vs NGF alone. °°°=p<0.001 vs (-)-(R)-RC-33•HCl; #=p<0.05 vs (+)-(S)-RC-33•HCl.*

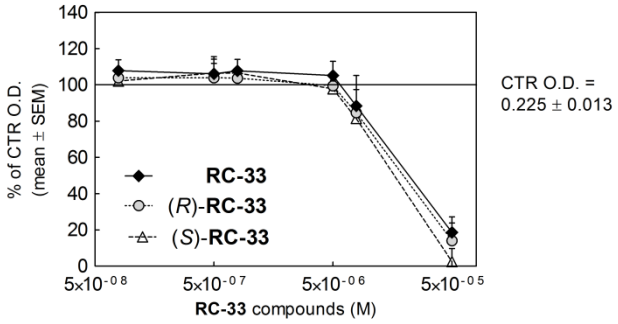


Figure 3.1.37 *Effect of σ_1 receptor ligands RC-33·HCl and enantiomers on HaCaT cell viability. Values are the mean ± SEM of at least three different experiments. ANOVA: p<0.0001; Tukey's post-hoc test:* ***=p<0.0001 CTR vs (R)-RC-33·HCl; (S)-RC-33·HCl and rac-RC-33·HCl.***

In vitro metabolic stability studies

To investigate the role of chirality on metabolic processes, the hepatic metabolic stability of RC-33·HCl enantiomers was tested in vitro using rat and human liver S9 fraction. We previously described rac-RC-33·HCl in vitro metabolic stability in several biological matrices, such as mouse and rat blood, rat, dog and human plasma, as well as rat and human liver S9 fractions, containing microsomes (metabolic enzymes of phase I) and cytosol (metabolic enzymes of phase II).[56] rac-RC-33·HCl showed high metabolic stability in all the investigated matrices, with the only exception of liver S9 fractions in the presence of NADPH. Therefore, in the present work we addressed our attention at evaluating the hepatic metabolism of (R)- and (S)-RC-33·HCl in rat and human liver S9 fractions both in the presence and in the absence of NADPH.

To this aim, our previously validated and reproducible method using an ultrafast liquid chromatography system interfaced with a photodiode array detector (UPLC/UV/PAD, Acquity, Waters) was applied.

Briefly, UPLC/UV/PAD analyses were carried out with a BEH Shield RP18 column, eluting with a mixture of water and methanol, containing 0.1% formic acid, in gradient at a flow rate of 0.5 mL min^{-1}. Calibration curves in the range of 1–200 mm (eight points of calibration) were determined, adding known concentrations of each enantiomer of RC-33·HCl to each considered biological matrix, previously deproteinized. In all cases calibration curves with a quadratic correlation coefficient (R^2) of 0.9999 were obtained. Quantification was performed by comparing the chromatographic peak areas for test solutions with those of external standard. The matrices could be directly analyzed by UPLC without clean-up steps.

Similarly to the racemate, both RC-33·HCl enantiomers underwent relatively weak non-oxidative metabolism, with degradation of ~30% within 4 h at 37° C, both in rat and human (Figure 3.1.38). In contrast, they were subjected to a relevant oxidative metabolism in both rat and human liver S9 fractions within the same frame time at 37° C. Notably, degradation of ~50% in both rat and human was evidenced for (R)-RC-33·HCl, while degradation of ~70% in both the biological matrices was observed for (S)-RC-33·HCl (Figure 3.1.39a,b). Moreover, as clearly shown in the semi-logarithmic plots shown in Figure 3.1.39 c,d, the oxidative degradation of both enantiomers follows first-order kinetics, with a half-life (t1/2) of 112 min in rat and 130 min in human for (S)-RC-33·HCl, and of 225 min in rat and 246 min in human for (R)-RC-33·HCl. Overall, results of the in vitro hepatic metabolic stability studies clearly show that RC-33·HCl enantiomers are similarly metabolized by non-oxidative processes in both rat and human; on the other hand, their degradation in the presence of NADPH was unequivocally proven to be enantioselective with a preference for the S enantiomer in both species.

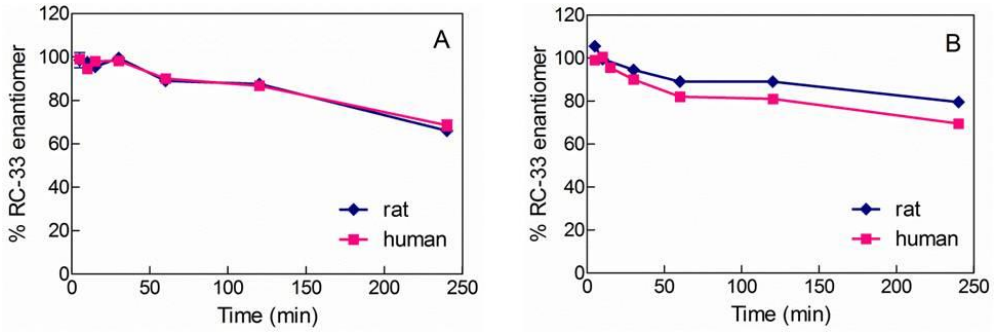


Figure 3.1.38 *Degradation time-courses of A) (S)-RC-33·HCl and B) (R)-RC-33·HCl in rat and human liver S9 fraction in absence of NADPH at 37°C. Data are expressed as the mean ± SD of two independent experiments.*

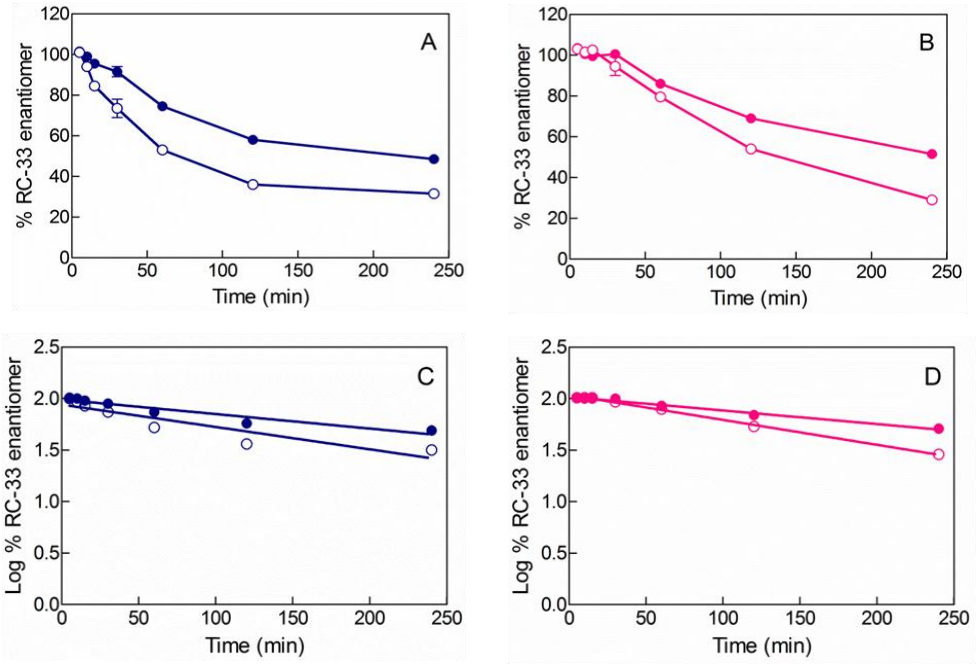


Figure 3.1.39 *Degradation time-courses of (S)-RC-33·HCl (open circles) and (R)-RC-33·HCl (full circles) in A) rat and B) human liver S9 fraction in presence of NADPH at 37°C. Data are expressed as the mean ± SD of two independent experiments. Enantiomers of RC-33·HCl showed a significant degradation in liver S9 fraction in the presence of NADPH, following a first order kinetic, as confirmed by the semilogaritmic plot C) for rat and D) for human.*

Conclusions

In the present work we addressed the role of chirality in the biological activity of RC-33, recently studied by us in its racemic form. RC-33 enantiomers were isolated by enantioselective semi-preparative HPLC, and their absolute configuration was assigned by applying an integrated CD analysis/chemical correlation strategy. Their in vitro binding affinities toward σ, opioid, and PCP binding site of the NMDA receptors were investigated, and their agonist/antagonist profiles at σ_1 receptor were derived. In synthesis, (R)-RC-33 and (S)-RC-33 showed nearly the same affinity for the σ_1 receptor and were equally effective as σ_1 receptor agonists, without toxic effects at the concentration shown to be effective. To provide a rational in silico explanation of the experimental evidence that the interaction between RC-33 enantiomers and the σ_1 receptor is non-stereoselective, the two enantiomers were docked into the putative binding site of our 3D receptor model, and their affinity toward the receptor was scored by MM/PBSA analysis. The results of our modeling investigations confirm that both enantiomers of RC-33 can be accommodated within the σ_1 binding site and establish the same network of stabilizing interactions with the target, supporting the non-stereoselective binding of RC-33 to the σ_1 receptor. Lastly, the hepatic metabolic stability RC-33 enantiomers was also tested in vitro using rat and human liver S9 fractions. (S)- and (R)-RC-33 were similarly metabolized by non-oxidative processes in both rat and human.

Interestingly, compared with the corresponding enantiomer, (S)-RC-33·HCl was preferentially metabolized in the presence of NADPH in both rat and human, thus proving that the oxidative metabolic processes possess the same enantiomeric preference in both species.

In summary, based on the overall results of the biological investigations, (R)-RC-33 emerged as the eutomer. Because σ_1 receptor agonists such as PRE-084 were recently shown to be promising candidates for a therapeutic strategy for ALS,[59] (R)- RC-33 represents the optimal candidate for the in vivo investigation.

Its pharmacokinetic profile and results of both shortand long-term investigations as drug candidate in an animal model of ALS will be reported in due course.

3.1.6 THE SIGMA ENIGMA: IN SILICO/IN VITRO SITE DIRECTED MUTAGENESIS STUDIES UNVEIL σ_1 RECEPTOR LIGAND BINDING

Several synthetic molecules belonging to different structural classes were thus found to bind to the σ_1 receptor. (+)-Pentazocine (PTZ) represents a very potent and highly selective σ_1 agonist, which is used as a "gold standard" in σ_1-ligand binding studies.

Membrane/protein model generation.

The CHARMM-GUI Lipid Builder web application[60,61] was used to insert our σ_1 3D model in a pre-equilibrated membrane. The membrane was built using 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn- glycero-3-phosphoethanolamine (POPE) and cholesterol in a 2:2:1 proportion. 125 and 115 lipid molecules constituted the top and bottom membrane leaflets, respectively. Both sides of the membrane were solvated by two layers of 0.150 M NaCl solution of TIP3P water molecules,[16] each extending 20 Å from the membrane lipid polar heads. The entire system was parameterized using the lipid11[62] and ff03[17] force fields for the lipids and the protein, respectively. After an initial relaxation (for details see Materials&Methods chapter), the systems was allowed to equilibrate for 50 ns adopting a semi-isotropic constant surface tension of 15.0 dyne/cm. Once an equilibrated value of area per lipid was reached, all remaining constraints were finally released, and additional 50 ns of MD NPT simulations were performed. At the end of the MD simulation, we taken the final snapshot and we separated the protein from the membrane. Then, we used this optimized protein structure as starting conformation for the subsequent docking studies.

Docking Studies.

The model structures of PTZ was sketched and geometrically optimized using Discovery Studio (DS, version 2.5, Accelrys). The optimized structure of PTZ was then docked into the σ_1 putative binding pocket by applying a consolidated procedure, already described in this thesis. All

docking experiments were performed with Autodock 4.3/Autodock Tools 1.4.6 on a win64 platform.

Molecular Dynamics simulations.

The ligand/receptor complex obtained from the docking procedure was further reinserted into the membrane and subsequently refined in Amber 12[63] as follow: each system was relaxed by 500 steps of steepest descent followed by 500 other conjugate-gradient minimization steps and then gradually heated to a temperature of 300 K in intervals of 50 ps of NVT MD, using a Verlet integration time step of 1.0 fs. The Langevin thermostat was used to control temperature, with a collision frequency of 2.0 ps^{-1}. The SHAKE method[18] was used to constrain all of the covalently bound hydrogen atoms, while long-range nonbonded van der Waals interactions were truncated by using dual cutoffs of 6 and 12 Å. The particle mesh Ewald (PME) method was applied to treat long-range electrostatic interactions. All simulations were carried out with periodic boundary conditions. The density of the system was subsequently equilibrated via MD runs in the NPT ensemble, with isotropic position scaling and a pressure relaxation time of 1.0 ps, for 50 ps with a time step of 1 fs. Each system was further equilibrated using NPT MD runs at 300 K, with a pressure relaxation time of 2.0 ps. Three equilibration steps were performed, each 2 ns long and with a time step of 2.0 fs. To check the system stability, the fluctuations of the rmsd of the simulated position of the backbone atoms of the σ_1 receptor with respect to those of the initial protein were monitored. All chemico-physical parameters and rmsd values showed very low fluctuations at the end of the equilibration process, indicating that the systems reached a true equilibrium condition. The equilibration phase was followed by a data production run consisting of 15 ns of MD simulations in the NVT ensemble. Only the last 10 ns of each equilibrated MD trajectory were considered for statistical data collections.

Free energy of binding analysis.

The binding free energy, ΔGbind, between PTZ and the σ_1 receptor was estimated by resorting to the MM/PBSA approach.[6] The free energy was calculated for each molecular species (complex, receptor, and ligand), and the binding free energy was computed as the difference..

A per residue binding free energy decomposition was performed exploiting the MD trajectory of each given compound/receptor complex, with the aim of identifying the key residues involved in the ligand-receptor interaction. This analysis was carried out using the MM/GBSA approach,[27,64] and was based on the same snapshots used in the binding free energy calculation.

Computational mutagenesis.

For the computational mutagenesis studies, mutation of the selected residues into alanine was performed on the membrane-bound equilibrated structure of the wild-type σ_1 receptor complex. Each mutated complex was equilibrated in a 0.150M NaCl TIP3P water solution to relax eventual conformational perturbations induced by residue mutation. Each system was then first subjected to 100 energy minimization steps, followed by further solvent equilibration for 100 ps of MD at 310 K. Finally, the entire system was equilibrated for 2 ns at 310 K. The subsequent MD simulation was carried out with the Amber 12 program as described above.

Results

To better understand the structure and ligand binding of the σ_1 receptor we have now carried out a coupled in vitro/in silico site-directed mutagenesis study of the receptor. Specifically, on the basis of preliminary in silico information 23 mutants of the σ_1 receptor were cloned, expressed, and their interactions with PTZ were determined experimentally (Figure 3.1.40c). As described below, all direct and indirect effects exerted by the mutant residues on the receptor/agonist interactions could be reproduced in silico and rationalized at the molecular level (Table 3.1.17).

At first we considered those σ_1 residues belonging to the protein putative ligand binding site. To the purpose, the interaction of PTZ with the wild-type (wt) σ_1 receptor 3D model was simulated using molecular dynamics (MD): the corresponding estimated drug/receptor affinity value K_i^c= 45 nM (Table 3.1.18) compares well with the experimentally measured value of K_i^e= 15 nM (19). Specifically, the following interactions were found to be essential for PTZ binding (Figure 3.1.41 a): i) a permanent salt bridge between the NH+ moiety of PTZ and the COO- group of D126 (average dynamic length (ADL) = 3.93 ± 0.09 Å); ii) a stable hydrogen bond between the carboxylate group of E172 and the hydroxyl substituent of PTZ (ADL = 1.98 ± 0.04 Å); iii) a T-stacking π-π interaction between the side-chains of Y120 and W121 and the heteroaromatic condensed rings of PTZ; and iv) highly stabilizing van der Waals and electrostatic interactions between R119, I128 and Y173 and the aliphatic/aromatic portions of the ligand. These receptor/ligand interactions were quantified via a per residue deconvolution of the free energy of binding (Figure 3.1.40 b,c).

Accordingly, the salt bridge and the hydrogen bond involving D126 and E172 are responsible for stabilizing contributions of -2.89 kcal/mol and -1.83 kcal/mol. Substantial van der Waals and electrostatic interactions are further contributed by R119 (-0.98 kcal/mol), Y120 (-1.49 (kcal/mol), W121 (-1.37 kcal/mol), I128 (-2.04 kcal/mol), and Y173 (-1.95 kcal/mol). T151 and V152 additionally contribute in stabilizing PTZ/σ_1 binding by -0.36 kcal/mol and -0.38 kcal/mol, respectively.

Table 3.1.17 *In silico site-directed mutagenesis of the selected σ_1 residues. In silico estimated free energy of PTZ binding (ΔGbind) of different mutants of the human σ_1 receptor as generated by computational mutagenesis. Errors are given in parenthesis as standard errors of the mean (SEM).* [a] *$\Delta\Delta$Gbind = ΔGbind,wt – ΔGbind,mutant.* [b] *By definition, a negative value of $\Delta\Delta$Gbind indicates a favorable contribution for the wt residue in that position and vice versa.*

σ_1 residue	ΔG_{bind} (kcal/mol)	[a,b] $\Delta\Delta G_{bind}$ (kcal/mol)
wt	-10.02 (0.05)	-
S101A	-8.16 (0.08)	-1.86
Y103A	-7.89 (0.09)	-2.13
F107A	-8.05 (0.05)	-1.97
Y120A	-9.34 (0.08)	-0.68
W121A	-8.48 (0.06)	-1.54
S125A	-8.57 (0.09)	-1.45
D126A	-7.01 (0.10)	-3.01
D126E	-9.13 (0.07)	-0.89
T127A	-8.78 (0.11)	-1.24
I128A	-7.55 (0.06)	-2.47
V145A	-9.89 (0.07)	-0.13
F146A	-9.42 (0.09)	-0.60
Y147A	-8.80 (0.06)	-1.22
E172A	-8.00 (0.05)	-2.02
E172D	-7.88 (0.06)	-2.14
Y173A	-7.32 (0.11)	-2.70
R175A	-8.00 (0.10)	-2.02
T181A	-9.99 (0.11)	-0.03
F184A	-10.55(0.07)	0.53
T189A	-9.87 (0.07)	-0.15
F191A	-9.13 (0.10)	-0.89
217Stop	-8.80 (0.06)	-1.22

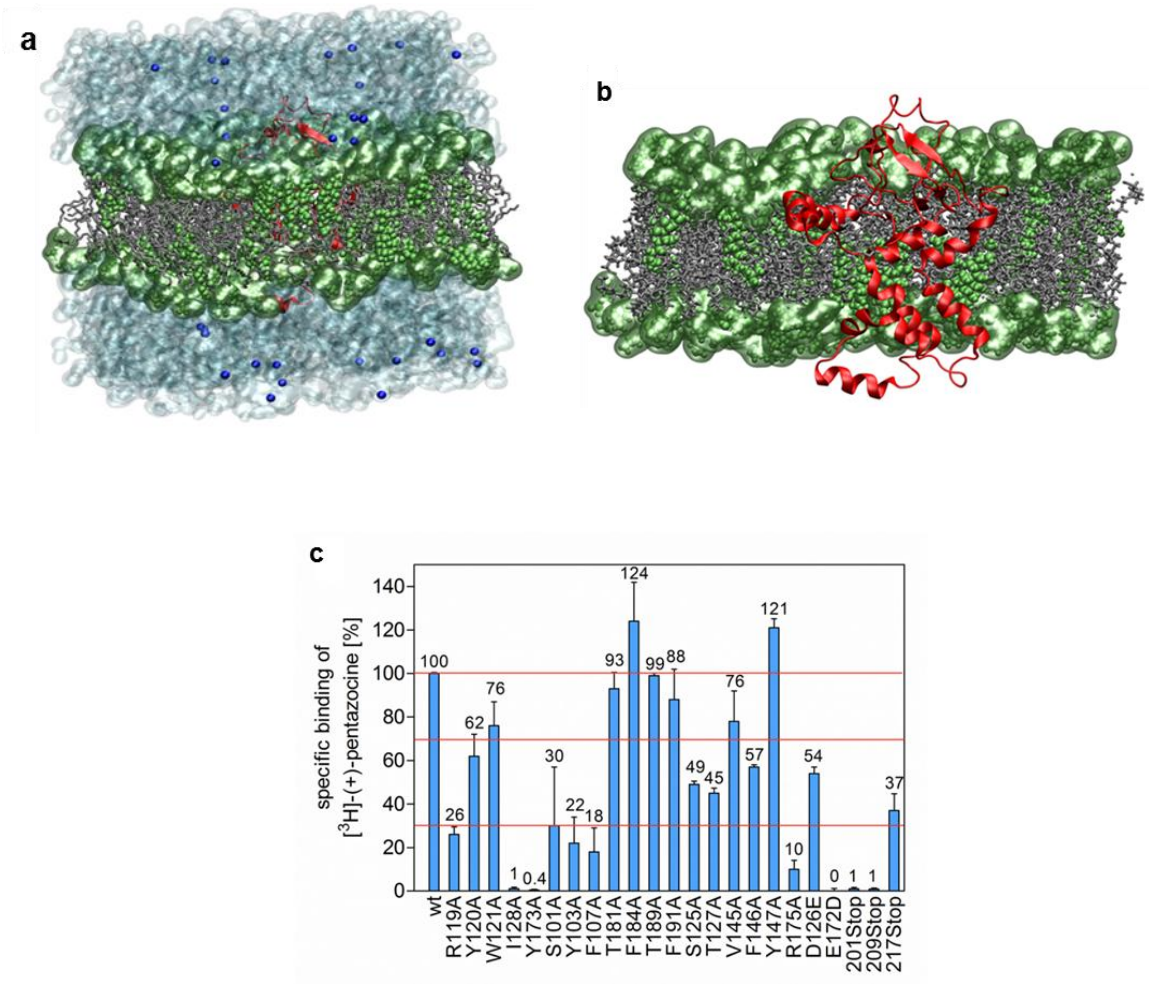
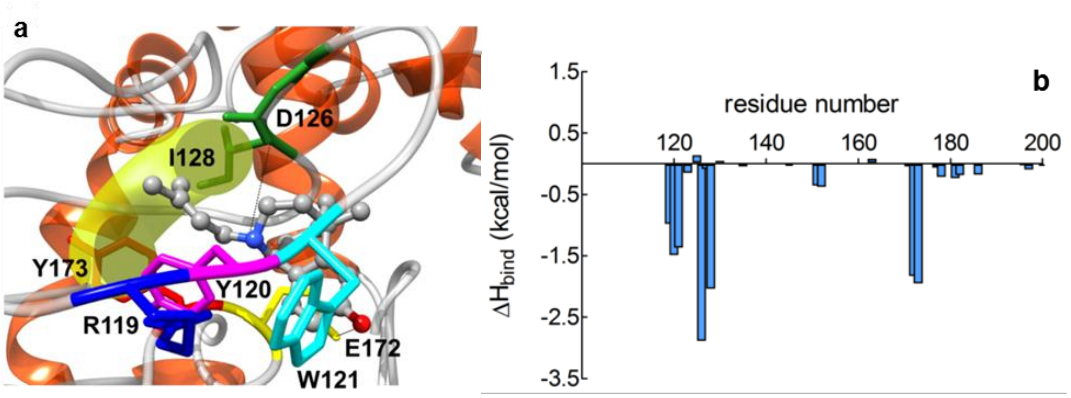


Figure 3.1.40 *Site-directed mutagenesis of the σ_1 receptor. (a) 3D structure of the σ_1 receptor embedded in a 1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine (POPC)/1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE)-cholesterol (2:2:1) membrane model. Water is evidenced via its light blue van der Waals surface, while some Cl- and Na+ ions are visible as dark blue spheres. The polar heads of POPC/POPE lipids are portrayed as forest green sticks-and-balls, their solvent accessible surface area being highlighted in transparent forest green. The lipid hydrophobic tails are shown as grey sticks. Cholesterol molecules are evidenced as light green CPK spheres, while the σ_1 receptor is depicted as a red ribbon. (b) Cross section of the system in (a) (water and ions not shown). (c) Specific [3H]-(+)-pentazocine (PTZ) binding (SB) at a concentration of 40 nM to different alanine mutants of the human σ_1 receptor ($n \geq 2$). SB is referred to the PTZ binding of the wild-type (wt) receptor (100 %, first column). **Significance level $p \leq 0.001$; *Significance level $p \leq 0.01$. Red lines: SB < 30 % indicates high influence of the particular amino acid on ligand binding; 30% ≤ SB ≤ 70% indicates moderate influence on ligand binding; SB > 70 % indicates no influence on ligand binding.*

Table 3.1.18 *In silico predicted affinity of PTZ for the wild-type σ_1 receptor. Binding free energy (ΔGbind) and its components PTZ in complex with the wt σ_1 receptor. All energies are in kcal/mol. Errors are given as the standard deviation from the mean. The calculated Ki values (nM) are also reported. The K_i^c value is obtained from the corresponding ΔGbind using the fundamental relationship ΔGbind = -RTln(1/K_i^c).*

ΔE_{VDW}	-36.31 ± 0.06
ΔE_{ELE}	-130.29 ± 0.10
ΔE_{MM}	-157.00 ± 0.12
ΔG_{PB}	141.88 ± 0.18
ΔG_{NP}	-2.84 ± 0.01
ΔG_{SOL}	139.04 ± 0.18
ΔH_{bind}	-27.56 ± 0.22
$-T\Delta S_{bind}$	17.54 ± 0.39
ΔG_{bind}	**-10.02 ± 0.45**
K_i^c	**45 nM**
K_i^e	**15 nM**



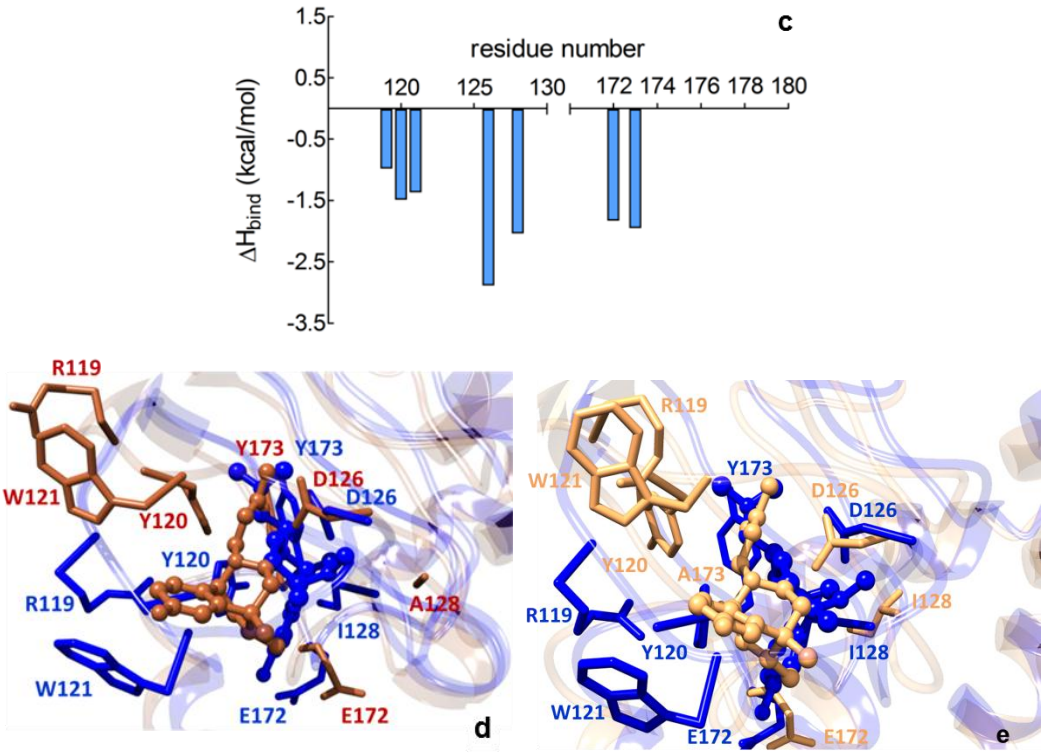


Figure 3.1.41 *In silico mutagenesis of σ_1 receptor binding site residues. (a) Equilibrated MD snapshot of the wt σ_1 receptor in complex with PTZ. The image is a zoomed view of the receptor binding site. The ligand is portrayed in colored sticks-and-balls (C, gray; N, blue; O, red). The protein residues mainly involved in the interaction with PTZ are highlighted as labeled colored sticks. Salt bridges and H-bonds interactions are shown as continuous and dotted black lines, respectively. The yellow shadow portrays the hydrophobic pocket generated by the side-chains of I128 and Y173. (b) Per residue energy decomposition for wt σ_1 receptor in complex with PTZ. Only σ_1 amino acids from position 100 to 200 are shown, as for all remaining residues the contribution to ligand binding is irrelevant. (c) Expanded view of (b) showing those σ_1 residues for which $|\Delta Hbind| \geq |0.5$ kcal/mol$|$. Comparison of the equilibrated MD snapshots of the wt σ_1 receptor (blue) with (d) I128A (sienna) and (e) Y173A (sandy brown) mutants in complex with PTZ. The ligand is portrayed in sticks-and-balls colored according to the respective protein complex. In panels a, d, and e hydrogen atoms, water molecules, ions, and counterions are omitted for clarity.*

R119, I128 and Y173 are essential residues for PTZ binding.

The results of the site-directed mutants of those residues identified in silico as crucial for PTZ binding are listed in Table 3.1.19. We see that mutating residues R119, I128, and Y173 within the σ_1 binding site into alanine leads to an almost complete loss of ligand binding, as predicted from simulation (Table 3.1.19, Figure 3.1.40 c, 3.1.41 d,e and Figure 3.1.42). Indeed, residues I128 and Y173 afford a significant contribution to PTZ/σ_1 complex (Figure 3.1.41 c), as their side-chains generate a hydrophobic pocket tailored to encase the methylbut-2-enyl moiety of

PTZ (Figure 3.1.41 d,e). Accordingly, the favorable hydrophobic interactions exerted by these two residues in binding PTZ are lost upon alanine substitution (I128: SB = 1%, $\Delta\Delta$Gbind = -2.47 kcal/mol; Y173: SB = 0.4%, $\Delta\Delta$Gbind = -2.70 kcal/mol, Figure 3.1.40 c). On the contrary, the affinity of the Y120A and W121A σ_1 mutants for PTZ was only partly reduced (SB = 62% and 76%, respectively). Interestingly, the relevant modeling analysis reveals that, although each of these two aromatic residues concur in stabilizing PTZ binding mainly via $\pi-\pi$ interactions (Figure 3.1.41 and Figure 3.1.42), the effect of mutating either of these two positions into alanine results in an apt rearrangement of the alternative residue side chain within the binding site and, hence, in a partial compensation of the receptor affinity loss ($\Delta\Delta$Gbind = -0.68 kcal/mol for Y120A and -1.54 kcal/mol for W121A, respectively, Table 3.1.19).

Table 3.1.19 *In vitro/in silico site-directed mutagenesis of σ_1 residues directly involved in PTZ binding. Specific PTZ binding of different alanine mutants of the human σ_1 receptor and in silico estimated free energy of binding (ΔGbind) of the same σ_1 alanine mutants as generated by computational mutagenesis. Errors are given in parenthesis as standard errors of the mean (SEM). [a] $\Delta\Delta$Gbind = ΔGbind,wt – ΔGbind,mutant. [b] By definition, a negative value of $\Delta\Delta$Gbind indicates a favorable contribution for the wt residue in that position and vice versa.*

σ_1 residue	**specific binding of PTZ [%]**	**ΔG_{bind} (kcal/mol)**	**[a,b] $\Delta\Delta G_{bind}$ (kcal/mol)**
wt	100	-10.02 (0.05)	-
R119A	26 (3.5)	-8.31 (0.05)	-1.71
Y120A	62 (10)	-9.34 (0.08)	-0.68
W121A	76 (12)	-8.48 (0.06)	-1.54
I128A	1.1 (0.6)	-7.55 (0.06)	-2.47
Y173A	0.4 (0.4)	-7.32 (0.11)	-2.70
D126E	54 (3)	-9.13 (0.07)	-0.89
E172D	0 (0.1)	-7.88 (0.06)	-2.14

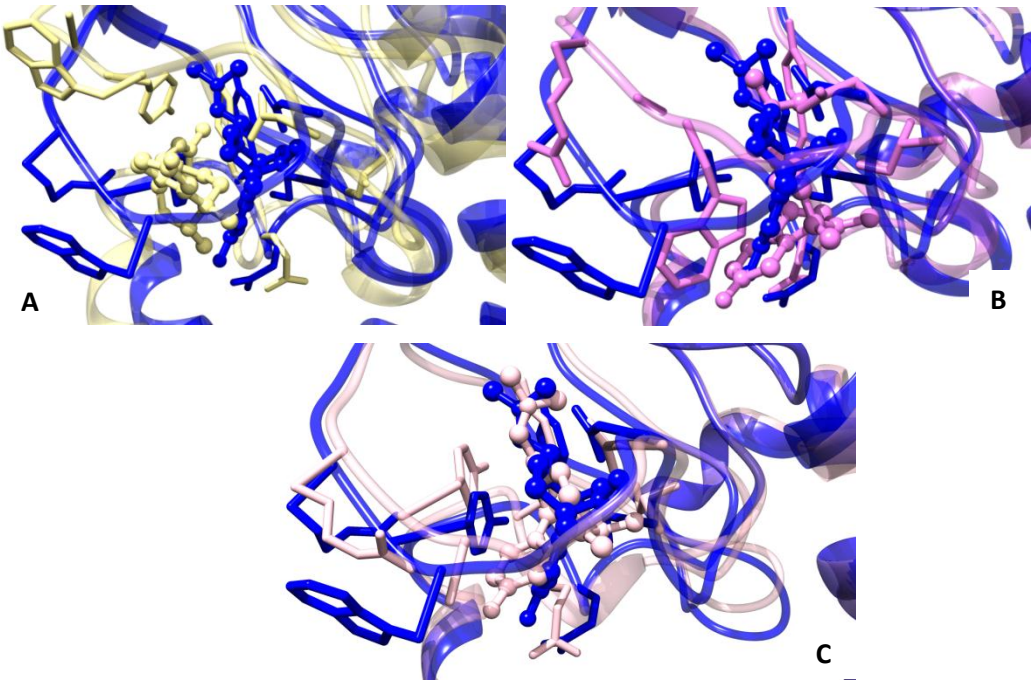


Figure 3.1.42 *In silico mutagenesis of σ_1 receptor binding site residues. Comparison of the equilibrated MD snapshots of the wt σ_1 receptor (blue) with (A) R119A (khaki), (B) Y120A (orchid) and (C) W121A (pink) mutants in complex with PTZ. The images are a zoomed view of the receptor binding site. The ligand is portrayed in sticks-and-balls colored according to the respective protein complex. In both panels hydrogen atoms, water molecules, ions, and counterions are omitted for clarity.*

D126 and E172 are strategically located within the σ_1 binding site.

The anionic amino acids D126 and E172 are known to be essential for ligand (i.e., haloperidol) binding,[5] and our combined study indeed confirms their role in binding PTZ (Figure 3.1.40 c and 3.1.41 b,c). To substantiate this prediction, however, we decided to swap these two residues by each other instead of replacing them by alanine in the corresponding site-directed mutagenesis experiments. Interestingly, the E172D σ_1 mutant was totally devoid of PTZ binding, while the D126E isoform preserved 54% of the wt affinity. These results clearly demonstrate the specific and strategic location of these two negative charges within the σ_1 binding site: while the elongation of the side chain from aspartate to glutamate at position 126 is somewhat tolerated ($\Delta\Delta$Gbind = -0.89 kcal/mol, Table 3.1.19), resulting in a mild rearrangement of the protein binding pocket which preserves the main network of interactions between the ligand and the protein, the corresponding reduction of the chain length in the E172D mutant fully abrogates the ability of the protein to bind the radioligand ($\Delta\Delta$Gbind = -2.14 kcal/mol, Table 3.1.19). In detail, the substitution of D/E at position 126 leads to a longer – and hence less stable – salt bridge between the -NH+ group ion PTZ and the –COO- group of E126 side chain (ADL = 4.85 ± 0.08 Å), leaving all other ligand/receptor

interactions almost unaltered (Fig. 3.1.43 a). Upon E172D substitution, however, the missing H-bond between position 172 and the –OH group on PTZ leads to a global readjustment of the protein binding site (Fig. 3.1.43 b) which ultimately leads to an overall decrease of the favorable contributions to ligand binding.

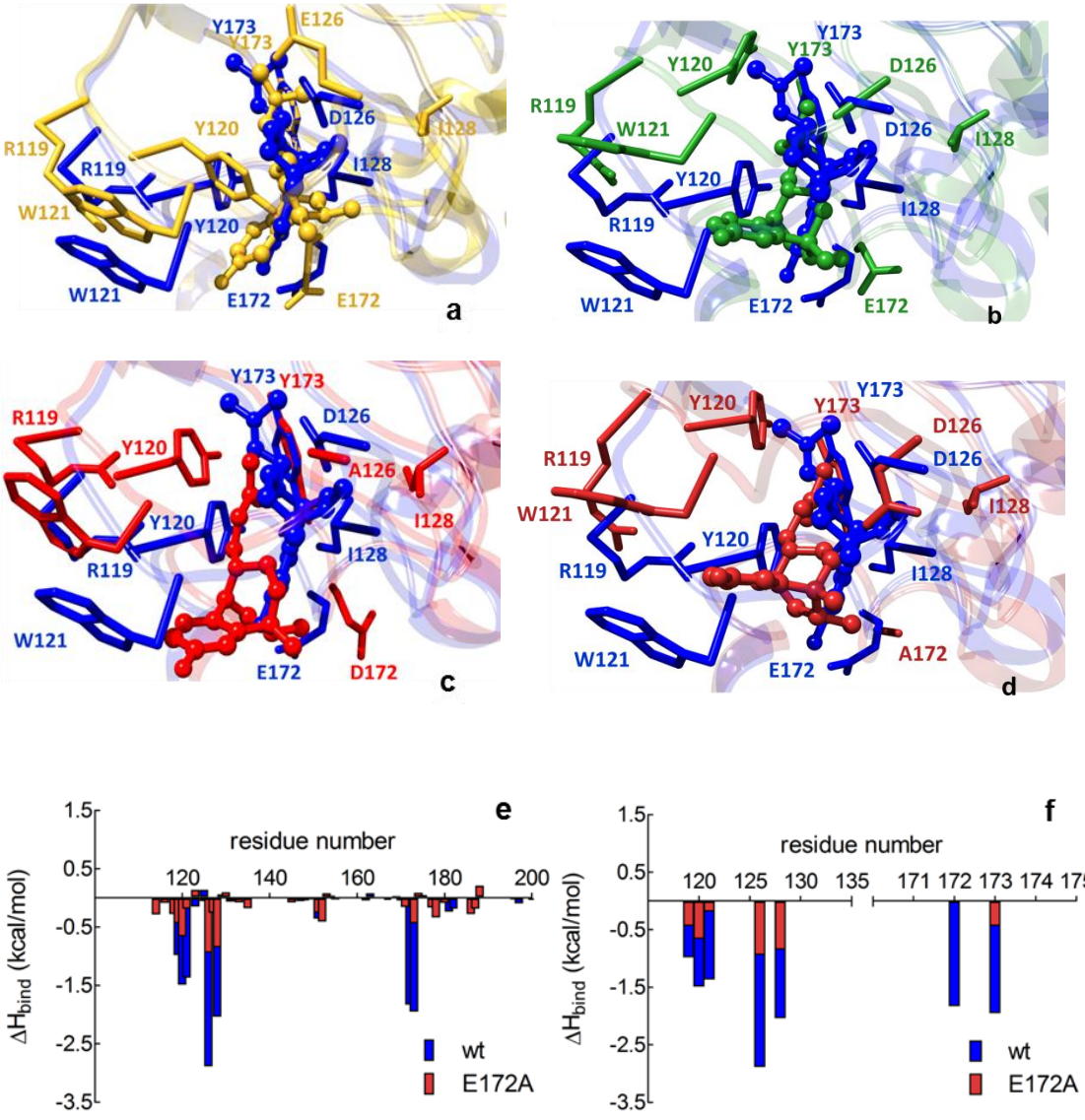


Figure 3.1.43 *In silico mutagenesis of σ_1 receptor binding site residues D126E, E172D, D126A and E172A. Comparison of the equilibrated MD snapshots of the wt σ_1 receptor (blue) with (a) D126E (goldenrod), (b) E172D (forest green), (c) D126A (red) and (d) E172A (firebrick) mutants in complex with PTZ. The images are a zoomed view of the receptor binding site. The ligand is portrayed in sticks-and-balls colored according to the respective protein complex. In panels a-c hydrogen atoms, water molecules, ions and counterions are not shown for clarity. (e) Comparison of σ_1 residue contributions to PTZ binding for the wt and the E172D mutant protein. (f) Expanded view of (e) showing those σ_1 residues for which $|\Delta Hbind| \geq |0.5$ kcal/mol$|$ in the wt receptor.*

To further corroborate this fundamental finding we also performed in silico mutagenesis of D126 and E172 into alanine. As expected, the prominent role exerted by D126 and E172 in binding PTZ is reflected in the highly unfavorable affinity predicted for the D126A and E172A mutants ($\Delta\Delta$Gbind = -3.01 kcal/mol and -2.02 kcal/mol, respectively, Table 3.1.17). The missing of the critical salt bridge and H-bond interactions in which these residues are engaged in the wt protein bound to PTZ results in a catastrophic loss of drug interaction when mutated to alanine (Figure 3.1.43 c-f). These predictions match the results previously reported by Seth et al.[5] on in vitro PTZ binding assays of the D126G and E172G σ_1 mutants; taken together, these results ultimately confirm the obligatory nature of the highly conserved D126 and E172 residues for the ligand binding function of the σ_1 receptor.

Polar residues are required in the SBDLI domain to maintain binding site geometry.

Next, we explored the effect of mutations on residues belonging to the SBDLI, and SBDLII σ_1 domains by performing a systematic substitution of the polar (S, T) and aromatic (F, Y) amino acids of these protein regions with alanine. The results are shown in Table 3.1.20. Mutating the polar and aromatic amino acids of the SBDLI (S101, Y103, and F107) into alanine both in silico and in vitro led to a considerable reduction of PTZ binding (SB = 18÷30%, $\Delta\Delta$Gbind = -2.13 ÷ -1.86 kcal/mol, Table 3.1.20, rows 1-3). A sensible explanation for these results is that all three alanine-mutated residues lie in proximity of the transmembrane domain of the σ_1 receptor (Figure 3.1.44 a) and, during the long MD simulation, they promote a substantial modification of both the entire binding site and the surrounding protein regions.

Table 3.1.20 *In vitro/in silico site-directed mutagenesis of σ_1 residues belonging to the SBDLI and SBDLII motifs. Specific PTZ binding of different alanine mutants of the human σ_1 receptor and in silico estimated free energy of binding (ΔGbind) of the same σ_1 alanine mutants as generated by computational mutagenesis. Errors are given in parenthesis as standard errors of the mean (SEM).* [a] *$\Delta\Delta$Gbind = ΔGbind,wt – ΔGbind,mutant.*

σ_1 residue	domain	specific binding of PTZ [%]	ΔG_{bind} (kcal/mol)	[a] $\Delta\Delta G_{bind}$ (kcal/mol)
S101A	SBDL I	30 (27)	-8.16 (0.08)	-1.86
Y103A		22 (12)	-7.89 (0.09)	-2.13
F107A		18 (11)	-8.05 (0.05)	-1.97
T181A	SBDL II	93 (7.5)	-9.99 (0.11)	-0.03
F184A		124 (18)	-10.55 (0.07)	0.53
T189A		99 (1.1)	-9.87 (0.07)	-0.15
F191A		88 (14)	-9.13 (0.10)	-0.89
201Stop		1.0 (0.6)	-	-
209Stop		0.9 (0.5)	-	-
217Stop		37 (7.7)	-8.80 (0.06)	-1.22

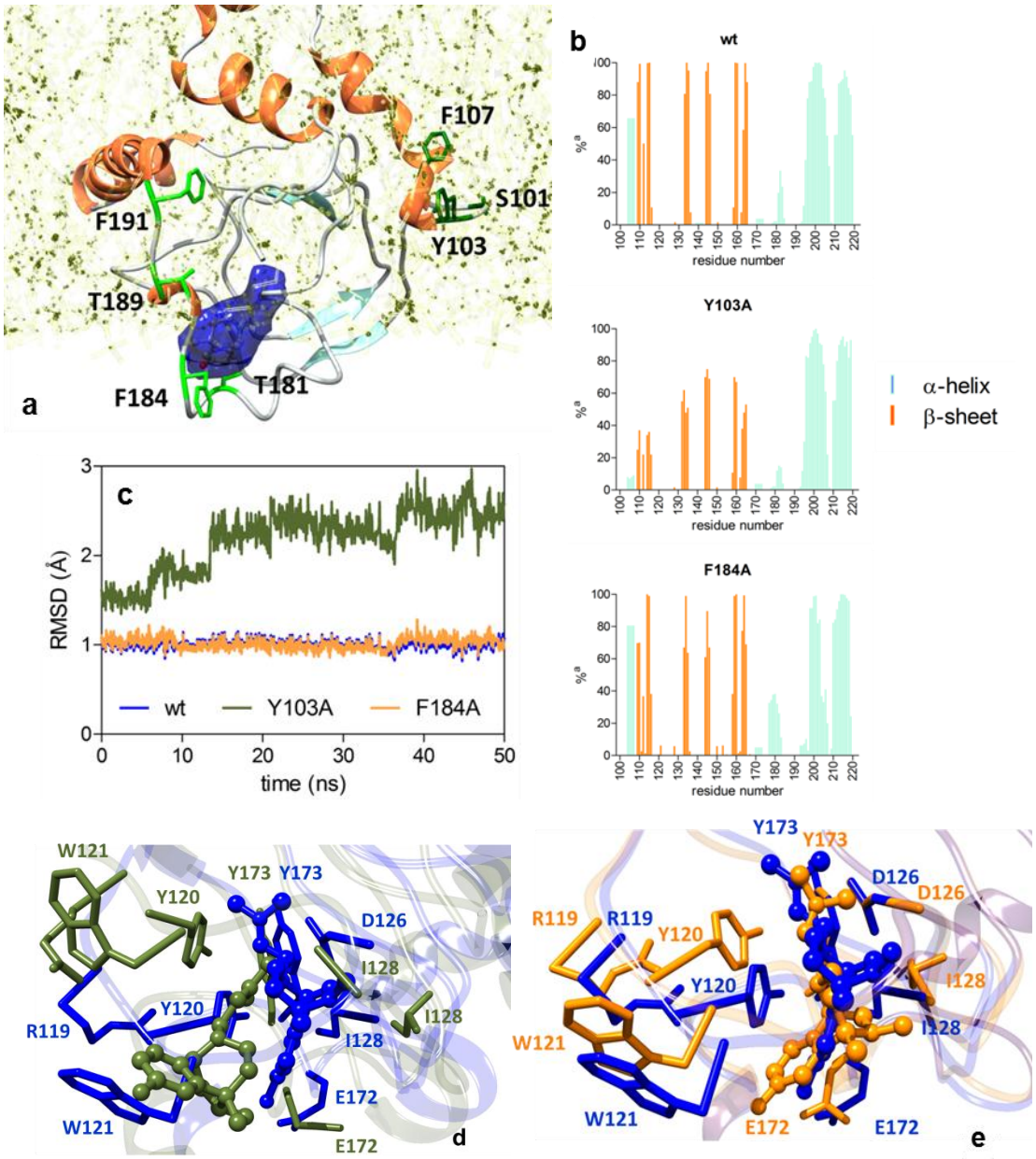


Figure 3.1.44 *In silico mutagenesis of σ_1 receptor residues at the SBDLI and SBDLII domains. (a) Zoomed view of the residues belonging to the σ_1 SBDLI and II domains, represented as forest and light green sticks, respectively. The ligand is in element-colored sticks-and-balls (C, gray; N, blue; O, red) with its van der Waals surface shown in blue. The membrane phospholipids are portrayed as semi-transparent light yellow balls-and-stick . (b) General secondary structure description of the σ_1 receptor (aa 100-223) during MD simulations of the wt, Y103A, and F184A PTZ complexes. [a]Percentage of specific secondary structure for each σ_1 residue. (c) RMSD of the coordinates of the heavy atoms of PTZ along the MD simulation compared with those of the initial structure. Comparison of the equilibrated MD snapshots of the wt σ_1 receptor (blue) with (d) Y103A (dark olive green) and (e) F184A (orange) mutants in complex with PTZ. The images are a zoomed view of the receptor binding site. The ligand is portrayed in sticks-and-balls colored according to the respective protein complex. In panels d and e hydrogen atoms, water molecules, ions and counterions are not shown for clarity.*

Taking the PTZ/σ_1 (Y103A) complex as a proof-of-concept, we observed a different rearrangement of the secondary structure of this protein portion with respect to the PTZ/wt assembly (Figure 3.1.44 b, top and central panels), ultimately resulting in an expansion of the receptor binding pocket from an average dynamic volume (ADV) of 509 ± 5 Å3 to 573 ± 7 Å3. Accordingly, the ligand binds the receptor more loosely, as its interactions with the protein residues lining the binding site become weaker (Figure 3.1.44c). This effect is clearly seen in the behavior of the root mean square deviation (RMSD) of the heavy atom of PTZ during the MD simulation (Figure 3.1.44d): after equilibration, the ligand assumes a stable conformation in the wt binding pocket with an average RMSD values of 1.01 ± 0.06 Å, while larger RMDS fluctuations (RMSD = 2.18 ± 0.35 Å) and longer time-to-equilibrium are observed for the Y103A mutated σ_1/PTZ complex.

Residues in the SBDLII domain scarcely contribute to ligand binding.

In stark contrast, analogous mutations in the SBDLII domain barely influenced PTZ binding if at all (SB = 93÷100%, $\Delta\Delta$Gbind = -0.59 ÷ -0.03 kcal/mol, Table 3.1.20). A comparison of the membrane-bound wt and SDBLII mutated σ_1 isoforms reveals that the presence of mutations at the SBDLII domain does not lead to substantial alteration of both the membrane and the protein binding site/overall structure. Indeed, in the representative example of the PTZ/σ_1 (F184A) complex, the fingerprint of the protein secondary structure is conserved (Figure 3.1.44b, top and bottom panels), and the average dynamic volume is essentially unchanged (507 ± 7 $Å^3$). Moreover, PTZ maintains a very stable conformation in the binding site during the entire MD simulation, as substantiated by the relevant RMSD deviation (1.01 ± 0.07 Å, Figure 3.1.44c). Lastly, all ligand/receptor interactions detected for the wt complex are maintained in this as well as in all other SBDLII mutated complexes examined with no significant differences in binding mode and strength (Figure 3.1.44e). Remarkably, in a previous study the SBDLII was postulated to be part of the σ_1 ligand binding site[7]. Our results clearly do not support this hypothesis. It has been further proposed that the SBDLII is responsible for anchoring the σ_1 receptors to the membrane and, in so doing, stabilizing the 3D structure of the protein. Once again, our combined in vitro/in silico experiments point in the opposite direction. Taken together, our findings lead to the conclusion that, while the SBDLI domain is part of the binding site of the σ_1 receptor, and as such mutations at this domain lead to a drastic decrease of receptor/ligand affinity, the SBDLII domain does not belong to the σ_1 ligand binding site. Accordingly, mutations in this protein domain exert only a marginal effect on ligand binding, if at all.

Even small deletions in the σ_1 C-terminal end abrogate ligand binding.

As briefly mentioned above, σ_1 ligand binding is abrogated when more than 15 amino acids are removed from the C-terminal end of the protein.[65] Thus, we further investigated this aspect by deleting 7, 15 and 23 amino acids, respectively, from the C-terminal end of the σ_1 receptor and determined the affinity of the truncated receptors for PTZ. As expected, elimination of 15 and 23 amino acids resulted in the loss of PTZ binding ability (Figure 3.1.40c, Table 3.1.20). More interestingly, also the removal of only 7 residues from the receptor C-terminal end led to a considerable reduction of PTZ binding (SB = 37%). Again, these experiments were rationalized by the corresponding in silico assays. While the major deletions resulted in a partially unfolded structure of the receptor missing a large portion of the ligand binding site, the affinity of the 7-residue truncated protein still preserved a low affinity for PTZ (Figure 3.1.45a), as quantified by the value of ΔGbind = -8.80 kcal/mol. The corresponding ΔΔGbind = -1.22 kcal/mol nicely correlates with the 63% affinity reduction with respect to the wt protein reported by in vitro mutagenesis. The 7 deleted amino acids are not directly involved in ligand binding; however, the missing sequence YLFGQDP results in a structural modification of the receptor that, like a domino effect, propagates along the protein backbone to the binding site (Figure 3.1.45b). This overall configuration rearrangement directly affects three σ_1 residues most important for ligand binding: R119, D126, and E172. Specifically, the interaction of D126 with PTZ becomes less favorable by 1.54 kcal/mol with respect to the wt isoform, while E172 decreases its contribution by -1.68 kcal/mol.

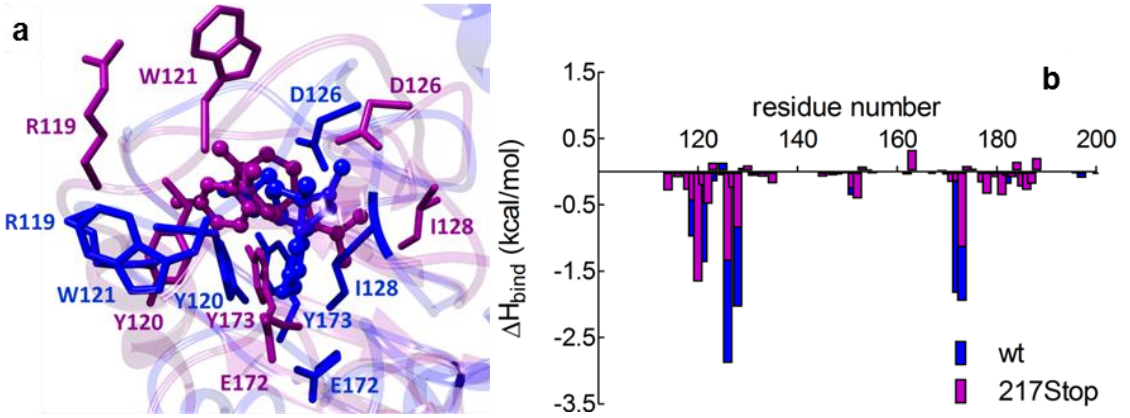


Figure 3.1.45 *In silico mutagenesis of σ_1 receptor residues: deletion of the 7-residue sequence YLFGQDP from the protein C-terminal domain. (a) Comparison of the equilibrated MD snapshots the wt (blue) and C-terminal 7 residue truncated σ_1 receptor (magenta) in complex with PTZ. The image is a zoomed view of the receptor binding site. The ligand is portrayed in sticks-and-balls colored according to the respective protein complex. Hydrogen atoms, water molecules, ions and counterions are not shown for clarity. (b) Comparison of σ_1 residue contributions to PTZ binding for the wt and the C-terminal 7-residue truncated protein.*

Distal residues also concur in shaping the σ_1 binding site.

Lastly, to determine whether other σ_1 receptor residues could play a critical role in binding PTZ, several alternative positions between the SBDLI and II protein domains were substituted by alanine. In silico mutagenesis results suggested that the hydrophobic V145 and the aromatic F146 and Y147 residues, when mutated to alanine, result in minor (if any) changes in the protein binding site conformation (Table 3.1.21). These results are confirmed by the corresponding in vitro experiments, showing that for these 3 residues the SB values range between 57% and 121%. However, the alanine mutants of the basic amino acid R175 and of the two polar residues of S125 and T127 show a moderate (S125A, T127A) to strong (R175A) influence on PTZ binding, as testified by the drastically less favorable $\Delta\Delta$Gbind values of -1.45 kcal/mol, -1.24 kcal/mol, and -2.02 kcal/mol, respectively (Table 3.1.21) and supported by the corresponding SB values (49%, 45%, and 10%). The two mutations S145A and T127A transform the environment in proximity of the negatively charged D126 from polar to hydrophobic which, in turn, decreases the strength of its salt bridge with PTZ (ADL = 4.76 ± 0.07 Å and 4.69 ± 0.10 Å for S145A and T127A, respectively).

Table 3.1.21 *In vitro/in silico site-directed mutagenesis of σ_1 residues belonging to the region between the protein SBDLI and SBDLII motifs. Specific PTZ binding of different alanine mutants of the human σ_1 receptor and in silico estimated free energy of binding (ΔGbind) of the same σ_1 alanine mutants as generated by computational mutagenesis. Errors are given in parenthesis as standard errors of the mean (SEM).* [a] *$\Delta\Delta$Gbind = ΔGbind,wt – ΔGbind,mutant.*

σ_1 residues	specific binding of PTZ [%]	ΔG_{bind} (kcal/mol)	[a] $\Delta\Delta G_{bind}$ (kcal/mol)
wt	100	-10.02 (0.05)	-
S125A	49 (1.5)	-8.57 (0.09)	-1.45
T127A	45 (2.3)	-8.78 (0.11)	-1.24
V145A	78 (14)	-9.89 (0.07)	-0.13
F146A	57 (0.3)	-9.42 (0.09)	-0.60
Y147A	121 (4.1)	-8.80 (0.06)	-1.22
R175A	10 (4.1)	-8.00 (0.10)	-2.02

The case of the R175A mutant is more complex, as experiments detect a drastically reduced affinity of this mutant σ_1 isoform for PTZ (SB = 10%). The calculated free energy of binding differs considerably from the wt σ_1 complex ($\Delta\Delta$Gbind=-2.02 kcal/mol) although this residue is not directly involved in PTZ binding. Importantly, however, along the entire MD course R175 forms a stable, bifurcated H-bond with Y120 and R114. These residues, in turn, stabilize the conformation of the Y173 for productive binding via another direct H-bond (Figure 3.1.46a). All these interactions translate into a -2.72 kcal/mol global favorable enthalpic contribution of

these 4 amino acids to ligand binding, as detailed in Table 3.1.22. As discussed above, both tyrosines Y120 and Y173 play a major role not only in direct PTZ binding but also in shaping the entire σ_1 binding site. Thus, replacing R175 by alanine results in an almost complete loss of the H-bond interaction network (Figure 3.1.46b): the aromatic interaction between Y120 and PTZ is lost, the entire binding site is enlarged, and the corresponding complex stabilization energy drops dramatically (-0.89 kcal/mol, Table 3.1.22).

Table 3.1.22 *Per-residue free-energy decomposition for wt and R175A mutant σ_1 receptor in complex with PTZ. All energy values are in kcal/mol. Data standard deviations are in the range ± 0.01 ÷ ± 0.10.*

wt	Y120	Y173	R175
R114	0.02	-0.91	-0.88
Y120	*	-0.26	-0.71
Y173	*	*	0.02
R175	*	*	*
$\sum\Delta H_{bind}$			**-2.72**
R175A	Y120	Y173	A175
R114	-0.01	-0.73	0.02
Y120	*	-0.03	-0.12
Y173	*	*	-0.02
A175	*	*	*
$\sum\Delta H_{bind}$			**-0.89**

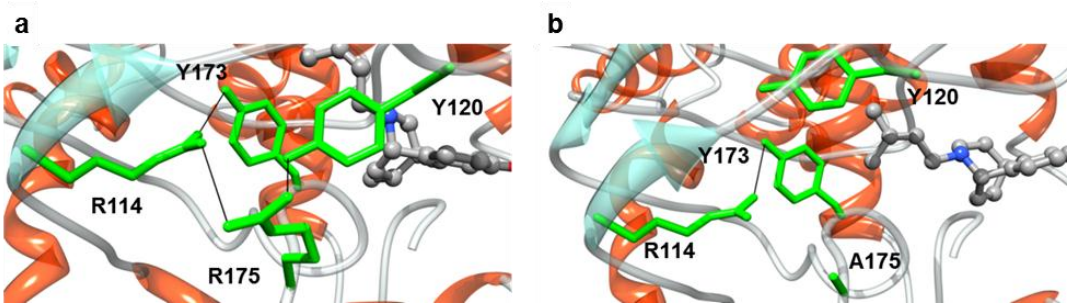


Figure 3.1.46. *In silico mutagenesis of σ_1 receptor residues: deletion of the 7-residue sequence YLFGQDP from the protein C-terminal domain. Details of the residues involved in a structurally stabilizing interaction of the (a) wt and (b) R175A mutant σ_1 receptor in complex with PTZ as obtained from equilibrated MD simulation snapshots. The protein backbone is portrayed as a transparent ribbon colored by secondary type (orange, α-helices; cyan, β-sheets; light gray, coils). The main residues involved in the interactions are shown as labeled green sticks. PTZ is depicted in sticks-and-balls and colored by element (C, gray; O, red; N, blue). In all panels, hydrogen atoms, water molecules, ions and counterions are omitted for clarity.*

Conclusion

The combined in vitro/in silico mutagenesis study reported in the present work confirmed some previous knowledge about the structural features of the σ_1 receptor and its binding site but also, and perhaps more importantly, qualified and quantified the role of several receptor

residues which play a prominent role in receptor-ligand binding. Specifically, among those residues belonging to the putative receptor binding pocket, replacing I128 and Y173 with alanine almost abrogates PTZ binding, consistently with a drastic reduction of stabilizing, hydrophobic interactions. The specific nature of the anionic residue E172 is critical for PTZ binding in that not only its replacement with alanine but also an exchange with a residue of similar nature but smaller side chain (E172D) results in the total failure of receptor ligand binding. The role of σ_1 amino acids belonging to the SBDLI and II domains has been differentiated and rationalized: while a critical role in maintaining the protein secondary structure of this protein portion with respect to the cellular membrane has been verified for those residues belonging to the first domain, mutagenesis performed on residues in the SDBLII region did not affect the receptor affinity for PTZ. The removal of small sequences in the C-terminal part of the protein (e.g., 7 residues) resulted in a substantial reduction of ligand binding activity, as these deletions result in a general reconfiguration of the receptor binding pocket, ultimately involving those residue which constitute the PTZ main anchor points. Finally, this combined approach unveiled the substantial role exerted by other σ_1 residues in ligand binding; for instance, the series of polar residues S125, T127, and R175 all concur in maintaining the high affinity of σ_1 receptor for PTZ, and their replacement with the small, apolar alanine result in a neat decrease of receptor ligand binding activity.

3.2 SELF-ASSEMBLY OF AMPHIPHILIC DENDRIMERS FOR FUNCTIONAL siRNA DELIVERY

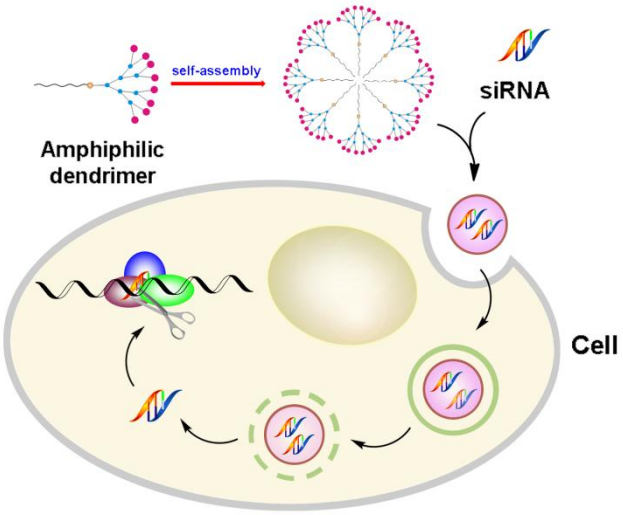
This work was carried out in collaboration with the group of Dr. Ling Peng, CINaM, Marseille (France)

Introduction

In this work we report the self-assembly of a series of amphiphilic dendrimers into nanosized supramolecular structures and their resulting capacity to effectively deliver siRNA.

Self-assembly is a fundamental concept in modern molecular science. It relies on non-covalent interactions to assemble molecular building blocks into supramolecular entities in a reversible, controllable, and specific manner, yet with relatively little synthetic input.[66,67] Of particular value is the ability of self-assembled nanosized superstructures to behave as more than the sum of their individual parts, and exhibit completely new properties. As an example, we present here our investigation on the self-assembly of amphiphilic dendrimers into nanostructures and the underlying mechanism for their delivery of siRNA therapeutics. Our

findings reveal that supramolecular nanostructures generated via the self-assembly of amphiphilic dendrons hold a great potential for future applications in siRNA delivery and gene silencing.



Over the past few years, our group has been actively engaged in developing dendrimers as nanovectors to deliver RNA therapeutics,[68,69] especially synthetic small interfering RNAs (siRNA), as a novel therapeutic strategy for the specific inhibition and knockdown of disease genes by RNA interference (RNAi).[70] The main challenge facing RNAi therapeutics is the safe and efficient delivery of the siRNA, the molecules that mediate RNAi.[71] An ideal delivery vector is one that assembles and compacts the siRNA within a complex, protects it from degradation, allows it to enter cells, and finally releases it from the endolysosomal compartment into the cytoplasm, where the RNAi machinery is located.[71] The most common non-viral vectors for siRNA delivery can be classified as lipids and polymers,[72-74] with cationic dendrimers (a special family of polymers) emerging as promising vectors by virtue of their well-defined structural architecture, intriguing features of multivalency, and high cargo payload within a nanosized volume.

Chemical synthesis

We have recently constructed a small amphiphilic dendrimer bearing a hydrophobic alkyl chain of 18 carbon atoms and a hydrophilic dendritic portion (**4** in Figure 3.2.1), which proved to be efficient for siRNA delivery by combining the advantageous features of lipid and dendrimer vectors.[68] Preliminary results revealed the importance of the hydrophobic alkyl chain in the amphiphilic dendrimers, as the dendrimer without the hydrophobic chain was devoid of delivery activity.[68] We surmised that the hydrophobic alkyl chain promoted self-assembly of the amphiphilic dendrimers to generate supramolecular structures resembling the structural definition of covalent dendrimers of high generations. In order to confirm this hypothesis and

study the impact of self-assembly of the amphiphilic dendrimer vectors on their resulting siRNA delivery capacity, we designed a series of dendrimers featuring hydrophobic tails of different lengths (**1-6** in Figure 3.2.1) or a hydrophilic pentaethylene glycol (PEG) chain (**7** in Figure 3.2.1). The hydrophobic portions are thus the alkyl chains of 12 to 22 carbon atoms in length, and the PEG tail is similar in length to the hydrophobic chain in **4** but is hydrophilic in nature, thus obviating dendrimer amphiphilicity. Dendron **8** is a small hydrophilic poly(amidoamine) dendrimer without any tail and served as a reference control. We systematically investigated these dendrimers for their self-assembly, their siRNA binding and delivery capacity using both experimental approaches and computational modeling. Our results demonstrate that the ability of these amphiphilic dendrimers to bind and deliver siRNA critically depended on their self-assembly which, in turn, was strongly controlled by the length of the hydrophobic alkyl chain. Consequently, the choice of an appropriate hydrophobic chain appears to be a key issue for successful siRNA delivery.

All dendrimers studied in this work (Figure 3.2.1) were synthesized via "click" chemistry using our protocol,[68] and fully characterized.

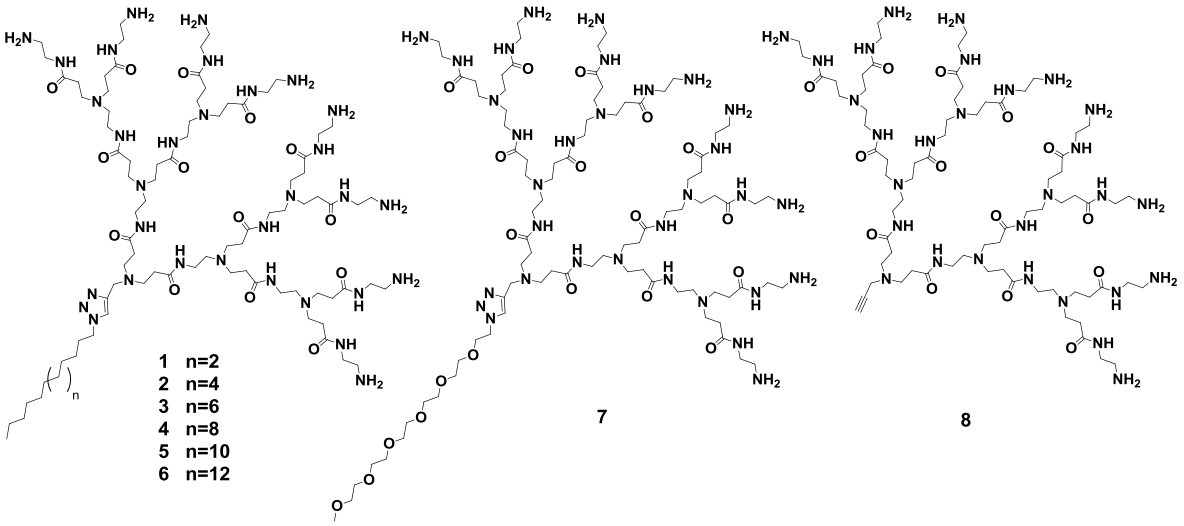
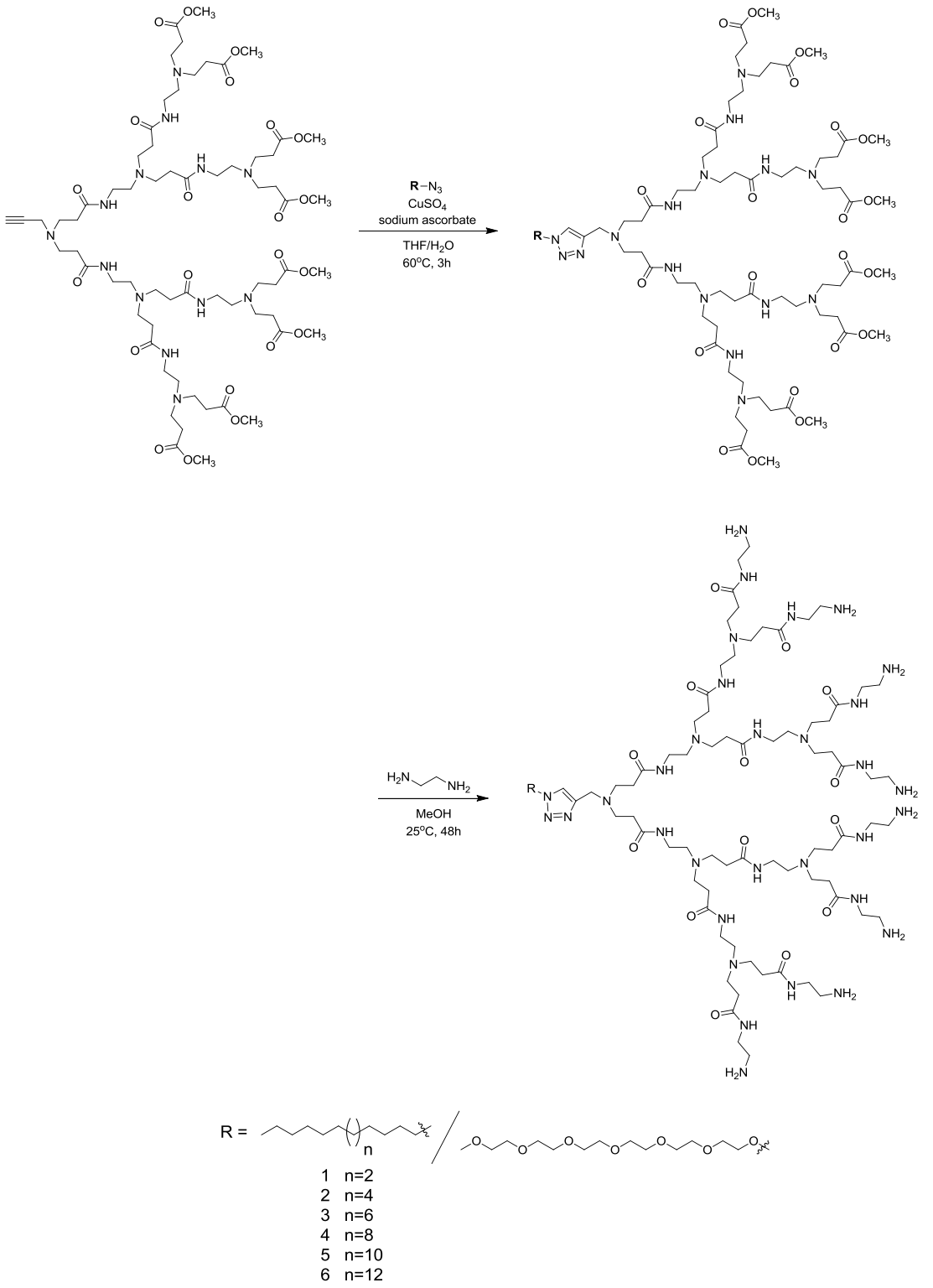


Figure 3.2.1 *Dendrimers investigated in this study: the amphiphilic dendrimers **1-6**, the dendrimer **7** bearing a hydrophilic pentaethylene glycol (PEG) chain, and dendron **8** without any chain.*



Scheme 3.2.1 *General procedure for the synthesis of dendrimer **1-7**.*

The self-assembly of these different dendrimers was first investigated using dynamic light scattering (DLS). We detected the predominant formation of nanoparticles around 7.0 nm in size in solutions containing dendrimers **3-6** respectively (Table 3.2.1), while no stable

nanostructures were observed with dendrimers **1** or **2**, the hydrophilic PEG-tail dendrimer **7** or the no-tail dendron **8**. Further results of transmission electron microscopic (TEM) imaging clearly revealed the formation of spherical nanoparticles with a diameter of about 7 nm (Figure 3.2.2), confirming the apparent self-assembly of dendrimers **3-6** indicated by our DLS data. The size of these nanostructures was in the range of typical micelles, thus acknowledging the tendency of these compounds to self-assemble into micellar supramolecular nanostructures. An increase in the alkyl chain length from C16 to C22 (**3** to **6**), was accompanied by an increase in the size of the micelles (D_{exp}) from 6.6 to 7.8 nm (obtained using DLS), consistent with the values (D_{cal}) predicted by related molecular simulations (Table 3.2.1 and Table 3.2.2). Additionally, the increasing alkyl chain length (**1** to **6**) was accompanied by a fall in the values of critical micelle concentration (CMC_{exp}) from 398 to 8.6 μM, indicating an inverse relationship between hydrophobic chain length and CMC_{exp}, i.e., the more effective packing of the longer hydrophobic chains results in better self-assembly. The relatively large values of CMC_{exp} for **1** and **2** explain the absence of micelle formation observed during DLS analysis, whereas for dendrimer **7** and dendron **8**, which are devoid of amphiphilicity, no CMC_{exp} (and, hence, no self-assembled micelles) could be determined.

Table 3.2.1 *The experimental values for micelle diameter (D_{exp}) and critical micelle concentration (CMC_{exp}) for the amphiphilic dendrimers **1-6** together with the calculated values for the packing parameter (P), free energy of micellisation (ΔG_{mic}), critical micelle concentration (CMC_{cal}) and micelle diameter (D_{cal})*

	D_{exp} (nm)	CMC_{exp} (μM)	P	ΔG_{mic} (kJ/mol)	CMC_{cal} (μM)	D_{cal} (nm)
1	-*	398±22	0.13	-38.7	415	5.6±0.1
2	-*	116±11	0.12	-44.4	127	5.9±0.1
3	6.6±0.1	49.9±3.3	0.11	-50.5	37.2	6.2±0.1
4	6.8±0.1	15.6±1.1	0.10	-56.4	11.4	6.4±0.1
5	7.3±0.1	11.2±0.6	0.10	-57.5	9.13	7.1±0.2
6	7.8±0.2	8.24±3.8	0.10	-57.8	8.50	8.1±0.2

Table 3.2.2 *Aggregation number N_{agg}, surface charge density σm and micellar diameter D_{calc} for the self-assemblies of dendrimers 1-6 as obtained from molecular simulations..*

Compound	N_{agg} (-)	σ_m (e/nm^2)	D_{calc} (nm)
1	5	0.41	5.6 ± 0.1
2	6	0.44	5.9 ± 0.1
3	7	0.46	6.2 ± 0.1
4	8	0.50	6.4 ± 0.1
5	10	0.50	7.1 ± 0.2
6	13	0.50	8.1 ± 0.2

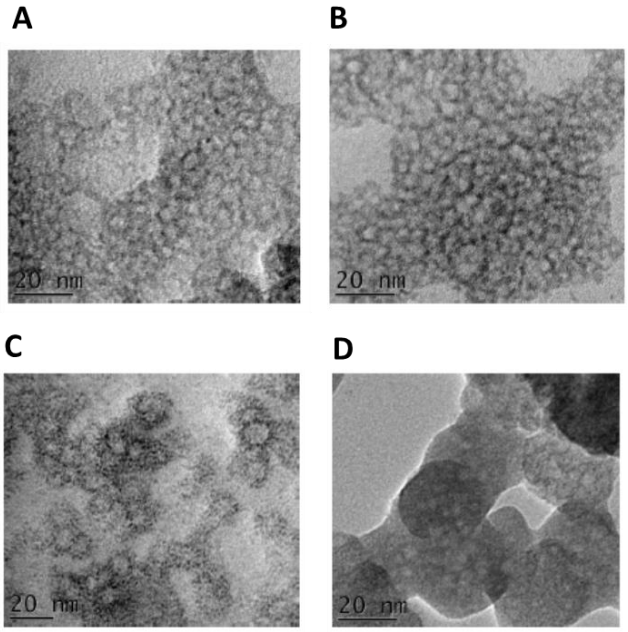


Figure 3.2.2 *TEM images of the self-assembled supramolecular micellar structures formed with **3** (A), **4** (B), **5** (C) and **6** (D), respectively, in phosphate buffer, at pH 7.4.*

Molecular modeling

We went on to inspect the self-assembly of these dendrimers using computational methods. For **1-6**, the estimated value of their critical packing parameter P, an effective ratio of the relative sizes of the polar and apolar domains of an amphiphilic compound, was well below 0.33 (Table 3.2.1), implying their tendency to self-assemble into spherical micelles.[75,76] This was confirmed by corresponding TEM imaging (Figure 3.2.2) and mesoscopic molecular simulation [77] (Figure 3.2.3 A-B). In addition, the associated free energy of micellization values ΔG_{mic} were largely negative, indicating that micellization was a spontaneous and highly favorable process for all the amphiphilic molecules **1-6**. The ΔG_{mic} did however decrease substantially (i.e., become less negative and, hence, less favorable) with decreasing chain length from **6** to **1** (Table 3.2.1). Accordingly, the predicted CMC values (CMCcal) follow the expected increasing trend as the hydrophobic character of the dendrimers decreases, in agreement with the experimental data. Since the hydrophilic dendron architecture was the same in all dendrimers, the main differential contribution to ΔG_{mic} - and hence to CMC_{cal} – must stem from the size of the hydrophobic component (see also SI for more details). Altogether, this combined experimental/theoretical approach allowed us to appreciate and understand the impact of the hydrophobic alkyl chain on the self-assembly of the amphiphilic dendrimers.

We next assessed the ability of these dendrimers to complex with siRNA molecules using fluorescent ethidium bromide (EB) displacement assay. The no-tail dendrimer **8**, the PEG chain dendrimer **7**, and dendrimer **1** bearing the shorter alkyl chain were all unable to effectively displace EB from siRNA (Figure 3.2.4A), presumably due to their inability to self-assemble under the adopted assay conditions. For the dendrimer series **2-6**, the C50 value decreased with increasing alkyl chain length, with dendrimer 6 being the most efficient at binding siRNA (Figure 3.2.4A and Table 3.2.2). Notably, these data were consistent with the results of siRNA binding obtained using gel shift assessment (Figure 3.2.4B), and confirmed the correlation between alkyl chain length and binding strength: the dendrimer bearing the longest alkyl chain was the most effective siRNA binder. In other words, these results concur to indicate that dendrimers endowed with self-assembly properties are the most effective at achieving siRNA binding.

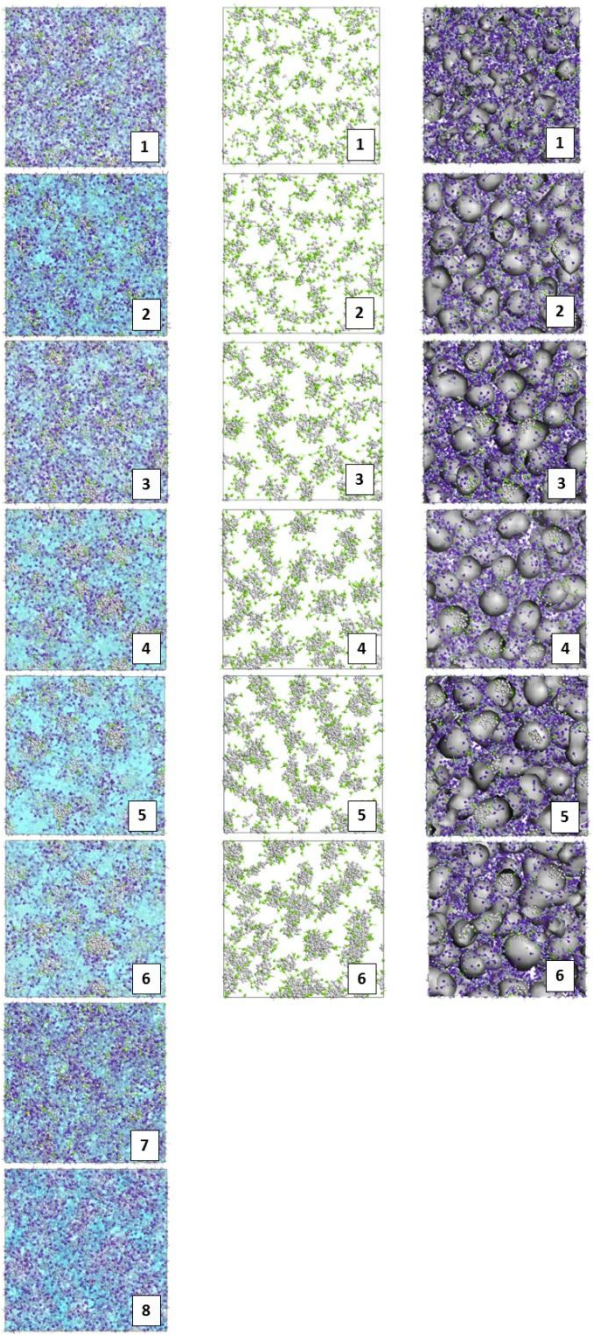


Figure 3.2.3 *Mesoscopic modeling of dendrimers **1-8** in panel A. Panel B shows only the micellar cores and panel C a zoomed view of the micelles formed by dendrimers **1-6**. Color legend: purple and lilac, hydrophilic dendron beads; light grey, hydrophobic tail beads; green, linker beads. Water and counterions are portrayed as a light blue field (panel A) or as a white field (panels B and C) for clarity.*

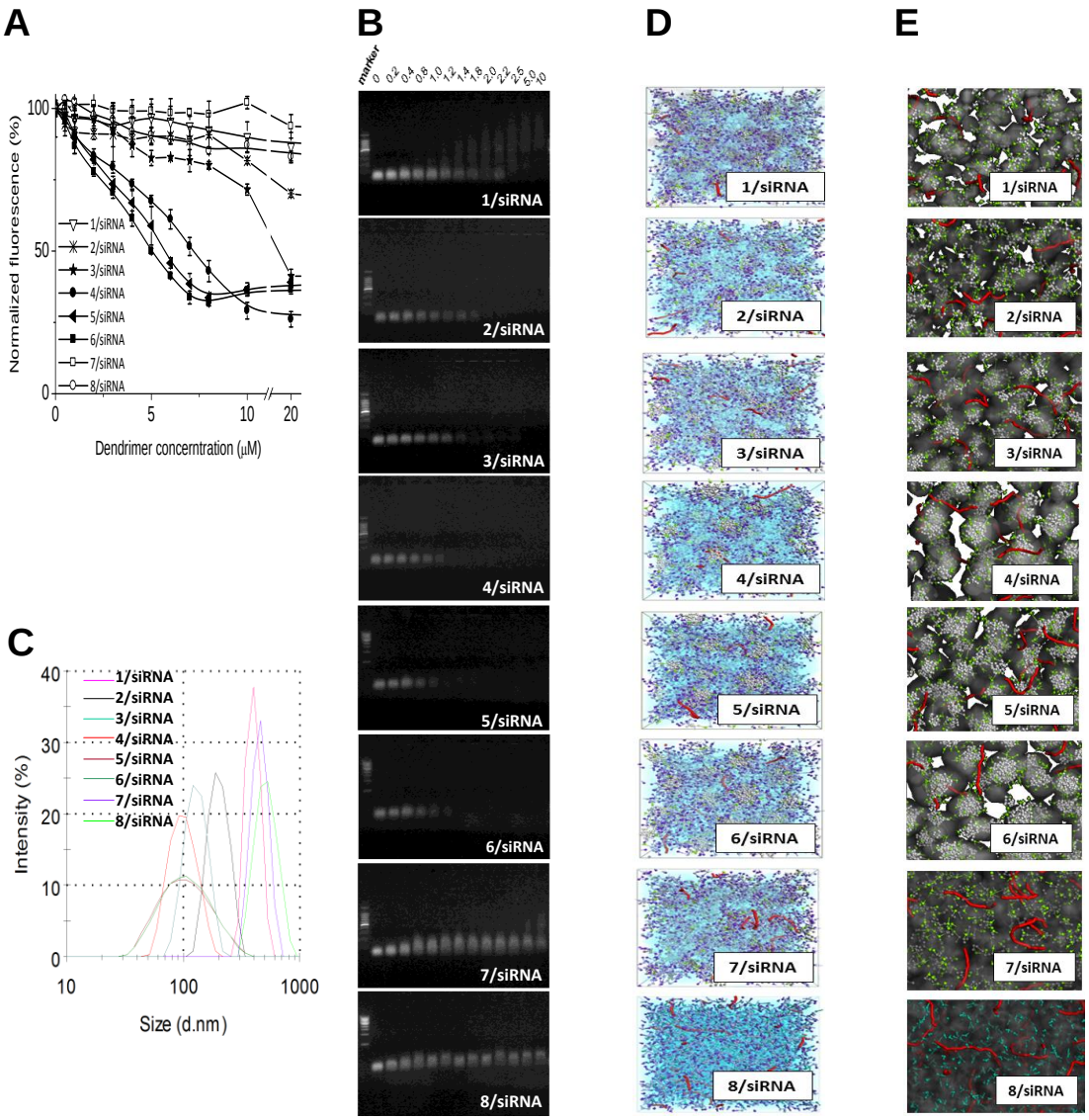


Figure 3.2.4 *siRNA binding revealed by (A) ethidium bromide displacement assay; (B) gel retardation of 200 ng siRNA in agarose gel with dendrimer at N/P charge ratios ranging from 0 to 10; (C) size and size distribution of siRNA/dendrimer complexes obtained from dynamic light scattering (DLS) analysis and (D) mesoscale modeling at N/P ratio 10 alongside (E) zoomed view of the siRNA/dendrimer system by dendrimers **1-8** respectively. Color legend for panels: siRNA molecules, thick red sticks.*

This concept was further established by DLS investigation of the siRNA/dendrimer complexes (Figure 3.2.4C). As shown in Table 3.2.3, dendrimers **3-6** bearing C16, C18, C20, and C22 alkyl chains, respectively, were able to form siRNA nanoparticles around 100 nm in size. Nevertheless, the nanoparticles formed with 5 and 6 were less uniform as indicated by the size distribution (Figure 3.2.4C) and the values of polydispersity index (PDI) (Table 3.2.3). On the other hand, dendrimers **1**, **2**, **7**, and **8** formed much larger sized aggregates with siRNA, implying an ineffective interaction with the nucleic acid fragment and, hence, its inefficient

compaction into complexes, the resulting dimensions of which are far too large for effective delivery.

Table 3.2.3 *siRNA/dendrimer binding data (C50) obtained from experiments of ethidium bromide displacement, size (D) and polydispersity index (PDI) of siRNA/dendrimer complexes obtained from dynamic light scattering analysis together with free energy of binding (ΔGbind), free energy of effective binding ($\Delta G_{bind,eff}$), number of effective charges (Neff) and effective-charge-normalized free energy of binding ($\Delta G_{bind,eff}/N_{eff}$) for siRNA/dendrimer complexes as derived from molecular dynamics simulations.* ⁎ *not available*

Complex	C_{50} (μM)	D (nm)	PDI	ΔG_{bind} (kcal/mol)	$\Delta G_{bind,eff}$ (kcal/mol)	N_{eff} (-)	$\Delta G_{bind,eff}/N_{eff}$ (kcal/mol)
1/siRNA	-⁎	474±12	-⁎	-8.23±0.91	-2.47±0.32	13±1	-0.19±0.03
2/siRNA	42.0±2.0	198±0.3	0.03	-13.4±1.1	-4.30±0.45	18±1	-0.24±0.03
3/siRNA	16.9±0.5	126±0.1	0.02	-17.1±1.3	-6.00±0.36	22±1	-0.27±0.02
4/siRNA	7.0±0.3	95.5±0.1	0.06	-22.6±1.2	-12.7±0.41	31±3	-0.41±0.04
5/siRNA	5.3±0.4	92.8±0.1	0.18	-46.6±1.6	-28.9±0.70	49±3	-0.59±0.04
6/siRNA	4.8±0.1	92.3±0.1	0.20	-57.8±2.3	-34.7±0.62	62±5	-0.56±0.05
7/siRNA	-⁎	501±1.0	-⁎	-	-	-	-
8/siRNA	-⁎	486±6.0	-⁎	-	-	-	-

Additional mesoscale molecular simulations were performed at this point to investigate the interactions between these dendrimers and the siRNA molecules. The morphology of the self-assembled micelles and siRNAs is presented in Figures 3.2.4D and 3.2.4E. As can be seen, at the same N/P ratio, from dendrimers **1** through to **6**, the siRNA molecules are more uniformly distributed and better encased within the micellar network, indicative of an efficient protection from the external environment. For the non-self-assembling systems like **7** and **8**, the siRNA molecules are less well protected from possible external access. Collectively, all

these findings clearly demonstrate that dendrimers displaying the highest self-assembly properties offer the best interaction with and optimal compaction of siRNA.

In vitro assay

In order to investigate the siRNA delivery capacity of our series of amphiphilic dendrimers, siRNA molecules targeting either heat shock protein 27 (Hsp 27) or translationally controlled tumor protein (TCTP) were tested in human castration-resistant prostate cancer PC-3 and C4-2 cells. As we can see, dendrimer **4** bearing a C18 alkyl chain clearly revealed the best siRNA delivery capacity, leading to potent gene silencing with no discernible toxicity.24 Interestingly, dendrimers with longer hydrophobic tails (5 and 6) were less effective, whereas those bearing shorter aliphatic chains (1-3), hydrophilic (PEG) chain (**7**) or no chain (**8**) failed to elicit any discernible gene silencing. These results could be sensibly rationalized in light of the previous discussion: because of the intrinsic incapacity of **7** and **8** to self-assemble, or the inability of **1-3** to generate stable micellar nanostructures at the concentrations employed in the transfection experiments, none of these five dendrimers was effective in siRNA delivery. In other words, the multivalency of the positive dendritic structure alone is not sufficient to efficiently bind, protect and deliver siRNA to cells. The reason why **5** and **6** are less effective than **4** in siRNA delivery and subsequent gene silencing might be ascribed to the fact that the siRNA/**5** and siRNA/**6** complexes are too stable to dissociate, thus impairing the siRNA release within the cell and ultimately decreasing the efficiency of gene silencing.

To verify this hypothesis, we then performed a quantitative characterization of micelle/siRNA interactions at a fully atomistic level to rank the affinity of each dendrimer micelle towards siRNA by molecular dynamics (MD) simulations based on the morphological information gathered by mesoscopic modeling. As presented in Table 3.2.3, the micelles formed by **4** possessed a total surface charge of +64, 31 of which were effectively engaged in complexing siRNA (Figure 3.2.6A), resulting in a $\Delta G_{bind,eff}/N_{eff}$ value of -0.41 kcal/mol (Table 3.2.3). On the other hand, the self-assembled nanocarriers generated by **5** and **6** were not only able to exploit 49 and 62 (out of their total 80 and 102 positive charges, respectively), to constantly bind siRNA (see Figure 3.2.6A), but they also did it more effectively, as testified by the higher favorable $\Delta G_{bind,eff}/N_{eff}$ values of -0.59 and -0.56 kcal/mol (Table 3.2.3).

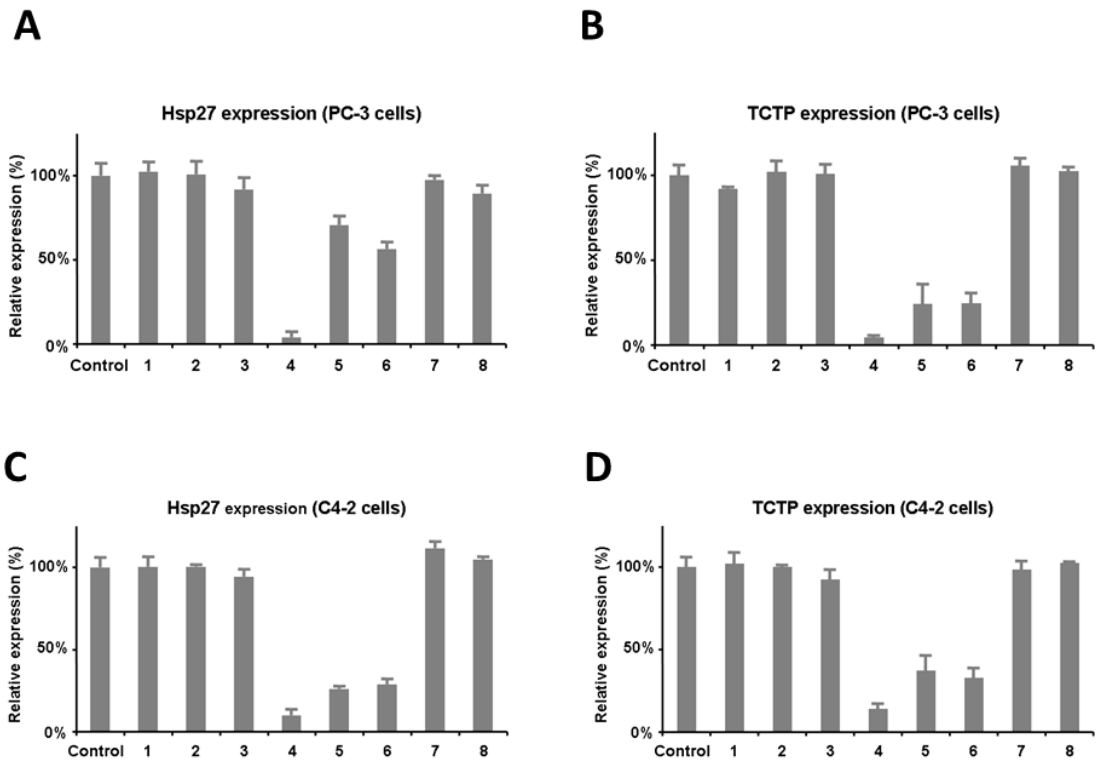


Figure 3.2.5. *Dendrimer-mediated siRNA delivery and gene silencing of heat shock protein 27 (Hsp27) and translationally controlled tumor protein (TCTP) in human prostate cancer PC-3 (A, B) and C4-2 (C, D) cells, respectively. Hsp27 and TCTP protein expression quantified using western blot 72 h post-treatment with the corresponding siRNA/dendrimer complexes at N/P ratio of 10.*

These higher values of $\Delta G_{bind,eff}/N_{eff}$ definitely characterize the two dendrimers **5** and **6** as the strongest siRNA binders of our series, in agreement with the corresponding experimental evidence obtained with the EB fluorescent assay and gel shift (Figures 3.2.4A and 3.2.4B). The siRNA binding affinity of micelles formed with **5** and **6** might indeed be too strong to allow intracellular siRNA release. It is tempting therefore to speculate that **5** and **6** achieve more effective siRNA binding - as indicated by both experiments and simulation - at the expense of effective siRNA release, as the complex simply becomes too “strongly bound” and cannot be broken apart.

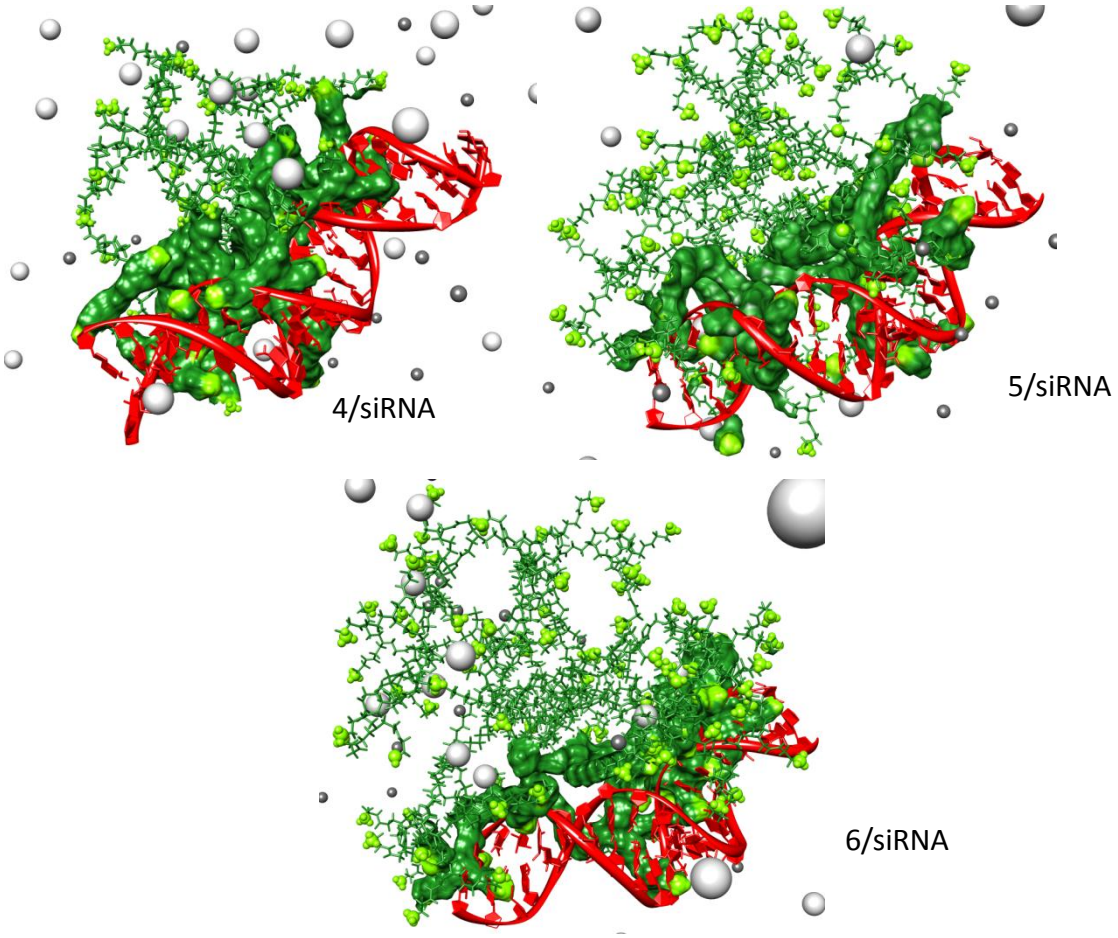


Figure 3.2.6 *(A) Atomistic MD simulations of amphiphilic dendrimer self-assembly in the presence of siRNA. In all panels, the dendrimers are shown as forest green, with the terminal charged amine groups highlighted as chartreuse green sticks-and-balls. The siRNA is represented as a red ribbon and some representative Cl- and Na+ ions and counterions are portrayed as light and dark gray spheres, respectively. Water is omitted in all panels for clarity. The binding interface region between each micelle and siRNA is also shown in each complex.*

To further support this finding, we examined the siRNA dissociation from the corresponding dendrimer complexes using the heparin-coupled ethidium bromide (EB) assay. In this method, heparin, a highly negatively charged polysaccharide, is employed for competing with siRNA for dendrimer binding. Displacement of siRNA by heparin from the dendrimer complex allows the intercalation of the fluorescent EB probe, which then emits a strong fluorescence. siRNA could be almost completely displaced from its complex with **4**, while it was only partly released from those complexes formed either with **5** or **6**. These results demonstrate that the self-assembled nanocarrier generated by **4** is bestowed with optimal siRNA binding strength and siRNA releasing ability, enabling it to safely carry the siRNA cargo along its journey to the cell and efficiently discharge the payload upon reaching its final destination.

Overall, these results fully illustrate that **4** mediates a uniquely effective siRNA delivery and gene silencing. The sensible molecular rationale for this is that an optimal balance has been reached between the hydrophobic chain length and the hydrophilic dendron structure in **4** providing this amphiphilic dendrimer molecule the ideal properties to efficiently self-assemble into supramolecular structures. Such structures are able to easily form nanocomplexes with siRNA before their disassembly during endosome release, ultimately resulting in the best siRNA delivery and the most potent gene silencing of the entire dendrimer series presented in this work.

Conclusion

In conclusion, we gained a unique and detailed insight into the ability of the amphiphilic dendrimers to self-assemble into nanosized micellar structures and consequentially interact with siRNA molecules for functional delivery using a powerful combination of experimental and theoretical approaches. Our dual investigation strategy revealed that an optimal balance between the hydrophobic alkyl chain length and the hydrophilic dendritic portion plays a crucial role in the self-assembly and thus in the excellent siRNA delivery activity displayed by the amphiphilic dendrimer **4**. The supramolecular nanostructures generated by **4** resemble somewhat the well-defined traditional covalent dendrimers of higher generation. The ability of carefully designed building blocks to spontaneously assemble into complex nanostructures underpins developments in a wide range of technologies, from materials science to molecular biology, highlighting huge potential for implementation in future applications.

3.3 INVESTIGATION OF THE ROLE OF S45F β CATENIN MUTANT AS MARKER OF AGGRESSIVENESS IN DESMOID TUMORS

This work was carried out in collaboration with Unit of Clinical Epidemiology and Trial Organization and Laboratory of Experimental Molecular Pathology, Istituto dei Tumori, Milan (Italy)

Introduction

Desmoid tumors (DTs) are rare mesenchymal lesions with fibroblastic proliferation characterized by a variable clinical course. In a clinical study involving 179 patients a high frequency of mutations (85%) in the gene encoding for β-catenin protein was found[78].

Interestingly, only three different point mutations in two different codons (41 and 45) have been identified in mutated samples: ACC to GCC in codon 41 (T41A; replacement of threonine by alanine); TCT to TTT in codon 45 (S45F; replacement of serine by phenylalanine), and TCT to

CCT in codon 45 (S45P; replacement of serine with proline)[79]; these two codons are targets for phosphorylation by GSK-3β and CK1, respectively. [80]

Among solid tumors it is unusual to demonstrate only three specific mutations in CTNNB1; most other neoplasms harboring exon 3 CTNNB1 mutations have wider mutational variability at multiple critical codons[81,82], suggesting that these specific CTNNB1 mutations may be critical in the development of desmoid-type fibromatosis and the type of CTNNB1 gene mutation could affect the levels β-catenin signaling, thereby impacting on desmoid formation and/or recurrence.

In summary three different mutants were identified: T41A, S45F and S45P. All this residues are phosphorylation/regulation sites for this protein. Specifically, phosphorylation of Serine 45 increases the affinity of β-catenin for GSK3, thus promoting its ubiquitination and degradation. Actually, only S45F mutant seems to play a fundamental role in tendency for local recurrence after complete surgical resection. Exploiting molecular simulation technique, we then studied the interactions of the wild-type (WT) and T41A, S45P, and S45F β-catenin mutants with GSK3. Of note, a proprietary 3D homology model of the full β-catenin structure was employed in this study.

Experimental section

Inclusion criteria and clinical database

Patients with primary sporadic DT histological diagnoses surgically treated at one of four sarcoma referral centers (Fondazione IRCCS, Istituto Nazionale dei Tumori, Milano, Italy; The University of Texas MD Anderson Cancer Center [UTMDACC], Houston, Texas, USA; Institute Gustave Roussy [IGR], Paris, France; Leiden Medical Center, Leiden, Netherlands) between April 1998 and March 2011 were included in this study. Of note none of the UTMDACC patients included in the current study were part of the previously published patient cohort[79].

Patients with FAP-related DT, recurrent disease at presentation, or incomplete surgical resection were excluded. Data retrieved included gender, anatomical site, tumor size, date of surgery, status of margins including microscopic positive (R1) and negative margins (R0), results of CTNNB1 mutational analysis, type of other treatments eventually administered before and/or after surgery, date of recurrence, treatment of recurrence, date and status at last follow up. Anatomical sites were classified as extra abdominal (including extremity and head/neck), abdominal/thoracic wall, and intra-abdominal. Tumor size was defined as the greatest DT dimension in the surgical specimen reported by the original pathologists. Surgical excisions were considered as macroscopically complete in the absence of gross residual disease. All macroscopically complete resections were classified according to the closest

surgical margin, which was microscopically categorized as positive (tumor within 1 mm from the inked surface) or negative (absence of tumor within 1 mm from the inked surface).

Non-surgical treatments were administered in the primary or recurrent phase of the disease on an individualized basis or as part of clinical trials and included radiotherapy, chemotherapy (frequently methotrexate and vinorelbine) and medical therapy- hormonal agents (tamoxifene, toremifene), NSAIDs (COX2 inhibitors).

Mutational analysis

Mutational analysis was performed on formalin-fixed-paraffin-embedded tissue in each participating institution using the same standardize technique. Microscopic dissection of 7 μm methylene blue-stained tissue sections (with >80% tumour cells) allowed the precise separation of neoplastic and normal tissues. Genomic DNA was extracted using the Qiamp DNA Kit (Qiagen, Chatsworth, CA, USA) as per manufacturer's instructions. CTNNB1 mutational analysis was made by polymerase chain reaction (PCR) using 100 ng of genomic DNA. Exon 3 of the CTNNB1 gene was amplified using the specific primers ex3F (5'-ATGGAACCAGACAGAAAAGC-3') and ex3R (5'-GCTACTTGTTCTTGAGTGAAG-3') and following the PCR conditions: 30 sec at 95°C, 30 sec at 56°C and 30 sec at 72°C for 35 cycles. The PCR products were purified by using exoSAP and sequenced with 3500 Dx Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Each sequence was performed at least twice, starting from an independent amplification reaction and evaluated by means of ChromasPro software. Three different mutations were identified at the codon 3 of the gene encoding for β-catenin, CTNNB1 (T41A, S45F, S45P). Patients without a specific mutation were classified as wild type (WT). Since the number of S45P cases was small, for purposes of the statistical analysis T41A and S45P were grouped together and termed "other".

Molecular modelling

All simulations were performed with AMBER 11 suite of programs[15]. Since the lack of the overall 3D structure of β-catenin protein in the Protein Data Bank (PDB), a validate homology modeling protocol was applied to the PDB codes = 2Z6H[83], 2G57[84] using as template importin α (PDB code= 3OQS[85]) and refined using different minimization rounds. Once obtained the complete WT structure, the GSK3-β receptor (PDB code = 3SAY) was docked through HADDOCK server[86] and the complex subjected to molecular simulation dynamics to minimize the energy. Then the involved mutations of β-catenin were introduced in the wild type structure following a well-validate procedure. Each mutant complex was then solvated and energy minimized using a combination of molecular dynamics techniques. The mutants β-catenin complex were generated taking into account different rotamers for each substitution and evaluating the one most probable.

We worked with a neutralized system, obtained by the addition of Cl- and Na+ in grids with the Coulombic potential around the protein largest positive or negative; sequentially, a correct number NaCl molecules were added to ensure physiological ionic strength (150mM). All the simulations were run in explicit solvent, immersing the complex in a TIP3P box[16] of water of a length of at least 10 Å in each dimensions.

In all simulation, long-range electrostatic interactions were treated by particle Mesh Ewald (PME). A dielectric costant of 1 was applied and a cutoff of 9Å for nonbonded interactions was set. 10 ns MD were run for data production.

MM/PBSA calculation[6] was performed, the affinity of wild type and mutated β–catenin with GSK3 (ΔGbind) was calculated as the sum of the van der Waals, polar solvation, electrostatic and entropic contributions. .

Results&discussion

In this series of 179 patients with DT surgically treated at four referral centers over a 10 years time span, CTNNB1 gene mutation was the only significant prognostic factor for relapse. Patients affected by DTs harboring S45F CTNNB1 mutation had a poorer RFS (recurrence free survival) as compared to those having DTs harboring T41A/S45P and those having DTs with a WT CTNNB1 gene (Figure 3.3.1). These data reproduce the previously reported MDACC results[79].

The impact of specific mutation on the pathogenesis of DTs and their predisposing role in local relapse remains uncertain. Only three different point mutations in two different codons (41 and 45) have been identified to date in desmoid-type fibromatosis. In particular, specific amino N terminal residues (Ser45 by CK1α, and Thr41, Ser 37, and Ser33 by GSK-3β in this sequential order) represent the phophorylation targets that mediate the degradation of β-catenin, a crucial step in maintaining low β-catenin cytoplasmic levels. In fact, deregulation of β-catenin activity is associated with cancer and other human diseases, due to its key involvement in the Wnt pathway. The identification of two different types of point mutation in the 45 position and the correlation between 45F and aggressive behavior suggests but does not prove a potential critical role of phophorylation mediated by GSK3. In this context, phophorylation in the position 45 amino acid residue by CK1α comprises the first step in β-catenin ubiquitination, a process that apparently cannot begin if a mutation in this specific residue is present.

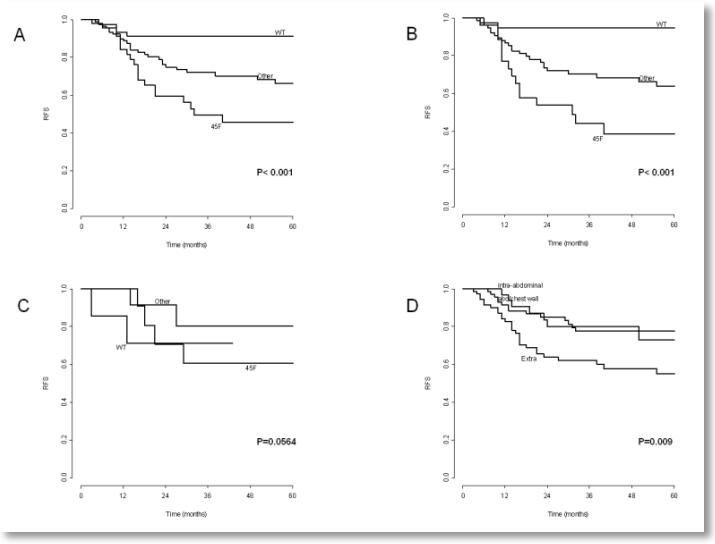


Figure 3.3.1 *RFS of patients grouped according to the specific mutation[80] (A); RFS of patients who underwent surgery alone (B) or surgery plus other treatments (C) grouped according to specific mutation; RFS of patients grouped according to specific anatomical site (D).*

In fact CTNNB1 mutations occur within this exon of the gene in the region encoding for sequence that contains different Serine/Threonine.

Molecular simulation techniques were applied to evaluate the importance of these residues for catenin ubiquitination, CK1 and GSK3 sequentially phosphorylate the amino terminal region of β-catenin resulting in its recognition by B Trcp, an E3 ubiquitin ligase subunit, and leading to ubiquitination and proteasomal degradation.

The available structures of β–catenin in the protein data bank (PDB code=2G57, PDB code=2Z6H) correspond, respectively, to the regions of the "phosphorilation motif" and the "Armadillo repeats domain" with C-ter part. An extensive study required preferentially the complete protein, for this reason we built the N- and, partially, C-terminal parts resorting to an homology modeling protocol well validated.

The first step of this procedure consists on the prediction of the secondary structure, employing the online available programs DISOPRED and PredictProtein, obtaining partially disordered regions involving the C-terminal part.

Then the obtained results were considerate under the light of the protein template identified; it requires an in-depth search, through MODWEBand SWISSPROTEIN, for the most suitable template. The output gave a considerable similarity of the specific portion required, more than 45%, with the protein importin-α, then the receptor/protein sequence was aligned and checked using the Build Homology model of DS.

The first structure was adjusted, inserting the already known regions, deriving from NMR study, and subjected to extensive MD simulation in explicit solvent to reach an energetically

reasonable conformation, monitoring different parameters (RMSD, Total Energy, Temperature) to evaluate the good success of dynamic (Figure 3.3.2A).

The reliability of the model was further verified through different tools (Whatif, Prosa), evaluating the scoring factor obtained and correlate it with the data present in literature. Figure 3.3.1 showed a positive structural output resulting from the Ramachandran plot, most of the residues, more than 95%, are located in favored regions while only a small portion stay in generously allowed regions and only 1% in less favorable area. Therfore, the validated 3D homology model of the full β-catenin structure was employed in this study (Figure 3.3.3).

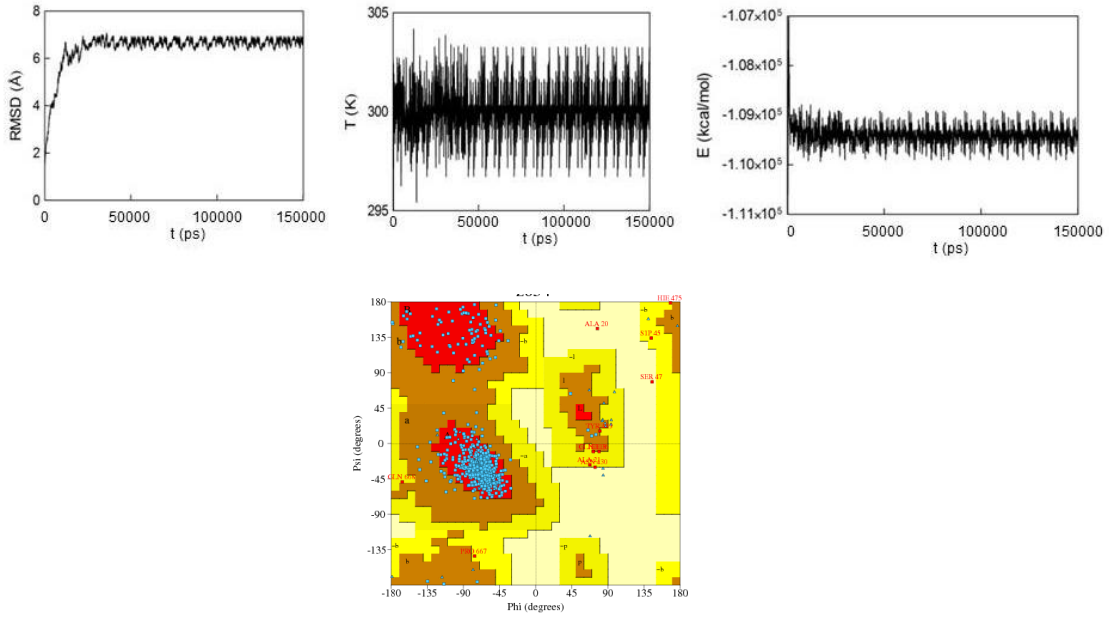


Figure 3.3.2 *A) (from left to right) RMDS of the coordinates of the backbone atoms of the homology model of the β catenin along the 15 ns MD simulation, temperature of the β catenin 3D model in water during the same MD simulation and time behavior of the total potential energy; B)Ramachandran plot generated from the MD refined 3D homology model of the β catenin.*

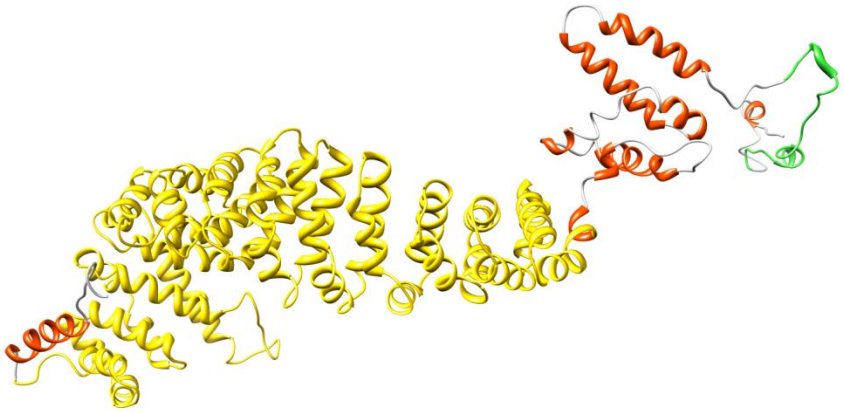


Figure 3.3.3 *Final 3D homology model structure of β-catenin; in green the phosphorylation motif, in red C- and N- terminal parts, in yellow armadillo repeats domain.*

As already known, all the residues mutated in β-catenin are phosphorylation/regulation sites for this protein. Specifically, phosphorylation of Serine 45 increases the affinity of β-catenin for Glycogen synthase kinase 3 (GSK3), thus promoting its ubiquitination and degradation. Actually, only S45F mutant seems to play a fundamental role in tendency for local recurrence after complete surgical resection. In order to verify whether the amino acidic changes caused by mutation in the exon 3 of CTTNB1 were actually responsible for the deregulation of β–catenin, the S45F and S45P mutations were introduced into the Wild Type (WT) β–catenin and investigated in complex with GSK3.

First of all the 3D structure of GSK3 was downloaded by PDB and the identification of the putative interaction site was carried out exploiting Haddock server. Different interesting point emerged: i) catenin bound GSK3 with P-Ser45, contained in the N-terminal and in particular in the phosphor motif; ii) GSK3 accepts the P-Ser45 in a pocket create by the loop region from Arg93 to Lys205[87]. Having localized the putative interactive zone in this region we docked the three different systems by Haddock (Figure 3.3.4). The resulting docked conformation was then subjected to MD simulation in explicit solvent and evaluated the different conformation for the three complexes (WT, S45P, S45F). The procedure for protein-protein interaction molecular modeling study was considered as in well described different works[35,88-90].

The all long simulation highlighted the importance of the side chains of the mutated Ser45 that seems to play a crucial role in binding and stabilizing GSK3 in the interaction zone. Actually, the side chain of F45 reaches deeply into the specific phospho-serine pocket, thus preventing the disruption of the complex and, contextually, hampering the kinase activity of GSK3 on T41, S33 and S37. S45 plays a fundamental role in the conformational properties and dynamics of the pocket to keep the dynamic equilibrium of this “sliding” mechanism.

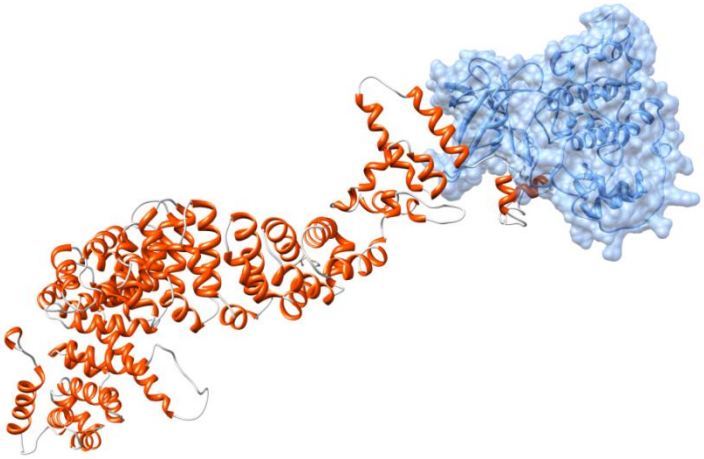
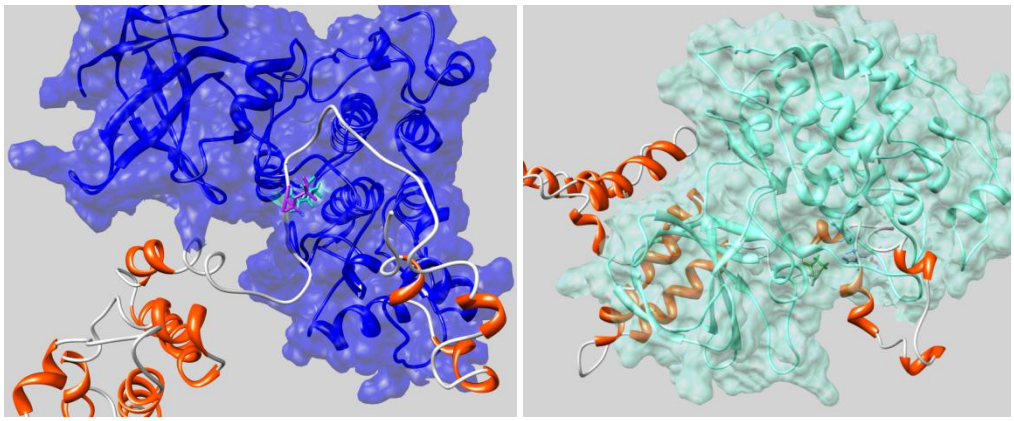


Figure 3.3.4 *Complex of WT β catenin (in red) in complex with GSK3 (in blue); in light blue is depicted the surface of GSK3.*

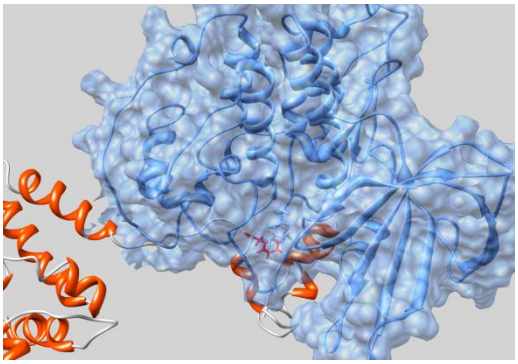
During 10 ns of MD simulation we observed that the steric hindrance of the phenylalanine side chain compared to serine reflects in a rearrangement of the equilibrium, altering the ubiquitination machine (Figure 3.3.5).

All these findings were confirmed by the thermodynamic data resulted from MMPBSA application, as showed in Table 3.3.1A ΔH of 45F has the strongest affinity for GSK3; on the other hand the presence of a proline seems to exert a minor perturbation to the destruction complex.

In particular, analyzing the deconvolution of energies per residues, we noticed that Phe45 and, gains about 6 kcal/mol compared to the WT (WT -4,8 kcal/mol to D473Y -3,2 kcal/mol) and in this mutant also the neighboring residues seems to affect the overall energies (Table 3.3.1B).



A B



C

Figure 3.3.5 *Modeled complex of GSK3 in complex with β catenin a) WT, b) S45P, and c) S45F, showing the key interaction zone proposed, highlighted the key residues involved in the mutations.*

	ΔH (kcal/mol)
WT	-57,44 (±0,92)
S45F	-116,00 (±0,99)
S45P	-75,74 (±0,91)

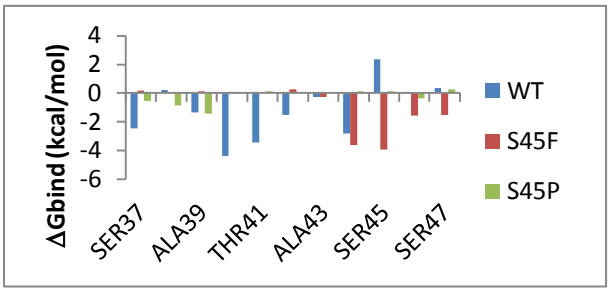


Table 3.3.1 *A) Enthalpic contributions ΔH (kcal/mol) for GSK3 in complex with β catenin wt, S45F and S45P. Errors are given in parenthesis as standard errors of the mean; B)Per residue binding free energy decomposition for GSK3 in complex with β catenin WT (in blue), S45F (in red), and S45P (in green). Only β catenin amino acids from position 37 to 47 are shown, since the interest in the phosphorylation motif.*

Conclusion

The clinical data confirmed that patients with S45F mutated primary and completely resected DTs exhibit a higher tendency for local recurrence. The detailed analysis of the binding free energy and its per residues deconvolution for all these supramolecular complexes shows that replacing Serine with Phenylalanine results in a stronger interaction of β-catenin with GSK3. The results of this in silico investigation support the clinical evidence for the high aggressiveness of the S45F mutant.

3.4 EFFECTS OF A DOUBLE MUTATION ON c-KIT RECEPTOR AFFINITY FOR IMATINIB[91]

Here we report for the first time a case of GIST whose molecular characterization revealed the presence of two mutations in KIT exon 11, one of which never described before, causing the double aminoacidic substitution W557G/Y578C. Based on the presence of KIT mutations in exon 11, the patient was subsequently treated with Imatinib, and responded well to this TKI-based therapy. In keeping, in silico (i.e., computer-based molecular simulations) and in vitro experiments demonstrated that the KIT receptor carrying the W557G/Y578C double mutation displays constitutive phosphorylation which can be switched-off upon Imatinib treatment.

Materials and Methods

Patient history

On November 2009, a 68-year-old man with a history of hypertension and atrial fibrillation owing to severe abdominal pain underwent an X-ray computed tomography (CT scan) that revealed hemoperitoneum due to an abdominal mass. An emergency surgery with resection of

the mass probably arising from stomach was carried out, providing a diagnosis of hemangiopericytomas.

Six months later, a CT scan showed multiple peritoneal lesions and the patient underwent to surgery of all peritoneal metastasis. Histological review of primary tumor was performed at Fondazione IRCCS Istituto Nazionale dei Tumori (Milan, Italy) with a diagnosis of GIST.

On September 2010, due to peritoneal progression of disease, the patient was administered 400 mg Imatinib per day. Disease assessment, including CT scan, was done before starting Imatinib. The patient underwent also a complete cardiac evaluation. Routine liver and renal function tests were within normal ranges. The last follow up visit was in November 2012.

GIST diagnosis

A paraffin-embedded formalin-fixed representative block of the primary tumor, previously diagnosed as hemangioperycitoma, was received by the Fondazione IRCCS Istituto Nazionale dei Tumori (Milan, Italy) for a second pathological opinion. Diagnosis of GIST was rendered based on tumor morphology, featuring a spindle and epithelioid cells growth, and on immunophenotyping, positive for CD117 (clone A4502, Dako) with a dot like decoration, DOG1 (clone-K9, Novocastra, Leica Biosystems), PKC theta (clone 27/PKCθ, BD Transduction, Franklin Lakes, NJ USA), PDGFRA (clone-sc338, Santa Cruz Biotechnology, Inc.) and focally for citokeratins (clone AE1/AE3, Dako). The mitotic count revealed 60 mitoses/HPF that, along with the tumor mass diameter (> 5 cm), categorized the tumor as a high risk GIST.[92]

KIT and PDGFRA molecular analysis

The mutation hot spots of both KIT and PDGFRA receptors were analyzed by direct sequencing of exons 9, 11, 13, and 17 of KIT, and exons 12, 14, and 18 of PDGFRA genes as previously described[93].

The PCR product of KIT exon 11 amplification was subcloned into a pGEMT vector (Promega) following manufacturer's instructions, and transformed into DH-5α E. Coli competent cells (Life Technologies). 30 bacterial ampicillin resistant colonies were selected, and their DNA was amplified for KIT exon 11 to verify the presence of the subcloned DNA. These PCR products were subsequently sequenced as previously described[93].

Computational details

All KIT three-dimensional (3D) mutant models (W557G, Y578C and the double mutant W557G/Y578C) were obtained starting from our optimized wild-type (WT) KIT model [93-96]. The validated structures of the Δ559 and T670I KIT mutants[97] were also considered as reference for an Imatinb responsive and a resistant TK isoform, respectively.

All computer simulations were performed using the AMBER 11 suite of programs on 256 processors of the IBM PLX-GPU supercomputer (CINECA supercomputer center, Bologna, Italy)

as previously reported[30,52,93,94,96]. The affinity of the mutant KIT isoforms for ATP and Imatinib, expressed as the value of the corresponding free energy of binding Δ Gbind, was estimated according to the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method[6] (see details in Material&Methods chapter).

Expression vector construction, transient transfection and Western blot analysis

KIT W557G and Y578C mutants have been obtained by site-directed mutagenesis of pCDNA3 expression vector containing a wild-type (WT) KIT[97] by Site-Directed Mutagenesis kit (Promega, Madsion WI) following manufacturer's instructions, using the following primers:

5'-CAACACAACTTCCTTGTGATCACAAATGGGAG-3' and

5'-CTCCCATTTGTGATCACAAGGAAGTTGTGTTG-3' for W557G mutation;

5'–CATGTATGAAGTACAGGGGAAGGTTGTTGAGG-3' and

5'-CCTCAACAACCTTCCCCTGTACTTCATACATG-3' for Y578C mutation.

These latter were employed to produce the W557G/Y578C double mutant using the KIT W557G expression vector as a template.

HEK293 cells were cultured and transfected as previously described.[98] Cells were harvested 48 hours after transfection following overnight serum starvation. Cell lysates were produced in RIPA modified buffer, and Western blot analysis was performed as previously described[98]. Anti c-KIT (clone H300) antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA); anti-phospho c-KIT, AKT and phospho AKT (Ser473) antibodies were from Cell Signal Technology (Beverly, MA, USA); ERK1/2 and phospho ERK1/2 antibodies were from Sigma-Aldrich (St. Louis, MO, USA).

Images were acquired by G-Box (Syngene) and densitometric analysis of the bands was performed by Image Quant software. Data are expressed as the ratio of phosphorylated/total c-KIT and normalized for untreated sample.

Patient evaluation after Imatinib treatment

Patient response to Imatinib treatment was assessed by whole-body CT scan after the first and the third month of treatment then every three months. Radiologic evaluation was done according to RECIST (Response Evaluation Criteria in Solid Tumors) criteria. The CT scan after the first month showed a complete response to therapy (Figure 3.4.1). During the treatment the patient underwent follow-up visit every month during the first three months of therapy, and subsequently, every three months. The toxicity, assessed thought blood-cell counts and blood chemical values, was analyzed every month.

Up To date, after 27 months of treatment, the patient on Imatinib (400 mg /daily) is still on response with no signs of toxicity.

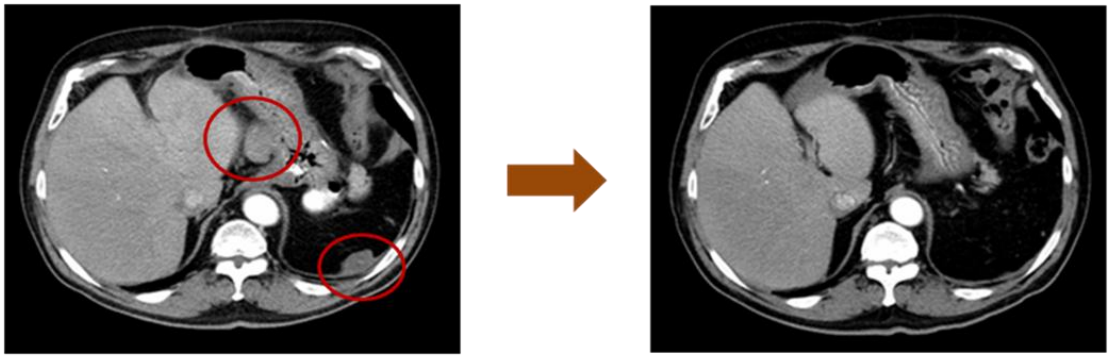


Figure 3.4.1 *Response to Imatinib. Peritoneal metastasis from gastric GIST treated with Imatinib: computed tomography scan (arterial phase after contrast medium). A RECIST complete response after 1 month of treatment with Imatinib was obtained and the therapy is ongoing.*

Molecular analysis of the GIST tumor

KIT gene sequencing of the primary tumor revealed a double mutation in exon 11 (T1669G and A1733G), present in heterozygous status and responsible for the W557G and Y578C substitutions on the KIT receptor. To assess whether both mutations were located on the same allele, the PCR fragment comprising the whole KIT exon 11 was subcloned into a pGEMT vector[99]. Sequencing of the subcloned DNA identified a co-presence of both W557G and Y578C KIT mutations in 26 out of the 30 E. coli colonies analyzed, thus confirming mutation heterozygosis. Accordingly, in the tumor the mutated KIT receptors were characterized by the presence of the double aminoacidic substitution. The W557G mutation, already detected in GISTs, is one of the most frequent point mutations in exon 11. Patients with GIST but also with melanoma [100] carrying this KIT mutation responded to Imatinib[101,102]. However, to our knowledge, in vitro data aimed to the analysis of intrinsic activity and Imatinib sensitivity of W557G KIT have not been reported so far. Furthermore, the Y578C mutation has never been described before, neither in GISTs nor in any other pathology (Sanger Institute, Cosmic database, www.sangerinstitute.com); notably, it is located closed to the KIT activating/Imatinib-resistant L576P substitution[93].

In order to determine the effect of the Y578C substitution, alone and in combination with the other detected mutation, W557G, on KIT receptor activity and sensitivity towards Imatinib, we performed a thorough study based on coupled in-silico/in-vitro experiments.

Molecular modeling

The upper half of Table3.4.1 lists the total free energy of ATP binding values that were calculated for the Δ559, T670I, W557G, Y578C, and the W557G/Y578C double KIT mutant receptors, respectively. The value of Δ Gbind for the Δ 559 KIT with respect to ATP was found to be Δ Gbind = - 24.20 kcal/mol, in agreement with our previous simulations[97] (Figure 3.4.2). As suggested by the ΔΔGbind (=ΔGbind,wt – ΔGbind,mut) values in the last row of the top half of Table 3.4.1, all KIT mutants showed an affinity for ATP comparable to that of the Δ559 reference protein, thus confirming the activating character of these mutations.

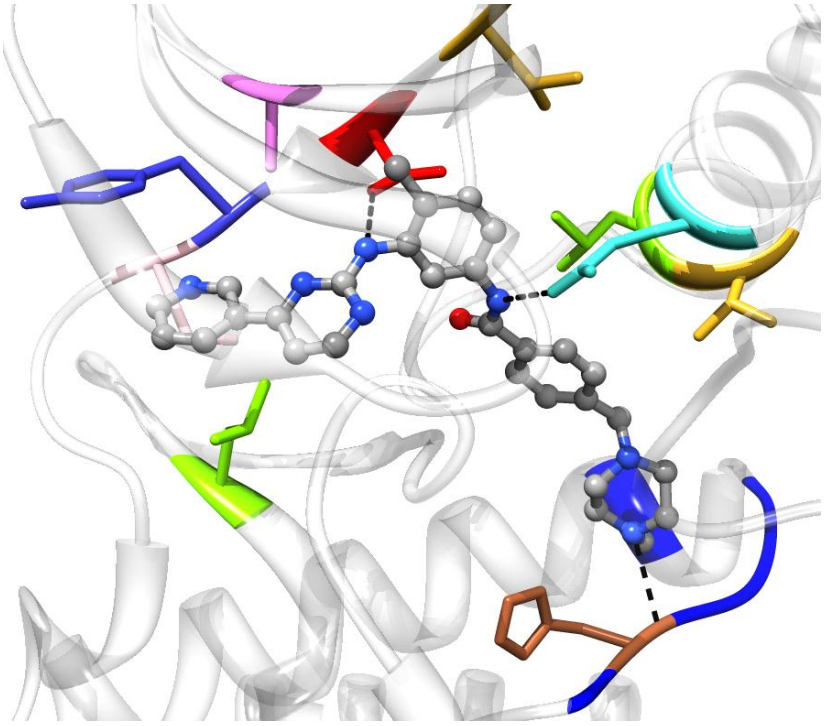
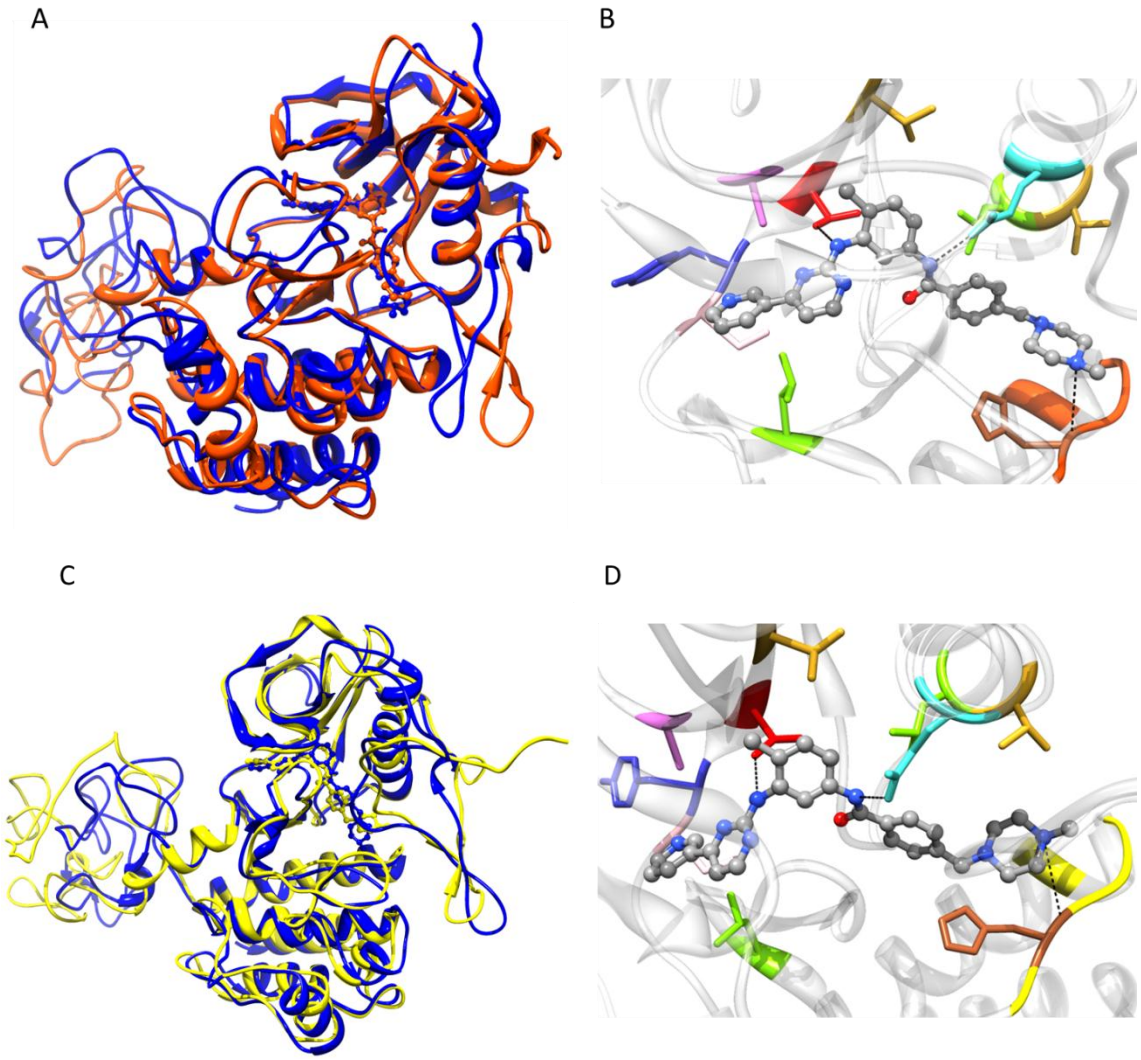


Figure 3.4.2 *Zoomed view of Imatinib in complex with Δ559 KIT. The inhibitor is depicted as atom-colored sticks-and-balls (color code: O, red; N, blue, C, gray). The main residues interacting with Imatinib are highlighted as colored sticks (A621, orchid; Glu640; turquoise; Val643, goldenrod; Thr670, red; Tyr672, medium blue; Cys673, pink; Leu799, chartreuse; his790, sienna, Leu 783, blue. H-bonds are evidenced by black broken lines. H atoms, water molecules, ions are counterions are omitted for clarity.*

As the affinity for Imatinib is concerned, the calculated free energy of binding values listed in lower part of Table 3.4.1 show that the double-mutant receptor is endowed with an affinity towards Imatinib (ΔGbind = -8.17 kcal/mol) slightly lower than that of the activating but Imatinib-responsive Δ559 isoform (ΔΔGbind = -0.98 kcal/mol) (see Figure 3.4.3, panels A and B). Conversely, the presence of the Y578C alteration seems to exert a minor perturbation to Imatinib binding, as the corresponding ΔGbind value (-8.89 kcal/mol) is comparable to that of the Δ559-mutated receptor (ΔΔGbind = -0.26 kcal/mol) (see Figure 3.4.3, panels C and D).

Intriguingly, the Y578C substitution involves a residue that is close to L576. KIT mutations at this position (e.g., L576P) have been shown not only to activate the receptor but also to confer a high degree of resistance to Imatinib[93,103]. Of note, the W557G KIT mutant is predicted to be utterly similar to the reference Δ559-mutated receptor in its affinity towards both ATP and Imatinib (see Figure 3.4.3, panels E and F), in agreement with previous findings[104], whereas the T670I is found very resistant to the TK inhibitor, as expected.



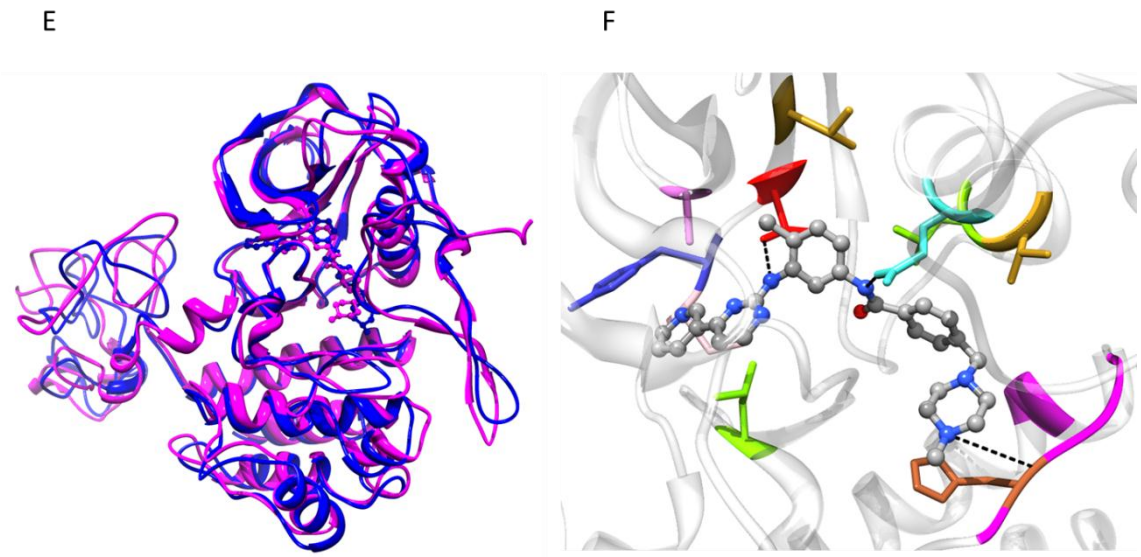


Figure 3.4.3 *(A) Superposition of MD equilibrated snapshots of Δ559 (blue) and the W557G/Y578C double mutant (orange) KIT in complex with Imatinib. (B) Details of Imatinib in complex with the W557G/Y578C double mutant KIT. (C) Superposition of MD equilibrated snapshots of Δ559 (blue) and the Y578C mutant (yellow) KIT in complex with Imatinib. (D) Details of Imatinib in complex with the Y578C mutant KIT. (E) Superposition of MD equilibrated snapshots of Δ559 (blue) and the W557G mutant (magenta) KIT in complex with Imatinib. (F) Details of Imatinib in complex with the W557G mutant KIT. In panels A, C, and E, Imatinib is portrayed in a stick-and-ball representation colored according to the protein isoform which is bound to. In panels B, D, and F, the inhibitor is depicted as atom-colored sticks-and-balls (color code: O, red; N, blue, C, gray). The main residues interacting with Imatinib are highlighted as colored sticks (A621, orchid; Glu640; turquoise; Val643, goldenrod; Thr670, red; Tyr672, medium blue; Cys673, pink; Leu799, chartreuse; his790, sienna, Leu 783, blue. H-bonds are evidenced by black broken lines. In all images, H atoms, water molecules, ions are counterions are omitted for clarity.*

Effect of W557G and Y578C mutations on KIT receptor activity

Next, we investigated the effect of W557G, Y578C and W557G/Y578C mutations on KIT receptor activity and Imatinib response in intact cells. KIT mutants were transiently transfected in HEK293 cells and KIT expression and phosphorylation in the absence or in the presence of Imatinib (1 and 5 μM) were analyzed by Western blot (Figure 3.4.4A). Δ559 and T670I KIT mutants were used as control for Imatinib sensitivity and resistance, respectively.

In untreated cells a delay of processing was observed for Δ559, W557G, and W557G/Y578C KIT mutants with respect to the others, as the 145 kDa fully glycosylated form was barely detectable with anti-KIT antibodies. However, such difference was abrogated by Imatinib treatment (Figure 3.4.4 A, top panel).

For all KIT mutants Western blot with anti-phospho-KIT antibodies showed constitutive phosphorylation of both 145 and 125 kDa forms. KIT phosphorylation levels of W557G, Y578C,

and W557G/Y578C KIT mutants were comparable to those of Δ559 mutant (Figure 3.4.4 A), as also observed by quantitative analysis. These data demonstrate that the Y578C mutation, alone or in association with W557G, induce constitutive KIT receptor activation, in agreement with the modeling results shown in Table 3.4.1. Quantitative analysis of relative KIT phosphorylation levels is reported in the graph of Figure 3.4.4 A. As expected, 1 and 5μM Imatinib treatment reduced Δ559 but not T670I mutant phosphorylation.

Imatinib treatment of cells expressing W557G mutant produced a reduction of receptor phosphorylation at both doses. Similarly, phosphorylation of the Y578C mutant receptor was decreased by treatment with both 1μM and 5μM Imatinib, indicating sensitivity of this mutation to the TKI, in accordance with in silico analysis prediction.

The double W557G/Y578C mutation displayed the same features of the single W557G and Y578C mutations: constitutive phosphorylation was reduced by at least 70% by Imatinib treatment, once again in agreement with the computer-based results. Moreover, as shown in Figure 3.4.4 B, KIT inhibition by Imatinib blocked the signaling pathways triggered by W557G/Y578C mutant receptor, as ERK1/2 and AKT phosphorylation were reduced, similarly to the effect shown for the Δ559 Imatinib sensitive KIT mutant (Figure 3.4.4 B)[98].

Table 3.4.1 *Binding free energies (ΔGbind) and binding free energy differences (ΔΔGbind)[a] for the Δ559, T670I, W557G, Y578C, and W557G/Y578C KIT receptors in complex with ADP and Imatinib. All values are in kcal/mol.*

ATP-binding	Δ559	T670I	W557G	Y578C	W557G/Y578C
ΔG_{bind}	-24.20 ± 0.82	-25.14 ± 0.94	-23.98 ± 0.22	-24.72 ± 0.52	-24.97 ± 0.69
$\Delta\Delta G_{bind}$	-	+0.94	-0.22	+0.52	+0.77
Imatinib-binding	Δ559	T670I	W557G	Y578C	W557G/Y578C
ΔG_{bind}	-9.15 ± 0.12	-6.38 ± 0.11	-9.02 ± 0.09	-8.89 ± 0.06	-8.17 ± 0.07
$\Delta\Delta G_{bind}$	-	-2.77	-0.13	-0.26	-0.98

[a] *According to its definition (ΔΔGbind = ΔGbind,wt – ΔGbind,mut), negative values calculated for ΔΔGbind indicate that the considered amino acid substitution at a given position of KIT is unfavorable in terms of ATP/Imatinib affinity, whereas positive ΔΔGbind values indicate that the considered mutation is favorable*

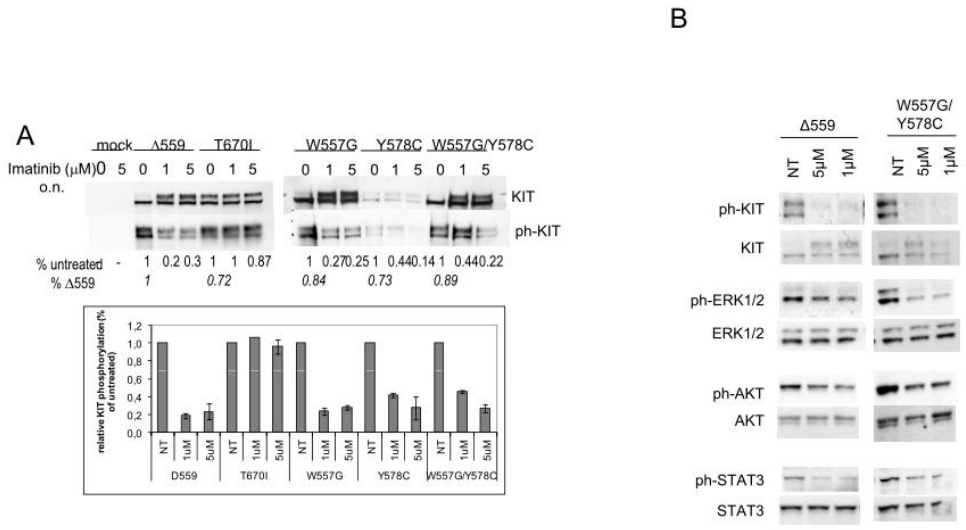


Figure 3.4.4 *Effect of W557G and Y578C mutations on KIT receptor activity and sensitivity to Imatinib. (A) Western blot analysis (top) of HEK293 cells transiently transfected with the indicated KIT mutant constructs, untreated (0) or treated with 1 or 5 mM Imatinib. Quantitative analysis of relative KIT phosphorylation (bottom) expressed as ratio between phosphorylated and total KIT levels and referred to untreated. The data represent the average of two independent experiments. (B) Effect of Imatinib on ERK1/1 and AKT phosphorylation induced by expression of Δ559 and W557G/Y578C KIT mutants in transiently transfected HEK293 cells. Arrows indicate the 145 and 125 kDa KIT receptor forms.*

Overall, these data suggest that each single W557C and Y578G mutation causes constitutive activation of the KIT receptor and is sensitive to Imatinib treatment. Moreover, the concomitant presence of the W557G and Y578C mutations on the same molecule does not alter the effect of single mutations on KIT receptor activation and sensitivity to Imatinib.

Conclusions

The data presented in this work indicate that the unusual combination of two KIT mutations, one of which reported for the first time and located close to the Imatinib insensitive mutation L567P, does not influence the overall sensitivity of KIT mutant towards Imatinib. In vitro analyses coupled with in silico studies once again proven to be a reliable and complete tool to predict drug affinity for mutated receptors. This is particular true in the light of the clinical response observed in the patient here reported after Imatinib treatment, and emphasizes the potent role of this jointed approach, able to yield consistent results (in a few hours as in silico analyses are concerned) in driving the choice for a suitable patient treatment in presence of unknown mutations whose drug sensitivity is lacking.

3.5 PRIMARY AND ACQUIRED RESISTANCE TO HEDGEHOG INHIBITOR VISMODEGIB THROUGH SMOOTHENED (SMO) RECEPTOR MUTATION IN BASAL CELL CARCINOMA

In collaboration with Head and Neck Medical Oncology Unit and Laboratory of Experimental Molecular Pathology, FondazioneIRCCS Istituto Nazionale dei Tumori, Milan (Italy)

Two patients enrolled in a phase II trial were evaluated to assess the safety of vismodegib in metastatic or locally advanced BCC, showing a pattern of primary and acquired resistance to the drug, respectively.

The first patient was a 82-years old woman presenting with a BCC metastatic to liver, lungs and bones. Histological confirmation of metastasis was obtained by liver biopsy. She received radiation therapy to lumbar (L2-L4) and cervical (C3) metastasis, then she started vismodegib 150 mg/die. After 2 months of continuous treatment CT scans showed disease progression at all sites compared (primary resistance).

The second patient was a 78-years old man presenting with a large (12 cm) and ulcerated lesion of the sovrascapular skin, that was biopsied and diagnosed for BCC (Figure 3.5.1, left). As a surgical approach would lead to substantiall morbidity, this patient was also set on a 150 mg/die vismodegib regimen. The lesion dramatically reduced after the first month, and a complete clinical response was obtained at the 6th month of treatment (Figure 3.5.1, left). However, at month 11, 2 subcutaneous nodules appeared within the area of the previous lesion, which were surgically removed. Histological exams showed a pattern of recurrent BCC (acquired resistance).

Methods

PTCH1 and SMO mutational analysis

Formalin-fixed paraffin-embedded (FFPE) tumor specimens of pre-treatment BCC liver metastasis (first case), and of bothpre-treatment primary tumor and recurrence arisen during vismodegib regimen (second case) were reviewed for tumor content and subjected to macrodissection. Genomic DNA was extracted using the Qiamp FFPE DNA kit (Qiagen, Chatsworth, CA), following manufacturer's instructions.

Exons 1-23 of PTCH1 and 8-10 of SMO were amplified by PCR using specific primers. All PCR primer sequences are available upon request. The PCR products were subjected to direct sequencing using an ABI Prism 3500 DX Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), and then evaluated by means of the ChromasPro software.

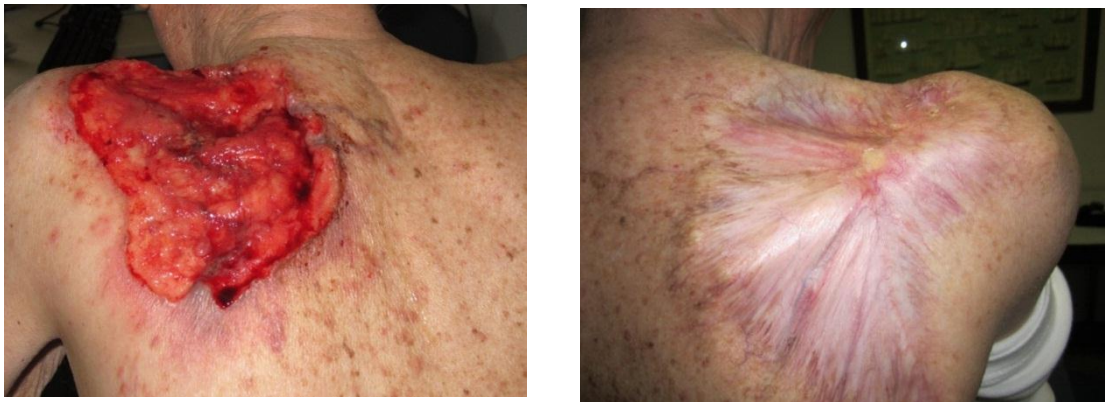


Figure 3.5.1 *Clinical appearance of the soprascapular BCC at diagnosis (left) and after 5 months of vismodegib (right).*

In silico experiments

The free energies of binding ΔGbind for the wild-type (WT) and mutant SMO/vismodegib complexes were obtained using an extensively validated procedure based on the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) methodology[30,52,91]. Accordingly, ΔGbind values are calculated for equilibrated structures extracted from the corresponding 50 ns molecular dynamics (MD) trajectories.

Steered molecular dynamics (SMD) simulations were performed to study vismodegib entry within the receptor binding pocket. With the adopted settings, 3.0 nm were covered in 1000 ps of SMD simulation.

Calculations were carried out using the AMBER 12 platform[63] with Amber ff03 and GAFF force field[105] on the EURORA supercomputer (CINECA, Bologna, Italy).

Molecular analysis

To explore possible mechanisms of Hh pathway activation, we firstly performed genomic sequencing of all 23 coding exons of PTCH1. In the case showing primary resistance, no PTCH1 mutation was observed by analyzing the pre-treatment BCC liver metastasis biopsy. By contrast, in the pre-treatment primary tumor biopsy of the second case, PTCH1 analysis revealed the presence of the nonsense mutation CAG>TAG creating a premature STOP codon (pQ64) in the exon 2 (Fig. 3.5.2 A). This PTCH1 mutation predicts translation to a heavily truncated protein and of primary importance since loss of function mutation in the PTCH1 gene have been associated with SMO activation. As expected, the same PTCH1 mutation was also observed in the recurrence samples taken obtained from the patient during vismodegib regimen (Fig. 3.5.2 B).

Next, we evaluated the mutational status of SMO - the direct protein target of vismodegib. In the case showing primary resistance, the pre-treatment BBC liver metastasis revealed the occurrence of the SMO missense mutation GGG>TGG, leading to the amino acid substitution G497W at exon 9 of the corresponding protein (Fig. 3.5.2 C). By contrast, in the pre-treatment primary tumor samples of the BCC second case, no SMO mutations were detected (Fig. 3.5.2D); however in the recurrence specimens obtained during vismodegib treatment SMO sequencing found the missense mutation GAC>TAC at exon 8, leading to the single amino acid substitution D473Y in the SMO receptor.

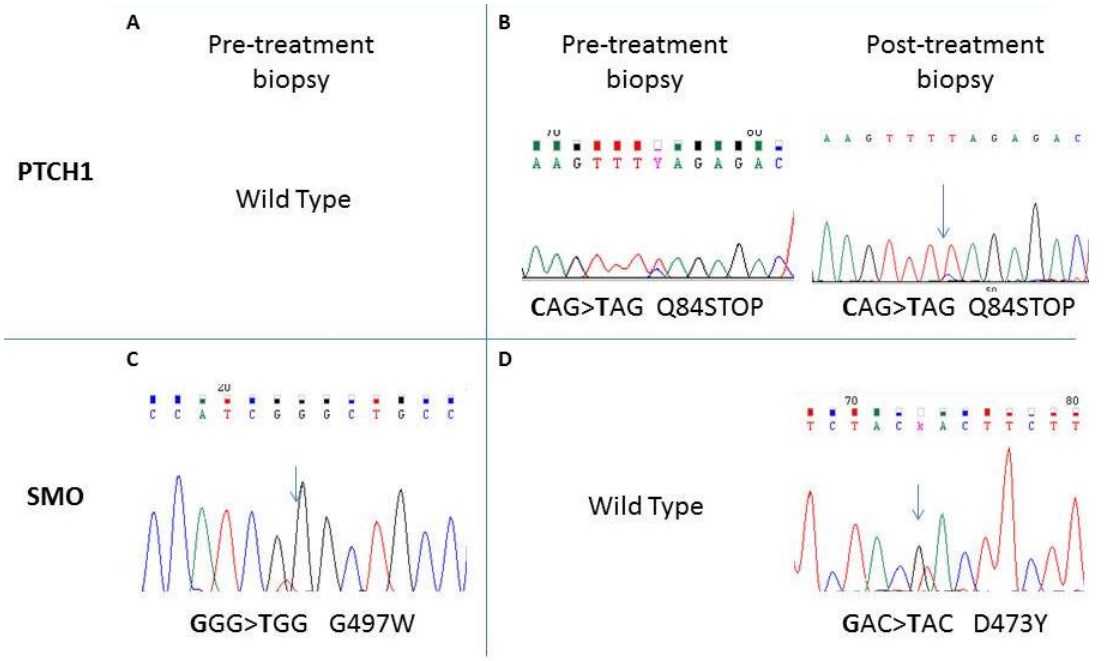


Figure 3.5.2. *PTCH1 and SMO mutational analysis. Pre-treatment BCC liver metastasis of the first case (primary resistance) showed PTCH1 wild type gene (panel A) and the SMO G49W mutation (panel C). Pre-treatment primary tumor and BCC recurrence (panel B) of the second case (acquired resistance) carried the nonsense Q84Stop PTCH1 mutation; the SMO D473Y mutation was observed only in the recurrence BCC sample (panel D).*

In silico experiments

Our in silico experiments revealed that, in the presence of the two SMO mutations detected in our patients, the estimated sensitivity of the receptor for vismodegib is reduced. In particular, for the SMO D473Y mutant isoform, the calculated IC50 value is 159 nM (ΔGbind = -9.27 ± 0.10 kcal/mol), resulting in an almost two order of magnitude lower affinity with respect to the WT protein (IC50 = 2.5 nM, ΔGbind = -11.75 ± 0.31 kcal/mol). Accordingly, a direct effect on drug binding is at play in the presence of the missense D/Y mutation at position 473 of the SMO receptor. In the case of the G497W point substitution, the calculated IC50 for vismodegib is also diminished albeit to a considerable lesser extent: IC50 = 69 nM and ΔGbind = - 9.78 ± 0.10

kcal/mol. Hence, another molecular mechanism of resistance should be at play in the case of primary resistance. Further, advanced simulations revealed that, in the presence of the G497W mutation, the drug pathway to the binding site is quite hampered, due a conformational rearrangement of the protein region surrounding the mutation. So, a higher energy is required to vismodegib to reach the mutant protein binding site, as discussed in detail below.

Discussion

To our knowledge, this study presents for the first clinical, molecular, and in silico evidences of primary and acquired resistance to vismodegib in BCC.

The G497W missense mutation in the SMO receptor characterizes the liver metastasis of the first patient showing a progression of disease after two months of vismodegib treatment. Even if less frequent than PTCH1 mutations in BCC[106], activating SMO mutations cause constitutive ligand-independent signal transduction thus making mutated SMO an oncogene in BCC[107]. In addition to a possible oncogenic function, we hypothesized that the mutation G497W likely represents the molecular mechanism associated with vismodegib primary resistance in BCC. To support this assumption we explored in silico the effects exerted by this mutation on the inhibitor binding to the mutated SMO protein.

Mutation G497W is in a distal position with respect to the protein drug binding site (Fig. 3.5.3). Accordingly, only a moderate direct effect on the protein affinity for the inhibitor can be anticipated. Indeed, in a sort of domino effect, a slight overall readjustment of the binding site residues is observed for the SMOG497W isoform. Accordingly, the estimated free energy of binding ΔGbind and the corresponding IC50(est) value for the SMOG497W/vismodegib complex are 9.78 ± 0.10 kcal/mol and 69 nM., respectively, which are somewhat but not dramatically higher than the corresponding values obtained for the WT/inhibitor complex (ΔGbind = -11.75 ± 0.31 kcal/mol and IC50(est)= 2.5 nM) (Table 3.5.1). Thus, other molecular mechanisms are likely contributing to the primary resistance of SMOG497W to vismodegib. A direct comparison of the WT/mutated protein structures in the area surrounding position 497 revealed that, in the presence of the tryptophan mutant residue, the entire region undergoes a conformational rearrangement, thus resulting in a narrowing of the protein drug entry site. For this reason, in the presence of the G497W mutation vismodegib might be less able to reach the protein binding pocket and, hence, less effective in its inhibitory activity. To support this mechanism we then devised and conducted steered molecular dynamics experiments of the binding process of vismodegib to the WT and SMO G497W isoforms. Figures 3.5.3 B and C show the different course of these binding processes. As we can see, not only quite a higher force, and hence energy, is required to vismodegib to reach its binding site in the presence of

the mutant residue (Figure 3.5.3 B) but also, and perhaps even more interestingly, contrarily to the WT case, for SMOG497W the drug has not yet reached its correct orientation within the protein binding pocket after the same time lag (Figure 3.5.3 C). In practical terms this fundamentally implies that, in time, the effective vismodegib concentration within the SMOG497W binding site is significantly lower than in the case of the WT receptor and this, is our opinion, constitutes a novel and effective mechanism of drug primary resistance.

In this respect, the identification of such a mutation could become a biomarker for vismodegib primary resistance, in a panorama where any predictive marker able to identify patients not responding to this drug is totally absent.

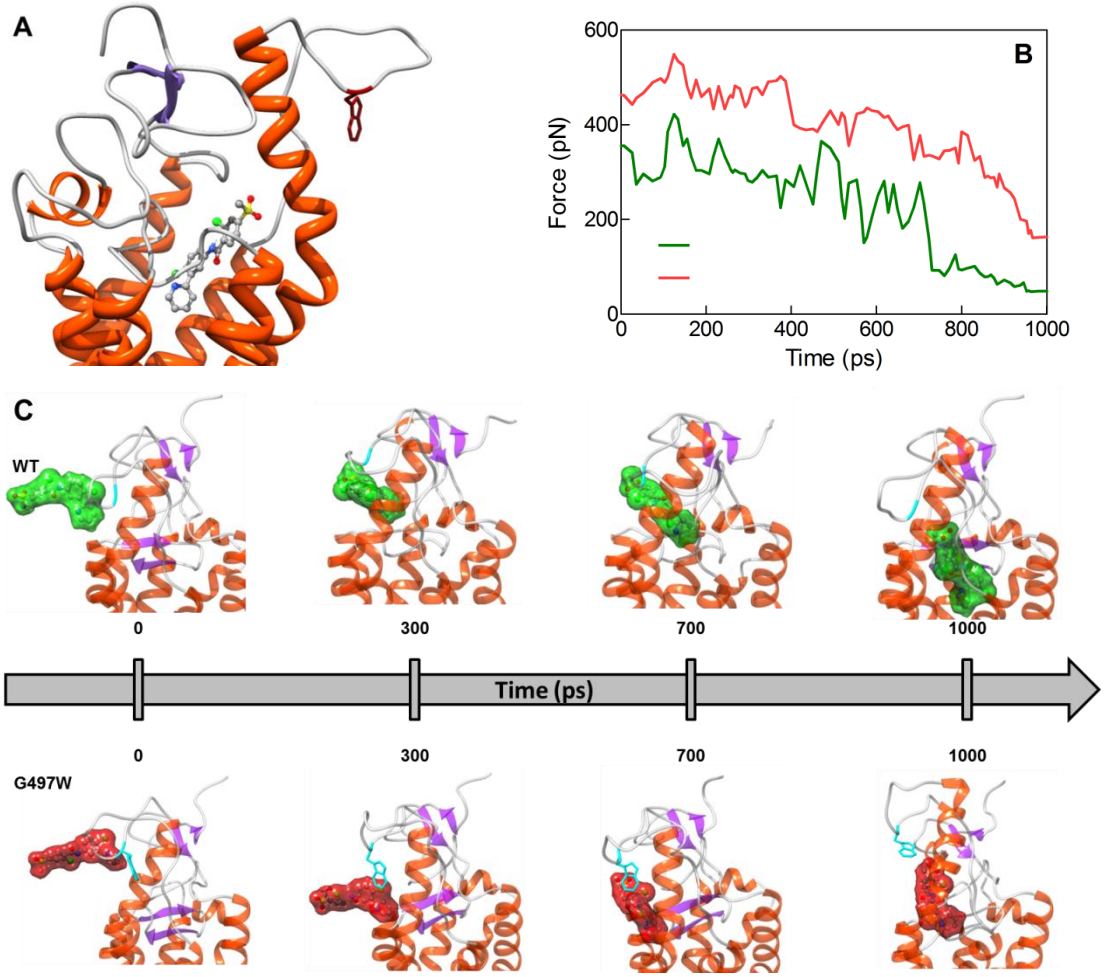


Figure 3.5.3 *(A) Zoomed view of the SMOG497W binding site in complex with vismodegib. The receptor is shown as a secondary-structure colored ribbon (orange, α-helices; purple, β-sheets; gray, coils). Vismodegib is portrayed as atom-colored sticks-and-balls (red, O; blue, N; green, Cl; S, sulfur; gray, C). Residue W497 is evidenced as dark red sticks. (B) Rupture force vs. time during the entry process of vismodegib within the wt (green) and SMOG497W binding site. (C) SMD snapshots of vismodegib entering the receptor binding pocket. The protein is shown using the same color scheme of panel A. Vismodegib is highlighted by its green/red van der Waals surface. In panels A and C, hydrogens atoms, ions, counterions and water molecules are omitted for clarity.*

Concerning the second patient, the PTCH1 inactivating mutation observed in the pre-treatment primary lesion along with absence of SMO mutation could explain the complete clinical response obtained after 6month of continuous vismodegib treatment. Essentially, loss of PTCH1 function by inactivating mutations relieves normal SMO inhibition and leads to the activation of Hh target genes such as GLI1.

Despite the presence of the premature STOP codon in PTCH1, 5 months later the patient developed BCC recurrence. Sequencing of PTCH1 in recurrence samples confirmed the presence of the same nonsense PTCH1 mutation. To characterize the mechanism of relapse we performed SMO sequencing and unveiled the appearance of the nucleotide mutation GAC>TAC coding for the missense D473Y substitution in the SMO transcript. This sequence alteration was not detected in any of the primary lesion samples examined. Notably, another amino acid substitution of the same SMO conserved aspartic acid (D473H) was reported and designated as the mechanism of vismodegib resistance in medulloblastoma[108]. Coherently, in an elegant in vitro study the same group characterized the role of D473 in SMO function by replacing aspartic acid with every other amino acid and showing that all mutant were resistant to vismodegib[109]. Our in silico data of SMO D473Y are in line with these findings and added convincing evidence about the pivotal role played by D473 in drug binding.

Our in silico analysis of the molecular effect of the SMO missense mutation D473Y clearly revealed a direct effect of the mutated residue on vismodegib binding. Indeed, the estimated affinity of the inhibitor for this mutated protein IC50 = 159 nM (ΔGbind = -9.27 ± 0.10 kcal/mol is almost two order of magnitude lower than the one measured/estimated for the WT counterpart (IC50(exp)= 3 nM, IC50(est) = 2.5 nM, respectively) (Table 3.5.1). Importantly, we discovered amino acid D473 is part of a peculiar triad of residues composed, besides D473, by R400 and H470. These triad forms is engaged in a small but fundamental network of hydrogen bonds which keeps the SMO binding pocket in the proper size and shape to tailor-fit vismodegib binding (Figure 3.5.4 A). In the presence of the D473Y mutation, a considerable effect on the binding site geometry is induced, required for the proper accommodation of the residue aromatic side chain, ultimately leading to the total disruption of the stabilizing hydrogen bond network. The inhibitor is hence shifted from its optimal position, the only surviving interaction with the receptor being an intermitted hydrogen bond with Y394 (Figure 3.5.4 B). These evidences are substantiated by the per residue energy deconvolution of the free energy of binding, as shown in Figures 3.5.4 C and 3.5.4 D. As we see, all receptor residues mainly involved in vismodegib binding undergo a severe reduction in the stabilizing contribution to inhibitor binding in the presence of the mutation (Figure 3.5.4 C). Analogously,

upon disruption of the hydrogen bond network among the SMO triad residues, the relevant, favourable contribution to receptor/inhibitor binding plummets drastically (Fig. 3.5.4 D).

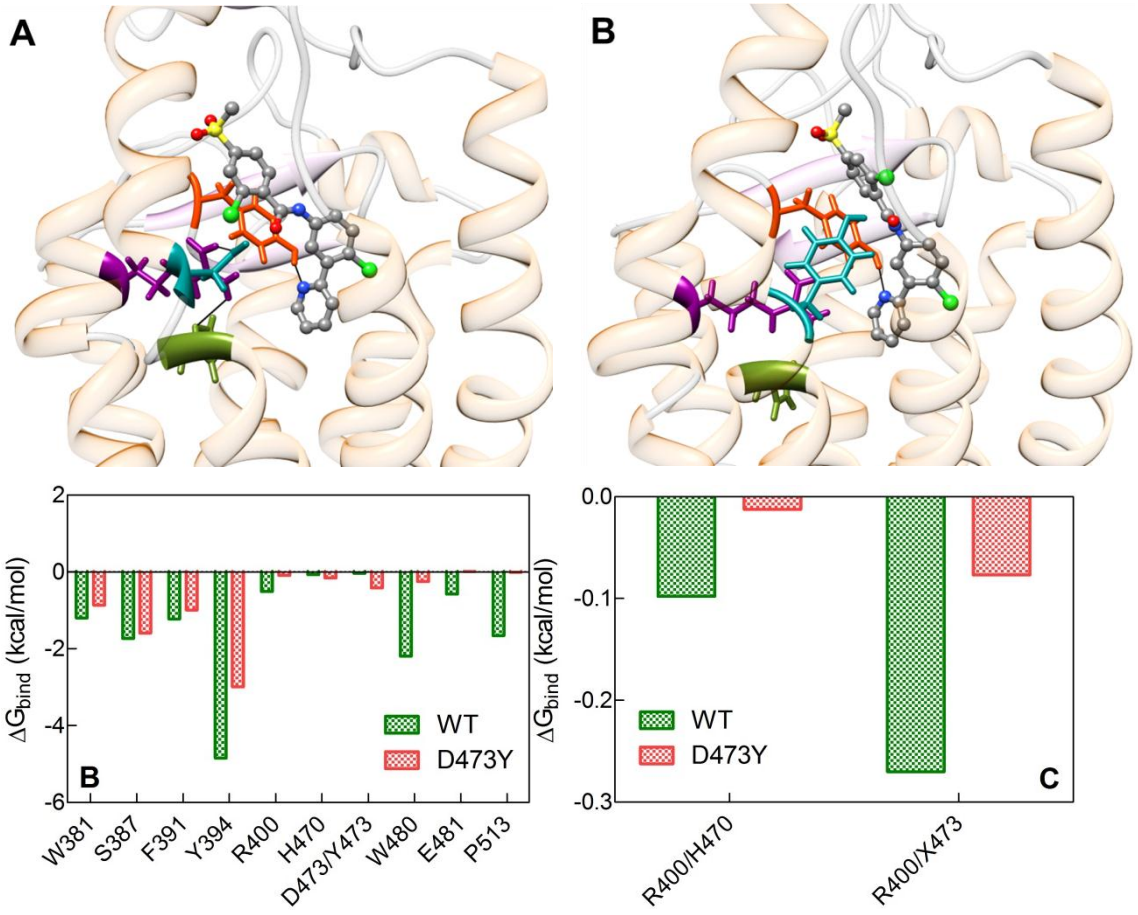


Figure 3.5.4 *Zoomed views of the WT (A) and SMOD473W mutant binding sites in complex with vismodegib. The receptor secondary structure is outlined as a semi-transparent ribbon (orange, α-helices; purple, β-sheets; gray, coils). Vismodegib is portrayed as atom-colored sticks-and-balls (red, O; blue, N; green, Cl; S, sulfur; gray, C). The triad of residues involved in the hydrogen-bond network are highlighted colored sticks: R300, dark magenta; H470, olive drab; D/Y473, dark cyan. Y394 is also shown as dark red sticks. H-bonds are evidenced as black lines. (B) Comparison of vismodegib binding energy contributions from WT (green) and D473Y (red) SMO residues. (C) Comparison between hydrogen bond network stabilization energies for SMO residues belonging to the WT (green) and SMOD473Y triad residues in the relevant vismodegib complexes. X denotes either D or Y residue at position 473.*

Table 3.5.1 *In silico estimation of free energy of binding (DG_{bind}) for WT and SMO^{D473H} mutant receptor in complex with vismodegib. Errors are given in parenthesis as standard errors of the mean (SEM).* [a] *$\Delta\Delta G_{bind} = \Delta G_{bind,wt} - \Delta G_{bind,mutant.}$ By definition, a negative value of $\Delta\Delta G_{bind}$ indicates that the WT/vismodegib complex is favored with respect to the mutated one, and vice versa.*

SMO complex	ΔG_{bind} (kcal/mol)	$\Delta\Delta G_{bind}$ (kcal/mol)	$IC_{50,calc}$ (nM)	$IC_{50,exp}$ (nM)
WT/vismodegib	-11.75 (±0.31)	-	2.5	3
D473Y/vismodegib	-9.27 (±0.10)	-2.48	159	-
W479	-9.78 (±0.10)	-1.97	68	-

Conclusion

Similarly to medulloblastoma, our data further raise the issue of overcoming the resistance related to mutation of the drug target. Under this perspective, efforts aimed at identifying second generation drugs showing potent activity against SMO mutants are mandatory. To this goal, a panel of compounds has already been screened in vitro and several promising antagonists able to inhibit tumor growth mediated by drug-resistant SMOs were selected in murine allograft model of medulloblastoma.

Another strategy worth pursuing is the inhibition of the Hh pathway through other mechanisms, the activity of itraconazole or arsenic oxide in blocking vismodegib-resistant BCC both in vitro and in vivo being prime examples[110].

An alternative approach for combating vismodegib resistance could consist in exploring other signaling pathways implicated in BCC in order to identify new therapeutic targets. Inhibition of the PI3K pathway seems already to be an exciting therapeutic possibility for treating resistant medulloblastoma patients[109,111].

It may be speculated that the emergence of such mutations causing drug-resistance could represent a clonal alteration of all BCC cells and/or it might interest only a fraction of the tumoral cells, thus justifying a prolonged regimen of vismodegib beyond progression, in combination with other therapeutic strategies. Future molecular studies on tumor heterogeneity are currently in progress in order to answer this question.

Considering that median duration of response to vismodegib in previous clinical trial was 7.6 months, the issue of acquired resistance is an important point to debate in order to exploit the activity of this drug to its full potential. In particular, it has been proposed that other

therapeutic schedules (i.e,. intermittent dosing) may alter resistance mechanism development[111]. Moreover, different schedules of administration might help in increasing the compliance, as in previous clinical experiences a patient percentage varying from 12% to 54% discontinued vismodegib due to the occurrence of adverse events[112]

There is also much debate about BCC rebound after vismodegib cessation for adverse events, with conflicting reports supporting both evidences. However, in patients with basal-cell nevus syndrome assuming vismodegib the rate of new BCC development after inhibitor cessation for adverse events was slower than in control group receiving placebo. This could imply the absence of mutation occurrence without drug pressure, at variance with tyrosine kinase inhibitors, sometimes experiencing disease flare after treatment cessation.

Our 2 cases of vismodegib resistance occurred in patients who had not received previously radiation therapy. It is possible that Hh inhibitor resistance outbreak could be higher in patients having received a radiation pressure, because of the increased susceptibility of cancer cells to genetic mutations.

In this report we discussed 2 different SMO mutations able to explain primary and secondary resistance to vismodegib developed in two distinct BCC cases. We are aware that several possible, alternativemechanisms can be at play to evade SMO inhibition at different levels, such as abnormal downstream pathway activation, with Gli2 or Cyclin D1 overexpression, or upregulation of PI3K-AKT-mTOR signaling pathway. However, the screening for the reported SMO mutations and the search for new therapeutic strategies to overcome these resistance mechanisms should constitute one fundamental step in for a better exploitation of Hh inhibitors in the near future.

Conclusions

In the late ’90s the discovery of a drug would have changed the world scenario on the approach to treat cancer; the ATP antagonist imatinib has been the first case of targeted therapy against chronic mieloyd leukemia (CML). This success triggered a series of studies that profoundly changed our understanding not only of cancer therapy but also of the classification of cancers, since the same drug, due to its intrinsic mechanism of action, beside inhibiting tyrosine kinase activity, was also found capable of inhibiting other enzymatic activities, such as those expressed by several RTKs including KIT, PDGFRA and PDGFRB.

However a couple of year later, even if, till today, it is the first line therapy for different kind of cancer, a target-associated resistance has also been observed; the gatekeeper residue of BCR-ABL1, T315, relapsed on treatment. A similar case was observed in a casa of melanoma treated by a RAF inhibitor, vemurafenib; a young man with BRAF-mutant melanoma, refractory to several therapeutic regimens, was treated with vemurafenib. After 16 weeks on the drug, the patient experienced widespread disease relapse, which by 23 weeks involved most previous sites of visceral and subcutaneous disease; vemurafenib was discontinued. Why this?

A detailed understanding of the mechanisms by which tumors aquire resistance to targeted anticancer agents should speed the development of treatment strategies with lasting clinical efficacy.

In this work we evaluated the possibility of applying molecular simulation techniques to understand mechanism behind resistance and identify new selective target.

In fact the systematic approach of molecular modeling and simulation in the selection of new drugs, mostly, replace the biologically based, time-consuming, large-scale screening of molecules, and become a very useful tool for the rapid ranking and selection of a smaller set of molecules to be tested in the laboratory by traditional functional assays, as we demonstrated for the development of selective ligands for σ_1 receptor. This new, in silico approach will undoubtedly contribute to paving the way towards truly personalized treatment in the foreseeable future, and we are confident that, through a strong interaction between in silico techniques and high throughput technologies, we will be able to design multi-drug approaches which are effective on the treatment of different kind of cancer.

Targeted therapies require new approaches to determine optimal dosing, to assess patient adherence to therapy, and to evaluate treatment effectiveness. The cost of these agents, which can exceed several thousand dollars per month, may become an important issue in health care economics.

Finally, our work on targeted therapy in GIST, BCC and desmoid tumors, particularly as regards the understanding the mechanism of the ‘‘resistance’’ phenomenon, has resulted in a novel multidimensional approach, which is also supported by an in silico dimension, to predict clinical results.

In fact, thanks to the availability of the crystallographic coordinates of some receptors and having available a well validate homology modeling procedure, all the relevant 3D structure could be determined. As shown in this thesis, thanks to molecular modeling – a combination of computer-based simulation techniques on the basis of thermodynamic parameters which are calculated on a three-dimensional protein model derived either from crystallography or by homology techniques – we are able to estimate the affinity of a given inhibitor towards its target receptor with great accuracy. Likewise, in the case of acquired resistance, the sensitivity and affinity of these drugs for the ‘‘resistant receptor’’ can be predicted in the same way.

One important aspect that has to be taken in consideration is the structural alterations induced by these mutations that can confer resistance to the drug. In particular, in the case of RTK inhibitors two distinct pathways could be indentified: i) a perturbation in the general conformation of the ATP pocket or ii) a transition from the closed form toward the activated form. The molecular mechanisms underlying these two pathways has been analyzed in this thesis to a finer level using molecular modeling and studying the energy contribution of the different structure for the affinity with drug or other protein.

Briefly, the last two decades have witnessed enormous progress in our knowledge and understanding of the molecular basis of neoplastic transformations. The complete sequencing of the human genome has represented one milestone on this path and, as it was anticipated at the beginning of the genome project, the completion of this enterprise brought particularly significant advancements to our understanding of cancer. Analyzing these cases we can point out the efficacy of a practice which leads us from bed to bench and back, i.e. the route which moves from the patient’s molecular profile to the modeling of the molecular target and the examination of its data, the identification of the best available drug, the evaluation of clinic response and the possibility, even in the presence of relapse and thus of resistance to the treatment, of restarting along this virtuous circle moving from the molecular profiling this time of the reactivated tumor which developed a secondary resistance, to new modeling and data mining which in turn lead to the design of a new treatment with a new drug.

The combination of computer-based molecular modeling and parallel assessment of biochemical and biological properties by experimental methods can open new avenue in this field of scientific research by providing powerful tools for rationale design of new drugs and to

help the clinician in finding the best treatment related to its personal genetic profile, a kind of “ad personam” therapy.

Bibliography

1 Laurini, E. *et al.* Homology Model and Docking-Based Virtual Screening for Ligands of the σ1 Receptor. *ACS Medicinal Chemistry Letters* **2**, 834-839, doi:10.1021/ml2001505 (2011).

2 Totowa, N. J. *Protein Structure Prediction: Methods and Protocols*. Davis M. Webster edn, (HUMANA PRESS 2000).

3 Evers, A., Hessler, G., Matter, H. & Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *Journal of medicinal chemistry* **48**, 5448-5465, doi:10.1021/jm050090o (2005).

4 Laurini, E. *et al.* A 3D-pharmacophore model for sigma2 receptors based on a series of substituted benzo[d]oxazol-2(3H)-one derivatives. *Bioorganic & medicinal chemistry letters* **20**, 2954-2957, doi:10.1016/j.bmcl.2010.03.009 (2010).

5 Seth, P. *et al.* Expression pattern of the type 1 sigma receptor in the brain and identity of critical anionic amino acid residues in the ligand-binding domain of the receptor. *Biochimica et biophysica acta* **1540**, 59-67 (2001).

6 Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate−DNA Helices. *Journal of the American Chemical Society* **120**, 9401-9409, doi:10.1021/ja981844+ (1998).

7 Pal, A. *et al.* Identification of regions of the sigma-1 receptor ligand binding site using a novel photoprobe. *Molecular pharmacology* **72**, 921-933, doi:10.1124/mol.107.038307 (2007).

8 de Costa, B. R. *et al.* Synthesis and evaluation of optically pure [3H]-(+)-pentazocine, a highly potent and selective radioligand for σ receptors. *FEBS Letters* **251**, 53-58, doi:http://dx.doi.org/10.1016/0014-5793(89)81427-9 (1989).

9 Moebius, F. F., Striessnig, J. & Glossmann, H. The mysteries of sigma receptors: new family members reveal a role in cholesterol synthesis. *Trends in pharmacological sciences* **18**, 67-70 (1997).

10 Laggner, C. *et al.* Discovery of high-affinity ligands of sigma1 receptor, ERG2, and emopamil binding protein by pharmacophore modeling and virtual screening. *Journal of medicinal chemistry* **48**, 4754-4764, doi:10.1021/jm049073+ (2005).

11 Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* **31**, 3784-3788 (2003).

12 Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291, doi:doi:10.1107/S0021889892009944 (1993).

13 Vriend, G. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56, doi:10.1016/0263-7855(90)80070-v (1990).

14 Sippl, M. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J Computer-Aided Mol Des* **7**, 473-501, doi:10.1007/bf02337562 (1993).

15 AMBER11 (University of California, San Francisco, 2010).

16 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:doi:http://dx.doi.org/10.1063/1.445869 (1983).

17 Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999-2012, doi:10.1002/jcc.10349 (2003).

18 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.

Journal of Computational Physics **23**, 327-341, doi:http://dx.doi.org/10.1016/0021-9991(77)90098-5 (1977).

19 Toukmaji, A., Sagui, C., Board, J. & Darden, T. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *The Journal of Chemical Physics* **113**, 10913-10927, doi:doi:http://dx.doi.org/10.1063/1.1324708 (2000).

20 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12-21, doi:doi:10.1107/S0907444909042073 (2010).

21 Luthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85, doi:10.1038/356083a0 (1992).

22 Cobos, E. J., Entrena, J. M., Nieto, F. R., Cendan, C. M. & Del Pozo, E. Pharmacology and therapeutic potential of sigma(1) receptor ligands. *Current neuropharmacology* **6**, 344-366, doi:10.2174/157015908787386113 (2008).

23 Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785-2791, doi:10.1002/jcc.21256 (2009).

24 Giliberti, G. *et al.* Synergistic experimental/computational studies on arylazoenamine derivatives that target the bovine viral diarrhea virus RNA-dependent RNA polymerase. *Bioorganic & medicinal chemistry* **18**, 6055-6068, doi:10.1016/j.bmc.2010.06.065 (2010).

25 Tonelli, M. *et al.* Antiviral activity of benzimidazole derivatives. II. Antiviral activity of 2-phenylbenzimidazole derivatives. *Bioorganic & medicinal chemistry* **18**, 2937-2953, doi:10.1016/j.bmc.2010.02.037 (2010).

26 Mehler, E. L. & Solmajer, T. Electrostatic effects in proteins: comparison of dielectric and charge models. *Protein engineering* **4**, 903-910 (1991).

27 Feig, M. *et al.* Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comput Chem* **25**, 265-284, doi:10.1002/jcc.10378 (2004).

28 Frecer, V., Kabelac, M., De Nardi, P., Pricl, S. & Miertus, S. Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. *Journal of molecular graphics & modelling* **22**, 209-220, doi:10.1016/S1093-3263(03)00161-X (2004).

29 Laurini, E. *et al.* Another brick in the wall. Validation of the sigma1 receptor 3D model by computer-assisted design, synthesis, and activity of new sigma1 ligands. *Molecular pharmaceutics* **9**, 3107-3126, doi:10.1021/mp300233y (2012).

30 Dileo, P. *et al.* Imatinib response in two GIST patients carrying two hitherto functionally uncharacterized PDGFRA mutations: an imaging, biochemical and molecular modeling study. *International journal of cancer. Journal international du cancer* **128**, 983-990, doi:10.1002/ijc.25418 (2011).

31 Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nature structural biology* **7**, 537-541, doi:10.1038/76744 (2000).

32 Sulea, T. & Purisima, E. O. Optimizing Ligand Charges for Maximum Binding Affinity. A Solvated Interaction Energy Approach. *The Journal of Physical Chemistry B* **105**, 889-899, doi:10.1021/jp0038714 (2001).

33 Kangas, E. & Tidor, B. Optimizing electrostatic affinity in ligand–receptor binding: Theory, computation, and ligand properties. *The Journal of Chemical Physics* **109**, 7522-7545, doi:doi:http://dx.doi.org/10.1063/1.477375 (1998).

34 Lee, L. P. & Tidor, B. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein science : a publication of the Protein Society* **10**, 362-377, doi:10.1110/ps.40001 (2001).

35 Gohlke, H., Kiel, C. & Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *Journal of Molecular Biology* **330**, 891-913, doi:http://dx.doi.org/10.1016/S0022-2836(03)00610-7 (2003).

36 Massova, I. & Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society* **121**, 8133-8143, doi:10.1021/ja990935j (1999).

37 Huo, S., Massova, I. & Kollman, P. A. Computational alanine scanning of the 1:1 human growth hormone–receptor complex. *Journal of Computational Chemistry* **23**, 15-27, doi:10.1002/jcc.1153 (2002).

38 Palmer, C. P., Mahen, R., Schnell, E., Djamgoz, M. B. & Aydar, E. Sigma-1 receptors bind cholesterol and remodel lipid rafts in breast cancer cell lines. *Cancer research* **67**, 11166-11175, doi:10.1158/0008-5472.CAN-07-1771 (2007).

39 Zampieri, D. *et al.* Synthesis, biological evaluation, and three-dimensional in silico pharmacophore model for sigma(1) receptor ligands based on a series of substituted benzo[d]oxazol-2(3H)-one derivatives. *Journal of medicinal chemistry* **52**, 5380-5393, doi:10.1021/jm900366z (2009).

40 Zampieri, D. *et al.* Substituted benzo[d]oxazol-2(3H)-one derivatives with preference for the sigma1 binding site. *European journal of medicinal chemistry* **44**, 124-130, doi:10.1016/j.ejmech.2008.03.011 (2009).

41 Meyer, C. *et al.* Pd-catalyzed direct C-H bond functionalization of spirocyclic sigma1 ligands: generation of a pharmacophore model and analysis of the reverse binding mode by docking into a 3D homology model of the sigma1 receptor. *Journal of medicinal chemistry* **55**, 8047-8065, doi:10.1021/jm300894h (2012).

42 Schlager, T., Schepmann, D., Wurthwein, E. U. & Wunsch, B. Synthesis and structure-affinity relationships of novel spirocyclic sigma receptor ligands with furopyrazole structure. *Bioorganic & medicinal chemistry* **16**, 2992-3001, doi:10.1016/j.bmc.2007.12.045 (2008).

43 Schlager, T. *et al.* Combination of two pharmacophoric systems: synthesis and pharmacological evaluation of spirocyclic pyranopyrazoles with high sigma(1) receptor affinity. *Journal of medicinal chemistry* **54**, 6704-6713, doi:10.1021/jm200585k (2011).

44 Meyer, C. *et al.* Exploitation of an additional hydrophobic pocket of sigma1 receptors: late-stage diverse modifications of spirocyclic thiophenes by C-H bond functionalization. *Organic & biomolecular chemistry* **9**, 8016-8029, doi:10.1039/c1ob06149f (2011).

45 Oberdorf, C., Schmidt, T. J. & Wunsch, B. 5D-QSAR for spirocyclic sigma1 receptor ligands by Quasar receptor surface modeling. *European journal of medicinal chemistry* **45**, 3116-3124, doi:10.1016/j.ejmech.2010.03.048 (2010).

46 Maier, C. A. & Wunsch, B. Novel spiropiperidines as highly potent and subtype selective sigma-receptor ligands. Part 1. *Journal of medicinal chemistry* **45**, 438-448 (2002).

47 Grosse Maestrup, E. *et al.* Synthesis of spirocyclic sigma1 receptor ligands as potential PET radiotracers, structure-affinity relationships and in vitro metabolic stability. *Bioorganic & medicinal chemistry* **17**, 3630-3641, doi:10.1016/j.bmc.2009.03.060 (2009).

48 Glennon, R. A. *et al.* Structural features important for sigma 1 receptor binding. *Journal of medicinal chemistry* **37**, 1214-1219 (1994).

49 Laurini, E., Da Col, V., Wunsch, B. & Pricl, S. Analysis of the molecular interactions of the potent analgesic S1RA with the sigma1 receptor. *Bioorganic & medicinal chemistry letters* **23**, 2868-2871, doi:10.1016/j.bmcl.2013.03.087 (2013).

50 Ablordeppey, S. Y., Fischer, J. B., Law, H. & Glennon, R. A. Probing the proposed phenyl-A region of the sigma-1 receptor. *Bioorganic & medicinal chemistry* **10**, 2759-2765 (2002).

51 Wunsch, B. The sigma(1) receptor antagonist S1RA is a promising candidate for the treatment of neurogenic pain. *Journal of medicinal chemistry* **55**, 8209-8210, doi:10.1021/jm3011993 (2012).

52 Pierotti, M. A., Tamborini, E., Negri, T., Pricl, S. & Pilotti, S. Targeted therapy in GIST: in silico modeling for prediction of resistance. *Nature reviews. Clinical oncology* **8**, 161-170, doi:10.1038/nrclinonc.2011.3 (2011).

53 Rossi, D. *et al.* Chemical, pharmacological, and in vitro metabolic stability studies on enantiomerically pure RC-33 compounds: promising neuroprotective agents acting as sigma(1) receptor agonists. *ChemMedChem* **8**, 1514-1527, doi:10.1002/cmdc.201300218 (2013).
54 Rossi, D. *et al.* Design, synthesis and SAR analysis of novel selective sigma1 ligands (Part 2). *Bioorganic & medicinal chemistry* **18**, 1204-1212, doi:10.1016/j.bmc.2009.12.039 (2010).
55 Rossi, D. *et al.* Identification of a potent and selective sigma(1) receptor agonist potentiating NGF-induced neurite outgrowth in PC12 cells. *Bioorganic & medicinal chemistry* **19**, 6210-6224, doi:10.1016/j.bmc.2011.09.016 (2011).
56 Rossi, D. *et al.* Identification of RC-33 as a potent and selective sigma1 receptor agonist potentiating NGF-induced neurite outgrowth in PC12 cells. Part 2: g-scale synthesis, physicochemical characterization and in vitro metabolic stability. *Bioorganic & medicinal chemistry* **21**, 2577-2586, doi:10.1016/j.bmc.2013.02.029 (2013).
57 Agranat, I., Caner, H. & Caldwell, J. Putting chirality to work: the strategy of chiral switches. *Nature reviews. Drug discovery* **1**, 753-768, doi:10.1038/nrd915 (2002).
58 Simard-Mercier, J., Jiang, J. L., Ho, M. L., Flynn, A. B. & Ogilvie, W. W. Single-isomer trisubstituted olefins from a novel reaction of (E)-beta-chloro-alpha-iodo-alpha,beta-unsaturated esters and amides. *The Journal of organic chemistry* **73**, 5899-5906, doi:10.1021/jo8008325 (2008).
59 Mancuso, R. *et al.* Sigma-1R agonist improves motor function and motoneuron survival in ALS mice. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics* **9**, 814-826, doi:10.1007/s13311-012-0140-y (2012).
60 Woolf, T. B. & Roux, B. Structure, energetics, and dynamics of lipid-protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins* **24**, 92-114, doi:10.1002/(SICI)1097-0134(199601)24:1<92::AID-PROT7>3.0.CO;2-Q (1996).
61 Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one* **2**, e880, doi:10.1371/journal.pone.0000880 (2007).
62 Skjevik, A. A., Madej, B. D., Walker, R. C. & Teigen, K. LIPID11: a modular framework for lipid simulations using amber. *The journal of physical chemistry. B* **116**, 11124-11136, doi:10.1021/jp3059992 (2012).
63 Amber12 (University of California, San Francisco, 2012).
64 Onufriev, A., Bashford, D. & Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B* **104**, 3712-3720, doi:10.1021/jp994072s (2000).
65 Kahoun, J. R. & Ruoho, A. E. (125I)iodoazidococaine, a photoaffinity label for the haloperidol-sensitive sigma receptor. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 1393-1397 (1992).
66 Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives*. (Wiley, 2006).
67 Lehn, J. M. Toward self-organization and complex matter. *Science* **295**, 2400-2403, doi:10.1126/science.1071063 (2002).
68 Yu, T. *et al.* An Amphiphilic Dendrimer for Effective Delivery of Small Interfering RNA and Gene Silencing In Vitro and In Vivo. *Angewandte Chemie International Edition* **51**, 8478-8484, doi:10.1002/anie.201203920 (2012).
69 Zhou, J. *et al.* Systemic administration of combinatorial dsiRNAs via nanoparticles efficiently suppresses HIV-1 infection in humanized mice. *Molecular therapy : the journal of the American Society of Gene Therapy* **19**, 2228-2238, doi:10.1038/mt.2011.207 (2011).
70 Davidson, B. L. & McCray, P. B., Jr. Current prospects for RNA interference-based therapies. *Nature reviews. Genetics* **12**, 329-340, doi:10.1038/nrg2968 (2011).

71 Whitehead, K. A., Langer, R. & Anderson, D. G. Knocking down barriers: advances in siRNA delivery. *Nature reviews. Drug discovery* **8**, 129-138, doi:10.1038/nrd2742 (2009).
72 Tseng, Y. C., Mozumdar, S. & Huang, L. Lipid-based systemic delivery of siRNA. *Advanced drug delivery reviews* **61**, 721-731, doi:10.1016/j.addr.2009.03.003 (2009).
73 Schroeder, A., Levins, C. G., Cortez, C., Langer, R. & Anderson, D. G. Lipid-based nanotherapeutics for siRNA delivery. *Journal of internal medicine* **267**, 9-21, doi:10.1111/j.1365-2796.2009.02189.x (2010).
74 Wagner, E. Polymers for siRNA delivery: inspired by viruses to be targeted, dynamic, and precise. *Accounts of chemical research* **45**, 1005-1013, doi:10.1021/ar2002232 (2012).
75 Israelachvili, J. N., Mitchell, D. J. & Ninham, B. W. Theory of self-assembly of lipid bilayers and vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **470**, 185-201, doi:http://dx.doi.org/10.1016/0005-2736(77)90099-2 (1977).
76 Nagarajan, R. Molecular Packing Parameter and Surfactant Self-Assembly: The Neglected Role of the Surfactant Tail†. *Langmuir* **18**, 31-38, doi:10.1021/la010831y (2001).
77 Jones, S. P. *et al.* Hydrophobically modified dendrons: developing structure-activity relationships for DNA binding and gene transfection. *Molecular pharmaceutics* **8**, 416-429, doi:10.1021/mp100260c (2011).
78 Goldblum JR, F. J. Ch. Desmoid-type fibromatoses., 83-84 (IARC Press, 2002).
79 Lazar, A. J. *et al.* Specific mutations in the beta-catenin gene (CTNNB1) correlate with local recurrence in sporadic desmoid tumors. *The American journal of pathology* **173**, 1518-1527, doi:10.2353/ajpath.2008.080475 (2008).
80 Colombo, C. *et al.* CTNNB1 45F mutation is a molecular prognosticator of increased postoperative primary desmoid tumor recurrence: an independent, multicenter validation study. *Cancer* **119**, 3696-3702, doi:10.1002/cncr.28271 (2013).
81 Willert, K. & Nusse, R. Beta-catenin: a key mediator of Wnt signaling. *Current opinion in genetics & development* **8**, 95-102 (1998).
82 Liu, F. & Millar, S. E. Wnt/beta-catenin signaling in oral tissue development and disease. *Journal of dental research* **89**, 318-330, doi:10.1177/0022034510363373 (2010).
83 Xing, Y. *et al.* Crystal Structure of a Full-Length ²-Catenin. *Structure (London, England : 1993)* **16**, 478-487 (2008).
84 Megy, S. *et al.* STD and TRNOESY NMR studies for the epitope mapping of the phosphorylation motif of the oncogenic protein β-catenin recognized by a selective monoclonal antibody. *FEBS Letters* **580**, 5411-5422, doi:http://dx.doi.org/10.1016/j.febslet.2006.08.084 (2006).
85 Mynott, A. V. *et al.* Crystal structure of importin-α bound to a peptide bearing the nuclear localisation signal from chloride intracellular channel protein 4. *FEBS Journal* **278**, 1662-1675, doi:10.1111/j.1742-4658.2011.08086.x (2011).
86 de Vries, S. J., van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nature protocols* **5**, 883-897, doi:10.1038/nprot.2010.32 (2010).
87 Cohen, P. & Frame, S. The renaissance of GSK3. *Nature reviews. Molecular cell biology* **2**, 769-776 (2001).
88 Wang, W. & Kollman, P. A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *Journal of Molecular Biology* **303**, 567-582, doi:http://dx.doi.org/10.1006/jmbi.2000.4057 (2000).
89 Wan, S. & Coveney, P. V. Regulation of JAK2 Activation by Janus Homology 2: Evidence from Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **52**, 2992-3000, doi:10.1021/ci300308g (2012).
90 Liang, S. *et al.* Exploring the Molecular Design of Protein Interaction Sites with Molecular Dynamics Simulations and Free Energy Calculations†. *Biochemistry* **48**, 399-414, doi:10.1021/bi8017043 (2008).

91 Conca, E. *et al.* Are two better than one? A novel double-mutant KIT in GIST that responds to Imatinib. *Molecular oncology* **7**, 756-762, doi:10.1016/j.molonc.2013.02.019 (2013).
92 Fletcher, C. D. *et al.* Diagnosis of gastrointestinal stromal tumors: A consensus approach. *Human pathology* **33**, 459-465 (2002).
93 Conca, E. *et al.* Activate and resist: L576P-KIT in GIST. *Molecular cancer therapeutics* **8**, 2491-2495, doi:10.1158/1535-7163.MCT-09-0662 (2009).
94 Tamborini, E. *et al.* Functional analyses and molecular modeling of two c-Kit mutations responsible for imatinib secondary resistance in GIST patients. *Oncogene* **25**, 6140-6146, doi:10.1038/sj.onc.1209639 (2006).
95 McAuliffe, J. C. *et al.* Unlucky number 13? Differential effects of KIT exon 13 mutation in gastrointestinal stromal tumors. *Molecular oncology* **2**, 161-163, doi:10.1016/j.molonc.2008.05.002 (2008).
96 Woodman, S. E. *et al.* Activity of dasatinib against L576P KIT mutant melanoma: molecular, cellular, and clinical correlates. *Molecular cancer therapeutics* **8**, 2079-2085, doi:10.1158/1535-7163.MCT-09-0459 (2009).
97 Negri, T. *et al.* T670X KIT mutations in gastrointestinal stromal tumors: making sense of missense. *Journal of the National Cancer Institute* **101**, 194-204, doi:10.1093/jnci/djn477 (2009).
98 Miranda, C. *et al.* KRAS and BRAF mutations predict primary resistance to imatinib in gastrointestinal stromal tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 1769-1776, doi:10.1158/1078-0432.CCR-11-2230 (2012).
99 Tamborini, E. *et al.* A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. *Gastroenterology* **127**, 294-299 (2004).
100 Rivera, R. S. *et al.* C-kit protein expression correlated with activating mutations in KIT gene in oral mucosal melanoma. *Virchows Archiv : an international journal of pathology* **452**, 27-32, doi:10.1007/s00428-007-0524-2 (2008).
101 Wardelmann, E. *et al.* Polyclonal evolution of multiple secondary KIT mutations in gastrointestinal stromal tumors under treatment with imatinib mesylate. *Clinical cancer research : an official journal of the American Association for Cancer Research* **12**, 1743-1749, doi:10.1158/1078-0432.CCR-05-1211 (2006).
102 Chen, L. L. *et al.* A missense mutation in KIT kinase domain 1 correlates with imatinib resistance in gastrointestinal stromal tumors. *Cancer research* **64**, 5913-5919, doi:10.1158/0008-5472.CAN-04-0085 (2004).
103 Antonescu, C. R. The GIST paradigm: lessons for other kinase-driven cancers. *The Journal of pathology* **223**, 251-261, doi:10.1002/path.2798 (2011).
104 Heinrich, M. C. *et al.* Primary and secondary kinase genotypes correlate with the biological and clinical activity of sunitinib in imatinib-resistant gastrointestinal stromal tumor. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 5352-5359, doi:10.1200/JCO.2007.15.7461 (2008).
105 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).
106 Reifenberger, J. *et al.* Somatic mutations in the PTCH, SMOH, SUFUH and TP53 genes in sporadic basal cell carcinomas. *The British journal of dermatology* **152**, 43-51, doi:10.1111/j.1365-2133.2005.06353.x (2005).
107 Xie, J. *et al.* Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90-92, doi:10.1038/34201 (1998).

108 Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572-574, doi:10.1126/science.1179386 (2009).
109 Dijkgraaf, G. J. *et al.* Small molecule inhibition of GDC-0449 refractory smoothened mutants and downstream mechanisms of drug resistance. *Cancer research* **71**, 435-444, doi:10.1158/0008-5472.CAN-10-2876 (2011).
110 Kim, J. *et al.* Itraconazole and arsenic trioxide inhibit Hedgehog pathway activation and tumor growth associated with acquired resistance to smoothened antagonists. *Cancer cell* **23**, 23-34, doi:10.1016/j.ccr.2012.11.017 (2013).
111 Buonamici, S. *et al.* Interfering with resistance to smoothened antagonists by inhibition of the PI3K pathway in medulloblastoma. *Science translational medicine* **2**, 51ra70, doi:10.1126/scitranslmed.3001599 (2010).
112 Sekulic, A. *et al.* Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *The New England journal of medicine* **366**, 2171-2179, doi:10.1056/NEJMoa1113713 (2012).

ACKNOWLEDGEMENTS

I would like to thank all the people that helped me during these three years and, in particular, of course, my supervisor Prof. Sabrina Pricl, my tutor Prof. Maurizio Fermeglia and all the people of MOSE lab: Erik, Domenico, Francesca and Paola.

I, really, appreciated the scientific collaboration with Istituto dei Tumori of Milan, in particular with Dr. Elena Tamborini, Dr. Federica Perrone, Dr. Silvana Pilotti, Dr. Paolo Bossi and Dr. Marco A. Pierotti and I thank for the financial support of my studies.

I am also very grateful to Dr. Ling Peng that gave me the possibility to spend a period in her group in Marseille and contribute to improve my knowledge in medicinal chemistry.

Last but not least I would thank Prof. Bernhard Wunsch and all his collaborators to our joint efforts in the enigmatic field of sigma receptors.