UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

DIPARTIMENTO DI CHIMICA

Doctoral School in Industrial Chemistry - XXXV Cycle

Computer-aided Molecular Design and Modeling of Tubulin Targeting Agents

Helena Pérez Peña

Supervisor: Prof. Stefano Pieraccini

Co-supervisors: Prof. Alexandre Varnek

Dr. Dragos Horvath

— Academic year 2022/2023 —

This PhD takes part of Marie Curie Actions. It was funded by the H2020-MSCA-ITN-European Joint Doctorate "*Tuning Tubulin Dynamics and Interactions to Face Neurotoxicity: a Multidisciplinary Approach for Training and Research*", named with the acronym **TubInTrain,** Grant agreement n°: 860070.

UNIVERSITÀ DEGLI STUDI DI MILANO

Université de Strasbourg

www.tubintrain.eu

MARIE CURIE ACTIONS

Acknowledgments

I would like to begin by expressing my immense gratitude to the person who provided me with the opportunity to pursue this Ph.D., my main supervisor, Prof. Stefano Pieraccini. Without him, I would not have had the chance to be part of the TubInTrain program and embark on this amazing experience. I appreciate his endless support during difficult times, as well as his shared happiness in celebrating successes together.

The second person who has played an essential role in the development of this Ph.D. thesis, and to whom I am infinitely grateful, is my second supervisor, Dragos. Words cannot adequately describe his unique personality, he is unlike anyone else, and there is never a dull moment when he is around. I am incredibly thankful to him for his invaluable scientific and personal guidance.

I also want to express my gratitude to Prof. Alexandre Varnek, who has always warmly welcomed me into his laboratory, and all the staff from the Chemoinformatics laboratory in Strasbourg, where I have experienced some of the best moments during my Ph.D. Thank you for constantly following our work and giving us great feedback and guidance.

From TubInTrain, special thank you to Prof. Passarella who was the person that made all this happen. Thank you for believing in me, for always cheering me up and appreciating my work. Benedetta, my dear, you are also a special one. Thank you for all the support during these three years. Future Dr. Sai Prashanth Santhapuri, you have passed from being a TubInTrain colleague, to a brother in a matter of a day. I am sure we will be friends for life after this Ph.D, I cannot thank you enough for everything. Zlata, my partner in crime, my house mate, my work buddy, my friend, always working hand in hand on whatever it took. I am super grateful for having had you by my side during this ride. Annie, the colleague I always wanted to have and the friend I am grateful to have, I appreciate the memorable times we shared, particularly the occasions when you took us out in Milan. Maxim, my computational buddy, probably the most helpful, selfless and kind person I have met in these three years. Thank you for always being there for me not matter what.

From Milan, all the members from Passalab, thank you for hosting me and including me in everything you did like if I was another lab member. Guille, gracias por estar a mi lado estos tres años, ayudándome siempre, eres una persona maravillosa.

My secondment in Strasbourg changed my life in so many ways and all thank to the people I was surrounded by that were amazing in all senses. William, you have to go first in this list, if it was not for you, my experience in Strasbourg would have probably been dramatically different. I deeply want to thank you buddy, you are amazing. My twins, Karina and Regina, Tagir, Yuliana, thank you for becoming such an important part of my life.

A special thanks to my family, since they always believed in me and supported me all the way. At my young age, I have lived in five different countries during my studies accumulating a total of five years living abroad, which is not always easy for a parent to deal with. Nonetheless, my parents always motivated me to do what is best for my future, prioritizing my personal development to their willingness to have me close to them in Spain. Thank you to my brother for supporting me during these university years. Gracias abuela, por ser mi fan número uno, te dedico esta tesis de doctorado en gran parte a ti, quien día a día me da una razón para seguir esforzándome y dando lo mejor de mí siempre. Nuestro mantra: poquito a poco se llega lejos.

Thank you to my childhood friends because no matter where I am, how much I travel, how absent I am, that every time I come back home, they are there for me regardless of anything.

A special thanks to my most beautiful coincidence during these three years of doctoral studies, the person who has been by my side during all the struggles of the last period of my Ph.D. and celebrating my successes, simplifying my life and making it much better, encouraging me to become a better version of myself every day, someone I look up to, my other half, Dr. Jean Joseph. Je t'aime.

Lastly, I want to thank myself for putting in the hard work and effort on a daily basis, never giving up and giving 110% of myself to this Ph.D. and everything related to it. Despite the adversities of starting the Ph.D. during the pandemic in Milan and the resulting mental health effects, constant changes of countries, continuous adaptation, creating a family wherever I went, and having to leave them each time I relocated, I have arrived at the moment of presenting the Ph.D. thesis to the world, so the scientific community can benefit from it.

I am extremely excited to have had the opportunity to complete this Ph.D. thesis, and I am deeply grateful to all those who have contributed to its success. Have a pleasant read.

Thank you very much. Grazie mille. Merci beaucoup. Vielen Dank. Muchísimas gracias.

Publications

TubInTrain Related Publications

H. Pérez-Peña, Abel, M. Shevelev, A. E. Prota, S. Pieraccini, D. Horvath, Computational Approaches to the Rational Design of Tubulin-Targeting Agents. *Biomolecules* **2023**, *13* (2), 285.

Z. Boiarska, H. Pérez-Peña, A.-C. Abel, P. Marzullo, B. Álvarez-Bernad, F. Bonato, B. Santini, D. Horvath, D. Lucena-Agell, F. Vasile, M. Sironi, J. F. Díaz, A. E. Prota, S. Pieraccini, D. Passarella, Maytansinol Functionalization: Towards Useful Probes for Studying Microtubule Dynamics. *Chemistry – A European Journal* **2022**, *29* (5).

P. Marzullo, Z. Boiarska, H. Pérez-Peña, A.-C. Abel, B. Álvarez-Bernad, D. Lucena-Agell, F. Vasile, M. Sironi, K. Altmann, A. E. Prota, J. F. Díaz, S. Pieraccini, D. Passarella, Maytansinol Derivatives: Side Reactions as a Chance for New Tubulin Binders. *Chemistry – A European Journal* **2021**, *28* (2).

N. Bisi, L. Feni, K. Peqini, H. Pérez-Peña, S. Ongeri, S. Pieraccini, S. Pellegrino, α-Synuclein: An All-Inclusive Trip Around Its Structure, Influencing Factors and Applied Techniques. *Frontiers in Chemistry* **2021**, *9*.

SARS-CoV-2 Related Publications

A. Citarella, D. Moi, M. Pedrini, H. Pérez-Peña, S. Pieraccini, C. Stagno, N. Micale, T. Schirmeister, G. Sibille, G. Gribaudo, A. Silvani, D. Passarella, C. Giannini, Discovery of a Novel Trifluoromethyl Diazirine Inhibitor of SARS-CoV-2 Mpro. *Molecules* **2023**, *28* (2), 514.

K. Hassanzadeh, H. Pérez-Peña, J. Dragotto, L. Buccarello, F. Iorio, S. Pieraccini, G. Sancini, M. Feligioni, Considerations around the SARS-CoV-2 Spike Protein with Particular Attention to COVID-19 Brain Infection and Neurological Symptoms. *ACS Chemical Neuroscience* **2020**, *11* (15), 2361–2369.

Submitted to *Organic & Biomolecular Chemistry*:

A. Citarella, D. Moi, M. Pedrini, H. Pérez-Peña, S. Pieraccini, A. Dimasi, C. Stagno, N. Micale, T. Schirmeister, G. Sibille, G. Gribaudo, A. Silvani, C. Giannini, D. Passarella, Synthesis of SARS-CoV-2 Mpro Inhibitors bearing a Cinnamic Ester Warhead with In Vitro Activity against Human Coronaviruses.

Abbreviations

CuAAC	Copper(I)-catalysed Azide-Alkyne Cycloaddition
DMSO	Dimethyl sulfoxide
DoF	Degrees of Freedom
FF	Force Field
GDP	Guanosine Diphosphate
GTP	Guanosine Triphosphate
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
MAP	Microtubule-Associated Protein
MD	Molecular Dynamics
MDA	Microtubule-Destabilizing Agent
MSA	Microtubule-Stabilizing Agent
MT	Microtubule
MTA	Microtubule Targeting Agent
PDB	Protein Data Bank
PIF	Pairwise Interaction Fingerprint
RMSD	Root Mean Square Deviation
S4MPLE	Sampler for Multiple Protein-Ligand Entities
SDF	Structure Data File
VS	Virtual Screening

Thesis Overview **1**
Résumé en Français **3**
Chapitre I : Introduction 4
Chapitre II : Éléments de théorie 7
Chapitre III : Modélisation moléculaire de petites molécules pour l'immobilisation de la tubuline 9
Chapitre IV : Fonctionnalisation du maytansinol pour développer des sondes moléculaires et étudier la dynamique des microtubules 12
Chapitre V : Développement assisté par ordinateur de nouveaux dérivés C7 du paclitaxel pour étudier l'effet sur la signalisation structurale des microtubules 15
CHAPTER I: Biological Introduction **16**
I-1. Introduction to Tubulin and Microtubules 16
I-2. Microtubule Dynamics 17
I-3. Microtubules in Cancer Treatment 19
I-4. Microtubules in Neurodegenerative Diseases 20
I-5. Microtubule Targeting Agents 22
I-6. Tubulin Binding Sites 23
CHAPTER II: Elements of Theory **28**
II-1. Computer-Aided Molecular Design 28
II-2. Substructure Search 29
II-3. Structure-Based Approaches 29
II-4. Molecular Dynamics Simulations 37
II-5. Root-Mean Square Deviation (RMSD) 40
II-6. System Selection for Structure-Based *in Silico* Experiments in the Field of Tubulin 41
CHAPTER III: Molecular Modeling of Small Molecules for Tubulin Immobilization **43**
III.1 Scaffold Search: VS of Commercial Libraries and In-House Designed Compounds **48**
III.1-1. Aim of the Work 48
III.1-2. Computational Details 48
III.1-3. Results and Discussion 52
III.1-4. Conclusion 71
III.2 Design and Modeling of Putatively Covalent Todalam-Site Binders **73**
III.2-1. Introduction 73
III.2-2. Aim of the Work 73

III.2-3. Computational Details 73
III.2-4. Results and Discussion 80
III.2-5. Conclusions 95
CHAPTER IV: Computer-Aided Maytansinol Functionalization to Develop Molecular Probes for Studying Microtubule Dynamics 97
IV.1 Design and Modeling of Long Chain Maytansinoids and Maytansinoid Conjugates .. 99
IV.1-1. Introduction 99
IV.1-2. Aim of the Work 101
IV.1-3. Computational Details 101
IV.1-4. Results and Discussion 104
IV.1-5. Conclusion 114
IV.2 Characterization of Short-Chain Maytansinoids 116
IV.2-1. Introduction 116
IV.2-2. Aim of the Work 116
IV.2-3. Computational Details 116
IV.2-4. Results and Discussion 119
IV.2-5. Conclusion 126
CHAPTER V: Computer-Aided Development of New C7-Derivatives of Paclitaxel to Study the Effect on the Structural Signaling of Microtubules 128
V-1. Introduction 128
V-2. Aim of the Work 132
V-3. Computational Details 132
V-4. Results and Discussion 137
V-5. Conclusion 142
General Conclusion and Future Perspectives 144
Bibliography 147

Thesis Overview

My research activity is in the frame of the TubInTrain European Joint Doctorate (EJD), which is a European research program formed by 13 early stage researchers (ESRs) from different disciplines and countries within the fields of biology and chemistry focused on the study of the breakdown of microtubules (MTs) associated with neurodegenerative diseases and neurotoxicity. Tubulin protein is the building block of MTs and is the target protein on which the work described in this Ph.D. thesis is focused.

Within the TubInTrain program, the objective of my project is to apply a series of computer-aided molecular design techniques to thoroughly develop new tubulin binding agents and to explore at an atomic level the dynamics of their interactions when in complex with tubulin. This will facilitate further studies of MTs and MT-associated proteins (MAPs) related to neurodegenerative diseases.

Chapter I provides an introduction of the biological system that is the main subject of this thesis: the MT and its building block tubulin. The structures and function of these important components of the cell are covered in detail, emphasizing their significance as therapeutic targets for the treatment of various diseases, such as cancer and neurodegeneration. The chapter also discusses the known tubulin binding sites and the effect of tubulin targeting agents on MT dynamics when bound to their tubulin site.

Chapter II is an introductory chapter to the theoretical principles of the applied computer-aided molecular design techniques, including docking techniques and molecular dynamics simulations. The utilized docking software S4MPLE is given special attention, as it was the main computational tool employed during the present Ph.D. work.

Chapter III details the investigation carried out on the recently discovered todalam site of α-tubulin. This chapter describes how we employed computer-aided molecular design approaches to efficiently search for and identify alternative chemotypes that target the todalam site. To test and characterize these new molecules, we established an interdisciplinary pipeline, which also included synthesis, X-ray crystallography, and *in vitro* assays of the computationally identified potential todalam-site binders. Through virtual screening we found several hits that were experimentally validated. This enabled further structure-based design of putative covalent todalam site binders

predicted to target αCys4 of α-tubulin. We conducted a virtual screening of building blocks with cysteine-binding warheads from commercial libraries and selected the best candidates based on the results of covalent and non-covalent docking experiments, as well as molecular dynamics simulations of the computationally designed compounds.

In Chapter IV is described the study performed on the maytansine binding site of tubulin. We computationally designed long-chain maytansinoids and maytansinoid conjugates. Using molecular docking, we predicted the binding mode of short-chain maytansinoids within the maytansine site. Our study reports the tolerance of the C3 position of maytansinol to the addition of bulky substituents without being sterically hindered. These findings provide a solid foundation for further exploration of maytansinoids as molecular tools for investigating MT dynamics.

Chapter V describes the work carried out on the taxane site of tubulin. In this chapter, molecular docking was performed to predict the binding mode of flutax-2 to β-tubulin and to identify four novel paclitaxel derivatives from a virtual library of 50 compounds, which were selected based on their chemical feasibility, reagent availability, and likelihood of exhibiting a binding mode similar to flutax-2. These derivatives offer an excellent opportunity to explore the impact of bulky substituents at the C7-OH site of paclitaxel and investigate whether there are steric requirements that affect MT lattice.

Résumé en Français

Cette thèse de doctorat fait partie du programme européen de doctorat TubInTrain en biologie et chimie, dans lequel 13 doctorants de différentes disciplines et pays unissent leurs efforts pour étudier les microtubules (MT) et leur dégradation dans les maladies neurodégénératives et la neurotoxicité. La complexité croissante de la recherche scientifique a rendu nécessaire la collaboration d'experts de divers domaines. La chimie computationnelle fait partie des acteurs clés de tels projets multidisciplinaires, car elle facilite l'intégration de modèles informatiques, l'analyse des données et la validation expérimentale, comblant le fossé entre la théorie et l'expérimentation.

Ce doctorat avait pour objectif de développer de nouveaux ligands à la tubuline en utilisant des techniques de conception et de modélisation moléculaire assistées par ordinateur pour accélérer la recherche et promouvoir la synergie entre les chimistes de synthèse, les biochimistes et les biologistes. Cette collaboration multidisciplinaire a conduit à l'obtention de solutions innovantes à des problèmes scientifiques complexes.

Dans cette thèse de doctorat, le Chapitre I présente les microtubules et la tubuline, leurs structures, leur dynamique, leurs fonctions ainsi que leur importance en tant que cibles thérapeutiques pour le traitement de maladies telles que le cancer et les maladies neurodégénératives. Le Chapitre II offre un aperçu des principales techniques de conception et de modélisation moléculaire assistées par ordinateur et utilisées au cours de mon activité de recherche, en mettant l'accent sur le logiciel de docking S4MPLE. Le Chapitre III décrit l'investigation du site todalam, récemment découvert dans la tubuline α. Ce chapitre détaille comment les approches de conception moléculaire assistées par ordinateur ont été utilisées pour identifier des chemotypes alternatifs ciblant le site todalam afin de les exploiter davantage pour concevoir des ligands covalents supposés à ce site. Le Chapitre IV détaille l'étude du site de liaison de la maytansine à la tubuline, où des maytansinoïdes à longue chaîne et des conjugués de maytansinoïdes ont été conçus et modélisés par calcul. Le Chapitre V se concentre sur le site taxane de la tubuline, où le docking moléculaire a été effectué pour prédire le mode de liaison de flutax-2 à la tubuline β et identifier de nouveaux dérivés de paclitaxel.

Les études informatiques présentées dans ce résumé de thèse ont été réalisées à la fois à l'Université de Milan (Italie) et à l'Université de Strasbourg (France). Les travaux sur la synthèse organique, la cristallographie par diffraction de rayons X et la biologie présentés dans ce document ont été

réalisés par d'autres collègues de TubInTrain basés à l'Université de Milan (Italie), l'Institut Paul Scherrer (Suisse) et le Consejo Superior De Investigaciones Cientificas (Espagne).

Chapitre I : Introduction

La superfamille des tubulines est un groupe de protéines que l'on trouve chez les eucaryotes, les bactéries et les archées. Chez les eucaryotes, il existe sept familles différentes de tubulines, les tubulines α et β étant les plus importantes car elles peuvent se combiner pour former des microtubules. Les MT jouent un rôle crucial dans de nombreux processus cellulaires tels que la signalisation cellulaire, la morphologie, la motilité, la croissance et la régulation du trafic à longue distance.

Les MT se développent en incorporant un dimère avec du guanosine triphosphate (GTP) dans les deux sites de liaison des nucléotides de manière tête-bêche, ajoutant toujours la tubuline α à la tubuline β exposée (Figure 1, à gauche). Ceci crée une structure polaire de MT avec la tubuline β à l'extrémité en croissance (extrémité plus du MT). L'intégration du dimère de tubuline dans la structure du MT déclenche un changement conformationnel (passage d'une géométrie courbée à une droite), suivi de l'hydrolyse du GTP dans le monomère β. Un "capuchon de GTP" situé à l'extrémité plus du MT et constitué de dimères avec du GTP dans les deux sites, stabilise l'extrémité contre la dépolymérisation. Lorsque le taux d'ajout de dimères dépasse l'hydrolyse du GTP, la croissance des MT se produit ; sinon, la dégradation des MT (catastrophe) s'ensuit.

Les protéines associées aux microtubules (MAPs), les modifications post-traductionnelles et les agents ciblant les MT (MTAs) modulent la dynamique du réseau de MT. Les MTAs à haute concentration exercent différents effets sur la dynamique des MT ce qui permet de les classer en deux catégories : les agents stabilisateurs de MT (MSAs) qui empêchent la dépolymérisation de la structure du réseau de MT et les agents déstabilisateurs de MT (MDAs) qui empêchent l'assemblage des dimères en MTs.

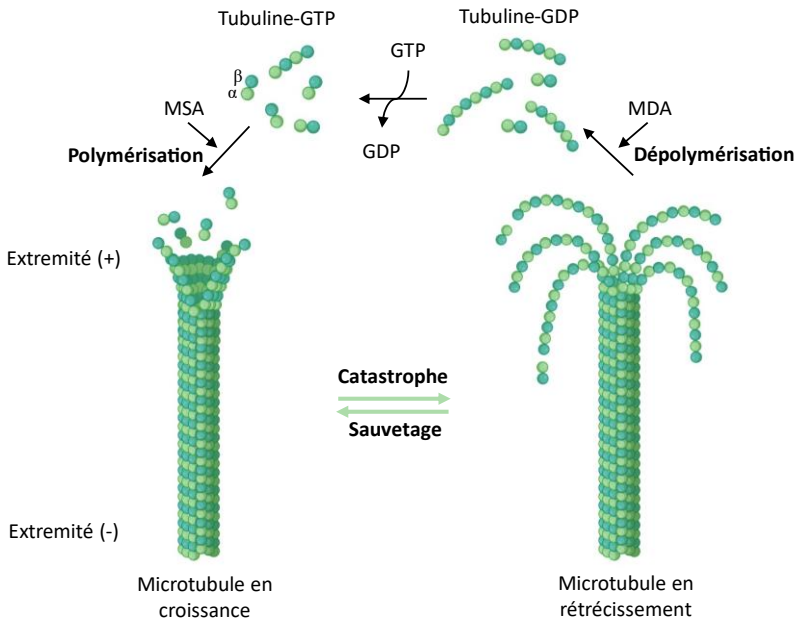


Figure 1. L'équilibre dynamique des MT implique une oscillation continue entre les phases de croissance et de rétrécissement. Bien que l'extrémité (-) du MT présente une certaine dynamique, la stabilité globale est principalement déterminée par les processus plus rapides à l'extrémité (+) du MT. La croissance du MT est favorisée par l'intégration de deux dimères de tubuline contenant du GTP dans l'extrémité (+), suivie de l'incorporation dans la structure, entraînant l'hydrolyse ultérieure du GTP. Un "capuchon de GTP" composé de dimères à GTP stabilise la structure du MT en croissance. Le remplacement de ces dimères du capuchon par de la tubuline-GDP initie la dépolymérisation.

Les MT jouent un rôle important dans le processus de mitose et sont impliqués dans de nombreux processus cellulaires. Les MT sont des cibles intéressantes pour la thérapie anticancéreuse car elles jouent un rôle clé dans le processus de division cellulaire, essentiel pour la croissance et la prolifération des cellules cancéreuses. Cependant, les MTAs manquent de spécificité et peuvent également nuire aux cellules saines, entraînant des effets secondaires graves. Dans les maladies neurodégénératives, un dysfonctionnement de la dynamique des MT neuronaux est un mécanisme sous-jacent ce qui fait des MT une cible prometteuse pour le traitement de ces troubles. Comprendre le rôle des MT dans la fonction neuronale et la maladie est important pour le développement de traitements efficaces pour les troubles neurologiques. Les approches computationnelles peuvent être appliquées pour concevoir et développer des MTAs en tant que sondes moléculaires, réduisant le temps, les coûts et les risques liés à la conception de nouveaux médicaments.

Les MTAs se lient aux MT et interfèrent avec leur fonction, et peuvent être utilisés comme agents thérapeutiques pour traiter le cancer et les maladies neurodégénératives. Les MTAs ont également des applications dans la recherche pour étudier la structure et la fonction des MT. Il existe huit

sites de liaison établis pour les MTAs sur la tubuline dont cinq sur la tubuline β (sites colchicine, taxane, vinca, peloruside/laulimalide et maytansine), un sur la tubuline α (site pironétine), un à l'interface intra-dimère (site gatorbuline) et un à l'interface inter-dimère (site todalam) (Figure 2).

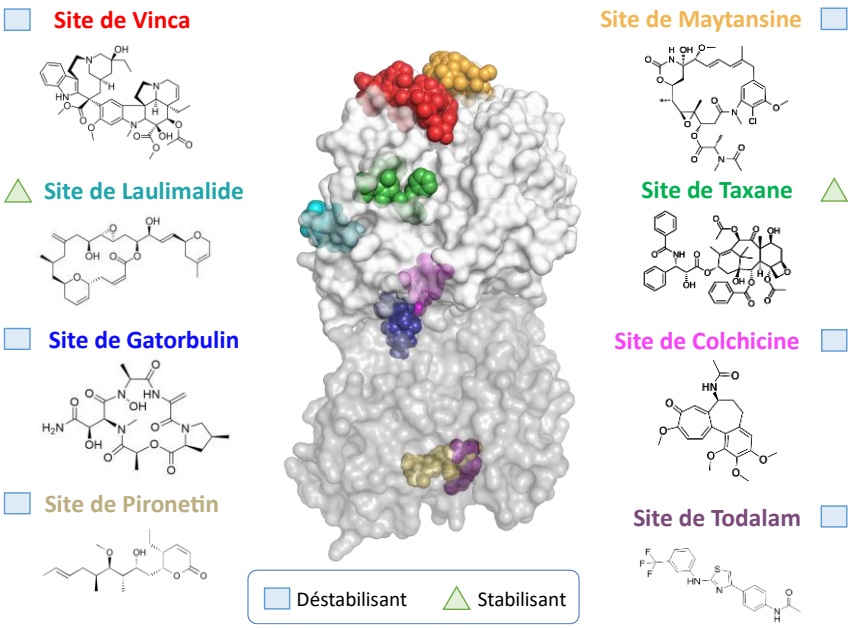


Figure 2. Le dimère de tubuline présente huit sites de liaison uniques, chacun mis en évidence avec leurs ligands respectifs sous forme de sphères colorées. La protéine est illustrée avec une représentation de surface transparente et les chaînes de tubuline α et β sont distinguées par des couleurs gris foncé et gris clair, respectivement. Les structures chimiques des ligands, qui ont inspiré le nom des sites de liaison, sont présentées à côté des étiquettes et suivent le schéma de couleurs de leur modèle sphérique.

Les MTAs exercent différents effets en fonction du site spécifique sur la tubuline auquel ils se lient. Le paclitaxel, un MSA qui se lie au site taxane, renforce la stabilité des MT en favorisant le passage d'une forme courbée à une forme droite. Les ligands du site colchicine bloquent la compaction de la poche formée par les brins βS8 et βS9 et les hélices βH8 et αH7, empêchant ainsi la transition courbe-droite nécessaire à l'assemblage des MT. Les ligands du site peloruside/laulimalide renforcent les interactions des dimères de tubuline entre les protofilaments voisins dans les MT. Les ligands du site vinca induisent un "coin" à la pointe du MT et favorisent l'assemblage de petits polymères hélicoïdaux de tubuline, réduisant ainsi la quantité de tubuline participant à l'assemblage. Les ligands du site maytansine bloquent l'assemblage des MT en inhibant l'ajout de nouveaux dimères de tubuline à l'extrémité en croissance. La pironétine se lie à une poche enfouie par fixation covalente à la Cys316, perturbant l'hélice αH8 et la boucle αT7. Les sites gatorbuline et todalam sont respectivement le 7ème et le 8ème site de liaison distinct sur le dimère de tubuline. Le todalam, le premier ligand de tubuline conçu rationnellement, entrave la

formation des MT en créant un coin dans la structure de l'oligomère de tubuline. Enfin, le mécanisme d'action de la gatorbuline est encore en cours d'investigation.

Les MTAs se sont révélés être de puissants modulateurs de la croissance cellulaire et sont d'une importance pratique exceptionnelle (en particulier les agents anticancéreux tels que le paclitaxel, la vinblastine et la maytansine). Le réseau européen ITN TubInTrain a consacré un ensemble de travaux à la découverte rationnelle de nouveaux ligands de la tubuline.

Chapitre II : Éléments de théorie

La conception moléculaire assistée par ordinateur est une approche computationnelle qui utilise la modélisation et les simulations informatiques pour concevoir et découvrir de nouvelles molécules. Le criblage virtuel est une technique couramment utilisée dans la conception moléculaire assistée par ordinateur qui consiste à rechercher des composés chimiques susceptibles de se lier à une cible thérapeutique, généralement des protéines. Une fois que les ligands potentiels ont été identifiés à l'aide de méthodes computationnelles, des approches expérimentales doivent être réalisées pour tester les meilleures molécules prédites. La recherche de sous-structures et le docking sont des exemples de méthodes de criblage virtuel basées sur les structures qui ont été utilisées dans cette thèse de doctorat pour explorer les modes de liaison et les affinités d'une bibliothèque virtuelle de composés. Ces outils computationnels sont essentiels pour identifier et hiérarchiser les candidats moléculaires potentiels, réduire le temps et le coût de développement de petites molécules et améliorer le taux de réussite des 'hits'.

Le docking moléculaire est une approche computationnelle utilisée pour prédire la structure et les interactions entre deux ou plusieurs molécules. La méthode peut prédire avec une certaine précision les meilleurs modes de liaison des complexes ligand-récepteur, leur affinité de liaison et les forces physiques impliquées dans leur interaction. Il existe trois principaux types d'approches de docking qui dépendent des degrés de liberté considérés : le docking rigide, le docking semi-flexible et le docking flexible.

Le principal logiciel de docking utilisé dans cette thèse de doctorat est S4MPLE (Sampler for Multiple Protein/Ligand Entities), un programme de sampling conformationnel polyvalent capable d'exécuter les trois types principaux de simulations de docking mentionnés précédemment. S4MPLE est un outil de modélisation moléculaire flexible qui utilise un algorithme génétique

hybride (GA) combinant des méthodes d'optimisation spécifiques à la modélisation moléculaire (optimisation lamarckienne) avec des stratégies d'échantillonnage évolutif traditionnelles (optimisation darwinienne). Développé par l'équipe du laboratoire de chemoinformatique de l'Université de Strasbourg, S4MPLE est écrit en Free Pascal et fonctionne en mode ligne de commande Linux. S4MPLE permet un contrôle total des degrés de liberté considérés, ce qui le rend adapté à diverses applications telles que l'échantillonnage conformationnel de petits composés organiques, les stratégies de conception de médicaments basées sur des fragments, le docking de composés de taille fragmentaire et de taille médicamenteuse, le docking de plusieurs ligands dans un même site de liaison et plus encore.

Les deux principaux défis du docking sont la notation (scoring) et la pose. S4MPLE aborde ces défis en utilisant une fonction d'énergie basée sur un champ de force et un algorithme génétique hybride qui imite des processus biologiques tels que le crossing-over et la mutagenèse. S4MPLE applique également une optimisation locale lamarckienne et une approche d'empreinte d'interaction pour contrôler la diversité de la population, garantissant que la diversité génétique au sein de la population d'individus en cours d'optimisation est maintenue.

La dynamique moléculaire (MD) est une technique de simulation informatique utilisée pour étudier le comportement temporel des systèmes biologiques. Les résultats d'une simulation MD sont une trajectoire qui montre les positions de tous les atomes du système au fil du temps. Pour effectuer une simulation MD il est nécessaire d’avoir une structure 3D initiale contenant les coordonnées de tous les atomes étudiés, un champ de force pour décrire les interactions physiques entre les atomes et un intégrateur pour résoudre les équations du mouvement.

Le champ de force est un élément essentiel de la simulation MD qui décrit les interactions physiques entre les atomes du système. Le choix d'un champ de force précis est crucial pour reproduire avec précision le comportement conformationnel et dynamique d'un système. De même, l'intégrateur est un algorithme mathématique qui résout les équations du mouvement pour un système d'atomes ou de molécules, et la précision des résultats de la simulation dépend du choix de l’intégrateur.

Les conditions périodiques aux limites (PBCs) sont une autre heuristique utilisée dans les simulations MD pour simuler les propriétés de masse avec un système de petite taille fini. Les PBCs supposent que le système en cours de simulation fait partie d'un système périodique plus

grand, où les atomes situés au bord de la boîte de simulation sont traités comme s'ils étaient voisins des atomes du côté opposé de la boîte.

La déviation quadratique moyenne (RMSD) est une mesure standard de la distance entre les structures 3D du même objet. Elle mesure la distance moyenne entre un groupe d'atomes et fournit une évaluation globale de la stabilité structurale du système à l'échelle de temps de la simulation. Dans le contexte des calculs de docking, la RMSD est utilisée comme mesure quantitative de la qualité du re-docking. Plus les valeurs de RMSD sont faibles, plus le système est stable dans les simulations. Dans le contexte des simulations MD, les chercheurs utilisent principalement cette mesure pour comparer les changements structuraux globaux du système contenant la tubuline induits par différents ligands moléculaires et leur capacité à se lier pour former un complexe stable avec la tubuline tout au long des simulations.

Chapitre III : Modélisation moléculaire de petites molécules pour l'immobilisation de la tubuline

L'étude des MTAs et de leurs affinités de liaison envers la tubuline dans des tests à haut débit est encore très difficile. Actuellement, les tests de déplacement par fluorescence sont limités par la disponibilité des sondes fluorescentes et la connaissance du site de liaison d'un ligand donné. Le consortium TubInTrain propose l'utilisation d'un test d'interférométrie guidée par ondes pour étudier la cinétique de liaison. Ce test peut être utilisé pour le criblage à haut débit et ne dépend pas de la disponibilité des sondes ni de la grande solubilité de la MTA dans les tampons. Cependant, le défi réside dans l'immobilisation de la protéine cible, en particulier pour la tubuline, qui possède de nombreux sites de liaison sur ses surfaces et est difficile à immobiliser sans bloquer les sites de liaison des petites molécules. Tout d'abord, il a été proposé d'utiliser la pironétine comme outil d'immobilisation dans les tests basés sur l'interférométrie guidée par ondes. Cependant, l'achat de pironétine était hors de notre portée, la synthèse de ce composé était très exigeante pour l'équipe de chimie de synthèse et les analogues de la pironétine conçus et testés ne se liaient pas au site de la pironétine. Par conséquent, nous avons concentré notre attention sur un nouveau site trouvé à proximité du site de la pironétine dans lequel des ligands covalents potentiels pourraient être conçus. Il s’agit du site todalam. C'est pourquoi, dans ce chapitre, nous discutons du développement récent de ligands du site todalam qui pourraient être optimisés en molécules potentielles pour immobiliser la tubuline de manière dirigée. L'étude vise à concevoir et valider

expérimentalement de nouveaux ligands du site todalam qui peuvent présenter un site réactif capable d'interagir avec l'αCys4 présent sur le site todalam. L'étude est divisée en deux étapes : (1) le saut de châssis et (2) la conception de ligands covalents potentiels basés sur les résultats de l'étape du saut de châssis.

L'objectif de cette étude n'était pas seulement d'explorer de nouveaux châssis et molécules se liant au site todalam, mais également de développer un pipeline de recherche pour leur découverte et optimisation. Le pipeline comprenait le criblage virtuel (VS) et la conception moléculaire, les simulations de dynamique moléculaire (MD), la synthèse, la cristallographie par diffraction de rayons X et les tests in vitro. Des méthodes assistées par ordinateur ont été utilisées dans les premières étapes du pipeline étant donné que les méthodes de recherche de sous-structure ont permis de filtrer efficacement les composés non pertinents pour la tâche en cours. Les expériences de docking et les simulations de MD ont ensuite été utilisées pour obtenir une compréhension plus détaillée des interactions potentielles entre les résultats virtuels et la biomolécule cible.

III.1 Recherche de scaffolds : VS des bibliothèques commerciales et des composés conçus en interne

L'objectif de cette étude était de trouver de nouveaux ligands de tubuline pouvant être utilisés pour immobiliser un dimère de tubuline pour des tests à haut débit, contournant ainsi la difficulté et le coût de la synthèse de MTAs structurellement complexes tels que la pironétine. L'accent était mis sur le site todalam.

Pour effectuer le saut de châssis autour du site todalam, nous avons réalisé un VS de la bibliothèque Enamine et des composés conçus en interne. La bibliothèque Enamine, qui contient environ 3 millions de composés commerciaux, a été criblée à travers plusieurs filtres : (1) recherche de sous-structure, (2) docking avec le logiciel PLANTS, et (3) re-docking des meilleurs candidats avec S4MPLE.

Tout d'abord, la stratégie de saut de châssis consistait à rechercher tout composé contenant un groupe acétaminophényl lié à un cycle aromatique moins spécifique. Les résultats nous ont permis de réduire le nombre de molécules d’environ 3 millions à une plage de quelques milliers. Par la suite, les molécules résultantes ont été dockées en utilisant PLANTS. Ce programme a été choisi pour un criblage rapide des molécules issues de la recherche de sous-structure, trouvant 590

ligands potentiels du site todalam. Enfin, le re-docking des 590 molécules sélectionnées a été réalisé en utilisant le logiciel de docking S4MPLE pour une analyse plus précise des molécules. Avec S4MPLE, nous avons effectué un docking flexible en sélectionnant les acides aminés à libérer lors des calculs et nous avons sélectionné des "points chauds" permettant aux modes de liaison du docking d'être plus précis. De plus, la bibliothèque de 176 molécules conçues en interne a été directement dockée en utilisant S4MPLE. Le todalam a également été re-docké avec S4MPLE puisqu'il était utilisé comme ligand de référence. Les mêmes paramètres S4MPLE que ceux expliqués ci-dessus ont été utilisés pour ces calculs de docking.

Les critères de sélection des molécules à partir des résultats de docking de S4MPLE reposaient principalement sur 3 aspects : (1) le score de docking, (2) le mode de liaison et (3) les interactions établies entre la tubuline et le ligand. Ainsi, les molécules qui avaient un meilleur score de docking que le site todalam, qui étaient capables de reproduire son mode de liaison et d'établir des interactions clés avec la tubuline ont été sélectionnées pour une analyse expérimentale ultérieure.

Par conséquent, des méthodes informatiques ont été utilisées pour identifier un nouvel ensemble de ligands prometteurs et chimiquement divers du site todalam. Ces ligands ont ensuite été validés expérimentalement par cristallographie par diffraction des rayons X et essais de polymérisation de la tubuline. Parmi les 15 molécules proposées par les méthodes informatiques, 12 ont été observées expérimentalement par liaison au site todalam avec un taux de réussite de 80 %. Les valeurs RMSD entre les structures cristallines des molécules identifiées et les résultats du docking étaient inférieures à 1,5 Å dans tous les cas, indiquant une bonne concordance entre les données informatiques et expérimentales. Ces 12 molécules capables de se lier au site todalam pourraient être regroupées en fonction de leurs structures en cinq groupes différents : (1) trois composés contenant un groupe urée ; (2) trois composés avec un groupe acetamide α-substitué (deux d'entre eux substitués par un atome d'oxygène et un par un atome de soufre) ; (3) trois composés contenant un groupe acrylamide ; (4) deux composés avec un cycle triazole et (5) un composé avec un groupe indole-2-carboxamide. Cette étude a permis d'établir les bases pour le développement ultérieur des ligands covalents.

Dans ce projet, nous avons montré comment les prédictions informatiques aident à la conception d'expériences pour les chimistes de synthèse et les biologistes. Cette synergie accélère la recherche en concentrant les ressources sur les candidats les plus prometteurs.

III.2 Conception et modélisation de ligands potentiellement covalents du site todalam

Le travail dans cette deuxième partie du projet repose sur des données expérimentales obtenues à partir de la cristallographie par diffraction de rayons X. Ce processus itératif de modélisation guidée par l'expérimentation renforce le pouvoir prédictif des modèles informatiques, permettant une meilleure prise de décision et des efforts expérimentaux plus ciblés.

En raison des défis de synthèse, les structures contenant des amides, des acrylamides et des triazoles ont été choisis pour une optimisation ultérieure en composés contenant un site réactif. En utilisant le pipeline établi, des petites molécules ciblant spécifiquement le site todalam ont été conçues de manière rationnelle et développées en composés contenant un groupement réactif dans le but de cibler le résidu αCys4. Pour cela, nous avons effectué plusieurs recherches de sous-structures réactives et, avec l'équipe de chimistes de synthèse, nous avons évalué les composés pouvant être obtenus à partir des réactifs récupérés. Les molécules sélectionnées ont été soumises à un docking covalent et non covalent. Ensuite, nous avons évalué leur score de docking, leur mode de liaison et les contacts établis au sein du site todalam. Parmi les 14 ligands covalents potentiels conçus par calcul, synthétisés et testés, neuf d'entre eux étaient liés au site todalam lors des expériences de cristallographie par diffraction de rayons X. Un de ces composés a montré une densité continue suggérant la présence d’une possible liaison covalente. Ce composé a été identifié comme le candidat le plus prometteur car il a permis de découvrir une longueur optimale pour un ligand covalent ciblant le groupe thiol du résidu αCys4 au sein du site todalam.

Les nouveaux ligands ainsi découverts du site todalam constituent une base pour le développement de ligands plus développés et plus sélectifs de ce site, ouvrant de nouvelles possibilités pour les futures recherches.

Chapitre IV : Fonctionnalisation du maytansinol pour développer des sondes moléculaires et étudier la dynamique des microtubules

La maytansine est un produit naturel à la cytotoxicité élevée qui a échoué lors des essais cliniques en raison d'une faible efficacité et d'une toxicité inacceptable. De nouveaux dérivés de la maytansine, conçus sous forme de conjugués anticorps-médicament (ADC), ont été efficaces dans le traitement du cancer. Le maytansinol, un précurseur des maytansinoïdes, a été utilisé pour préparer des maytansinoïdes naturels et semi-synthétiques, qui sont une nouvelle classe de

médicaments anticancéreux émergents. L'acylation du maytansinol est une étape cruciale dans la préparation des ADC des maytansinoïdes ou des nanoparticules. La conception de nouveaux maytansinoïdes pourrait également servir de sondes moléculaires pour mieux comprendre leurs relations structure-activité ou pour identifier de nouveaux ligands. La structure cristallographique à haute résolution du complexe formé par la β-tubuline et la maytansine fournit la base pour la conception rationnelle de dérivés de la maytansine. Dans ce chapitre, nous présentons de nouveaux maytansinoïdes obtenus à la suite de la réaction d'acylation du maytansinol et d'autres transformations structurelles non décrites précédemment. Ces molécules ont été conçues par calcul et testées expérimentalement.

IV.1 Maytansinoïdes à longue chaîne et conjugués de maytansinoïdes

La génération de sondes moléculaires à base de maytansine est difficile en raison de la complexité de la structure du produit naturel. De nouvelles études sur la chimie des dérivés de la maytansine pourraient guider la fixation de marqueurs fluorescents ou de radionucléides. Par conséquent, en raison de la proximité observée dans la structure cristalline avec ID PDB 4TV8 entre la molécule de maytansine et le site nucléotidique échangeable (site E) occupé par GDP/Mg^{2+} dans la β-tubuline, nous avons étudié de nouveaux maytansinoïdes fonctionnalisés en C3 et capables d'interagir avec Mg^{2+} ou d'atteindre le site de liaison du GDP. À cet effet, des maytansinoïdes à longue chaîne et des conjugués de maytansinoïdes (contenant du maytansinol et des mimétiques nucléotidiques) ont été conçus. Dans le cas de ces dernières molécules, la partie maytansinol agirait idéalement comme un point d'ancrage en maintenant la partie nucléotide modifiée à proximité du site E grâce à un espaceur flexible, favorisant ainsi l'échange de nucléotides. Différents connecteurs vers des chélateurs de magnésium ou des mimétiques nucléotidiques ont été sélectionnés pour leur capacité à interagir favorablement avec Mg^{2+} et la protéine tout en couvrant la distance requise.

Dans notre enquête, nous avons également examiné la tolérance des maytansinoïdes vis-à-vis de l'introduction de groupes volumineux et de leur impact sur la liaison à la tubuline, ainsi que les effets possibles que ceux-ci pourraient avoir sur la tubuline ou les MTs. Par conséquent, après la synthèse des molécules initialement conçues par calcul, des études cristallographiques par diffraction de rayons X ont été réalisées pour examiner le mode de liaison des maytansinoïdes à la

β-tubuline. Des tests d'affinité de liaison et de cytotoxicité ont également été effectués pour évaluer l'efficacité des nouveaux maytansinoïdes.

Nos résultats ont montré que l'ajout de substituants volumineux au squelette de la maytansine en position C3-O n'affectait pas son mode de liaison. Les maytansinoïdes ont conservé un arrangement spatial fondamental avec la β-tubuline, arrangement qui n'a pas été altéré par les différents substituants. Nous avons conclu que la position C3-O est un emplacement idéal pour les modifications du squelette de la maytansine, permettant l'incorporation de chaînes longues et flexibles ou même de groupes volumineux. Cela pose les bases pour explorer davantage la position C3 des maytansinoïdes en intégrant des modifications fonctionnelles telles que des fluorophores spécifiques ou des chélateurs et qui pourraient servir de sondes pour étudier la dynamique des MTs *in vitro*. En outre, ces modifications pourraient inclure d'autres types d'entités chimiques, ouvrant la voie au développement de composés bivalents à base de maytansinol capables de cibler des protéines spécifiques d'intérêt, telles que les PROTACS.

IV.2 Caractérisation *In Silico* des maytansinoïdes à chaîne courte

Le processus de synthèse des conjugués de maytansinoïdes a présenté de nombreux défis pour l'équipe de chimie de synthèse. En conséquence, plusieurs sous-produits ont été générés. La disponibilité de ces composés a suscité notre curiosité pour étudier comment les modifications de leurs structures moléculaires affectaient leur mode de liaison en complexe avec la tubuline et leur activité biologique.

Notre recherche, menée au sein du consortium TubInTrain, visait à caractériser la liaison des maytansinoïdes à chaîne courte à la tubuline à l'aide de méthodes computationnelles et expérimentales. Nous avons réalisé des simulations de docking pour prédire le mode de liaison de sept maytansinoïdes à chaîne courte et avons validé ces prédictions à l'aide de la cristallographie par diffraction de rayons X. Nos résultats ont révélé que les interactions par liaison hydrogène entre le C1-O et l'atome d'azote de βVal181 et le C24-O et les chaînes latérales de βLys105 et βAsn102 au sein du site de liaison de la maytansine déterminaient le mode de liaison des maytansinoïdes étudiés. De plus, l'introduction de substituants en position C3 n'a pas modifié l'arrangement tridimensionnel prédit du noyau de la molécule et a eu peu d'impact sur l'environnement entourant le site de la maytansine.

Ces résultats fournissent de nouvelles perspectives pour la conception moléculaire basée sur la structure de dérivés de maytansine qui pourraient potentiellement cibler des caractéristiques jusqu'alors inexplorées du site de liaison de la maytansine provenant de la β-tubuline et de ses environs.

Chapitre V : Développement assisté par ordinateur de nouveaux dérivés C7 du paclitaxel pour étudier l'effet sur la signalisation structurale des microtubules

Le flutax-2 est un dérivé fluorescent du paclitaxel qui permet de déterminer les paramètres thermodynamiques et cinétiques du processus de liaison taxane-tubuline, d'analyser la conformation du ligand et de localiser les MT du centrosome et du pôle du fuseau en tant que principales cibles des taxanes pour induire la mort cellulaire.

En présence de flutax-2, les biologistes de TubInTrain ont observé un comportement atypique dans le processus d'expansion du réseau des MTs à travers des expériences de diffraction de rayons X sur fibres. Pour étudier plus en détail ce résultat intéressant, le groupe de chimie de synthèse, en collaboration avec l'équipe de biologie, a proposé de réaliser la synthèse d'une série de dérivés du taxol présentant des substituants volumineux dans la même position que le flutax-2 (position C7), dans le but de comprendre quelle est l'exigence stérique qui produit ce comportement anormal du réseau des MTs.

Cinq liens possibles et 10 amines/alcools d'intérêts pouvant être couplés à la structure du paclitaxel nous ont été présentés. Nous avons combiné ces structures moléculaires en utilisant l'outil Reactor de ChemAxon, obtenant un total de 50 molécules à tester. Notre objectif était de trouver parmi ces 50 composés ceux qui étaient susceptibles d'adopter un mode de liaison similaire à celui du Flutax2 lorsqu'ils sont liés à la tubuline. En effet, le but n'était pas de trouver un ligand à haute affinité mais un ligand susceptible de produire le même effet que le Flutax2 dans les microtubules. Pour cela, le docking des 50 molécules a été réalisé en utilisant S4MPLE et les résultats ont ensuite été analysés à l'aide d'empreintes de docking pour les classer en fonction de leur ressemblance avec le mode de liaison du flutax-2. Actuellement, un des quatre meilleurs candidats proposés a été synthétisé et est en cours de tests biochimiques.

CHAPTER I: Biological Introduction

I-1. Introduction to Tubulin and Microtubules

The tubulin superfamily is a group of guanosine triphosphatases (GTPases) that are a type of proteins that hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP) and a phosphate group. Their activity is regulated by the binding and hydrolysis of GTP. They are found in eukaryotes, bacteria, and archaea. Within eukaryotes, there are seven different families of tubulin: α, β, γ, δ, ε, ζ, and η-tubulins.[1]

The most significant tubulin families in eukaryotes are the α and β-tubulin, as they can combine to form heterodimers that make up MTs. The γ-tubulins can be found in centrosomes and spindle pole bodies, where they play a crucial role in the formation and orientation of MTs. δ and ε-tubulins are involved in the regulation of centriole function, but their specific roles in cells and their mechanism of action are not yet fully understood. ζ-tubulins are not present in all eukaryotes and they are found in the basal body and the centriolar region of trypanosomes. η-tubulins have been found to be associated with γ-tubulins at the basal body.[1]

The sequence identity between the α and β-tubulin monomers is of ~40%, the sequence of each consisting of approximately 450 amino acids, and have very similar three-dimensional (3D) structures as shown in the very first tubulin data sets obtained with electron diffraction in 1998 (PDB ID 1TUB, 3.7 Å)[2] and 2001 (PDB ID 1JFF, 3.5 Å).[3] Based on these structures of the monomers, different secondary structure elements were identified and named, as well as the GTP and GDP binding sites in α and β tubulin, respectively (Fig. I-1).

α,β-tubulin dimers are the building blocks of MTs that are cylindrical hollow polymers, which in cells are formed by mainly 13 protofilaments (Fig. I-2). MTs play an essential role in numerous cellular processes such as cell signaling, morphology, motility, growth, and long-distance trafficking regulation (e.g., along axons and dendrites), making them an attractive target for cancer and neurodegeneration research.[4,5]

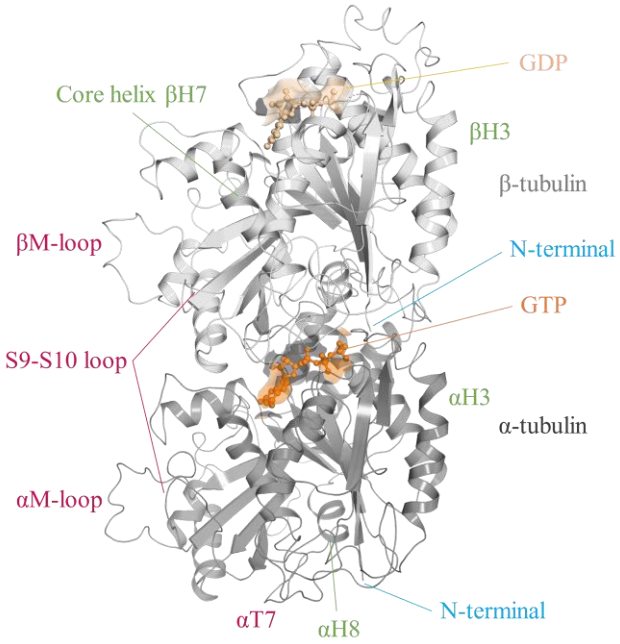


Figure I-1. 3D structure with PDB ID 1TUB of the first experimentally solved α,β-tubulin heterodimer in complex with nucleotides GTP (dark orange) and GDP (light orange) in the α (dark gray) and β-tubulin (light gray) monomers, respectively. The labels of the marked loops are colored in dark pink, the ones of the helices in green and the ones of the N-terminal amino acids in blue.

I-2. Microtubule Dynamics

From a structural point of view, MTs are highly dynamic polymers of α,β-tubulin. MTs have an enzymatic function of GTP hydrolase, a reaction that is the source of free energy for the conformational changes required for MT function: GTP-bound and GDP-bound are distinct, interconvertible states of tubulin. Each tubulin monomer is able to bind to GTP. The hydrolysis of GTP is limited to the β-monomer (exchangeable, E-site), providing energy for the conformational changes required for the formation of MTs. Within the α-monomer GTP is always retained (non-exchangeable, N-site). The α,β-tubulin heterodimers are bound to each other by non-covalent bonds, so they can always be in a bound/unbound equilibrium. Within cells, MTs are maintained in what is termed 'dynamic equilibrium', alternating between growth and shrinkage phases in which the α,β-tubulin heterodimers at the two ends of the MT are continuously added and released, allowing them to perform their various physiological activities. The transition from polymerization to depolymerization is called 'catastrophe', while the opposite event is called 'rescue' (Fig. I-2).[6,7]

Almost 40 years after the first mechanism of MT formation was proposed,[8] the details about this biological process still remain an ongoing topic of discussion; the main steps as understood today

are outlined here. Nucleation of MTs occurs in cells at MT organizing centers (MTOC) such as the γ-TuRC complex.[9-11] On the basis of this template structure, MTs grow by adding a dimer carrying GTP in both nucleotide binding sites in a head-to-tail fashion, always adding α-tubulin onto exposed β-tubulin. Thus, the MT is formed as a polar structure and exposes β-tubulin at the growing end (MT plus end). Incorporation of tubulin dimers into the MT lattice is accompanied by a conformational change of the dimer from a curved structure to a more rigid straight structure (curved-to-straight transition), which is then followed by GTP hydrolysis in the β-monomer.[12] Only at the plus end of the MT a 'GTP-cap' consisting of dimers that contain GTP in both sites is sustained, which is thought to stabilize the end against depolymerization. As the rate of addition of new dimers is faster than the rate of GTP hydrolysis, the MT grows, while in the opposite case, it shrinks (catastrophe).[13]

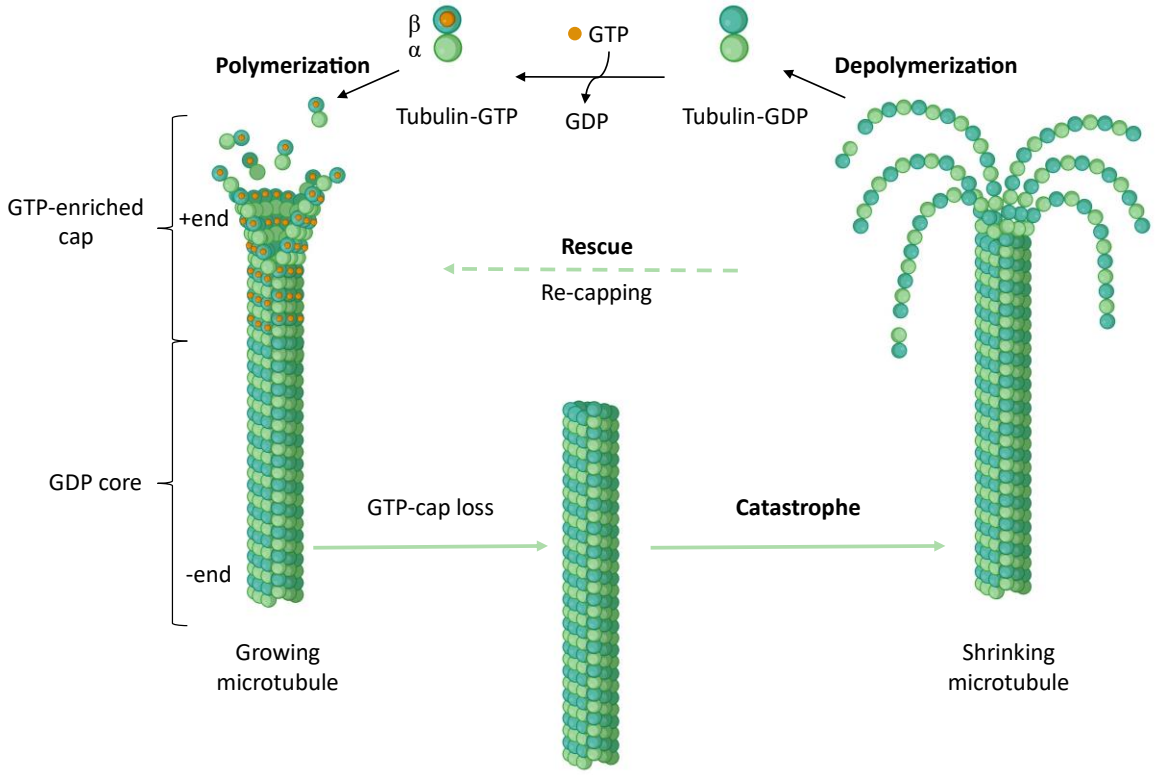


Figure I-2. Dynamic equilibrium of microtubules. MTs are constantly alternating between growth and shrinkage phases, while the (–) end of the MT displays some dynamics, the overall stability is governed by quicker processes at the MT (+) end. The growth of a MT is facilitated by incorporation of two GTP-containing tubulin dimers into the (+) tip of the MT, which leads to subsequent GTP hydrolysis. On top of the growing MT a 'GTP-cap' consisting of GTP dimers stabilizes the structure. The exchange of these capping dimers against GDP tubulin leads to depolymerization.

Microtubule Associated Proteins (MAPs), post-translational modifications, and MT targeting agents (MTAs) modulate the dynamics of the MT network. MTAs at high concentrations exert different effects on MT dynamics, which are used to categorize them into two classes: MT stabilizing agents (MSAs) that lead to an increased stability of the present MT by promoting assembly or stabilization of structure of the MT lattice, and MT destabilizing agents (MDAs) which prevent the assembly of dimers into MTs.[7]

I-3. Microtubules in Cancer Treatment

MTs play a crucial role in the process of mitosis, which is the process by which a single cell divides into two identical daughter cells. MTs help form the mitotic spindle, which is responsible for separating chromosomes during cell division.[14] The mitotic spindle is made up of MTs that extend from opposite poles of the cell and help position the chromosomes in the center of the cell. As the cell begins to divide, the MTs of the mitotic spindle pull the chromosomes apart and position them on opposite sides of the cell. This ensures that each daughter cell receives a complete set of chromosomes (Fig. I-3). In cancer cells, the mitotic process is often dysregulated, leading to uncontrolled proliferation and tumor growth.[14]

Therefore, any perturbation of the MT network severely affects cell survival, thus making MTs attractive targets for cancer therapy, since they play a key role in the cell division process, which is essential for the growth and proliferation of cancer cells. Presently, several MTAs such as vinca alkaloids and taxanes are used to treat different types of cancer. MTAs can promote cancer cell apoptosis through several independent mechanisms. These MTAs can work by disrupting the assembly and stability of MTs, thus inhibiting cell division and arresting the cell cycle in the mitotic phase. This can lead to the death of cancer cells and the suppression of tumor growth.[15]

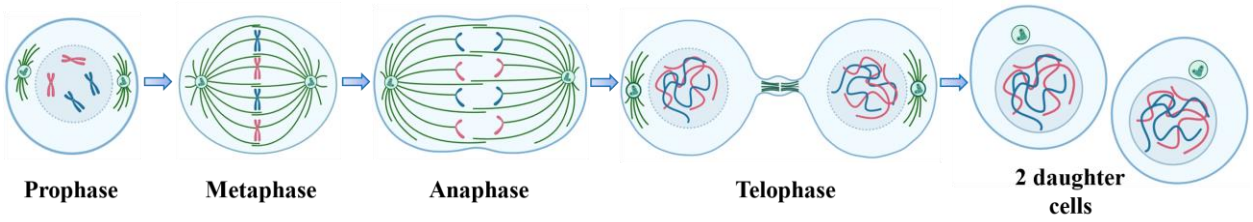


Figure I-3. Schematic representation of the four phases of mitosis: prophase, metaphase, anaphase, and telophase. Green lines represent MTs, dark blue and red lines represent the genetic material in different forms during mitosis, light blue external ovals surrounding each cell represent the cell membrane, light blue dotted internal ovals determine the nucleus of each cell, and green cylinders contained in green spheres represent the centrosomes.

While MTAs exhibit selectivity towards cancer cells, they lack specificity. Their mechanism of action involves targeting rapidly dividing cells, and thus they can affect not only cancer cells but also normal cells that undergo frequent division. As a result, healthy cells such as the ones present in the gut, hair follicles, and white blood cells are also affected, leading to severe side effects.[16]

It is important to weigh the potential benefits and risks of using MTAs in the treatment of cancer. While they may be effective in fighting the disease, they can also cause harm to healthy cells. As such, MTAs should be considered as a treatment option only when no other drug is available, particularly given the potentially fatal nature of cancer.

I-4. Microtubules in Neurodegenerative Diseases

Neurodegenerative diseases are one of the biggest global health and social care challenges together with cancer. Current drugs help manage the symptoms but there are no pharmacological therapies that can overcome injury and disease of the nervous system.

Recently, neuroscientists have found growing evidence pointing to neuronal MT dynamics dysfunction as an underlying mechanism in neurodegenerative diseases, defining the MT as a promising target for treating these disorders.[6,7]

MTs are involved in the transport of proteins and organelles along the length of the axon, which is a long, thin extension of the neuron that is responsible for transmitting electrical signals to other neurons (Fig. I-4). There are two types of intracellular transport in neurons: anterograde transport, which moves cargo from the cell body toward the axon terminal, and retrograde transport, which moves cargo from the axon terminal back toward the cell body. The motor proteins kinesin and dynein primarily perform the anterograde and retrograde transport, respectively. Both types of transport rely on MTs for proper functioning, therefore, inhibition of MTs function can disrupt the transport process by causing MTs to become unstable or by preventing them from growing or shrinking. This can lead to the accumulation of transported materials at the site of inhibition, causing cellular stress and potentially leading to the degeneration of neurons.

However, the structure and dynamic behavior of MTs associated with neurodegenerative diseases and neurotoxicity remain to be elucidated. Therefore, there is room for novel potential molecular strategies that can focus on modulation of the MT by targeting interactions between tubulin units

or between tubulin and MAPs (e.g., kinesin and dynein). The study of the mode of action of binding agents that can target the interactions mentioned above will not only help in the validation of MTs as a therapeutic target, expanding our understanding of the influence of MT breakdown associated with neurodegenerative diseases, but it can also bring us closer to determining the viability of a MT-targeting strategy as a treatment for these diseases.

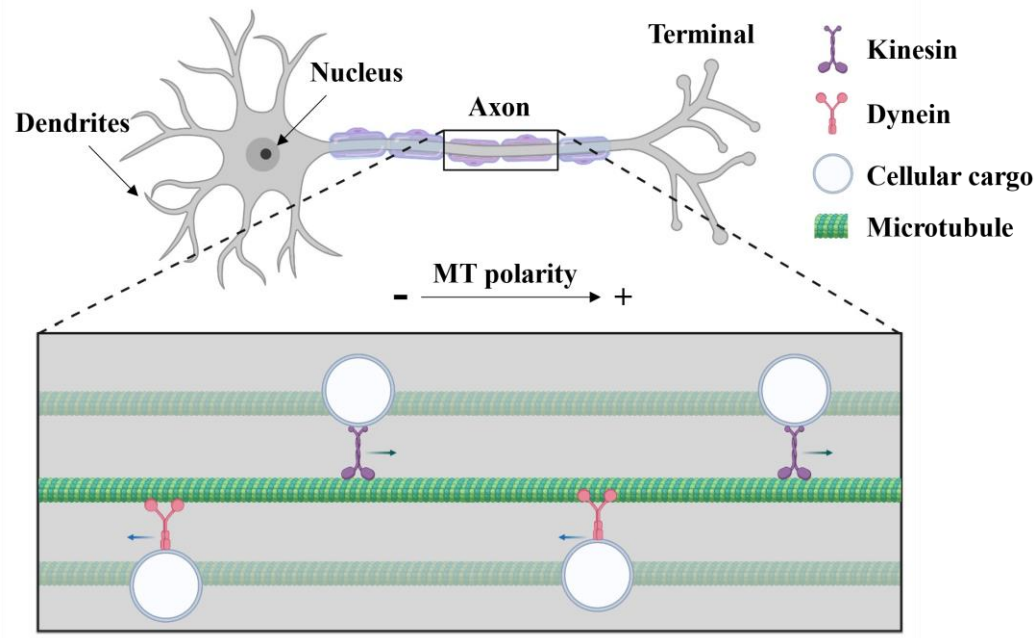


Figure I-4. Schematic representation of the main components of a neuron and an example of cargo transport performed by the MAPs kinesin and dynein.

In summary, MTs are essential for the correct transport of cellular components in neurons, and inhibiting their activity can potentially contribute to the degeneration of neurons. Understanding the role of MTs in neuronal function and disease is important for the development of effective treatments for neurological disorders; hence, continuing the study of MTs as biological targets of neurodegenerative diseases is highly necessary.

Part of my project focuses on the use of computer-based techniques to investigate this problem by designing and developing MTAs to serve as molecular probes. There are several computational approaches that can be applied at different stages of the drug discovery and development pipeline. The use of these tools reduces the time, costs, and risks involved in the design of new drugs for the many colleagues who collaborate on the TubInTrain project across multiple disciplines and countries.

I-5. Microtubule Targeting Agents

MTAs are molecules that specifically bind to MTs and interfere with their function. MTAs can be used as research tools to study the function of MTs and the role of MTs in various cellular processes. They can also be used as therapeutic agents to treat diseases that are caused by dysregulation of the dynamics of MTs, such as cancer and neurodegenerative diseases, as mentioned above.

Amoroso identified the first mitotic toxin (MTA) in 1935,[17] when he observed the anti-cancer properties of colchicine, a substance found in plants of the *Colchicum* genus. Through the research on colchicine, Borisy and Taylor identified tubulin in 1967.[18,19] In the following decades, eight binding sites for MTA have been identified and characterized, with the most recent discoveries occurring in the last two years.[20,21] From a medicinal chemistry point of view, MTAs are known to be more complex, have higher structural diversity (Fig. I-5), and have different interaction profiles than other anticancer drug categories.[22] This opens up new possibilities for scaffold design and optimization.

To understand the mode of action of MTAs, much effort has been dedicated to solving high-resolution MT and ligand–tubulin complex structures. Up to 2021, seven distinct binding sites for small molecules had been found on the α,β-tubulin dimer and at the inter-dimer interface. They have been thoroughly characterized by X-ray crystallography. In 2021, a combination of crystallographic fragment-based screening and molecular dynamics (MD) simulations evidenced 11 completely novel putative binding pockets and sites, occupied by 56 chemically diverse fragments.[23] Some of these fragments were used in a straight-forward fashion to develop a lead-like molecule from non-cytotoxic building blocks called todalam.[21]

To date, a large number of known tubulin binders have been derived from natural products, while only one (todalam) has been developed by rational structure-based drug design. Several of these tubulin binders show promising *in vitro* profiles, while presenting unacceptable off-target effects when tested in patients.[24] Therefore, there is a continuing demand for the discovery of safer and more efficient MTAs. Because tubulin structural data is readily available, the use of computer-aided design techniques can be a key element to focus on the relevant chemical space and guide the design process of new MTAs.

I-5.1. MTAs as Molecular Probes

MTAs can be used as molecular probes to study the structure and function of MTs, to track the interactions of MTs with other cellular components, and to perturb MT functions in cells, allowing researchers to study the effects of MT disruption on cellular processes. MTAs can be labeled with fluorescent dyes or other types of markers, such as radioisotopes.[25-27]

I-6. Tubulin Binding Sites

Extensive work has been done on determining the binding mode of MTAs. Here, there is an overview of the eight established binding sites and their mode of action in modulating the MT dynamics is provided. Five of these sites are located on β-tubulin: the colchicine, taxane, vinca, peloruside/laulimalide, and maytansine sites. The pironetin site is on α-tubulin, the gatorbulin site is at the intra-dimer interface and the todalam site at the inter-dimer interface (Fig. I-5).

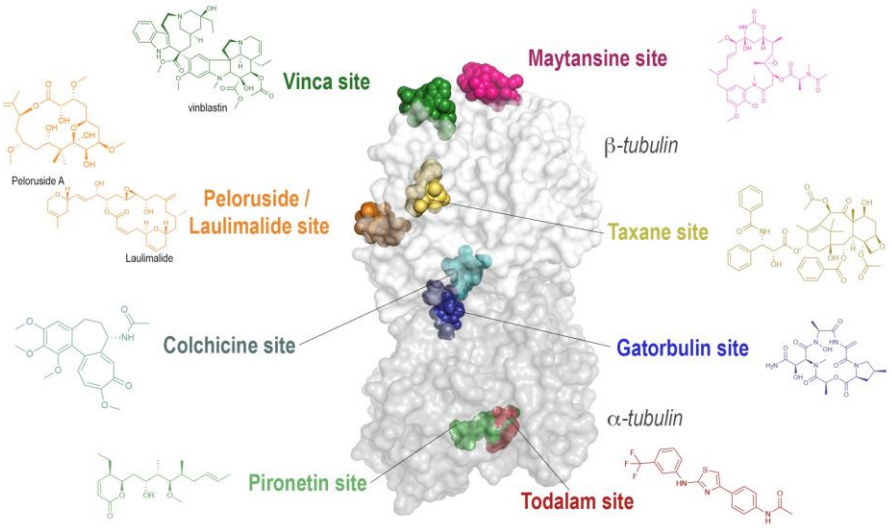


Figure I-5. The eight distinct binding sites are highlighted on one tubulin dimer with all of their representative ligands in colored sphere representation. The protein is shown in a transparent surface representation and α-tubulin and β-tubulin chains are colored dark and light gray, respectively. The chemical structures of the ligands after which the binding sites were named are indicated next to the labels and colored following the color code of their sphere model.

MTAs exert different effects according to the specific site on tubulin to which they bind: MSAs bind to the taxane and peloruside/laulimalide sites, whereas MDAs bind to the colchicine, vinca, maytansine, pironetin, gatorbulin, and todalam sites. The specific effects of these agents are determined by the way in which they alter the conformation of tubulin when they bind, although the exact mechanism of action is not always fully understood.

I-6.1. Taxane Site

The most prominent member of MTAs is paclitaxel, sold as a blockbuster drug under the name Taxol®, which is an MSA that binds to an exposed pocket on β-tubulin. Taxane-site ligands can enhance the stability of MTs, by promoting the curved-to-straight transition, as does paclitaxel,[28,29] or by direct structural stabilization of the βS7-βH9 loop (M-loop) a key structural element forming inter-dimer lateral contacts in MTs (Fig. I-1),[3] as does epothilone A or zampanolide.[30] The M-loop in its natural state is structurally disordered. However, when a ligand binds to the taxane site, it causes the M-loop to become ordered into a helical structure. This strengthens longitudinal contacts between tubulin units and stabilizes microtubule polymerization (Fig. I-6).[7] In Chapter V the design of new taxol derivatives is described.

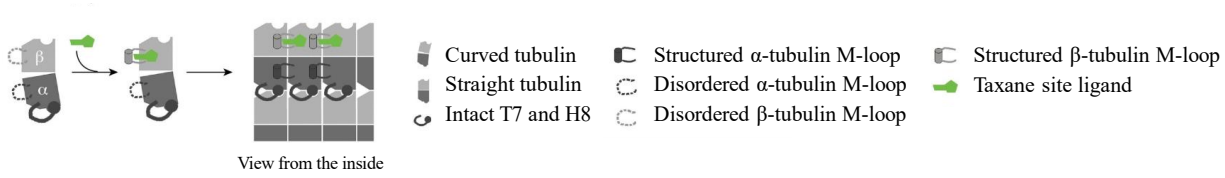


Figure I-6. Schematic representation of the mechanism of action of taxane-site ligands.[7]

I-6.2. Colchicine Site

In the group of MDAs, colchicine-site ligands are present with a great variety and a high number of different scaffolds. They bind in a buried pocket at the intra-dimer interface of α and β-tubulin, flipping the βT7 loop out of its native position. By occupying this binding site, they effectively prevent the curved-to-straight transition necessary for MT assembly by blocking the compaction of the pocket formed by the strands βS8 and βS9, and by the helices βH8 and αH7. Therefore, the presence of a ligand at the colchicine site inhibits polymerization by simply blocking the conformational change of tubulin (Fig. I-7).[31,32]

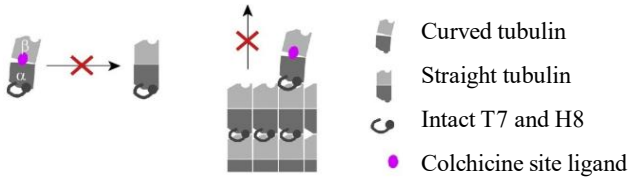


Figure I-7. Schematic representation of the mechanism of action of colchicine-site ligands.[7]

I-6.3. Peloruside/Laulimalide Site

Peloruside/laulimalide-site ligands strengthen the interactions of tubulin dimers across neighboring protofilaments in MTs by binding to a pocket near the lateral protofilament interface (Fig. I-8). Furthermore, these agents have been described to allosterically stabilize the M-loop to some extent.[33,34]

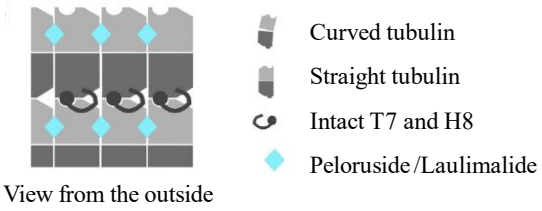


Figure I-8. Schematic representation of the mechanism of action of the peloruside/laulimalide-site ligands.[7]

I-6.4. Vinca Site

Another well-known group of MDAs are vinca alkaloids, which bind at the longitudinal interface between tubulin dimers. Vinca-site ligands induce a 'wedge' at the tip of the MT and thus prevent the straightening of the dimers (Fig. I-9).[35] Additionally, they promote the assembly of small helical tubulin polymers, thereby effectively reducing the amount of assembly-competent tubulin. Vinca-site ligands have also been observed to interfere with GTP hydrolysis by blocking the proper alignment of catalytic residues, further hindering the polymerization process.[36,37]

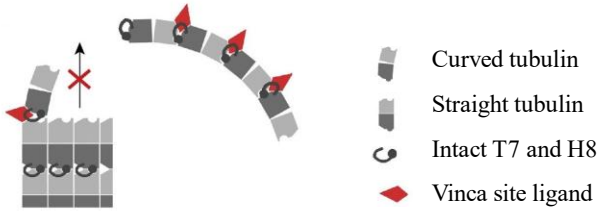


Figure I-9. Schematic representation of the mechanism of action of vinca-site ligands.[7]

I-6.5. Maytansine Site

The group of maytansine-site ligands blocks the assembly of MTs by inhibiting the addition of new tubulin dimers to the growing end. This is achieved by binding to the exposed site of β-tubulin and then effectively blocking the site that should accommodate the αH8 and αT7 loop of the binding tubulin dimer.[38] Ligands bound at this site not only block further growth of MTs but are also capable of fully blocking the formation of smaller tubulin oligomers, at high concentrations,

effectively keeping tubulin within the dimeric state (Fig. I-10). This is the site for which in this Ph.D thesis new maytansinoids were designed and will be discussed in Chapter IV.

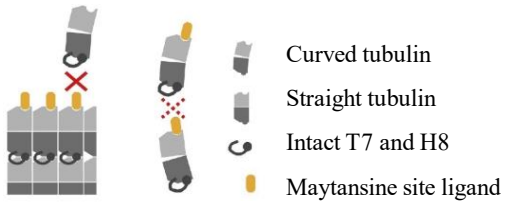


Figure I-10. Schematic representation of the mechanism of action of maytansine-site ligands.[7]

I-6.6. Pironetin Site

Currently, the only ligand known to exclusively bind to α-tubulin is pironetin, which binds to a buried pocket by covalent attachment to Cys316.[39,40] The binding of pironetin perturbs the helix αH8 and the αT7 loop, which are essential to establish longitudinal interactions between heterodimers along the protofilament. Thus, similar to maytansine, pironetin prevents the interaction of these elements with the neighboring tubulin and fixes tubulin in an assembly-incompetent state. Furthermore, pironetin also prevents growth at the minus end of the MTs, which exposes the α-tubulin surface that harbors both the helix αH8 and the αT7 loop and thus eventually promotes the disassembly of already formed MTs (Fig. I-11).[39]

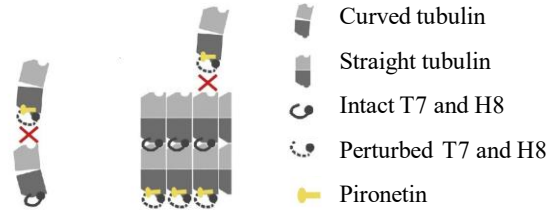


Figure I-11. Schematic representation of the mechanism of action of pironetin-site ligands.[7]

I-6.7. Gatorbulin Site

Recently, the 7th and 8th distinct binding sites on the tubulin dimer have been described. Gatorbulin, a cydodepsipeptide isolated from marine cyanobacteria, was found to bind to the intra-dimer interface adjacent to the well-known colchicine binding site.[20] Its mechanism of action is not yet well described.

I-6.8. Todalam Site

Todalam,[21] the first rationally designed tubulin binder, which emerged from a crystallographic fragment screen,[23] binds to the inter-dimer interface at a site located between the maytansine site on β-tubulin and the end of the pironetin pocket on α-tubulin. Both compounds are thought to hinder the formation of MTs by a mechanism similar to that of the vinca site ligands, by creating a wedge in the tubulin-oligomer structure (Fig. I-12). As observed for vinblastine, todalam was also shown to promote the formation of ring-like tubulin oligomers, further decreasing the pool of tubulin available for polymerization. The main research of this Ph.D. thesis focuses on the design and development of new todalam site binders (Chapter III).

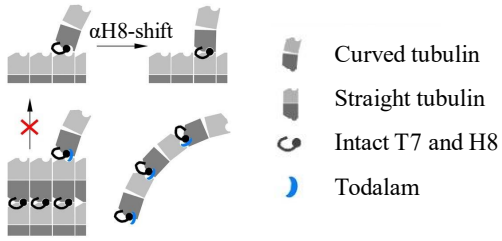


Figure I-12. Schematic representation of the mechanism of action of todalam-site ligands.[21]

CHAPTER II: Elements of Theory

II-1. Computer-Aided Molecular Design

Computer-aided molecular design is a computational approach that uses computer modeling and simulations to design and discover new molecules. It involves the use of various algorithms and software tools to predict the binding mode and affinity of small molecules to a protein target and optimize their structure to improve their physicochemical properties. Computer-aided methods constitute an essential part of the discovery and development of molecules, and they are used in conjunction with experimental techniques such as synthetic chemistry, biochemical and biological assays.[41] *In silico* approaches are the fastest, most economical, efficient, and simplest tools used to identify new binders to a certain protein site and analyze the protein-ligand biomolecular interactions.[42] Therefore, *in silico* methods help researchers identify and prioritize potential molecular candidates, reduce the time and cost of small molecule development, and improve the hit success rate. Computational tools have been widely applied in ligand-based and structure-based drug design campaigns to search for promising hit-to-lead structures (ligands) with desirable properties for protein targets by performing virtual screening (VS) of compound libraries.[1]

II-1.1. Virtual Screening

VS is a computational technique that consists in the search for chemical compounds that could bind to a drug target, usually proteins. VS allows to automatically evaluate very large libraries of compounds and filter them using computer programs into a manageable number of molecules that can be synthesized, purchased, and experimentally tested. VS can be used to select in-house database compounds for screening, to choose compounds that can be purchased externally, and to choose which compounds should be synthesized next. There are plenty of free databases of commercially available compounds that can be used as a starting point for a VS campaign, such as the Enamine[43] and ZINC[44] compound libraries, which are the ones used in the present Ph.D. thesis.

As the accuracy of the methods has increased, VS has become an integral part of the drug discovery process.[1] Nonetheless, once a set of compounds has been computationally identified as potential binders, experimental approaches need to be performed to test the best predicted molecules.

II-2. Substructure Search

In the present Ph.D. thesis, substructure search was the VS strategy used as a first stage filtering method to find new binders at the todalam site in Chapter III. The substructure query permits to specify open sites or atom lists for certain positions in a chemical structure. Substructure search is used to identify compounds that match the initial query with substitutions at open-valence positions. In this way, the user can select from a large set of molecules those that present a substructure that matches the molecule of interest. This can be done using computer algorithms and software designed to identify all molecular graphs containing one of the subgraphs/fragments specified by the query.[45] The resulting hits from the search can be analyzed and evaluated for their potential as molecular candidates.

II-3. Structure-Based Approaches

Most biological processes are based on the association of biomolecules which, in turn, are based on intermolecular interactions. Biological macromolecules such as DNA, RNA, or protein molecules (receptors) require, in most cases, association with small bio-substrates (ligands) to carry out their biological function. Structure-based VS relies on knowledge of the 3D structure of the biomolecular targets extracted from experimental methods (e.g., X-ray crystallography, NMR spectroscopy, cryo-electron microscopy) or computational-based predictive methods such as homology modelling.

In addition to structure-based pharmacophore screening, one of the most commonly used structure-based VS methods which involves the exploration of the binding modes and affinities of a virtual library of compounds is docking. Due to the extensive experimentally solved structural data available on tubulin and MTs, computer-aided structure-based molecular design approaches allow efficient exploration and identification of alternative chemotypes capable of targeting different tubulin sites.[1] As such, Mangiatordi *et al.* used protein-ligand docking to further filter the results of a prior pharmacophore screening and prioritize remaining compounds, the latter containing 31 new colchicine site-targeting agents with *in vitro* anti-proliferative properties.[46] Similarly, Guo *et al.* reported protein-ligand docking as a step that allowed them to discover eight confirmed cytotoxic agents targeting the colchicine binding site.[47] In addition, Zhou *et al.* used protein-ligand docking to highlight five virtual hits found by pharmacophore screening as the most

promising ones, their cytotoxic action related to binding at the colchicine site was later confirmed *in vitro*.[48]

II-3.1. Docking Experiments

Molecular docking is a computational approach that can predict the structure of the intermolecular complex formed by two or more constituent molecules given their 3D structure in atomic detail, in addition to a thorough description of the interactions established between them at the molecular level. Accordingly, it can predict the best binding modes of the ligand-receptor complex, their binding affinity, and the physical forces involved in their interaction.[49-52]

Presently, there are three main docking approaches, depending on the considered degrees of freedom (DoF) (Tab. II-1).

Table II-1. Three main types of docking simulation depending on considered degrees of freedom.

	Rigid docking	Semi-flexible docking	Flexible docking
Ligand	rigid	flexible	flexible
Target	rigid	rigid	partly flexible

The challenge still remains in the development of algorithms that can screen a database of flexible compounds while simulating the conformational changes of the binding site of the target induced by the binding of the flexible compounds. This phenomenon is called 'induced fit'. However, even the flexible algorithms do not account for 'induced fit'. The docking software mainly used in this Ph.D. thesis is S4MPLE[53,54], a conformational sampling program that can perform the three main docking simulation types reported in Tab. II-1.

II-3.2. S4MPLE: Sampler for Multiple Protein/Ligand Entities

S4MPLE (Sampler for Multiple Protein/Ligand Entities) is a flexible molecular modelling program based on a hybrid genetic algorithm (GA) that embeds molecular modelling specific optimization approaches (Lamarckian optimization) into classical evolutionary sampling strategies (Darwinian optimization). S4MPLE is developed by the Chemoinformatics Laboratory team of the University of Strasbourg, written in Free Pascal, and used in Linux command-line mode. S4MPLE allows full control of the considered DoF of the system under study, and therefore it can be used to perform:

- Conformational sampling of small organic compounds and peptides (1 entity)
- Fragment-based drug design strategies by addressing growing/linking of fragments (2 or more entities)
- Repositioning of protein loops in protein homology models (1 entity)
- Docking of fragment-size and drug-size compounds into a full rigid or partially flexible binding site (2 entities)
- Docking in the presence of mobile protein loops (2 or more entities)
- Covalent docking (2 entities)
- Docking in the presence of water molecules (at least 3 entities)
- Docking of several ligands simultaneously into a same binding site (at least 3 entities)

The two major problems in docking are the scoring and the posing, and the algorithms used differentiate the different molecular docking softwares.

II-3.2.1. Scoring: Force Field and Energy Function

The scoring step in docking is a quantitative measure of the fit of a ligand into the active/binding site for every pose. In S4MPLE, the energy function is based on a Force Field (FF). In molecular mechanics, a molecular system is treated as a mechanical system composed of atoms that move upon stretching, bending, and torsion motions (bonded interactions) and interact with each other *via* van der Waals and electrostatic forces (non-bonded interactions). To model all these molecular effects, different functions are used. To model the length and angle of chemical bonds, harmonic functions are needed, while torsional angles are modelled using sinusoidal functions. On the other hand, van der Waals interactions are modelled by the Lennard-Jones potential, and electrostatic interactions are modelled by the Coulomb law. Presently, S4MPLE does not implement any free energy scoring function, and the best poses are selected with respect to their FF energies, which are used to calculate the potential energy (E_{pot}) of a system under study. S4MPLE uses the existing AMBER[55,56] FF that is specifically parameterized for proteins and nucleic acids (to parametrize the target protein), together with its organic molecule generalization GAFF[57] (to parametrize the ligand molecule/s), a specific continuum solvation model, and a pair-based contact term. The native terms of the AMBER/GAFF energy scheme are also called Core FF (Fig. II-1).

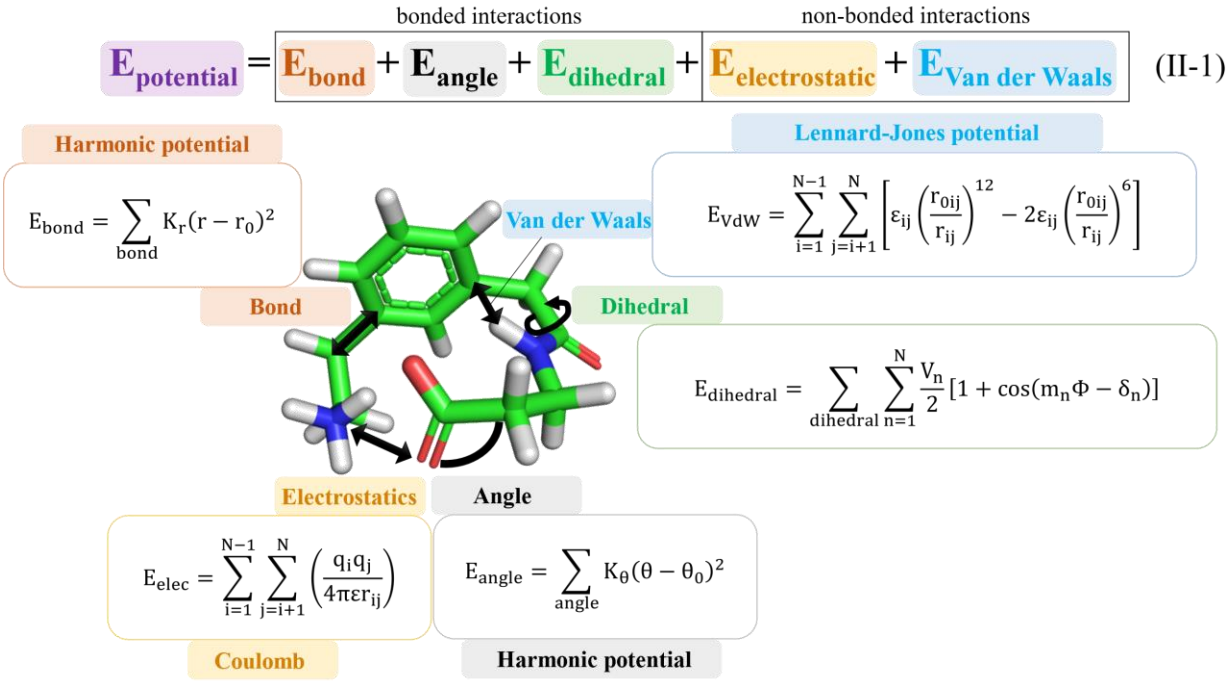


Figure II-1. Native summation terms of the AMBER[55,56] force field.

S4MPLE includes additional terms to obtain a more accurate scoring function. These additional terms consist of an implicit water-solvent model and a pair-based contact term that were calibrated by the developers of S4MPLE.

Pair-based desolvation penalty with a global constant k[58,59] inspired by Gilson and Honig's continuum solvent model that relies on functions of inter-atomic distances[60]:

Desolvation

$$desolv(i,j) = k \cdot \frac{\left(q_i^2 + q_j^2\right)}{{d_{i,j}}^4}$$

i,j : atoms
q : partial charges
d : distance
k : desolvation factor

Inspired by Gilson and Honig's continuum solvent model

(II-2)

The linear distance-dependent relative dielectric constant (ε_r) is used in the Coulomb term:

Dielectric constant in the Coulomb term

$$ECoulomb_{ij} = \frac{332 Q_i Q_j}{\varepsilon_r d_{ij}}$$

i,j : atoms
q : partial charges
d : distance
εr : relative dielectric constant

(II-3)

Pair-based contact term.[58] This term rewards favorable interactions such as hydrophobic close carbon contacts in space (C-C) and hydrogen-bond contacts (HB). The force constant k_{ij} depends on the type of contact, since it is different for the contacts between C-C and the H-B ($k_{C\text{-}C} \neq k_{H\text{-}B}$):

C-C and H-B Contacts

$$Econtact_{ij} = k_{ij}C_{ij} = k_{ij}\left[0.5 + 0.5 \cdot \cos\left(\pi \frac{d_{ij}^2 - d_{min}^2}{d_{max}^2 - d_{min}^2}\right)\right] \quad \text{(II-4)}$$

i,j : atoms
k : force constant
d : distance
C : contact strength

II-3.2.2. Posing

In S4MPLE, the search in space of possible poses and conformations is done by a hybrid GA that mimics Darwinian evolution and relies on a population of individuals. As implied by the name, the solution, structure, or problem must be encoded as a genome. In protein-ligand docking, the position, orientation and conformation of the ligand (state variables) are encoded for the stochastic search (Fig. II-2). The goal is to find low-energy geometries by exploring the energy landscape *via* the quantitative measure given by the fitness function.

The DoF are the dihedral angles (θ) of the studied system and are encoded in loci. The θ of each individual are encoded on 1 chromosome.

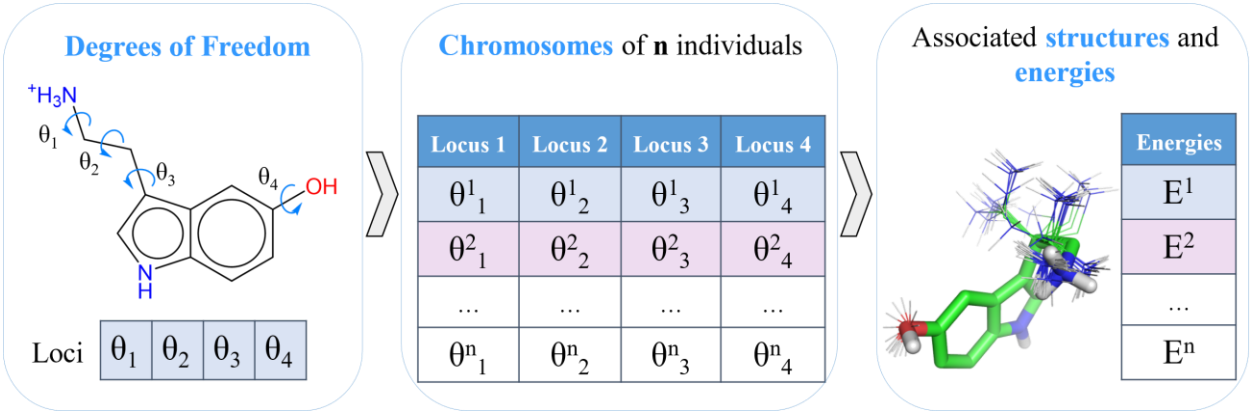


Figure II-2. Schematic representation of the strategy followed by the GA of S4MPLE to encode the geometry of each individual.

II-3.2.2.1. Genetic Operators: Crossing-Over and Mutagenesis

S4MPLE is able to explore the variable space through GA by mimicking biological processes such as crossing-over (CO) and mutagenesis. S4MPLE applies these genetic operators to random

chromosomes within the population to perform conformational sampling. This process is carried out as described in Fig. II-3 and Fig. II-4.

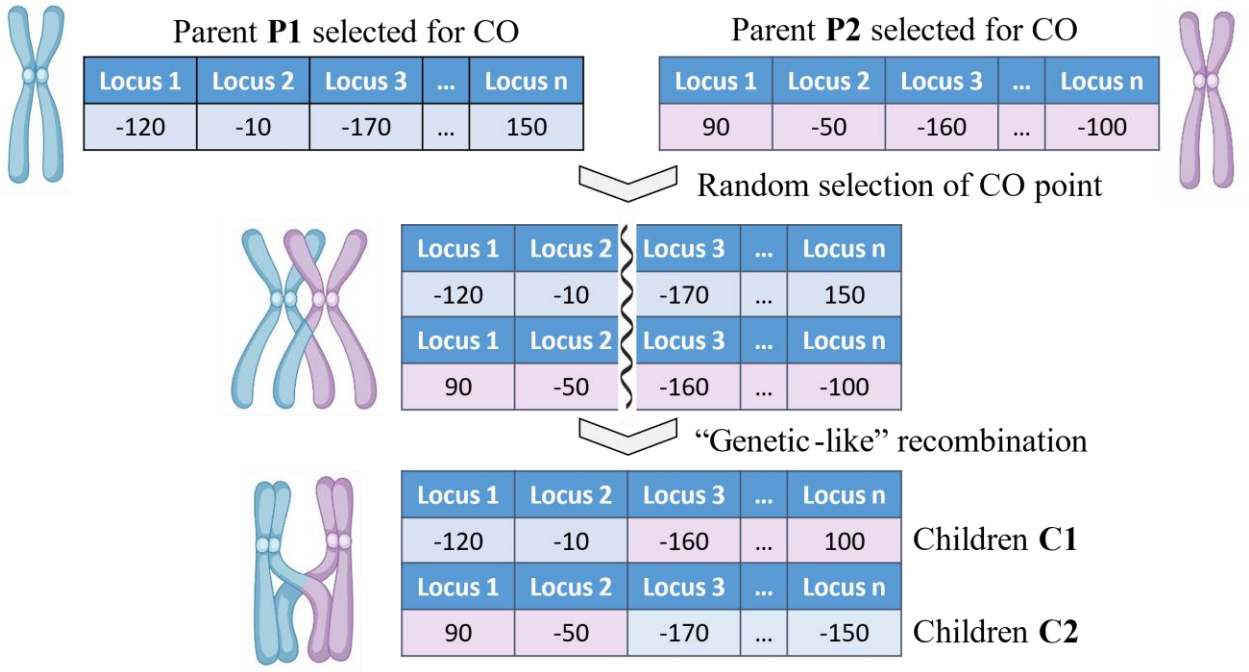


Figure II-3. The GA crossover strategy to randomly explore diverse geometries.

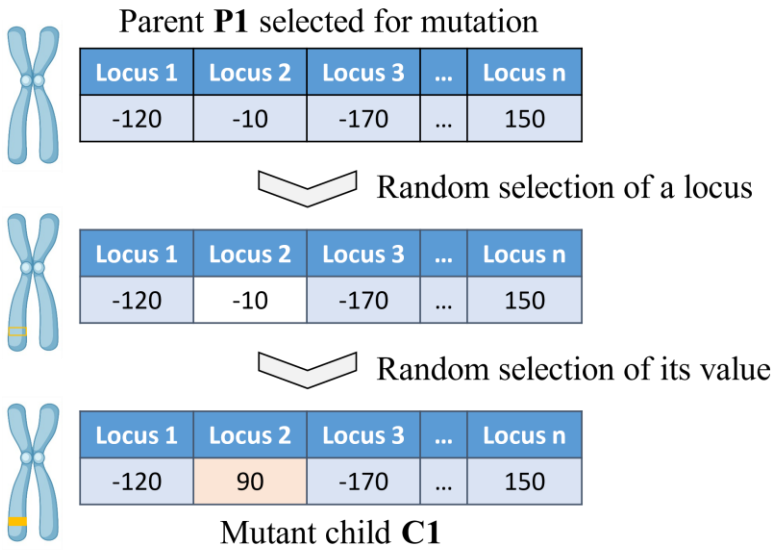


Figure II-4. GA mutation strategy to randomly explore diverse geometries.

II-3.2.2.2. Lamarckian Local Optimization

During the evolutionary sampling procedure, genetic searches are combined with gradient-based local optimizations, also known as 'Lamarckian' local optimization. At the last step of all genetic operators, there is a random number of conjugate gradient minimization iterations between 20 and 50.

II-3.2.2.3. Control of Population Diversity

The population diversity control strategy is very important in any population-based heuristic to avoid premature convergence due to the accumulation of minor mutants of a dominant local energy minimum. S4MPLE applies a pair-based interaction fingerprint (PIF)[53,54] approach to control the conformational similarity of the population under study, which monitors two types of interactions: close carbon contacts based on the C-C distance (hydrophobic contacts), and hydrogen bonds. The PIF encodes the geometry of the system by monitoring its contacts (Fig. II-5). If the Hamming distance between the fingerprints of two geometries of the system under study is lower than the established threshold defined by the user, these two geometries are considered equivalent. In this way, if two geometries share the same set of contacts, they are considered redundant and discarded. The PIF is also applied in the diversity step of the GA to avoid inbreeding. Inbreeding refers to a situation where the genetic diversity within the population of individuals being optimized is reduced over time, leading to a lack of new and diverse genetic material. This can happen when the same individuals, or individuals that are genetically similar to each other, are repeatedly selected as parents to produce the next generation of individuals. The result is that the population may converge on a suboptimal solution, limiting the search for a globally optimal solution.

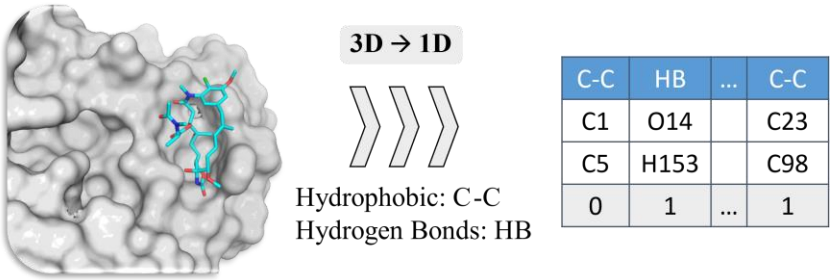


Figure II-5. Schematic representation of the strategy used by S4MPLE to encode C-C and HB interactions of the studied system.

II-3.2.3. Workflow of S4MPLE

The S4MPLE workflow to select the best solutions, as shown in Fig. II-6, begins by building a random starting population. The population is made by individuals whose phenotype (3D structure) is encoded in the genotype (chromosomes). Then, parents P are selected to apply the genetic operations (CO and mutations) to obtain children C, which are later evaluated and filtered by the results of the fitness function. The Lamarckian component, named after Jean-Baptiste Lamarck because it shares similarities with the idea of acquired traits being passed on to offspring in Lamarckian evolution, constitutes a partial minimization over a random number of iterations and it is embedded in each genetic operator. This process is applied to avoid near-systematic rejection of new conformers obtained from brute recombinations or mutations rather than waiting for a genetic operator to randomly discover a better conformation, which would be a more time-intensive approach. As a final step, a child (C) that is fitter than one of its parents will replace it. The workflow stops when the algorithm fails to come up with a new so far best conformer over a given number of generations.

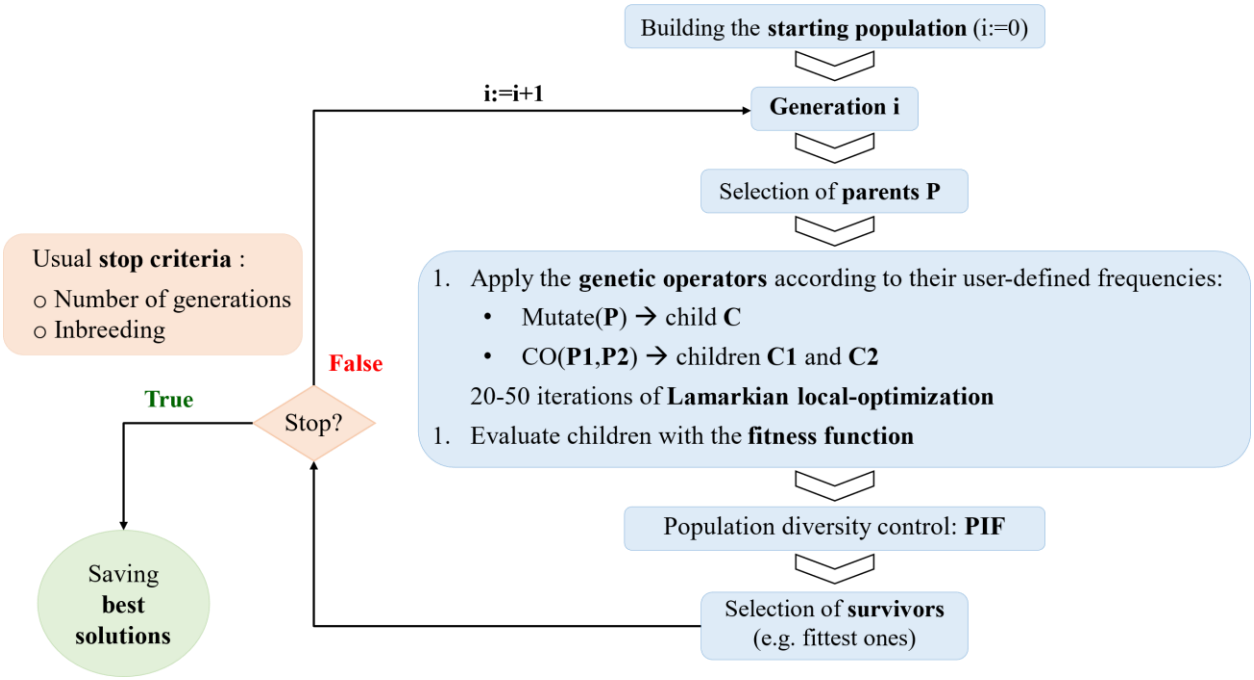


Figure II-6. Workflow of S4MPLE[53,54]

II-4. Molecular Dynamics Simulations

Molecular Dynamics (MD)[61] is a computational simulation technique that allows exploration of the behavior of a biological system over time. This is of great importance for research because biomolecules are entities whose atoms are in constant motion. In this way, by using MD, time-dependent processes in molecular systems can be monitored to facilitate the analysis of their structural, dynamic, and thermodynamic properties. MD is an ideal tool to explore the functional role of proteins, to examine the binding of a therapeutic agent to a protein and its effect, or for sampling the structural folding process, among other applications.[62]

The result of an MD simulation is a trajectory which includes a series of positions of all atoms contained in a system in space over time.

Three elements are essential to perform MD simulations:

- An initial 3D structure containing the coordinates in the space of all atoms under study.
- A potential able to describe the studied system.
- An integrator.

Given these three elements, the MD trajectory is obtained by calculating the time-dependent (*dt*) coordinates (*r*) and velocities (*v*) of all atoms of the studied system and the force (*f*) acting on it by integrating Newton's equations of motion:

$$\frac{d\vec{v}_i}{dt} = \frac{1}{m_i}\vec{f}_i \qquad \frac{d\vec{r}_i}{dt} = \vec{v}_i \tag{II-5}$$

Where $\vec{f}_i$ is the force acting on the *i-th* atom and m_i is its mass. $\vec{f}_i$ is calculated as the negative gradient of the potential energy function.

II-4.1. Force Field

A FF is needed to solve these equations. A FF is made by the equations (functional form) and by the set of parameters needed for the energy calculation. The FF describes the physical interactions of the atoms constituting a certain molecular system. Therefore, the selection of an accurate FF to reproduce the conformational and dynamic behavior of a system is essential. Thus, the FF is used to calculate the essential total potential energy of the system in the MD simulations (Fig. II-1). The choice of the FF that best suits the system under study is important because the quality of the MD

simulation results depends on the quality of the energy function used to treat the interactions among atoms in the system.

II-4.2. The Integrator

The integrator is a mathematical algorithm that is used to solve the equations of motion for a system of atoms or molecules. This allows the behavior of a molecule to be simulated over time and to study how it responds to different conditions. The integrator solves (approximately) the equations of motion to calculate the positions and velocities of the atoms at each time step.[63,64] This information is then used to update the positions and velocities of the atoms, and the process is repeated until the desired amount of time has elapsed to describe the time evolution of (all the particles of) the system. Different types of integrator can be used depending on the type of simulation and the level of accuracy that is desired. Some common types of integrators include the leapfrog[65] algorithm (Fig. II-7) which is the one used for the MD simulations discuss in Chapter III.

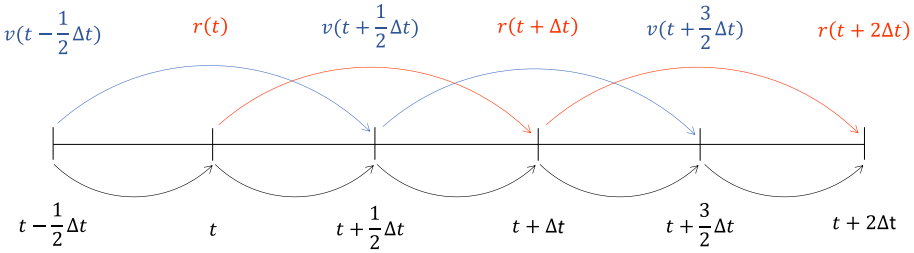


Figure II-7. Graphical representation of the leapfrog[65] algorithm.

II-4.3. Periodic Boundary Conditions

Periodic boundary conditions (PBCs) are a computational technique used in MD simulations to simulate bulk properties with a finite small size system. PBCs assume that the system being simulated is part of a larger, periodic system where the atoms at the edge of the simulation box are 'wrapped around' and treated as if they were neighbors of the atoms on the opposite side of the box. This allows the simulation to continue indefinitely without the need to consider the effects of the edge of the system on the molecules being modeled. PBCs are used to simulate an infinite system by treating the system as if it were periodic, with the simulation box repeating itself in all dimensions (Fig. II-8).

One of the most time-consuming parts of an MD simulation is the calculation of the non-bonded energies. In principle, non-bonded interactions should be calculated for every pair of atoms in the system that are infinite if we use the PBCs. To overcome this problem, the *minimum image convention* is applied. This technique is based on the idea that, when the simulation box is treated as if it were periodic, the distances between the atoms should be calculated using the shortest possible distance between them. Therefore, in this way, each atom can 'see' only one image of every other atom in the system. In order to optimize the computational efficiency of simulations, it is frequently necessary to introduce a cutoff value to limit the calculation of interactions beyond a certain distance. This cutoff also serves the purpose of preventing particles from interacting with their periodic images, as it is constrained to a maximum distance of half the length of the shortest side of the simulation box.

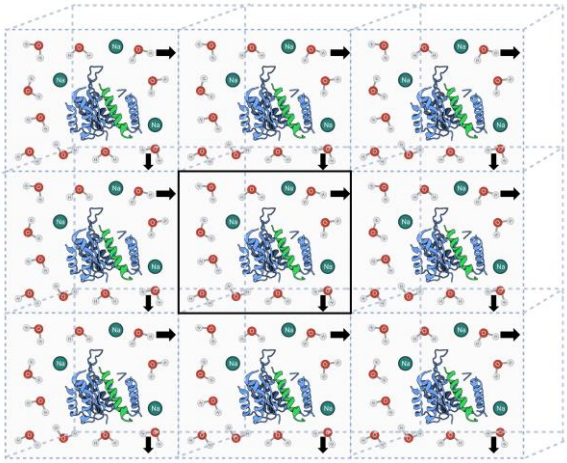


Figure II-8. Graphical representation of the concept of periodic boundary conditions.

II-4.4. Classical MD Simulations Workflow

MD simulations sometimes provide valuable information that is not accessible from experiments, allowing us to formulate new hypotheses. Furthermore, computational progress, both in the algorithms developed and in the power of computers, allows the study of biological macromolecules of larger dimensions, on longer timescales, and the predictions inferred from these simulations are increasingly reliable, making MD simulations a very valuable computational approach in the field of drug design.[1] The workflow of a classical MD simulation can be seen in Fig. II-9.

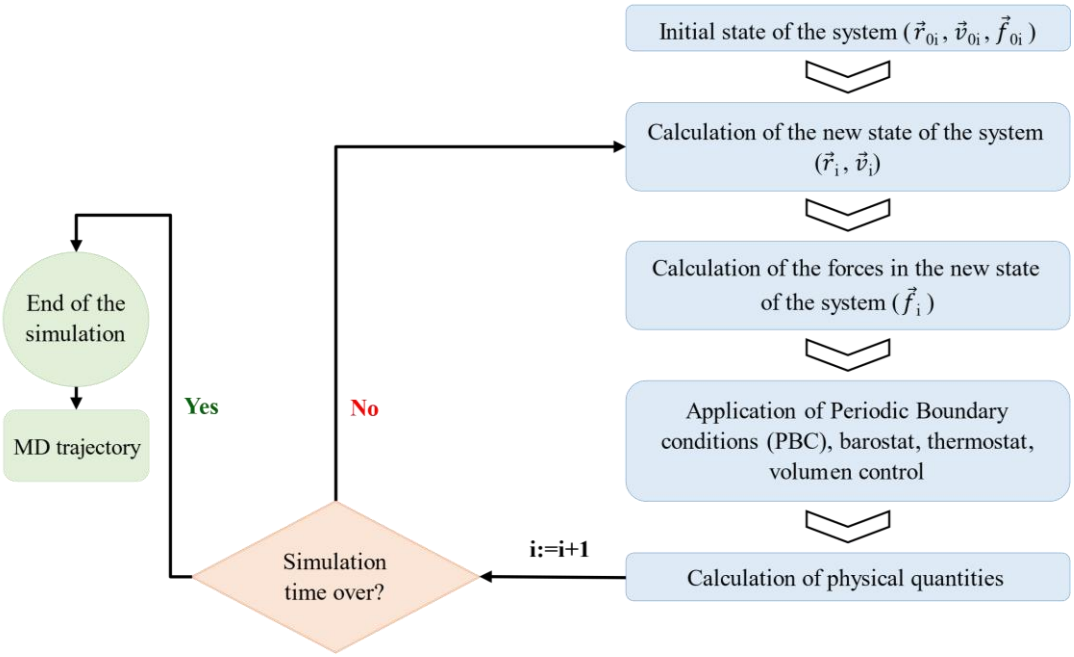


Figure II-9. Workflow of classical MD simulations.

II-5. Root-Mean Square Deviation (RMSD)

The root mean square deviation (RMSD) is a standard measure of the structural distance between 3D structures of the same object: it measures the average distance between a group of atoms. In the context of docking calculations, RMSD is used as a quantitative measure of re-docking quality. The RMSD function inputs the experimental (*e*) and computed (*c*) coordinates of the ligand atoms (*n*) and outputs an average distance between them as a square root of the average sum of positional differences for each atom *i*, following formula II-6. A re-docking RMSD score usually lower than 2.0 Å is considered to signify the validity of configured docking protocol.

In the context of MD simulations, RMSD values provide an overall evaluation of the structural stability of the system on the simulation time scale. Essentially, the RMSD values help to evaluate the changes in the position of selected atoms of a particular molecule throughout the simulation compared to their position in the initial structure (first frame). The RMSD values are measured to monitor stability by quantifying the displacement changes of the 3D system during the course of the MD simulations. The lower the RMSD values, the more stable the system is in the simulations. Researchers in the field of tubulin and MTs use this measure mainly to compare the global

structural changes of the tubulin-containing system induced by different molecular binders and their ability to bind to form a stable complex with tubulin throughout the simulations.

$$\mathrm{RMSD}(e,c) = \sqrt{\frac{1}{n}\sum_{n=1}^{i}\left((e_{ix}-c_{ix})^2 - \left(e_{iy}-c_{iy}\right)^2 - (e_{iz}-c_{iz})^2\right)} \qquad \text{(II-6)}$$

II-6. System Selection for Structure-Based *in Silico* Experiments in the Field of Tubulin

When choosing the protein structure in which to perform docking experiments, one should select a high-resolution structure with the least number of missing atoms from the binding site of the target protein or a robust model of it. This is because the accuracy of the protein docking prediction is dependent on the quality of the protein binding site structure. Since tubulin has several binding sites, in Tab. II-2 we provide a summary of the best structures to use for docking experiments, depending on the binding site of interest.[1]

Table II-2. List of high-resolution tubulin crystal structures by binding site.[1]

Binding site	PDB ID	Resolution (Å)
Apo	5NQU	1.8
	3RYC	2.1
	4I55	2.2
Taxane site	4I4T	1.8
	5LXT	1.9
	6SES	2.0
Laulimalide	4O4H	2.1
	4O4J	2.2
Maytansine	4TV9	2.0
	6JFM	2.1
	4TV8	2.1
Colchicine	6S8K	1.5
	6ZWB	1.7
	7Z2P	2.0
	5M7E	2.0
	6TH4	2.1
Vinca	5YIZ	1.8
	5J2T	2.2
	5JH7	2.3
Pironetin	4TUY	2.1
	5FNV	2.6
Todalam	5SB3	2.2
	5SB6	2.3
Gatorbulin	7ALR	1.9

When selecting the system to be simulated using MD, it is also important to take into consideration the assembly of tubulin into protofilaments and MT structures. If the binding site studied is far from any tubulin inter-dimer interface (e.g., the colchicine site) or is considered to completely prevent the interaction of two dimers (e.g., the maytansine site or the pironetin site), a dimer can serve as a model for tubulin binders. However, if the binding site is present at the longitudinal inter-dimer interface (e.g., vinca, todalam, gatorbulin, laulimalide) or the lateral axes (e.g., taxanes), a more complex system may need to be considered. For example, the taxane site is known to be located between two adjacent β-tubulin units and scientists such as Castro-Alvarez *et al.* opted to study a 'tetramer' model to investigate binders at the taxane site.[66] The choice of the system size is a trade-off between the accuracy of the site and the computational effort needed.

CHAPTER III: Molecular Modeling of Small Molecules for Tubulin Immobilization

As described in Chapter I, MTAs are of great interest both as anti-cancer drugs or probes for neurodegeneration as well as other diseases. Up to date, high-throughput determination of binding affinities of MTAs has been limited to fluorescent displacement assays,[67] and thus limited to the availability of a fluorescent probe for the given site, and even the knowledge of the binding site of a given ligand. One potential solution to this problem might be the set-up of a surface-based assay such as Surface Plasmon Resonance SPR or a wave-guided interferometry assay to screen for binding kinetics, which can be used for high-throughput screening, and it does not depend on the availability of probes or a high solubility of the MTA in buffers.[68] This type of assay could provide valuable information not only on the tubulin-binding kinetics of small molecules but also on the study of the interaction of tubulin with other proteins. LeadXpro, a company which is also part of the TubInTrain project, successfully applies such wave-guided interferometry-based assays to determine binding affinities of ligands to their targets which we envisioned to exploit for the set-up of our assays. This method uses Grating-Coupled Interferometry and presents advantages over other approaches such as Surface Plasmon Resonance. Measurements of interactions of multiple immobilized proteins enhances both the sensitivity and the signal-to-noise ratio. The problematic of this wave-guided assay is that it requires the immobilization of tubulin on a chip surface, as represented in Figure III-1.

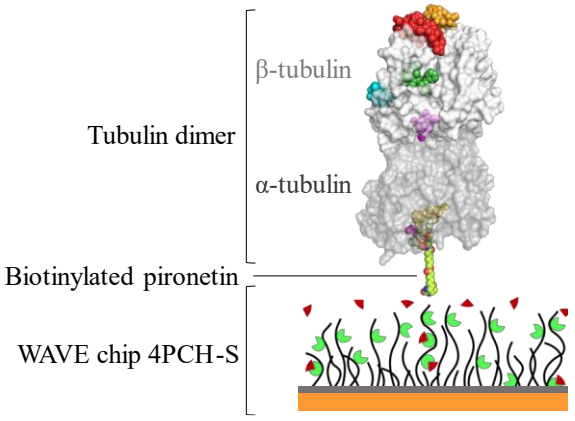


Figure III-1. Schematic representation of an α,β-tubulin dimer attached to the surface of a WAFE chip 4PCH-S through a pironetin molecule attached to a biotinylated linker.

A very basic problem common to all surface-based assays is the immobilization of the ligand, in this case tubulin, onto the matrix in a manner that allows for interactions to occur. For tubulin, this is especially challenging as it is most often isolated from natural sources, thus not allowing the introduction of affinity tags and has many binding sites on its surfaces. Therefore, an undirected immobilization e.g. *via* amine coupling could lead to a blockade of small molecule binding sites. One specific and mild way of immobilization often used in these cases are antibodies. However, antibodies have a very large size to perform tubulin immobilization on the chip surface, and these proteins reduce the sensitivity of the assay. This is because the assay has a general issue that can be described in two parts: (1) as the size of the molecule on the surface increases, the measurement of binding becomes more challenging because the signal becomes weak, and (2) the strength of the generated signal is directly correlated to the sizes of the interacting species as you are measuring differences in optical density that occur upon ligand binding. Therefore, when a small molecule is immobilized and a larger molecule binds, the signal is very high. The problem is that tubulin, which has a mass of 100 kDa, is already very large. The goal is to measure the interaction of molecules that are usually 300-600 Da, so adding another protein to the tubulin could push the assay to the limit of detection.

To address this need in the TubInTrain network, we planned to first design small molecules that we can use to covalently modify tubulin in a directed manner. The ultimate goal would then be to modify the selected molecules to immobilize the protein onto the surface in an oriented and controlled manner (e.g., by biotinylation).

In 2016, Prota *et al.* solved the crystal structure of the first α-tubulin covalent binder called pironetin and characterized its binding mode.[39] Pironetin was found covalently bound to αCys316.[40] In our research, we proposed the utilization of the molecule pironetin to be used as an immobilization tool in wave-guided interferometry-based assays. Additionally, since most binding sites are located on β-tubulin (as illustrated in Fig. III-1), the immobilization of the tubulin dimer through the pironetin site would leave all the other tubulin sites available to perform experiments. However, the current high cost and difficulty of synthesis associated with pironetin, specifically due to its numerous chiral centers, prompted us to seek an alternative within the TubInTrain project. Our initial objective was to develop more cost-effective derivative forms of pironetin that could be attached to a linker for immobilization purposes.

The synthetic chemistry team initiated the study by creating a collection of 12 derivative forms of pironetin that were able to be synthesized using the reagents readily available in their laboratory. (Fig. III-2).

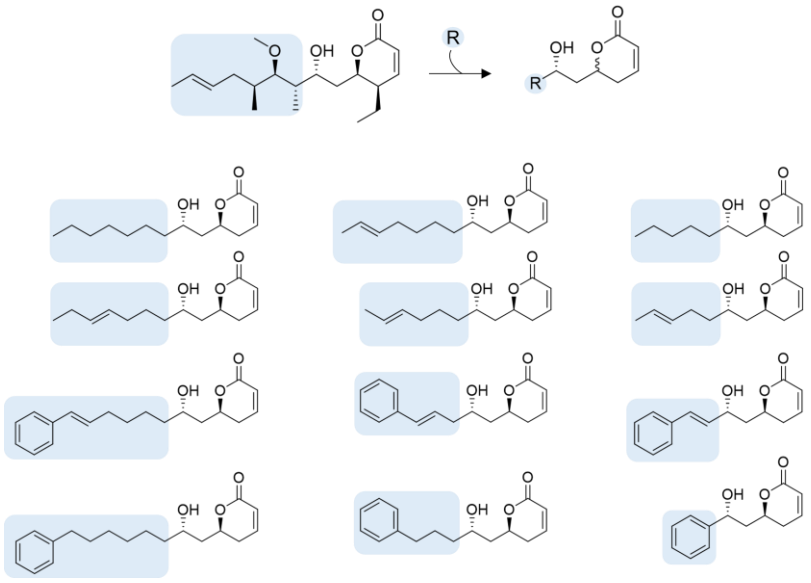


Figure III-2. 2D representation of pironetin and the pironetin derivatives designed to be synthesized. The chemical group of each derivative that differs from that of the parent compound is colored light blue.

To support the design of the synthetic chemistry team and to evaluate the potential utility of synthesizing such molecules we performed non-covalent docking of the suggested pironetin derivatives with Audock Vina[69,70] in the pironetin binding site of the structure mentioned above solved by Prota *et al.* with PDB ID 5LA6.[39] The results of this *in silico* investigation indicated that the optimal binding modes of the various pironetin derivatives involved positioning the lactone ring in a manner similar to the parent compound pironetin suggesting that these molecules could potentially interact covalently with αCys316. However, after being synthesized and experimentally tested *via* X-ray crystallography, the results showed that none of the designed pironetin derivatives were bound to the pironetin site. In an attempt to identify and design other alternative pironetin site binders, the chemoinformatic group of the University of Strasbourg carried out different VS campaigns. All of them were unsuccessful, which led us to search for an alternative tubulin binding site that would serve us to achieve our goal.

In the last two years, a milestone has been set towards the design and development of novel and easily accessible molecules with the publication of a combined crystallographic and computational fragment screen, in which Mühlethaler *et al.* identified more than 50 chemically diverse fragments

binding to well-known and new binding sites on tubulin (Fig. III-3, A). On this screen, three fragments were found to bind to the same novel site located at the inter-dimer interface between α and β tubulin in close proximity to the pironetin site (Fig. III-3, B).[23] In a subsequent publication, these fragments were combined and optimized into the first known rationally designed MTA, dubbed todalam, which is capable of modulating the dynamics of MTs (Fig. III-3, CD).[21] It was noticed that in the todalam site, the binders are proximal to residue αCys4 (Fig. III-3, C). This inspired us since covalent todalam-site binders could serve as molecular probes to immobilize tubulin. Together, these recent publications pave the way toward more rationally designed and easily accessible molecules, leaving room for the exploration and optimization of novel todalam site binders.

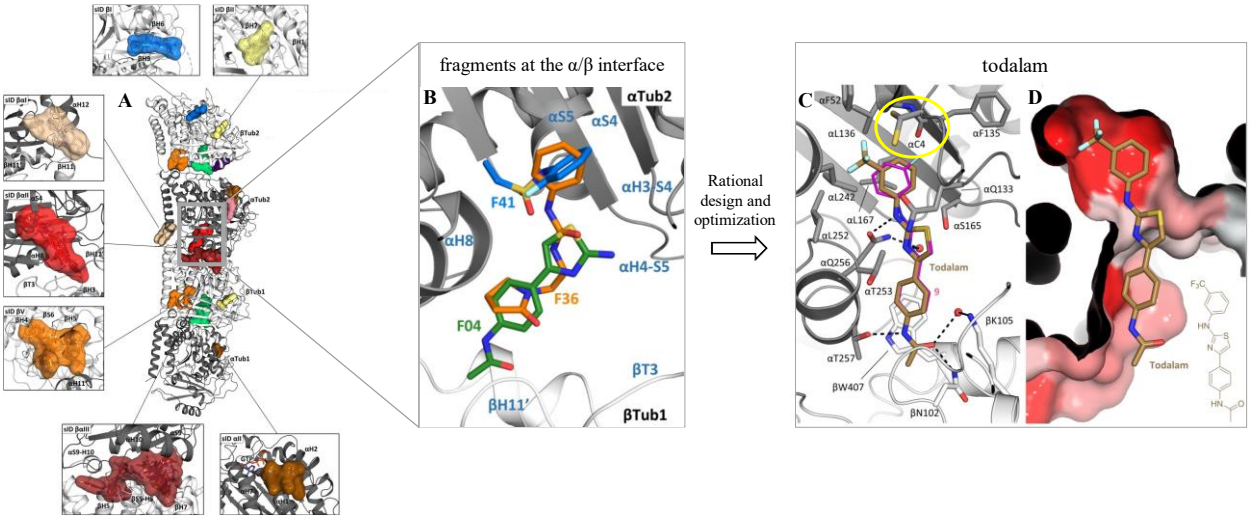


Figure III-3. A) Fragment-binding sites in tubulin determined by X-ray crystallography.[23] B) Superposition of the binding poses of fragment 04 (green), fragment 36 (orange) and fragment 41 (blue). C) Binding mode of todalam in the tubulin-todalam complex. D) Surface representation of the todalam site colored with a gradient from a red (hydrophobic) to a white (hydrophilic) environment. The todalam molecule is shown in the bottom right. These images are extracted from Mühlethaler *et al.*[21]

In the TubInTrain consortium, we pursued interdisciplinary research to explore the todalam site and design new ligands that could further be optimized into covalent binders by bringing together different fields of study (Fig. III-4). Five different research groups collaborated on this project. Due to the extensive structural data available on the todalam site, computer-aided structure-based molecular design approaches have allowed efficient exploration and identification of alternative chemotypes that can target this novel site. First, during my stay in the chemoinformatics group at the University of Strasbourg under the supervision of Dr. Horvath, I performed VS campaigns based on the substructure replacement of todalam and docking, which were successfully applied.

The best hits were subjected to MD simulations for docking validation and refinement. This step of the pipeline was carried out in collaboration with Anne-Catherine Abel at the University of Milan. This facilitated the flow of information between the two laboratories specialized in computational chemistry within the TubInTrain consortium. I collaborated with Anne-Catherine Abel on the analysis and interpretation of the results from the MD simulations she conducted. The selected binders were either purchased or synthesized by the chemistry group of Prof. Passarela (University of Milan). These computationally identified potential todalam site binders were later validated by observing their tubulin-bound X-ray structures at the Paul Scherrer Institute (PSI) in Switzerland and by *in vitro* MT-polymerization assays at the Consejo Superior de Investigaciones Científicas (CSIC) in Spain.

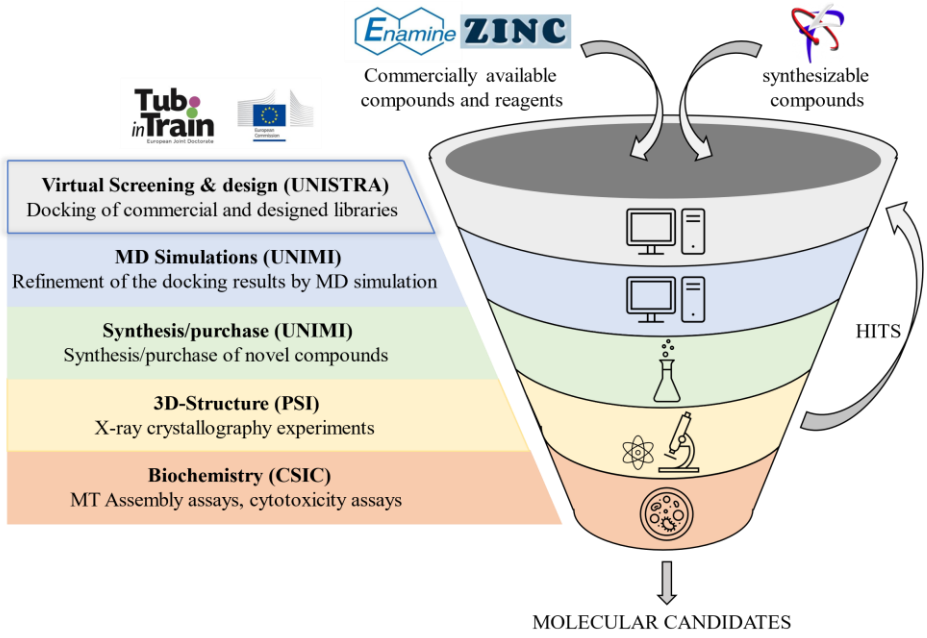


Figure III-4. TubInTrain workflow followed to computationally design and experimentally validate new todalam site binders.

This study was divided into two distinct stages: (1) scaffold hopping, aimed at identifying new chemical structures that can bind to the todalam site, and (2) the design of potential covalent binders based on the outcomes from the scaffold hopping step. Initially, five new scaffolds that are capable of binding to the todalam site were computationally identified and experimentally characterized, providing the foundation for subsequent computational structure-based design of todalam site binders that feature a warhead that can interact with the αCys4 present at the todalam site.

III.1 Scaffold Search: VS of Commercial Libraries and In-House Designed Compounds

III.1-1. Aim of the Work

The aim of this work was to perform VS based on todalam-targeted substructure replacement and docking, in order to discover alternative chemotypes (scaffold hopping) capable of targeting the todalam site.

III.1-2. Computational Details

III.1-2.1. Standardization of the Enamine Compound Library

To perform scaffold hopping around todalam, we performed VS of the Enamine[43] library and in-house-designed compounds. For this purpose, the Enamine compound dataset was first standardized. The standardization process helps to ensure that the compounds in the library are of high quality and are equally accurately represented in terms of their physical and chemical properties while avoiding the presence of duplicates. Standardization was carried out according to the approach implemented on the VS Web Server of the Laboratory of Chemoinformatics at the University of Strasbourg[71] using ChemAxon Standardizer.[72] This procedure includes: (1) dearomatization and final aromatization (heterocycles such as pyridone are not aromatized); (2) dealkalization; (3) conversion to canonical SMILES; (4) removal of salts and mixtures; (5) neutralization of all species, except nitrogen(IV); and, (6) generation of the major tautomer according to ChemAxon.

The Enamine library, which contains ~3M purchasable compounds, was filtered through several steps: (1) substructure search, (2) docking with PLANTS[73], (3) re-docking of the best candidates with S4MPLE[53,54], and (4) researcher selection criteria.

III.1-2.2. Substructure Search

First, the scaffold hopping strategy consisted in the search for any compound containing an acetaminophenyl group linked to a less specific aromatic ring head (Fig. III-5). The goal was to find analogues in which the central thiazole group (not seen to be involved in any specific interactions with the protein site) is replaced by other, unspecified 'linkers' of compatible length.

This substructure search was performed by running an in-house developed ChemAxon-powered Java script.[74] The substructure search workflow was carried out as described below:

1. Definition of the **substructure search query** (Fig. III-5). Design of a SMARTS query using the ChemAxon MarvinSketch tool.[75]
2. **Fast screening phase**. At this stage of the process, the query and database fingerprints are compared. Structures that fulfill the criteria for a good match are subsequently submitted to an atom-by-atom search (ABAS).
3. **Atom-by-atom search**. The ABAS algorithm works by systematically scanning a chemical structure and examining each individual atom and bond within it. The ABAS algorithm is used to compare the structures of multiple compounds and identify common features.
4. **Hit selection phase**. The cases in which all atoms and bonds of the substructure query completely overlap with a set of atoms and bonds of a compound from the Enamine library, the latter is considered a match and it is added to the hit list.
5. **Size filtering**. To avoid possible bulky molecules that cannot fit in the pocket due to steric effects, a filtering of the hit list by a maximal number of heavy atoms of 30 was performed.

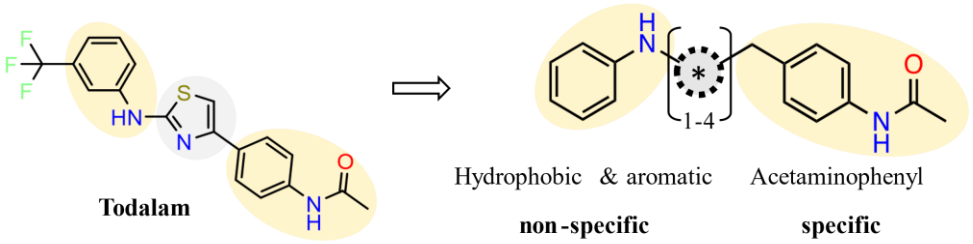


Figure III-5. Left: 2D structure of todalam. Right: SMARTS query used to perform a substructure search in the Enamine library.

III.1-2.3. Receptor Preparation for Docking

The crystallographic structure used for the docking studies was obtained from the PDB entry with ID 5SB7, which contains todalam in complex with tubulin. The process of preparing the binding site of the target protein structure for docking started by selecting the neighboring residues located <12 Å away from the todalam structure and deleting the rest of the atoms present in the structure, such as the crystallographic waters, cofactors, the structure of todalam, and other small molecules. Then, the binding site was hydrogenated. In this case, it was not necessary to perform the (de)protonation of any residue side chain according to the experimental pH (6.8). The preparation

of the todalam binding site was performed using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.[76] The protein site file was stored as 'ref.mol2' in Sybyl MOL2 format in a site directory, together with the fixed_atom list (here, all protein atoms were considered fixed, except for those shown in Fig. III-6A).

III.1-2.4. Docking with PLANTS

Subsequently, the resulting molecules obtained from the substructure search were docked using the PLANTS docking software, in which the posing algorithm is based on a class of stochastic optimization algorithms called ant colony optimization (ACO) with ChemPLP[73] as scoring function. Due to the high speed of PLANTS, this program was selected for rapid screening of the molecules output by the substructure search to further select those that presented a docking score equal to or lower to that of todalam. For this docking process, the default settings were kept. Todalam was re-docked with PLANTS to evaluate the performance of the docking algorithm and to use it as a reference ligand to compare its docking score with the ones obtained for the screened compounds.

III.1-2.5. Re-Docking with S4MPLE

The selected molecules from the VS performed in PLANTS were re-docked using the in-house developed docking software S4MPLE[53,54] for a more precise analysis of the binding modes acquired by the selected molecules at the todalam site. Todalam was also re-docked with S4MPLE since it was used again as a reference ligand.

S4MPLE allows the performance of flexible docking by selecting the amino acids to be freed during the calculations and, the identification of 'hot spots' that allows the initial docking binding modes to be more precise. S4MPLE tries to find clash-free geometries by roto-translating the ligand into a position that best satisfies the contacts of the ligand with the 'hot spots' selected at the binding site. The docking protocol used in this study consisted in the performance of flexible docking by keeping unfixed the protein side chains of specific amino acids during the calculations which are shown in Fig. III-6A. The atoms that were selected as 'hot spots' or preferential anchoring points at the todalam site are shown as spheres in Fig. III-6B.

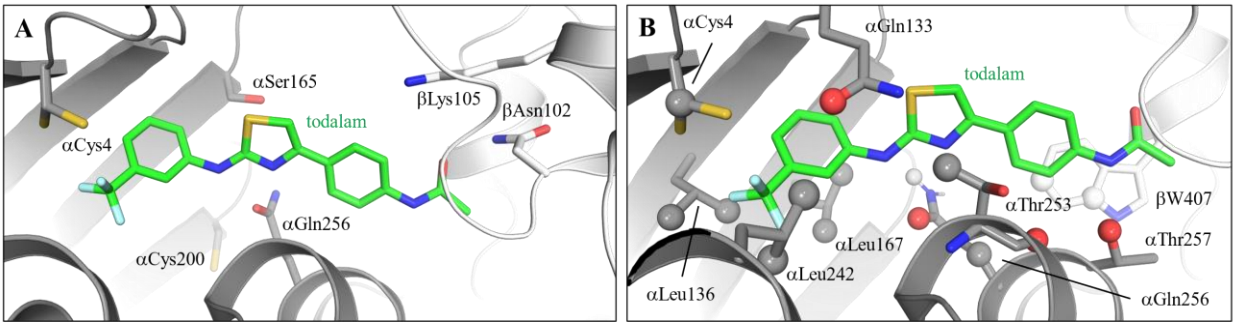


Figure III-6. A) Stick representation of the amino acids that were selected to be unfixed and, B) spheric representation of the atoms selected as 'hot spots' within the todalam binding site (gray) during the docking calculations using S4MPLE. Todalam is represented as green sticks, shown to highlight important amino acids and atoms for the binding of todalam to tubulin. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively.

III.1-2.5.1. Ligand Preparation for Non-Covalent Docking with S4MPLE

The preparation of the ligand for non-covalent docking was carried out by a Web service of the Laboratory of Chemoinformatics at the University of Strasbourg.[77] This server allows the submission of a list of <1000 SMILES strings and sequentially processes each compound with the following steps:

1. **Standardization** of the compound as for the Enamine library (see section III.1-2.1.).
2. Full **hydrogenation** and **protonation state** setup according to the pKa predictor modules of ChemAxon.[74]
3. Detection of **saturated ring systems** and automated suggestions of '**broken bonds**'. The 'broken bonds' may be necessary to be determined to help S4MPLE autodetect key intramolecular degrees of freedom that can account for the torsional axes to which genetic operators (CO and mutation) will be applied.
4. **2D to 3D structure conversion** using the conformer plugin of ChemAxon.[74]
5. **Assignment of partial charges** by the Gasteiger method.[78]
6. **GAFF**[57]**parameter assignment**, with occasional creation of novel GAFF parameters added to the GAFF parameter library.

The internally designed focused library of 176 triazole-containing molecules was directly docked using S4MPLE with the above-described settings.

III.1-2.5.2. Final Selection Criteria

The criteria of selection of the molecules from the S4MPLE re-docking results were mainly substantiated in 3 aspects: (1) docking score, (2) binding mode, and (3) tubulin-ligand established interactions. Therefore, those molecules with a better or equal docking score to that of todalam, capable of mimicking the todalam binding mode and establishing key interactions with tubulin, were selected for further experimental analysis.

III.1-3. Results and Discussion

To create a new library of todalam site binders, we first collected the available structural information to better direct rational identification of potential hits. By using information about an already known ligand as a starting point, we can focus our efforts on synthesizing and testing molecules that are expected to have a higher likelihood of success, rather than starting from scratch and trying to perform *de novo* design of a molecule. The 2D structure of todalam has been divided in this Ph.D. thesis into three different moieties (Fig. III-7, left):

- The **anchor point** is represented by the acetaminophenyl group.
- The **central linker** consists of an aminothiazole group.
- The **hydrophobic head** contains a hydrophobic trifluoromethylbenzene group.

These three moieties that make up the structure of todalam lie in different parts of the todalam site in the α,β-tubulin inter-dimer interface and are surrounded by distinct residues with various physico-chemical properties (Fig. III-7, right).

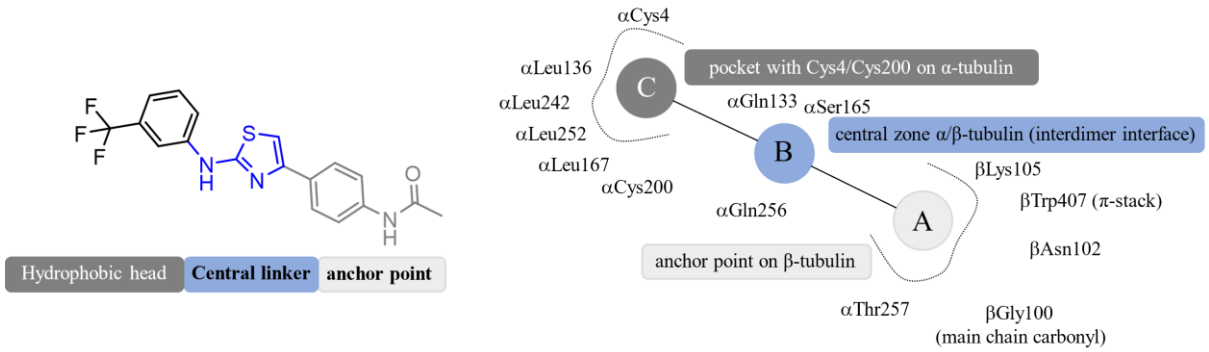


Figure III-7. Left: Chemical structure of todalam and its schematic division into three moieties: anchor point, central linker, and hydrophobic head. Right: Schematic 2D representation of the location of the different moieties of todalam within its binding site, highlighting the most important neighboring residues.

III.1-3.1. Anchor Point Analysis

To provide a better foundation for the design and identification of todalam site binders, we first verified the ability and specificity of the anchor point to bind to the pocket. For this, additional experiments were carried out by the X-ray crystallography group at PSI. They performed soaking experiments using the tubulin system T_2R-TTL with the ligand N-(4-aminophenyl)acrylamide (ZB57). The T_2R-TTL tubulin crystallization system (Fig. III-8) is formed by a protein complex composed of two bovine brain α,β-tubulin heterodimers, the rat stathmin-like protein RB3, and the chicken tubulin tyrosine ligase (TTL).

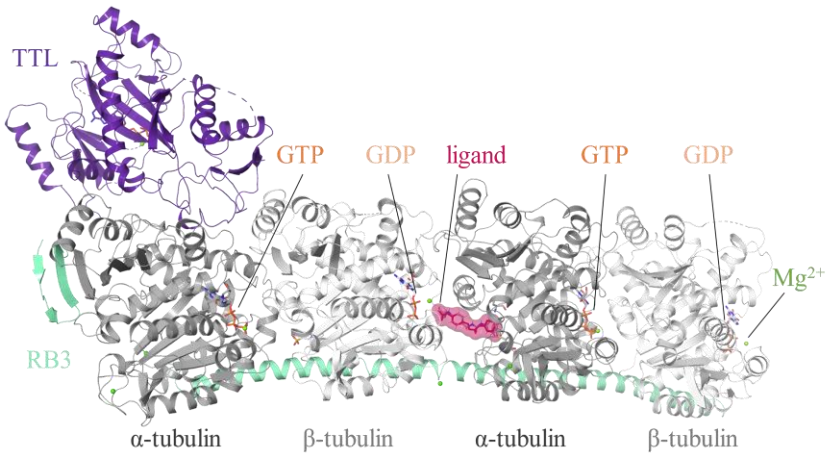


Figure III-8. Overall view of the T_2R-TTL complex structure bound to todalam (PDB 5SB7). The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively. TTL is colored in violet and RB3 is represented in light green ribbon, respectively. The protein residues are shown in ribbon representation while the bound todalam site ligand is represented in a pink surface.

The results showed that the binding mode acquired by ZB57 within the todalam site can be perfectly superimposed on that obtained for fragment 04 (PDB ID 5S4O)[23] and both are able to establish the same key interactions at the site (Fig. III-9, left). Fragment 04 is one of the fragments identified in the todalam site after the fragment screening campaign performed by Mühlethaler *et al.* and that was part of the rational design of todalam (Fig. III-3). These results showed that the anchor moiety specifically binds to the 'A' section of the todalam site, as depicted in Fig. III-7. Furthermore, the presence of an acetyl or acrylamide group in the anchor point does not appear to affect its binding mode or interactions. The anchor moiety in todalam establishes a hydrogen bond with residue βAsn102 and another hydrogen bond with residue αThr257 as well as parallel-

displaced π-π stacking interactions with βTrp407 (Fig. III-9, right), which are conserved as well in the ZB57 and fragment 04 conformation.

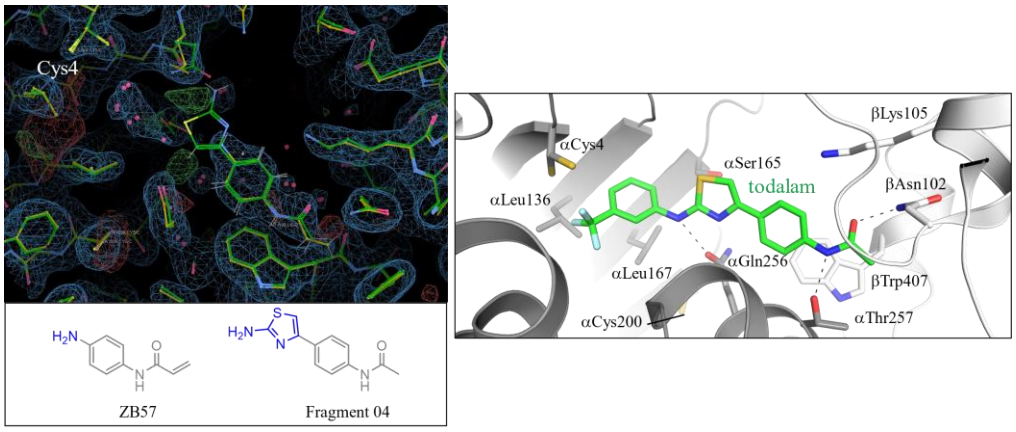


Figure III-9. Left: Superimposition of electron density maps of compounds ZB57 and fragment 04 bound to the todalam site of tubulin in their corresponding T_2R-TTL complexes. The direct electron density is contoured at 1 sigma and colored blue, the mFc-DFc map is contoured at +/- 3 sigma and colored green and red, respectively. Protein residues are indicated by line representation, the carbon atoms of the structures T_2R-TTL-ZB57 and T_2R-TTL-frag04 are colored yellow and green, respectively. Oxygen, nitrogen and sulfur atoms are colored red, blue and dark yellow, respectively. Right: Binding mode of todalam in the tubulin-todalam complex (extracted from PDB 5SB7). Hydrogen bond interactions formed between todalam and tubulin are displayed as dashed black lines. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, fluorine atoms in cyan, and sulfur atoms in golden. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively.

III.1-3.2. Central Linker

In the central area of the todalam site in which the linker of the todalam structure is buried, there are two different glutamine residues present (αGln133 and αGln256) and a serine residue (αSer165). Both glutamine residues present two hydrogen bond donors (HBDs) and two hydrogen bond acceptors (HBAs), and the serine residue two HBDs and one HBA (Fig. III-10). These residues could be exploited to promote binding of the newly designed molecules to the todalam site *via* the formation of protein-ligand hydrogen bonds, thereby fixing the molecules within the site.

III.1-3.3. Analysis of the Hydrophobic Head

The area of the todalam site opposite that of the anchor point, represented in Fig. III-7 as 'C', presents a high content of non-polar amino acids (αLeu136, αLeu167, αLeu242, αLeu252) (Fig. III-10). Therefore, an unspecific hydrophobic and aromatic group, also contained in the molecular structure of todalam, would be a well-suited substructure feature to add to the search for new todalam-site binders.

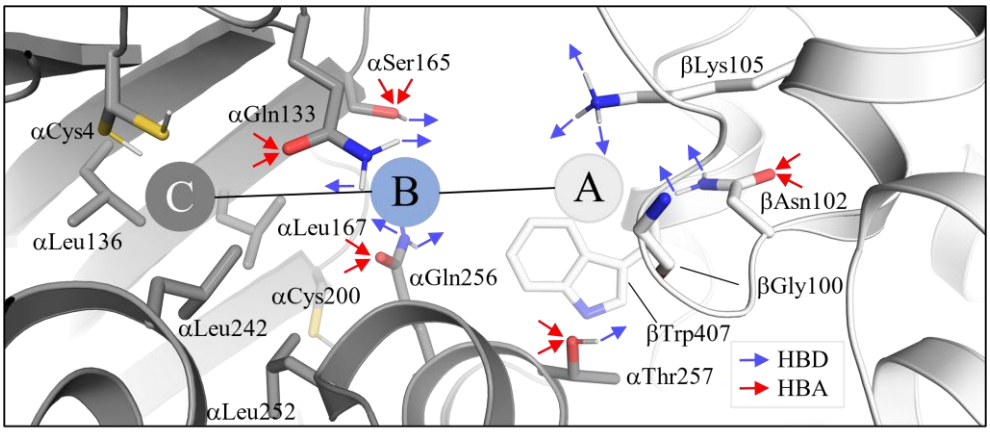


Figure III-10. 3D representation of the key residues present in the todalam binding site marking with arrows the atoms that could act as HBD (blue arrow) and HBA (red arrow). Furthermore, the schematic representation of the different moieties of todalam within its binding site shown in Fig. III-7 is also included in this figure to clarify their location in the 3D structure. Side chain of residues involved in key interactions are shown as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, and sulfur atoms in golden. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively.

III.1-3.4. VS of the Enamine Library

The substructure search of the Enamine library reduced the number of molecules to screen from ~3M to ~3K molecules. This approach allowed us to filter out compounds by a 1000-fold reduction of the dataset in a very fast and efficient manner. Subsequently, these ~3K compounds were docked using PLANTS docking software, finding that 590 of them had a docking score equal to or lower than that obtained for todalam. The 590 best candidates were submitted to a more focused and throughout analysis, starting by re-docking them using S4MPLE. Since the todalam docking score predicted by S4MPLE was ~ -67 kcal/mol, we selected all the molecules that had a docking score equal or lower than -67 kcal/mol, resulting in a total of 50 compounds from the 590 re-docked with S4MPLE. From these 50 selected molecules, we performed another filtering step that was dependent on the binding mode of the compounds with respect to that of the crystal structure of the todalam complex with tubulin and its established interactions with tubulin. Following this protocol (Fig. III-11), from the ~3M screened Enamine molecules, 12 were selected as potential hits to be further purchased and experimentally tested (Fig. III-12).

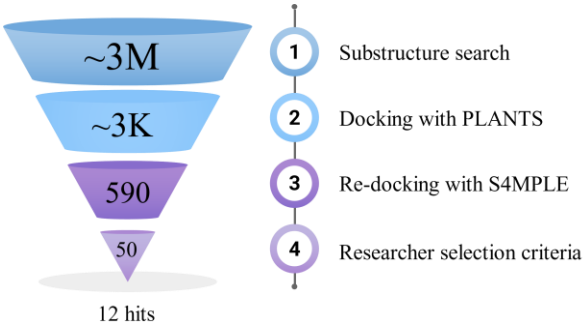


Figure III-11. VS funnel representing the number of molecules present in each step of the VS campaign and the computational approach used in each step.

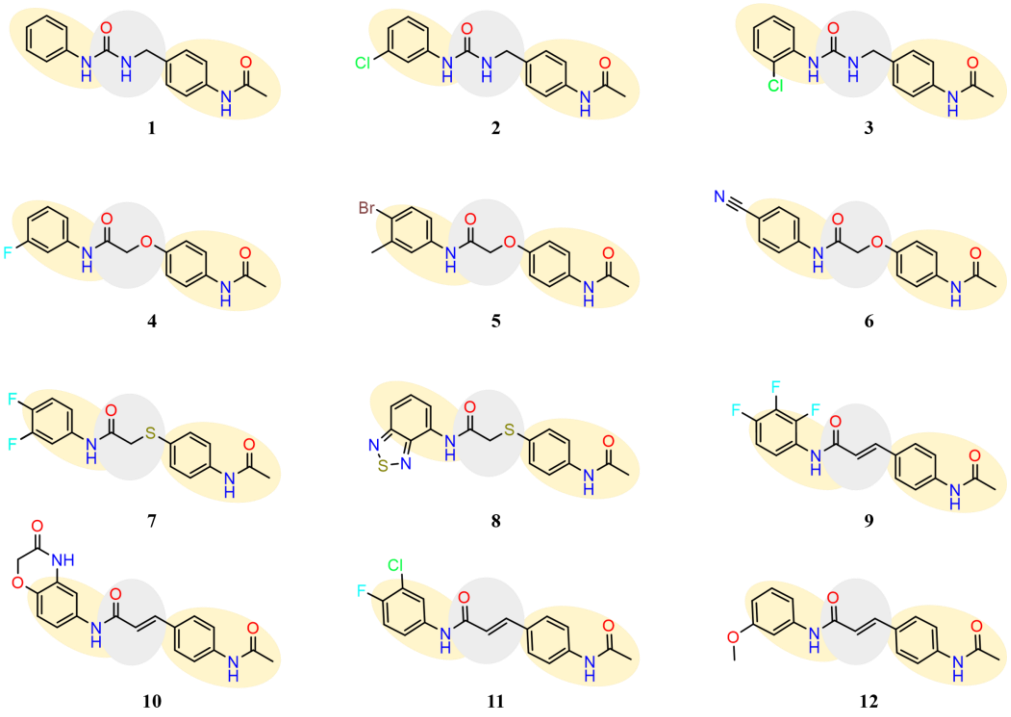


Figure III-12. 2D structure of the top 12 hits obtained after the VS of the ~3M purchasable Enamine compounds. The yellow and gray ovals represent the chemical groups found in the search related to the SMART query established from the todalam structure (Fig. III-5).

As mentioned above, in the selection process, it was important to identify those compounds capable of binding to the todalam site in a mode similar to that of todalam. To simplify the analysis of these binding modes, Fig. III-13 shows the superimposition of all the best conformers of the selected hits that share the same scaffold, with respect to the crystallographic structure of todalam. As can be seen, all hits are able to similarly reproduce the binding mode of todalam and, therefore, were selected.

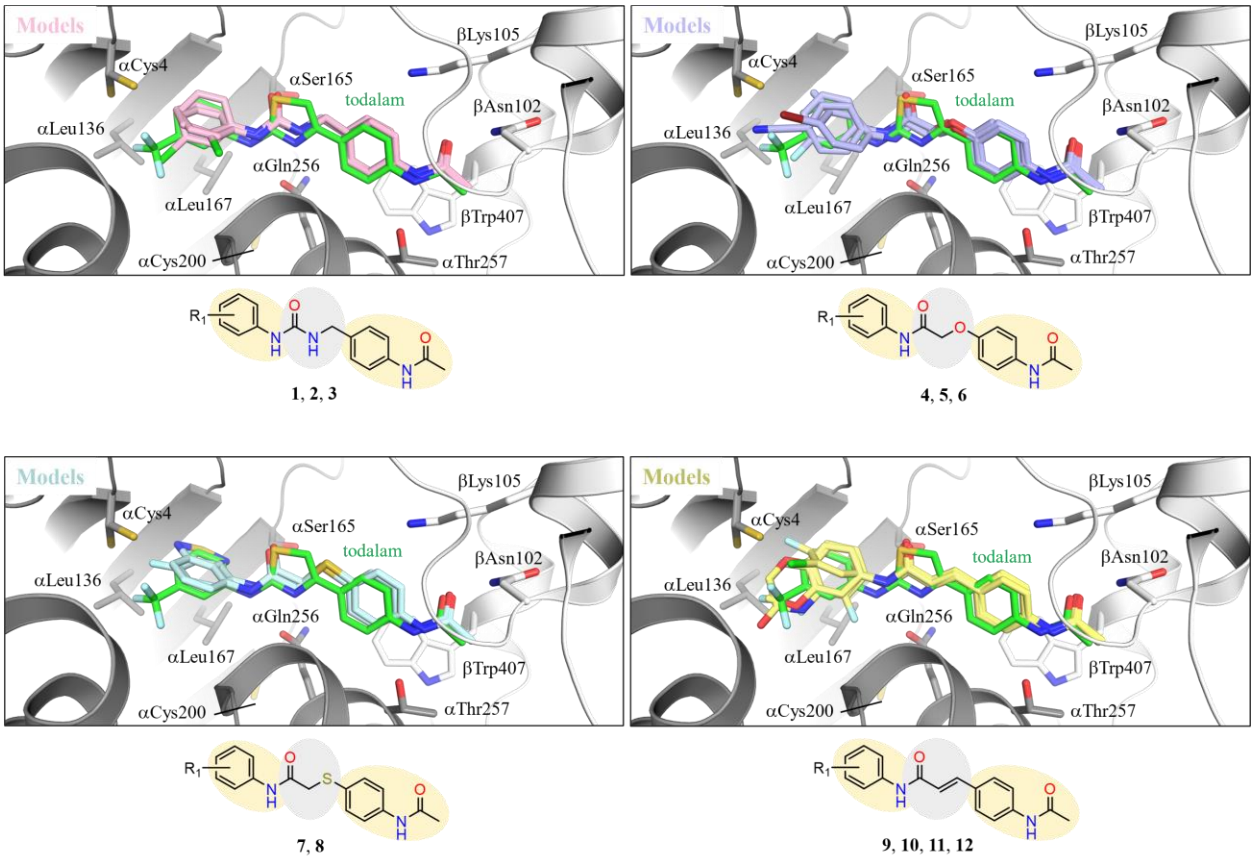


Figure III-13. Superimposition of the selected hits from the Enamine library with todalam (green). The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, chlorine atoms in green, fluorine atoms in cyan, and bromine atoms in brown. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively.

The 12 selected hits can be clustered into three different groups according to the chemical group present in the central linker of the molecules (Fig. III-14): (1) three hits contain an urea group; (2) five hits an α-substituted amide, three of them are substituted with an oxygen atom and two with a sulfur atom; (3) four hits contain an acrylamide group (Fig. III-12).

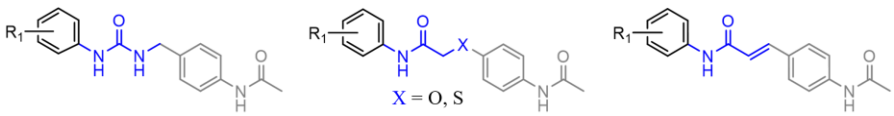


Figure III-14. Schematic representation of the 2D generic structure of the three different scaffolds representing the 12 selected hits. The atoms of the anchor point are colored light gray, those of the central linker in navy blue, and those of the hydrophobic head in black.

Furthermore, analysis of the intermolecular interactions formed between each selected hit and the residues located at the todalam binding site was performed. Those molecules that present the same central linker establish very similar interactions with tubulin. Therefore, to simplify the analysis of the interactions, only one molecule containing each scaffold is shown in Fig. III-15.

According to the analysis of the interactions that can be observed in Fig. III-15, in all the selected hits, the main interactions of the anchor point with tubulin are conserved. The hydrogen bonds established between their anchor point and the polar side chains of the tubulin residues βAsn102 and αThr257 are maintained, as well as the parallel-displaced π-π stacking interactions with the aromatic side chain of βTrp407.

The three molecules containing an urea group in the central linker of the structure are able to form three hydrogen bond interactions with tubulin in the 'B' section of the molecules. This is because the urea group contains two HBD (-NH) that are located in close proximity to the oxygen atom of the carbonyl group of the side chain of residue αGln256 which, as shown in Fig. III-10, is able to form two hydrogen bond interactions acting as an HBA. Furthermore, the carbonyl group of the urea group that can behave as an HBA is facing the hydroxyl group of the side chain of αSer165 which acts as an HBD.

In the case of the hits that contain an α-substituted amide or an acrylamide in the central linker of the structure, they are able to establish the same two hydrogen bond interactions with tubulin through the amide group of these central linkers. As in the case of the urea, the amide group interacts through the amino group that acts as a HBD with the carbonyl group of the polar the side chain of residue αGln256, whereas the carbonyl group of the amide establishes a hydrogen bond with the hydroxyl group of the polar side chain of αSer165.

The aromatic rings of all selected hits fit nicely in the hydrophobic pocket of the todalam site formed by several leucine residues, also described as section 'C' (Fig. III-7). They lay close to αCys4 but not sufficiently to reach its thiol group and interact covalently with it. Nevertheless, this was not the goal in this first step of the project, being our aim setting the basis for further development of these scaffolds into warhead-containing todalam site binders that can covalently react with αCys4.

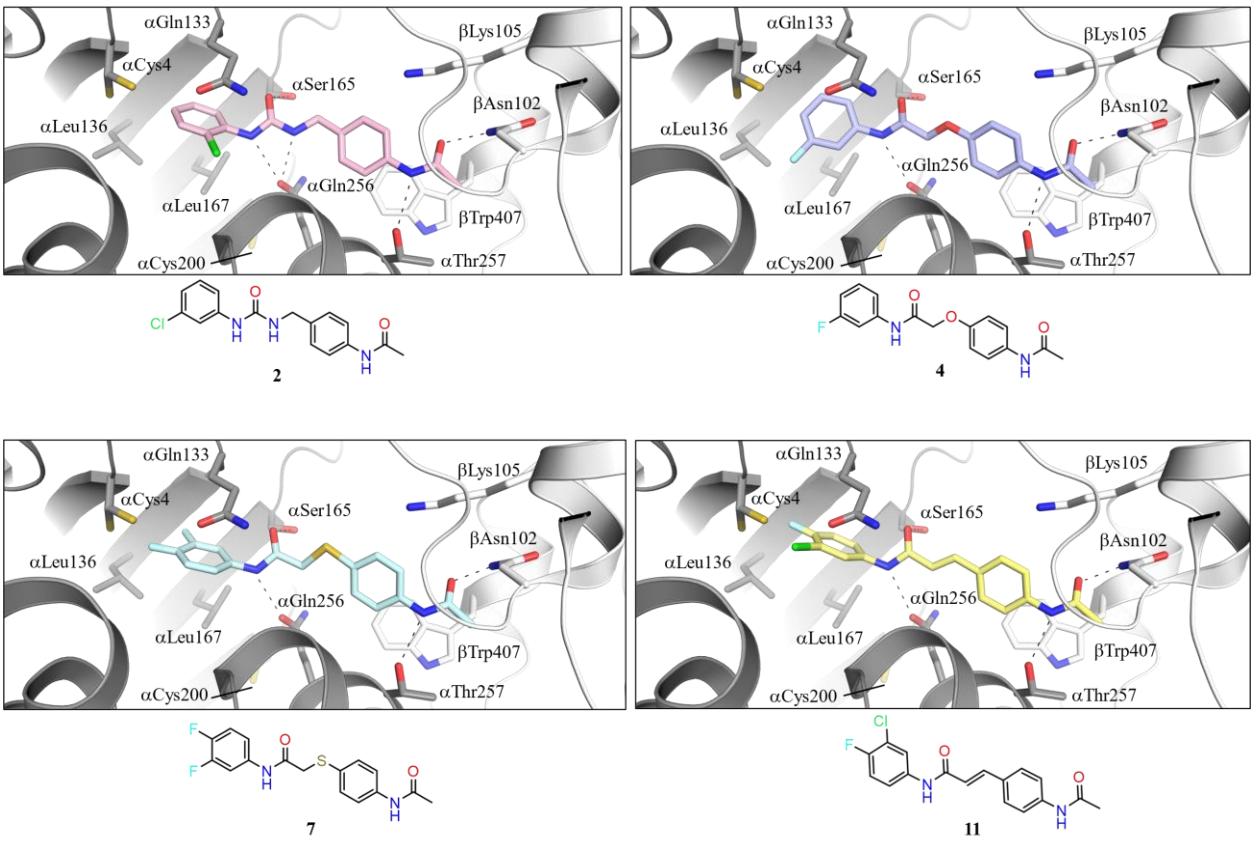


Figure III-15. Close-up of the interaction between four of the 12 hits and tubulin (gray). The α-tubulin and β-tubulin are represented in dark and light gray ribbon, respectively. The interacting residues and ligand are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, chlorine atoms in green and fluorine atoms in cyan. Hydrogen bonds are displayed as black dashed lines.

Overall, according to our *in silico* studies, all the selected hits form very similar interactions when in complex with tubulin, with the urea scaffold as the one that allows the formation of the highest number of hydrogen bonds (Fig. III-15).

III.1-3.5. VS of the In-House Design Compound Library

The synthetic chemistry group designed a library of 176 compounds containing all the possible molecules that could be synthesized in one or two steps using the reagents available in their laboratory. The central linker of many of the compounds consisted of a triazole ring, since the chemistry team planned to perform 'click chemistry' to produce the molecules. After the VS campaign of this set of 176 compounds that was carried out by flexible docking with S4MPLE, 3 hits were selected as potential todalam site binders (Fig. III-16).

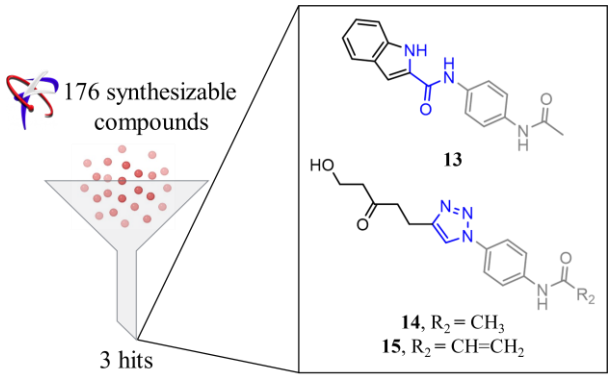


Figure III-16. 2D structure of the 3 selected hits out of the 176 in-house designed molecules. The atoms of the anchor point are colored light gray, those of the central linker in navy blue, and those of the hydrophobic head in black.

To analyze the results of these 3 hits, we also superimposed the binding mode of the best conformer of each hit obtained with S4MPLE on that of todalam (Fig. III-17) and later investigated the interactions formed by the hits within the todalam site (Fig. III-19). The binding pose of the selected hits within the todalam site closely resembles the one described for todalam, although not as precisely as the hits obtained from the Enamine library, but still acceptable (Fig. III-17). Moreover, we observed once more that the presence of $-CH_3$ or $-CH=CH_2$ in the anchor point did not compromise the binding mode of the molecule (Fig. III-17, right) as already shown before in Fig. III-9.

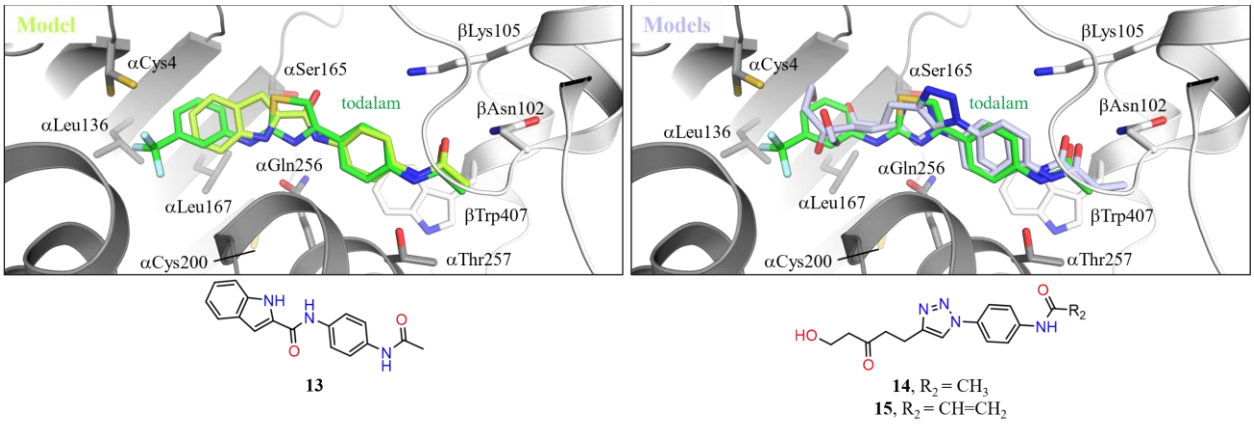


Figure III-17. Superimposition of the potential hits from the in-house designed library with todalam (green). The α-tubulin and β-tubulin are represented in dark and light gray ribbon, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, and fluorine atoms in cyan.

The central linker of the hits could be divided into 2 groups: (1) two hits contain a triazole ring, and (2) one hit with an indole-2-carboxamide group (Fig. III-18).

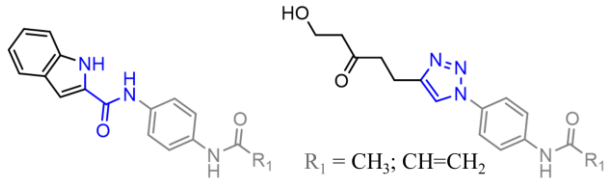


Figure III-18. Schematic representation of the 2D generic structure of the 2 different scaffolds representing the 3 selected hits. The atoms of the anchor point are colored light gray, those of the central linker in navy blue, and those of the hydrophobic part in black.

In Fig. III-19, we can examine the spatial relationships between the atoms of each compound and the residues in the binding site. In the evaluation of ligand-protein interactions, we observed the specific interactions established between the hits and the protein in the docking pose. This involved looking again at hydrogen bonding patterns and hydrophobic interactions to see if they are consistent with those of todalam, while bringing some chemical diversity.

The anchor point was able to be accommodated in the same area of the todalam binding site, establishing the same interactions as those formed by todalam and the selected hits from the Enamine library. The triazole group forms a hydrogen bond with the hydroxyl group of the side chain of αSer165, whereas the indole-4-carboxamide group interacts through its two amino groups by hydrogen bonding with the carbonyl group of the side chain of αGln256.

The aromatic ring as well as the aliphatic chain of the triazole derivatives nicely accommodate in the hydrophobic pocket of the todalam site also referred to as section 'C' in Fig. III-7. These results provided new structural information since it was predicted that substituents of both aliphatic and aromatic nature can adapt in the section 'C' of the todalam site.

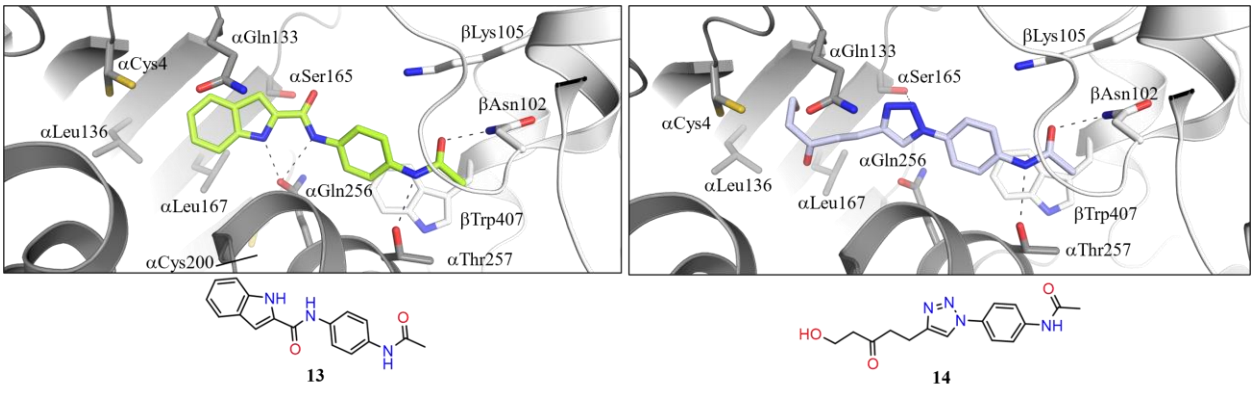


Figure III-19. Close-up of the interaction between two of the 3 hits and tubulin (gray). The α-tubulin and β-tubulin are represented in dark and light gray ribbon, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red and nitrogen atoms in navy blue. Hydrogen bonds are displayed as black dashed lines.

III.1-3.6. Docking Validation *via* MD Simulations

As a second computational selection process, the best candidate molecules from docking studies were submitted to MD simulations to study the established interactions with the protein and to validate their stability within the binding site. MD simulations of the docked ligand-receptor complex were performed in collaboration with Anne-Catherine Abel in the laboratory of Prof. Pieraccini. This gave some insights into the dynamic behavior of the complex and how well the docking pose captures the dynamic binding mode of each selected hit.

MD simulations were performed using the gpu-accelerated engine GROMACS2021.[79,80] The input files were prepared using the AmberTools2021[81] suite with Leap. The antechamber[82] tool was used for ligand parametrization and Gaussian16[83] to calculate its partial charges. The command-line software acpype[84] was used to convert the Amber topology and coordinate files to GROMACS format. Subsequently, the parameters of the partial charges of the GTP and GDP nucleotides, the bond and torsion angles, in addition to the atom types, were retrieved from the AMBER Parameter Database.[85] The Mg^{2+} and Ca^{2+} ions present in the input-crystal structure with PDB ID 5SB7 were preserved. The structure was completed by superposition with PDB ID: 5S5D and superimposed onto the docked structure to generate the tubulin-ligand complex. The selected FFs were ff14SB[86] (protein) and GAFF[57] (ligand). The molecules were explicitly solvated in 10.0 Å buffer with TIP3P[87] water molecules using additional Na^{+} ions to neutralize the charges of the system. Protein complexes were energy minimized using the steepest steps descent algorithm, followed by 100 ps temperature equilibration and 100 ps pressure equilibration. During the equilibration steps Cα atoms were positionally restrained. Simulations were run for 100 ns with a step size of 2 fs at 300 K under constant pressure and temperature and periodic boundary conditions. The leapfrog[65] algorithm was used and bonds were constrained, a 1.5 Å cutoff was used to treat Particle Mesh Ewald (PME)[88] interactions.

RMSD and h-bond numbers and distances were calculated using the GROMACs tools *gmx rms*, *gmx hbond* and *gmx distance* superposing the structures over Cα-atoms and calculating values for every 100th frame (0.1 ns, averages calculated over 1000 values). PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, was used to visualize the simulations and interactions as well as to generate the shown graphics.

Due to the position of the todalam site at the inter-dimer interface, special considerations were done for the set-up of the simulated system. To use computational resources in a most efficient manner, we explored the possibility of simulating an artificial α,β-tubulin dimer instead of two adjacent tubulin dimers to reduce the overall size of the system from 160 000 atoms to 84 000 atoms increasing the simulation speed by a factor of 2 (Fig. III-20).

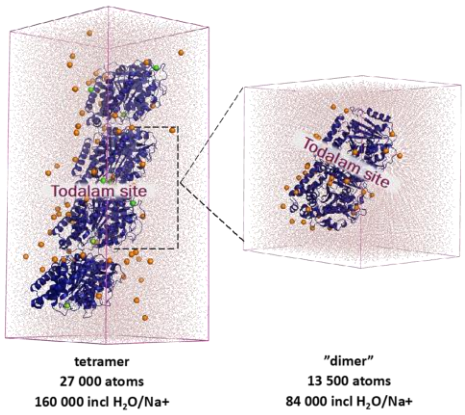


Figure III-20. Tubulin tetramer and artificial dimer submitted to MD simulations for system refinement shown as navy blue ribbon structures. Water molecules are shown as red dots. Ions Mg^{2+} and Ca^{2+} are represented as green spheres and Na^{+} as orange spheres. The box in which each system is embedded is represented as a pink box.

We carefully evaluated the implications of this artificial set-up on the binding site by comparison of simulations of one ligand in both possible systems. For instance, the hydrogen bond distances of the ligand to the anchor residues mentioned above were calculated in the tetramer and in the artificial dimer. The results did not indicate any disturbance of the binding mode of the molecule throughout the simulations due to the artificial system setup. Therefore, this artificial dimer was the system of choice for simulating the docking hits.

Then, for the analysis of each hit, the following criteria were used to rank and select the evaluated molecules: plateau of the RMSD values of the ligand under study, average number of established hydrogen bonds, persistence of anchoring-hydrogen bonds to residues αThr257, αGln256 and βAsn102 apart from additional hydrogen bonds depending on the compound. For instance, here are some examples of the analysis made for compounds **2** and **3**. The RMSD values of each compound were calculated along the time of the simulations in order to analyze the overall stability of the docking conformations within the todalam site in 100 ns (Fig. III-21).

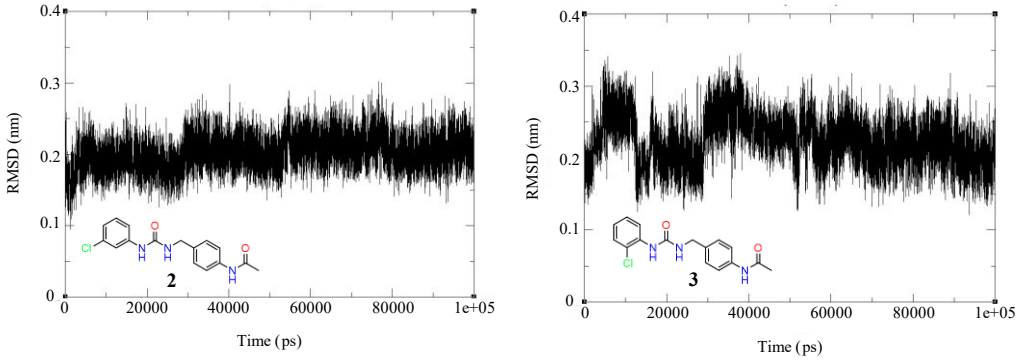


Figure III-21. RMSD values of compounds **2** and **3** after fitting the protein Cα for every protein ligand complex during the simulations.

Additionally, the persistence of hydrogen bonding interactions between atoms in the docking conformer and the todalam site and other relevant protein-ligand atoms was evaluated by calculating the average distance between each pair of atoms over the course of the simulation, as shown in Fig. III-22.

hbond	Compound 2 (av. dist. in Å)	Compound 3 (av. dist. in Å)
O1 – Asn102	2.89	2.92
N1 – Thr257	3.11	3.03
N2 – Gln256	3.05	3.20
N3 – Gln256	2.85	3.37
O2 – Ser165	3.38	4.09
O2 – Gln133	4.97	5.45

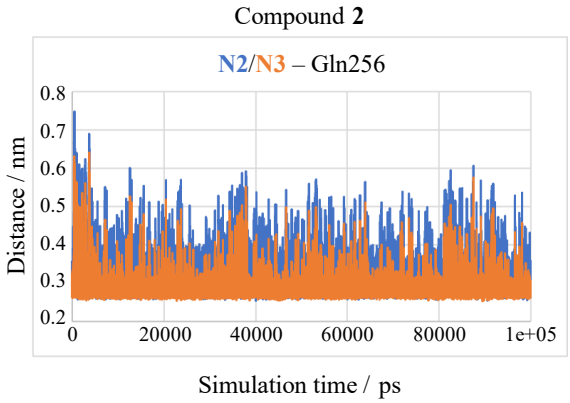


Figure III-22. Left: Table summarizing the average distance between each described pair of atoms. Right: Example of a graph displaying the distance in nm between atom N2 and N3 of compound **2** with respect to Gln256 for each frame in the simulations.

In this way, the docking poses were validated using MD simulations by comparing the predicted binding pose of each hit to the resulting binding poses obtained from the simulation. One way to do this was to calculate the RMSD values between the predicted and simulated binding poses. In all cases, the RMSD values were in average lower than 3 Å, which indicated a good agreement between the predicted and simulated binding poses, and suggests that the predicted binding pose is likely to be a good approximation of the true binding pose. Additionally, it was also possible to evaluate the stability of each protein-ligand complex during the MD simulation by measuring the fluctuations of the intermolecular distances, and by analyzing the formation of the hydrogen bonding network.

III.1-3.7. Experimental Validation

Although the molecules sharing the same scaffold establish very similar interactions with tubulin, several molecules of each scaffold were chosen to be tested experimentally to investigate the effect of these specific substituents on the properties of the molecules. For example, by comparing molecules with the same scaffold but with different substituents, it is possible to determine how these substituents affect the reactivity, solubility, and other characteristics of the molecule. The strategy of testing a number of related molecules was implemented in order to obtain structure-activity data and establish an analogue series, which is typically used to improve the primary hit into a lead compound.

III.1-3.7.1. X-Ray Crystallography Assays

All the 15 computationally selected todalam-site candidates were shipped to Paul Scherrer Institut (PSI) in Villigen, Switzerland, in order to determine their ability to bind to the todalam site of tubulin. X-ray crystallography experiments were performed by Zlata Boiarska, using the T_2R-TTL tubulin crystallization system (Fig. III-8) applying a soaking protocol.

Following this approach, 12 out of the 15 compounds submitted to X-ray crystallography experiments were seen to be bound to the todalam site. The bound compounds belonged to all found scaffolds: urea, α-substituted amide, acrylamide, triazole and indole-2-carboxamide. Some examples of the results from X-ray crystallography are shown in Fig. III-23. In particular, the crystallographic structures of compounds **2**, **4** and **14** are displayed with their respective electron densities.

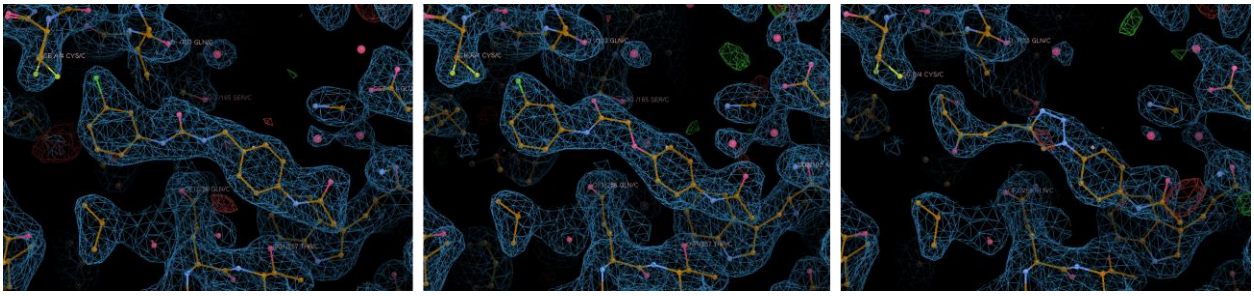


Figure III-23. Electron-density maps of compounds **2** (left), **4** (middle) and **14** (right) bound to the todalam site of tubulin in their corresponding T_2R-TTL complexes. The SigmaA-weighted 2mFo - DFc (deep blue mesh) and mFo-DFc (green mesh) maps are contoured at +1.0 and +3.0, respectively.

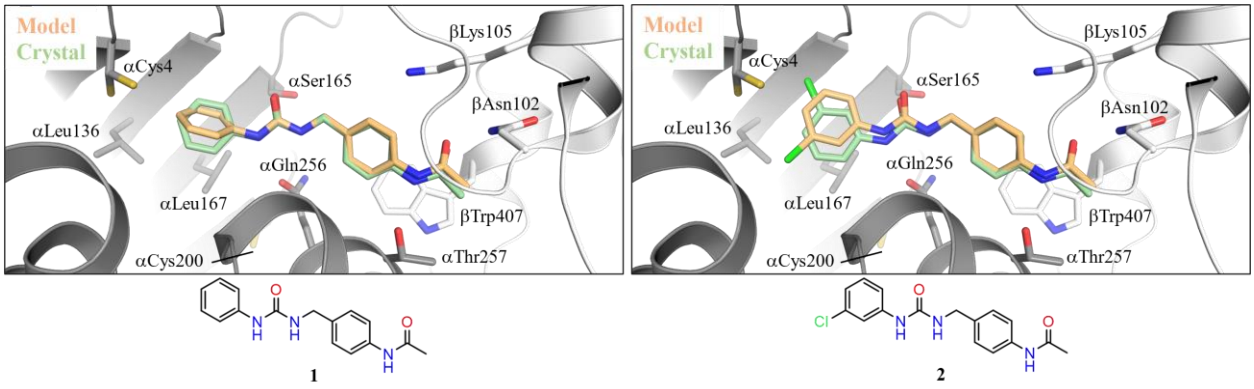
One way to validate the docking results of a computational study is to compare the predicted binding pose of the ligand (the molecule being docked) with the experimentally determined

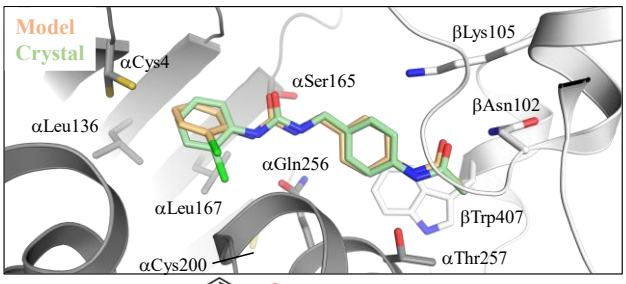
binding pose in the crystal structure. This was envisaged in a first step by calculating the RMSD values for the corresponding atoms between the crystal structure and the docked structure of each ligand. This allowed us to quantitatively assess the structural differences between the two conformers and obtain an objective measure of the accuracy of the docking results. A lower RMSD value indicates a better match between the predicted and crystal binding poses, and therefore a more accurate docking prediction. In general, an RMSD value below 2.0 Å is considered to be good, while values below 1.0 Å are considered to be excellent. In our case, all the hits presented RMSD values lower than 2.0 Å while two of them presented RMSD values lower than 1.0 Å (Tab. III-1) which showed the high accuracy of the software S4MPLE.

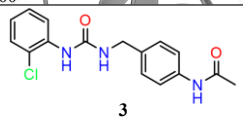
Table III-1. RMSD values calculated between the predicted and crystal binding poses of each compound.

Compound	1	2	3	4	5	7	9	11	12	13	14	15
RMSD (Å)	0.63	1.47	0.78	1.54	1.78	1.25	1.70	1.53	1.69	1.76	0.89	1.75

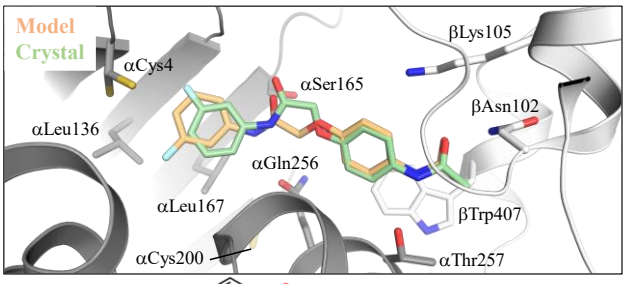
It is also useful to perform a visual inspection of the docking prediction and the crystal structure to visually compare the binding poses. This was done using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, which allowed us to superimpose the two structures (model *vs* crystal) and assess the overall fit. In Fig. III-24, a good overlap between each model and the corresponding crystal structure can be observed, especially considering the overall shape and conformation of each ligand.

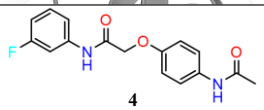




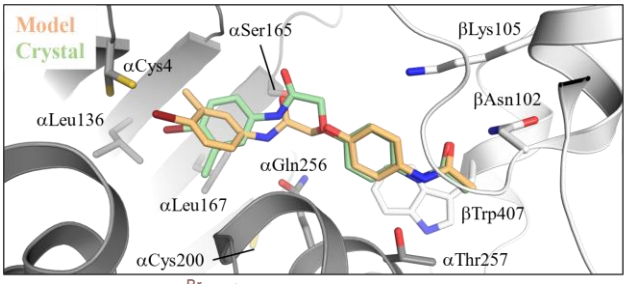


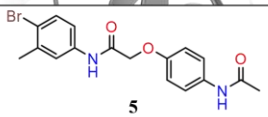
3



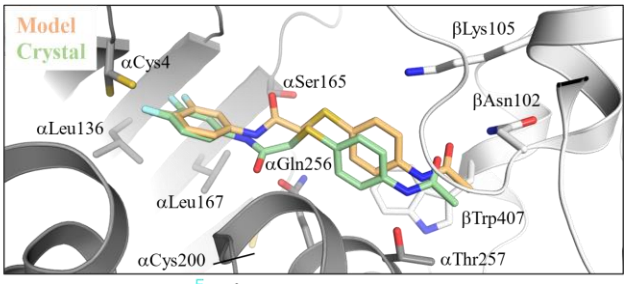


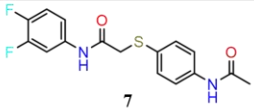
4



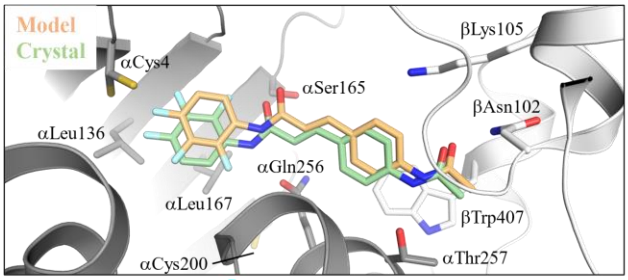


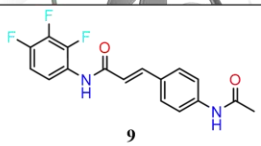
5



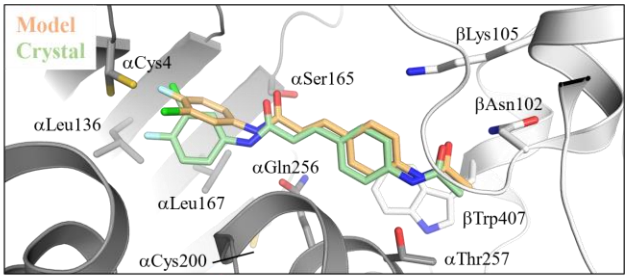


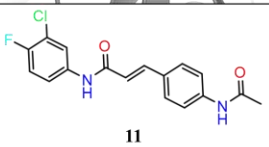
7



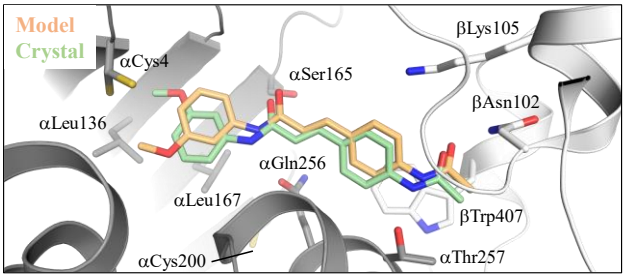


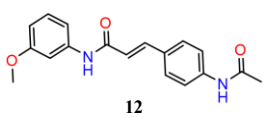
9



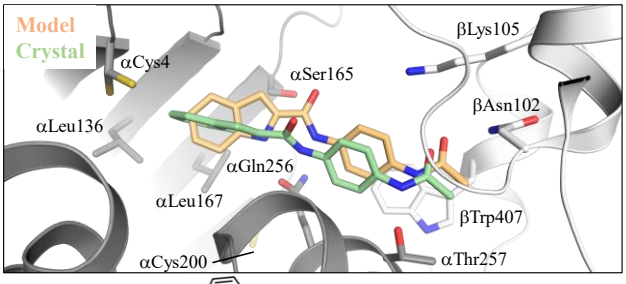


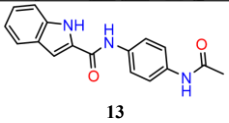
11





12





13

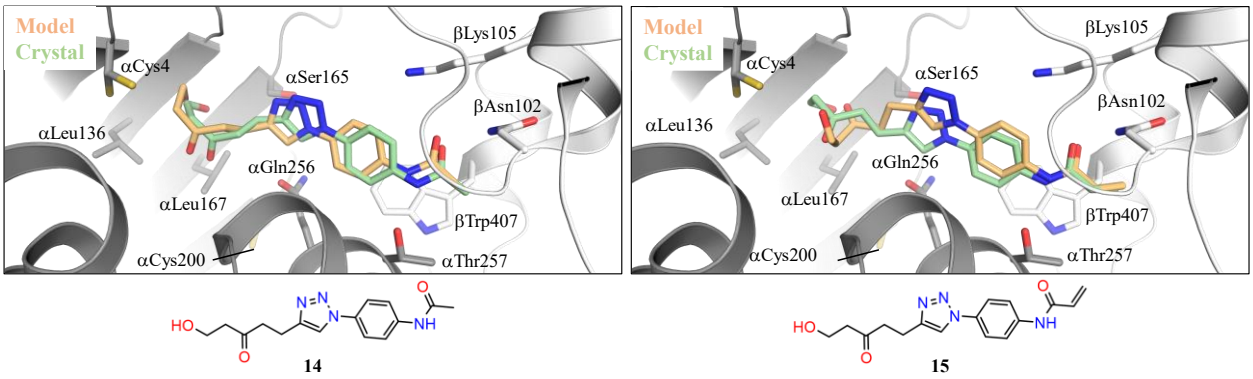


Figure III-24. Superimposition of the predicted (light orange) and crystal (light green) binding poses of each compound. The α-tubulin and β-tubulin are represented in dark and light gray ribbon, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, chlorine atoms in green, fluorine atoms in cyan, and bromine atoms in brown.

After analyzing the established interactions of the solved crystal structures of the selected hits with the residues of the todalam site, we detected very similar interactions to that of the predicted conformers. In Tab. III-2 there is a summary of the number of hydrogen bonds established by each scaffold within the todalam site, which are in accordance with the ones we expected from docking and MD simulations.

Table III-2. Number of hydrogen bonds established by each experimentally tested scaffold with the todalam site residues.

			X = O, S			R_2 = $CH_2CH=CH_2$
Number HB	3	5	4	4	3	3

In this summary, we observe that ligands presenting an urea scaffold have been shown to establish a higher number of hydrogen bonds within the binding site compared to other ligands, making them attractive candidates for the next generation of todalam-site binders.

III.1-3.7.2. Biochemical Polymerization Assays

Encouraged by the positive results of our docking studies, we submitted the hit compounds for biological evaluation. They were sent to Centro de Investigaciones Biológicas Margarita Salas (CIB-CSIC) in Madrid, Spain, for *in vitro* biochemical testing to determine their effect on tubulin polymerization dynamics. The 15 ligands were tested by Francesca Bonato by incubating them at 37°C with tubulin purified from calf brains and by following the absorption at 350 nm, which monitors the formation of MTs or other aggregates in solution. The polymerization dynamics was

monitored measuring the absorbance at 350 nm for at least 4600s. Tubulin with DMSO (the solvent used as vehicle), podophyllotoxin (PODO, a reference destabilizer binding to the colchicine-site), pironetin (PIR, a reference destabilizer targeting α-tubulin), and todalam (TOD) were used as controls. The major tubulin isotypes present in the tubulin preparation were tubulin βII (58%) and tubulin βIII (25%), the other 17% contained other β-tubulin isotypes. For polymerization enhancers, the increase in the amount of polymerized tubulin is accompanied by an increase in the maximum slope as well as the plateau of the curve. Moreover, the higher number of nucleation events results in a decreased lag time, which is often too short to be detected in these experiments. On the other hand, for strong polymerization inhibitors such as PODO or pironetin, no increase in absorbance is observed indicating the complete prevention of MT formation. Todalam is a weak inhibitor of tubulin assembly, therefore its polymerization curve is characterized by a decrease in the slope and in the plateau of the curve, as well as in a longer lag time which indicates fewer nucleation events occurring in solution.

Compounds **1**, **2**, **4**, **5**, **7**, **9**, and **13** had a moderate to negligible effect on tubulin polymerization, while compound **3** showed a promising destabilizing effect in comparison to the DMSO control (Fig. III-25).

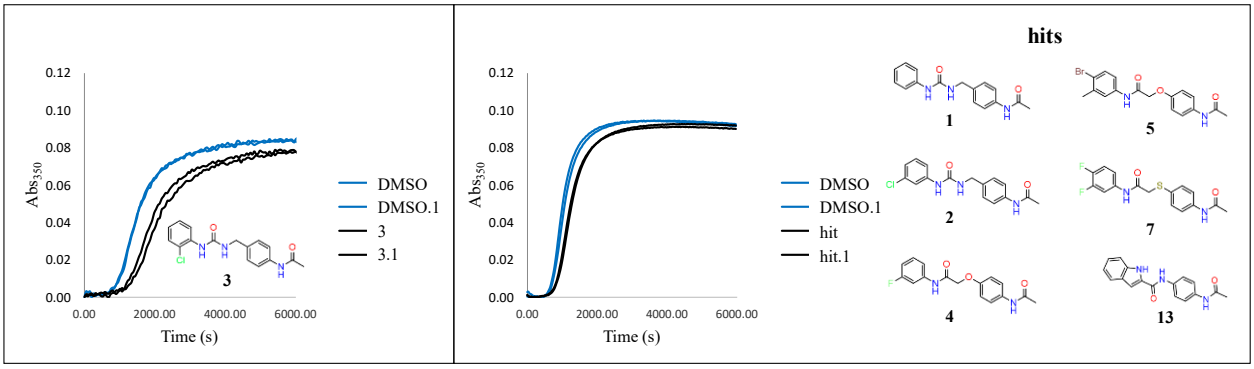


Figure III-25. Results from the biochemical polymerization assays of tubulin in the presence of controls and the studied compounds. All polymerization experiments were done with 27.5 µM of each compound (vs. 25 µM of tubulin).

Moreover, compounds **11**, **12**, **14** and **15** presented a tubulin destabilizing effect comparable to that of todalam (Fig. III-26).

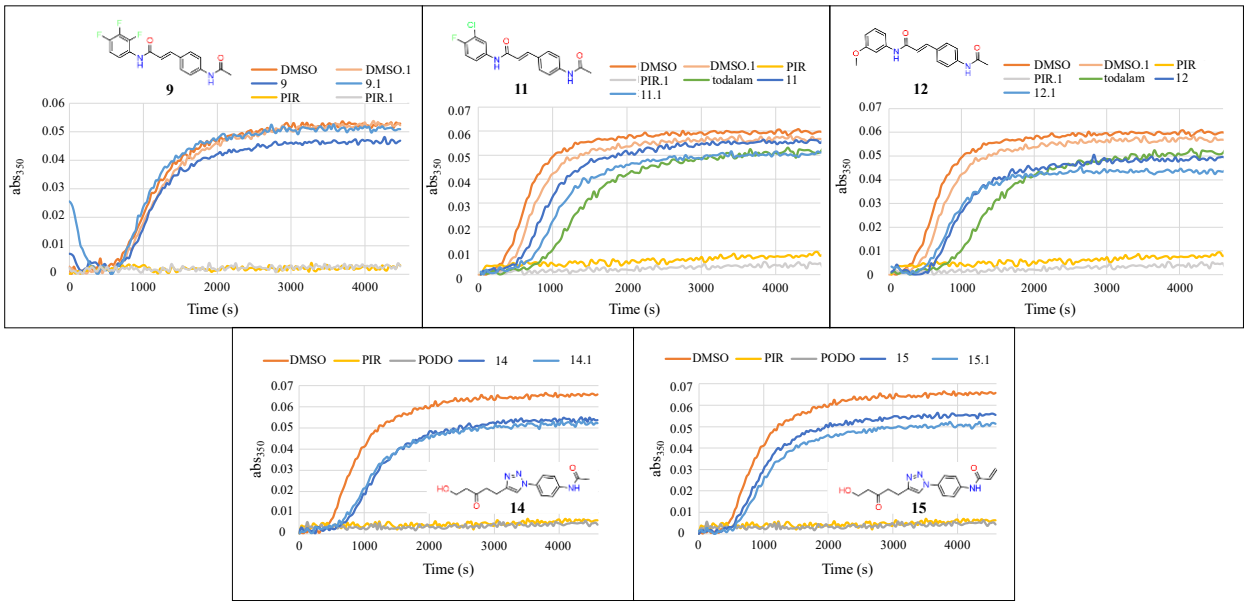


Figure III-26. Results from the biochemical polymerization assays of tubulin in the presence of controls and the studied compounds. All polymerization experiments were done with 27.5 µM of each compound (vs. 25 µM of tubulin).

III.1-3.8. Docking of a Pironetin Analogue to Guide Structure Refinement

After identifying a number of molecular hits, the crystallography team revisited the experiments they had conducted on the pironetin derivatives presented in the introduction. They discovered that one of the synthesized pironetin derivatives was present in the todalam site. However, at the time they performed these initial analyses, they did not have much structural data on the todalam site and the electron density maps were difficult to interpret due to poor definition. They were uncertain as to whether there was one or two molecules binding in the site and it was difficult to determine their orientation. Attempts to fit the molecule into the electron density maps failed, so we suggested that performing docking experiments of the molecule in the todalam site could help guide structure refinement. This would provide valuable information as the tested pironetin derivative contains a lactone ring, which was designed to covalently interact with αCys316 in the pironetin site. If we could predict where this lactone ring lies within the todalam site, this structural information could later be used for the design of covalent todalam site binders.

The docking experiments were carried out using the software S4MPLE, following the same protocol used for the presented hits. The best conformers obtained with S4MPLE fit the electron density maps and suggested that probably different conformers of the same molecule were present in the site (Fig. III-27).

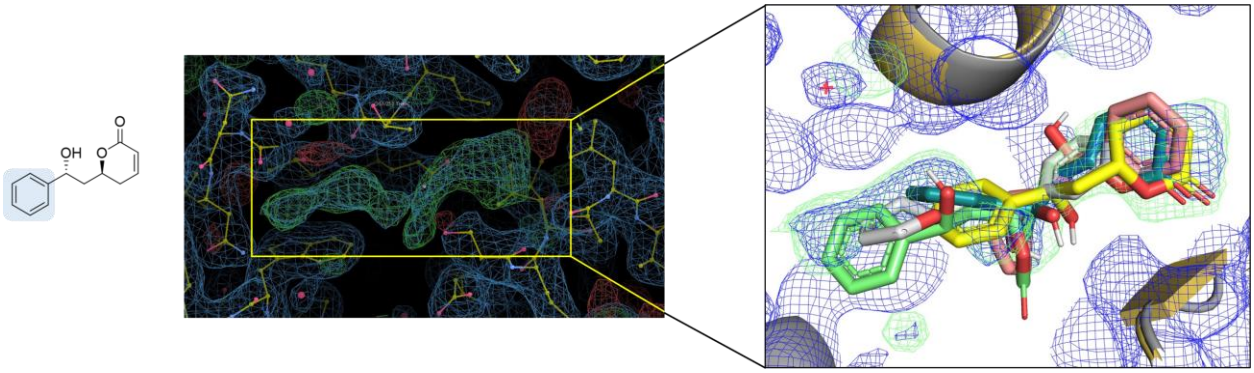


Figure III-27. Left: Electron density maps obtained from X-ray crystallography of the pironetin derivative whose 2D structure is represented. The color of the omit map is green/red. Right: Best conformers obtained from docking experiments performed using the S4MPLE software, superimposed in the electron density maps.

To date, we still do not have the final structural refinement of this molecule within the binding site. Nevertheless, it was a good case study in which, once again, S4MPLE showed its potential as a docking software and the power of computational tools to support experimental work.

III.1-4. Conclusion

This research aimed to bypass the difficulty and expense of synthesizing structurally complex MTAs, such as pironetin, for tubulin immobilization by studying new, simpler tubulin ligands that target the todalam site, a recently discovered binding site close to the pironetin site. A new set of promising chemically diverse todalam site binders was successfully identified by using a combination of computational methods including *in silico* substructure search, docking approaches, and MD simulations, which were later experimentally validated through X-ray crystallography. Of the 15 molecules proposed by computational methods, 12 were experimentally observed to bind to the todalam site, resulting in a hit rate of 80%. Furthermore, the RMSD values between the crystal structures of the hit compounds and the docking results were found to be less than 1.5 Å, indicating a good agreement between the computational and experimental data. This supports that S4MPLE correctly predicted the binding mode of the majority of hit compounds to the todalam site. However, it is important to note that a low RMSD value alone does not necessarily indicate that a docking result is biologically relevant or accurate, that is why biological experiments were necessary. The results of these experiments showed that the urea scaffold was able to establish the highest number of hydrogen bond interactions with the todalam binding site (Tab. III-2). Additionally, compound **3** (urea scaffold), **11**, **12** (acrylamide scaffold) and, **14** and **15** (triazole scaffold) presented a tubulin destabilizing effect in the biochemical polymerization

assays. Moreover, it was observed that the presence of an anchor point, such as an acetanilide or acrylanilide moiety, was necessary for effective binding to the todalam site, and the presence of an extra terminal alkene did not compromise the binding mode of the molecule. Furthermore, a variety of five different scaffolds were discovered (Fig. III-28). The hydrophobic head of the molecules allowed for the presence of both aliphatic and aromatic substituents because they were confirmed to bind to the hydrophobic pocket of the todalam site.

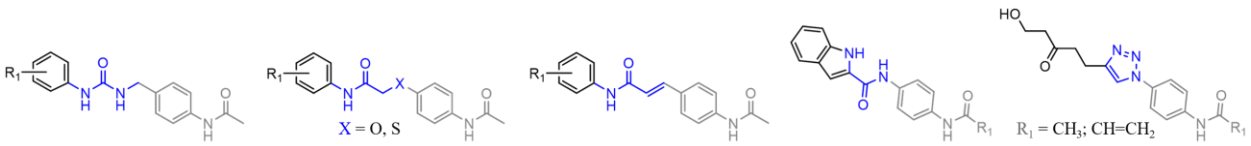


Figure III-28. 2D structure of the five scaffolds selected as potential todalam binders after the scaffold hopping campaign.

In this first part of the project, we set the basis for further development of the identified scaffolds into covalent binders. On the basis of the good agreement of computational and experimental data, we were confident that we could proceed with the use of computational approaches to develop covalent todalam site binders.

III.2 Design and Modeling of Putatively Covalent Todalam-Site Binders

III.2-1. Introduction

In this second stage of the investigation, we focused on the optimization of previously discovered scaffolds into potential todalam site ligands capable of covalently binding to αCys4. We conducted three VS campaigns of fragment-like compounds with Cys-binding warheads from commercial libraries which could serve as building blocks to synthesize the desired optimized molecules. The term 'warhead' is used to refer to a functional group that could potentially bind covalently to the thiol group of a cysteine residue. The top-scoring designed covalent binders of this computational search were then filtered based on the non-covalent binding pose (conservation of ligand-site contacts undergone by todalam and the new found binders). The stability of the tubulin-ligand complex within the todalam binding pocket was assessed by MD simulations. To date, some of the computationally identified candidates were synthesized and further optimized, subjected to X-ray crystallography experiments and *in vitro* MT-polymerization assays.

III.2-2. Aim of the Work

The aim of this work was to develop new covalent tubulin binders targeting the todalam site based on the previously discovered scaffolds. Subsequently, these molecules could serve as molecular tools for tubulin immobilization.

III.2-3. Computational Details

To design warhead-containing todalam site binders on the basis of the results of the scaffold-hopping VS, we performed three different substructure search approaches. Our main goal was to find building blocks that would allow the chemistry group to synthesize the computationally designed molecules.

III.2-3.1. Substructure Search of the ZINC, Enamine and SciFinder Databases

The initial substructure search through the ZINC library of ~800 million molecules aimed to identify building blocks containing a terminal primary amine or isocyanate group and a warhead

(a halide or a nitrile group). This was for the chemistry team to generate urea-containing potential covalent binders by one-step synthesis. These two groups should be connected by 1-4 atoms of any type. We selected the approximate size of the building block we were looking for by examining the available space within the todalam binding site. For this, we designed the queries shown in Fig. III-29 using the chemical structure drawing tool ChemAxon MarvinSketch.[75]

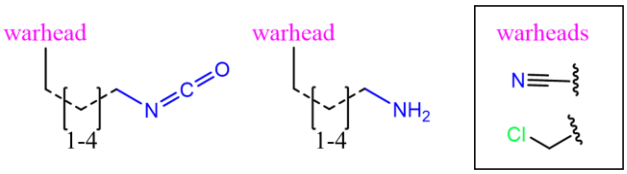


Figure III-29. SMARTS queries used to perform a substructure search in the ZINC library.

The ZINC library was previously downloaded and submitted to the standardization procedure in the same way as the Enamine library (see section III.1-2.1.). Additionally, the substructure search was conducted using the same method as that used in the scaffold hopping step, utilizing an in-house developed ChemAxon-powered Java script.

In the second substructure search, we expanded the criteria for identifying molecules by using a larger number of SMARTS queries and a different script. This script was developed by Maxim Shevelev at the University of Strasbourg. In this search, we aimed to find commercially available molecules that have one of the following terminal groups: a primary amine, an acrylic acid, or a terminal alkyne. These molecules should also have any of the desired warhead groups, which are separated from the terminal groups by 1-4 intermediate atoms of any type (Fig. III-30). To ensure that the molecules are linear and to avoid branches, we restricted the number of hydrogen atoms that could be attached to the intermediate atoms. We also allowed the use of carbonyl, hydroxyl, and primary amine-attached carbon atoms as intermediate groups because we reasoned that additional hydrogen bond donor and acceptor groups could facilitate ligand binding. The queries allowed bonds of any type between the intermediate atoms. This time, we screened two libraries of chemical compounds: the already standardized ZINC library and the Enamine library. To check if a molecule met the specific criteria, we applied the VF2 graph isomorphism algorithm[89,90], which is implemented in the RDKit python library.[91] This algorithm compares the structure of each molecule in the libraries to the pattern specified by the SMARTS query.

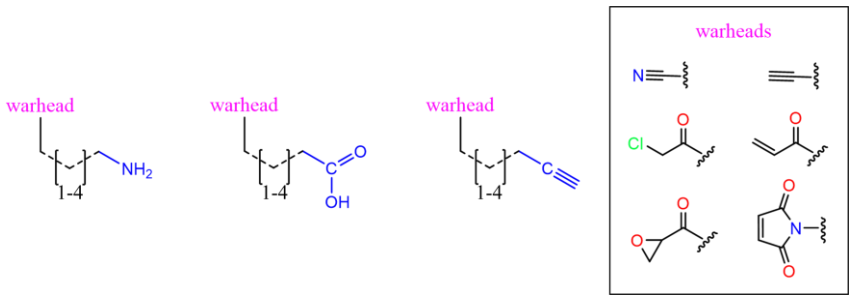


Figure III-30. SMARTS queries used to perform a substructure search in the ZINC and Enamine library.

The third substructure search was based on the use of the SciFinder web engine to search for suitable purchasable building blocks within the SciFinder database.[92] To perform the substructure search in SciFinder, we first specified the different queries representing the building blocks that we were searching for, the same ones we designed for the second substructure search (Fig. III-30). This was done by drawing the query using their built-in chemical structure editor. In the advanced search options, we selected a filter that allowed us to retrieve only the building blocks that were commercially available.

III.2-3.2. Modeling based on Docking Experiments using S4MPLE

After searching for substructures, we used the list of identified building blocks to create new molecules using ChemAxon MarvinSketch.[75] These compounds contained a warhead group and some of the central linkers that were experimentally tested in the first stage of the project. They were subjected to non-covalent and covalent docking calculations using S4MPLE.[53,54] We used the same binding site structure for the docking studies as in the scaffold search section (see section III.1-2.3.). However, for the covalent docking studies, we made a small change by replacing the αCys4 amino acid with a glycine residue. This amino acid substitution was made for technical reasons, which will be explained later. During the docking simulations, the position of the atoms in the glycine residue were fixed.

III.2-3.2.1. Non-Covalent Docking with S4MPLE

The non-covalent calculations with S4MPLE were performed in the same way as those explained in section III.1-2.5.1. ChemAxon MarvinSketch was used for the generation of the molecules to be docked based on the building blocks retrieved from the performed substructure searches

containing the different central linkers we were interested in, and the acetaminophenyl anchor point was used in all of the designed todalam site binders.

III.2-3.2.2. Covalent Docking with S4MPLE

S4MPLE cannot rigorously simulate covalent docking by FF-based approaches because FF simulations are not typically equipped to handle covalent bond breaking and formation, and instead rely on a predetermined table of bonded and non-bonded interactions. Moreover, the FF parametrization strategies differ for the protein (AMBER) and the ligand (GAFF), although these two FF have been developed to work together, it is technically more difficult to program a simulation in which the two moieties (protein and ligand) are covalently bound to each other. Therefore, a covalent docking trick was employed by S4MPLE. It consisted of the extraction of the mercaptan fragment from the αCys4 side chain by removing the Cα-Cβ side chain bond. The mercaptan fragment required three explicit hydrogens on the carbon atom since the dangling valency needed to be satisfied to avoid GAFF parametrization artifacts. After this, the mercaptan group was formally detached from the protein (the initial Cys4 was substituted by a glycine), but kept at its PDB coordinates – yet within the covalent binding distance to the amino acid backbone atoms of Gly4 (ex-αCys4). The explicit hydrogen atom bound to the sulfur atom was also deleted to satisfy the free valence required for it to be functionalized and replaced with bonds to the building block. The sulfur atom was then marked with an atom map flag using ChemAxon MarvinSketch to determine it as a connector atom. In these covalent docking calculations, also the acetaminophenyl group of todalam was kept as another anchor point in the same location as it appears in the PDB structure 5SB7 and its preparation followed the mercaptan one.

III.2-3.2.2.1. Building Block to Synthon Conversion

The substructure searches from ZINC and Enamine databases retrieved real building blocks. However, S4MPLE needs atom-marked synthons (modified versions of the building blocks) as input. These synthons are provided to S4MPLE as a file of SMILES strings with atom-mapped connectors to facilitate their binding to the anchor moieties and form the desired molecules (Fig. III-31).

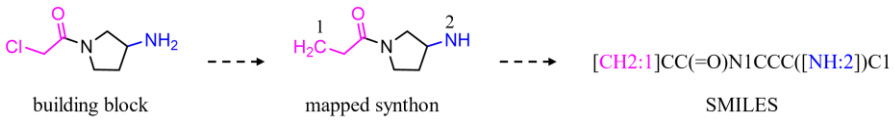


Figure III-31. Visual representation of the process of converting a building block obtained from a substructure search into a mapped synthon, with the corresponding mapped SMILES string also displayed for the covalent docking preparation step.

Both the synthons from the found building blocks and anchor fragments (mercaptan and acetaminophenyl group) have special atom markings for free valence. The synthon of the mercaptan anchor has a single connector atom mapped as '1', the synthon of the acetaminophenyl anchor is mapped as '2', while the synthons of the building blocks retrieved from the substructure searches have two connectors '1' and '2' so they can be specifically bound to the corresponding anchor atoms (Fig. III-32).

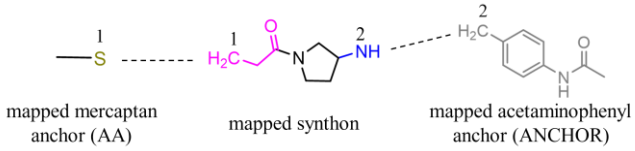


Figure III-32. Visual representation of the fragment linking process in S4MPLE, where potential linkers such as the mapped synthon depicted in this figure are searched for to connect the αCys4 sidechain (depicted as a mercaptan fragment with connector atom mapped as '1') to the anchor point (depicted as an acetaminophenyl fragment with connector atom mapped as '2') in an optimal manner. The dashed lines show how each fragment will link to each other to form the final ligand.

The covalent docking workflow carried out in S4MPLE can be divided into six different steps: (1) building block insertion; (2) in-site building block sampling; (3) ligand sampling; (4) RMSD assessment; (5) free-ligand sampling; and (6) ranking (Fig. III-33).

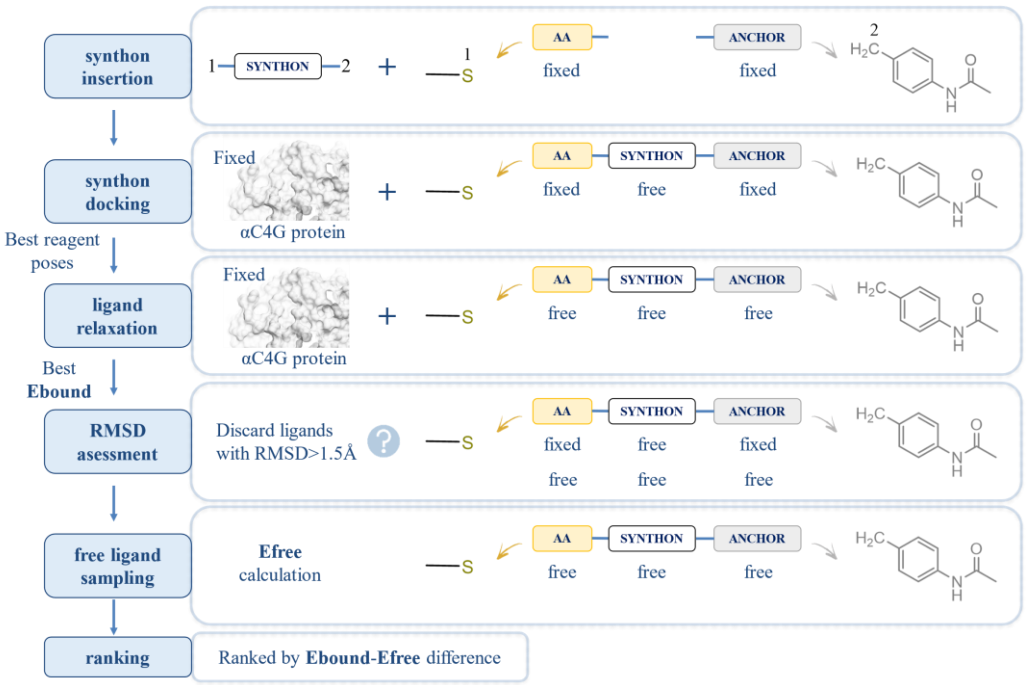


Figure III-33. Covalent docking protocol followed by S4MPLE.

III.2-3.2.2.2. Building Block Insertion

To prepare the input files for the covalent docking with S4MPLE, the web service of the Chemoinformatic group at the University of Strasbourg was used.[93] It requires the submission of a .tar archive of the directory containing the anchor data and the SMILES file of the moieties to be coupled with the anchors. The workflow of the script implemented in the web service is similar to that explained in section III.1-2.5.1., except that the generation of the 3D structure of the final ligand first requires the establishment of bonds between the building blocks and anchor atoms. This is followed by positioning the initial random building block conformation into the coordinate system of the fixed anchor atoms, in search of geometries ensuring a minimal strain within the building block connected to its anchors. This is performed by a preliminary and fast evolutionary optimization by S4MPLE using default harmonic force constants for bond stretching and angle bending and a default van der Waals repulsive term. This initial geometry for the docking simulations has the building block already pre-aligned with respect to the anchor moieties, and herewith implicitly to the protein site. The GAFF typing and the calculations of the partial charges are run for the described pre-constructed ligand structure.

III.2-3.2.2.3. In-Site Building Block Sampling

This step involves a calibration step of the contact fingerprint-based redundancy threshold: Random geometries are generated and the mean, minimal, and maximal dissimilarity scores of their contact fingerprints are monitored. On the basis of these observations, a dissimilarity threshold below which two conformers are considered redundant is derived. The threshold is evaluated for each molecule, because the size of the contact fingerprint may highly vary, and one cannot extrapolate a same cutoff under these circumstances.

Then, the non-fixed linker is sampled for 500 evolutionary generations, in search of its optimal orientation with favorable contacts to the protein site, all while respecting the covalent constraints imposed by bonds to the fixed anchor fragments. The best energy level obtained in this simulation is stored in a 'local best energy' list. This 500-generation simulation is then repeated, leading to a presumably different 'local best energy', concatenated into the 'local best energy' list, which is ordered ascendingly. As soon as the first three top energy values in this list are within 1 kcal/mol, it is considered that the simulation 'reproducibly' found the so far best energy optimum. Otherwise, up to 500-generation simulations are repeated, after which the process is force-stopped in spite of failure to 'converge' in the above sense.

III.2-3.2.2.4. Ligand Relaxation

The 20 top poses from the above sampling step are re-submitted to S4MPLE for energy minimization of the fully free ligand with the unfixed anchoring atoms. However, if this relaxation step triggers a complete repositioning of the ligand, at the cost of significantly dragging anchor fragments away from their 'native' positions, the resulting pose must be rejected.

III.2-3.2.2.5. RMSD Assessment

In order to analyze whether unfixing of the anchor fragments causes the ligands to move from their initial positions, S4MPLE is used to calculate the heavy-atom RMSD values between initial (fixed-anchor moieties) and energy-minimized ligand poses. The RMSD is calculated within the fixed reference frame of the protein site (ligand structures are not re-overlaid to minimize it). If energy minimization causes the ligand to drift away by a RMSD > 1.5 Å, the pose is discarded. The lowest minimized energy of non-discarded poses is then taken to represent *Ebound*.

III.2-3.2.2.6. Free-Ligand Sampling

Then, we sample the completely free ligand for 200 generations and we obtain after this step its lowest free-state energy value *Efree*. We rank the ligands by the difference of *Ebound-Efree*. We do the sampling of the free-ligand to calculate the energy difference between the best energy of the standalone compound with respect to the best energy of the docked compound. That is the base of the scoring function of S4MPLE.

III.2-4. Results and Discussion

In the present work, the goal was to design an MTA that can react covalently with αCys4 at the todalam binding site. To design new warhead-containing todalam site binders able to reach αCys4 and react with it to form a covalent bond, we needed to understand three main points: (1) how reactive αCys4 was; (2) what warheads to use to increase the reactivity of the designed binder towards αCys4; (3) which of the discovered scaffolds would be most effective in guiding the warhead to αCys4 while being suitable for synthesis.

III.2-4.1. Structural Analysis of αCys4

Cysteine is an amino acid that contains a thiol group and is one of the least abundant amino acids within proteins. Due to the large atomic radius of sulfur and the low dissociation energy of the S-H bond, the thiol group of cysteine has a strong nucleophilic character: among the 21 standard amino acids, it is the most nucleophilic.[94]

First, we analyzed how the reactivity of αCys4 was influenced by the local environment at the todalam site. Cysteine is frequently selected as the target for covalent inhibitors, particularly in the case of cysteine protease catalytic cysteines, where deprotonation facilitated by the triad plays a crucial role (as demonstrated in the development of Pfizer's nirmatrelvir).[95] Tubulin is not a cysteine protease and in the environment of αCys4, we could not observe the presence of a catalytic triad that would help deprotonate the thiol group.[96,97] In the search for warhead-containing building blocks, we prioritized those containing HBA groups that, from the docking experiments, were predicted to be oriented toward the thiol group of αCys4. Any HBA of the building block strategically placed to assist the electrophilic attack on the -SH group might facilitate the release of the hydrogen atom from the thiol group, thereby promoting the desired covalent reaction in this

manner. However, detailed modeling of the reactivity of warheads would require expensive quantum-mechanical approaches and were not envisaged in this work. The selection of putatively covalent binders was based on a computational estimation of their ability to fit the binding site and comply with the sterical constraints of binding to Cys (see above), plus an empirical evaluation of warhead-containing building blocks: availability, cost, compatibility with the synthesis conditions of final ligands, estimated reactivity (based on chemistry known-how and common sense, as very reactive acylating or alkylating groups would non-specifically attack other nucleophilic tubulin residues).

III.2-4.2. Scaffold Selection

From the scaffold hopping step of the project, we had computationally identified and experimentally validated 12 different todalam site binders that could be grouped into five different scaffolds (Fig. III-28). Following the identification of potential scaffolds through the scaffold search, a collaborative effort was undertaken with the organic chemistry team to determine the optimal central linker to look for warhead-containing building blocks that would suit the synthesis of the final products.

The purpose of this research was to identify a central linker that would enhance the orientation of the warhead toward the αCys4 residue to increase the probability of protein-ligand covalent bond formation. Generally, ligands that form a higher number of hydrogen bonds with a protein target are considered more favorable as these interactions can be strong and specific, thus contributing to the stability of the protein-ligand complex. To this end, in the initial phase of the development of potential covalent binders for the todalam site, the focus was on designing molecules with a central linker that would facilitate the formation of the highest number of hydrogen bonds with the tubulin protein target. Our previous study had shown that molecules with an urea group as their central linker were able to form a greater number of hydrogen bonds with tubulin (Tab. III-2) than those that presented other central linkers. As a result, this scaffold was selected for the design of warhead-containing binders in the initial phase of the study.

Another important factor that was considered when designing the molecules was their synthetic feasibility, which refers to the practicality and cost-effectiveness of synthesizing the molecules. To minimize costs and time, we searched for readily available building blocks which could reduce the

number of synthetic steps required to obtain the final molecule. To perform the search for building blocks presenting a warhead that could form the desired molecules, it was important to first understand the chemical reactions that the chemistry team was planning to perform for the synthesis of the chemical compounds. The synthetic chemistry group proposed a synthetic approach in which, by reacting an isocyanate and a primary amine, it was possible to obtain an urea. Therefore, our computational strategy consisted of finding commercially available isocyanate or primary amine-containing building blocks in which a warhead was present in the opposite site of these groups. For this, we performed a substructure search of the SMARTS queries shown in Fig. III-34 in the ZINC library. In parallel, efforts in the synthetic chemistry laboratory were focused on the synthesis of two different molecules: (1) an acetaminophenyl (anchor point) that would be attached to a primary amino group and (2) an acetaminophenyl (anchor point) attached to an isocyanate. By purchasing the warhead-containing building blocks identified through computational means and synthesizing the anchor component, we were planning to combine these two parts in order to synthesize a warhead-containing todalam binder with an urea central linker in one step.

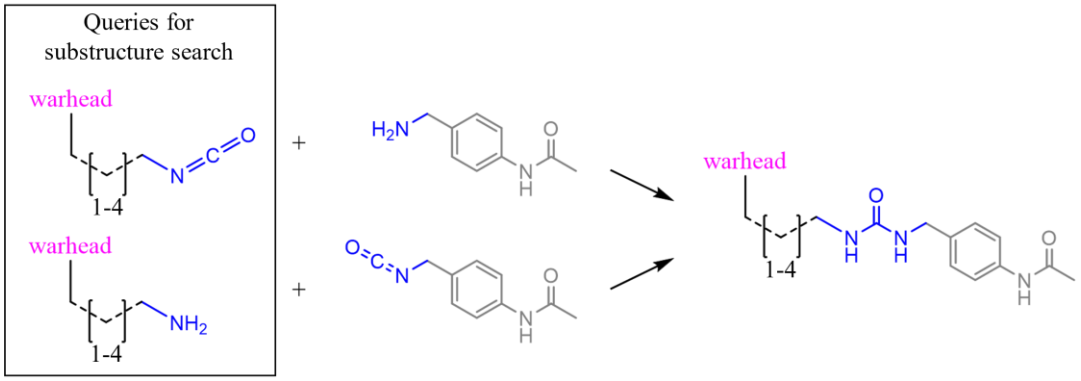


Figure III-34. Planned synthetic strategies for the synthesis of warhead-containing todalam site binders presenting an urea group in the central linker part of the molecule. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink. On the left are highlighted with a square the SMARTS queries used to perform the substructure search.

III.2-4.3. Warhead Selection

In the initial phase of the search for substructures, a decision was made to focus on specific types of warheads. Molecule **6**, identified during the scaffold search, featured a nitrile group. Although initial experimental evaluations indicated that this molecule did not bind to the todalam site, likely as a result of the bulky nature of the nitrile group in the para position, further investigation was

deemed worthwhile. This was due to curiosity to evaluate whether the nitrile group when present in the meta position would be able to bind or if in a different molecule it would be able to react with αCys4. Therefore, we proposed molecules **16-M** and **16-P** for further synthesis and experimental evaluation, to gain more structural information on the todalam site binders (Fig. III-35).

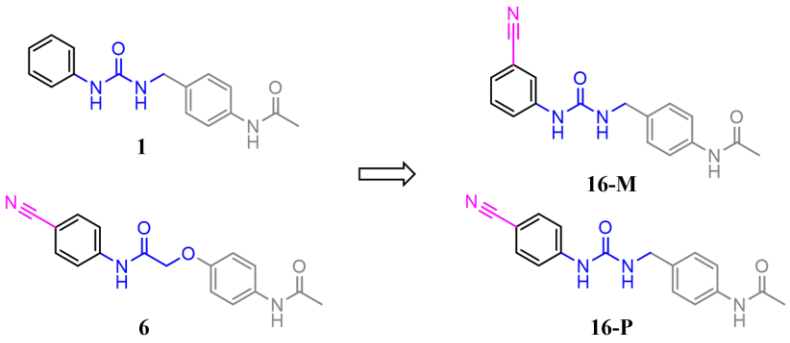


Figure III-35. Proposed warhead-containing todalam site binders with an urea central linker **16-M** and **16-P** based on the previously tested molecules **1** and **6**. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink.

With this in mind, the nitrile group was included in the substructure search as a potential warhead due to its ability to react with cysteine via nucleophilic addition. Furthermore, to evaluate alternative warheads, an α-haloketone and a methylene halogen were also included in the substructure search, both of which possess the ability to react with αCys4 through nucleophilic substitution (Fig. III-36).

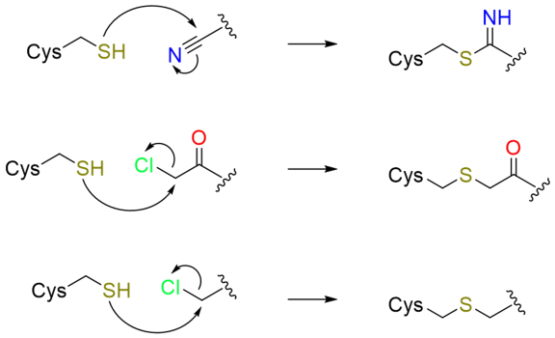


Figure III-36. Nucleophilic groups capable of covalently interacting with a cysteine: nitrile, α-haloketone and methylene halogen.

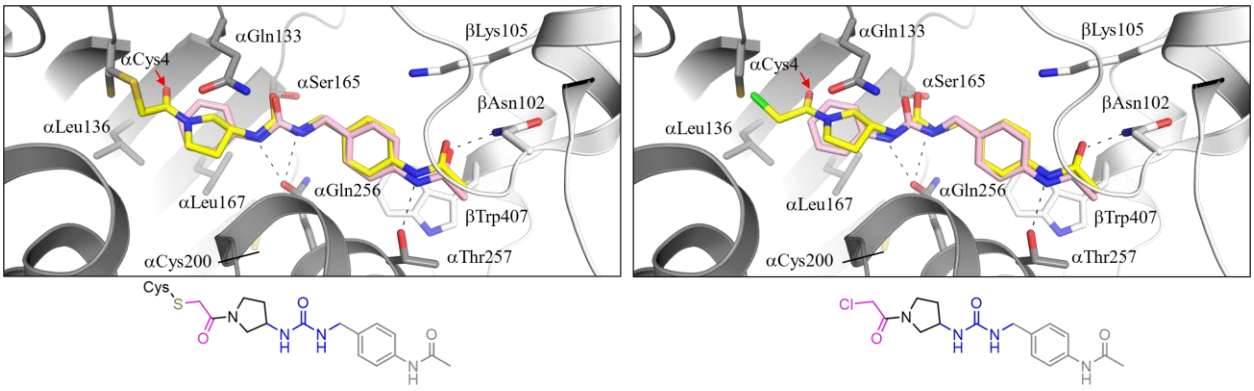
III.2-4.4. Urea-Containing Todalam Site Potential Covalent Binders

The substructure search conducted using the ZINC library resulted in the identification of a few hundred of potential warhead-containing building blocks available for purchase. On the basis of

these building blocks, the complete structure of the suggested tubulin binders was modeled. The compounds were later submitted to non-covalent and covalent docking experiments using S4MPLE following the steps described in section III.2-3.2.1. and section III.2-3.2.2., respectively. After filtering the molecules whose score in the non-covalent docking experiments was lower or equal to the docking score of todalam (-67 kcal/mol), there were 105 molecules that were later submitted to covalent docking experiments. The criteria for prioritizing the best hits were the following:

- Binding pose matching contacts of the crystal structure of the urea-containing todalam site binder **1** and the hits based on the identified ZINC building blocks.
- The presence of HBA close to the thiol group of αCys4 to enhance the reaction.
- Non-covalent docking results with a docking score <-67 kcal/mol.
- Non-covalent docking of the molecules "spontaneously" places the warhead next to αCys4 at a distance of ~3 Å from its thiol group.
- Covalent docking with a docking score <-67 kcal/mol.

Fig. III-37 illustrates some examples of the visual evaluation of the results retrieved from the covalent and non-covalent docking experiments performed on some of the selected hits.



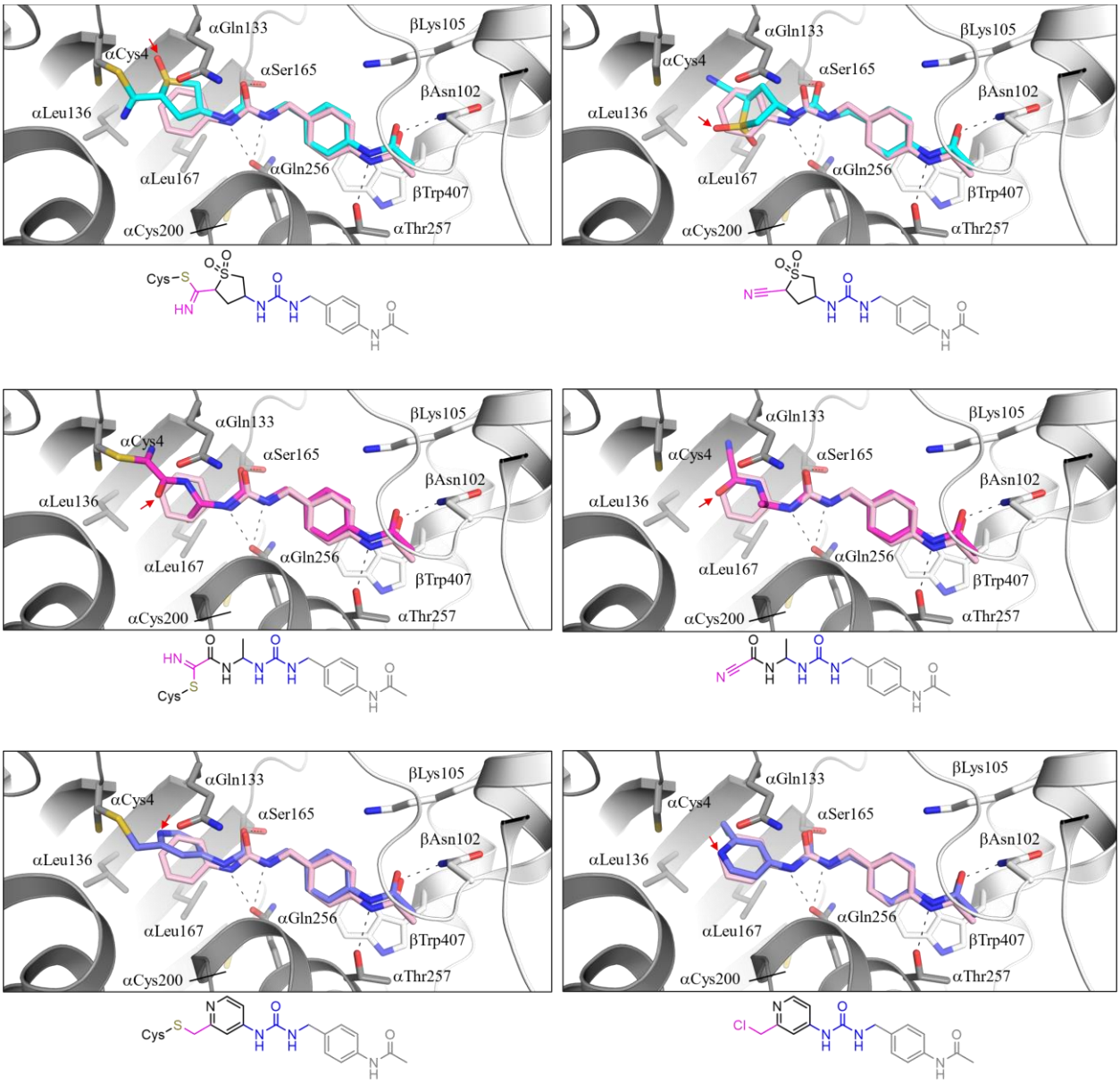


Figure III-37. Close-up of the interaction between four of the 105 hits and tubulin (gray). Hydrogen bonds are displayed as black dashed lines. The best covalent (left) and non-covalent (right) conformers of each hit are displayed together with the crystal structure of compound **1** (light pink). The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, chlorine atoms in green and sulfur atoms in gold. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbons, respectively. HBAs are marked with a red arrow.

After this filtering, 12 molecules were selected as potential todalam site covalent binders that were generated based on the building blocks shown in Fig. III-38, obtained from the substructure search.

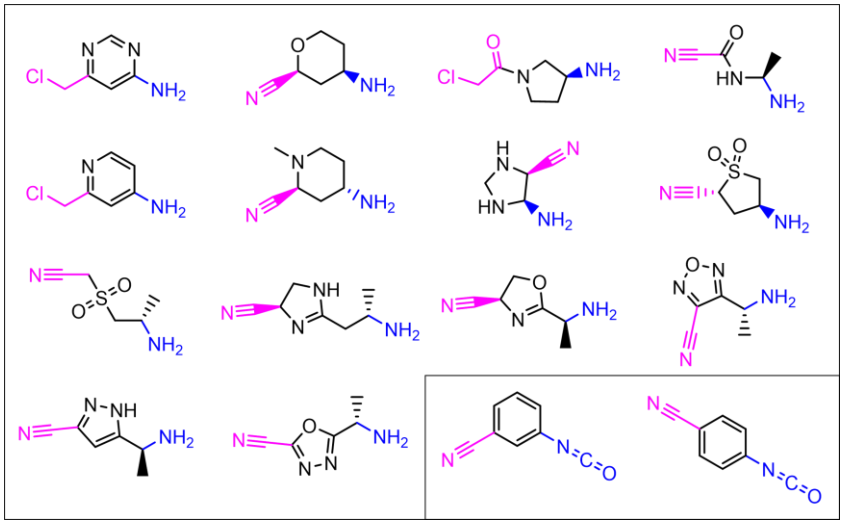


Figure III-38. Building blocks selected for the synthesis of potential todalam site covalent binders. The chemical group represented in blue is the one through which the molecules would be connected to another building block to form the resulting urea-containing molecules. In pink is represented the warhead present in each building block.

To validate the accuracy of the non-covalent docking pose for the 12 selected hits, MD simulations were conducted following the same protocol as the one explained in section III.1-3.6. The bound conformation of all but one molecule was observed to be stable within the binding site during 100 ns, keeping their hydrogen bonds in place. In the analysis of the results, the distance between the warhead and the thiol group of αCys4 was included, considering it good if the average distance during the simulations oscillated around ~3 Å. Therefore, these 12 molecules were presented to the synthetic chemistry group to proceed with their synthesis.

Following these initial searches, the organic chemistry team decided to abandon the synthesis of the anchor point attached to the isocyanate group because of safety concerns. However, they were successful in synthesizing the acetaminophenyl (anchor point) group attached to a primary amine. As a result, only the building blocks identified in the substructure search that contained an isocyanate group were purchased for further synthesis of compounds **16-M** and **16-P** shown in Fig. III-35. Then, these molecules were tested experimentally *via* X-ray crystallography. Of the two molecules evaluated, only the one containing the nitrile group in the *meta* position was observed to be bound in the crystals, however, it was not found to be covalently bound to the todalam site. On the contrary, the molecule featuring the nitrile group in the *para* position was not observed to be bound. The todalam site is relatively restricted in terms of space, and the design of todalam site binders requires a high degree of caution and thoroughness because even small

variations, such as the addition of an extra carbon atom or the positioning of substituent groups, can greatly influence the ability of a molecule to bind to the site.

Since warhead-containing building blocks attached to an isocyanate group presented a good approach for the formation of urea-containing molecules, we manually searched for more building blocks with these characteristics in SciFinder. However, the search was not successful, since building blocks containing an isocyanate group and a warhead were either not available or too expensive with respect to the fixed standard (<200 €/gram). Therefore, we needed to rethink our strategy to design warhead-containing todalam site binders.

III.2-4.5. Scaffold Optimization for Synthesis Feasibility

To optimize the synthetic feasibility of these molecules, our strategy was to simplify the chemical structure of the compounds, using readily available starting materials, and streamlining the synthesis process. In this way, we tried to balance the competing considerations of the number of established hydrogen bonds between the ligand and the protein and the synthetic feasibility when designing a new molecule.

The identification of urea and amide-containing molecules as hits at the todalam site allows for the development of new todalam binders through the utilization of versatile and adaptable synthetic chemistry approaches. The chemistry team proposed the formation of amide bonds through the use of two building blocks, one containing a terminal primary amine and the other containing a terminal carboxylic acid (Fig. III-39).

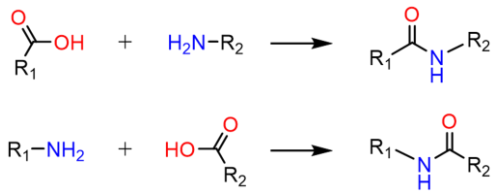


Figure III-39. Synthetic pathway for the synthesis of amide bonds.

Based on these synthetic insights, the previously acquired structural information of the new todalam site binders, a comprehensive analysis of the interactions established by the hits, and supplementary non-covalent docking studies, we proposed a new chemical group for the central linker of todalam site binders (as illustrated in Fig. III-40). The synthesis of molecules containing

the amide central linker would allow the utilization of the already synthesized acetaminophenyl group (anchor point) attached to a primary amine.

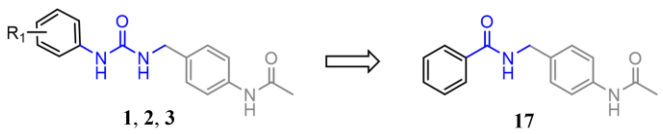


Figure III-40. Optimization of the urea scaffold for synthetic feasibility.

The proposed molecule was successfully synthesized and subsequently evaluated by X-ray crystallography, revealing that it binds to the todalam site in the predicted binding mode (Fig. III-41). Therefore, this new central linker served as a basis for the future design of potential covalent binders.

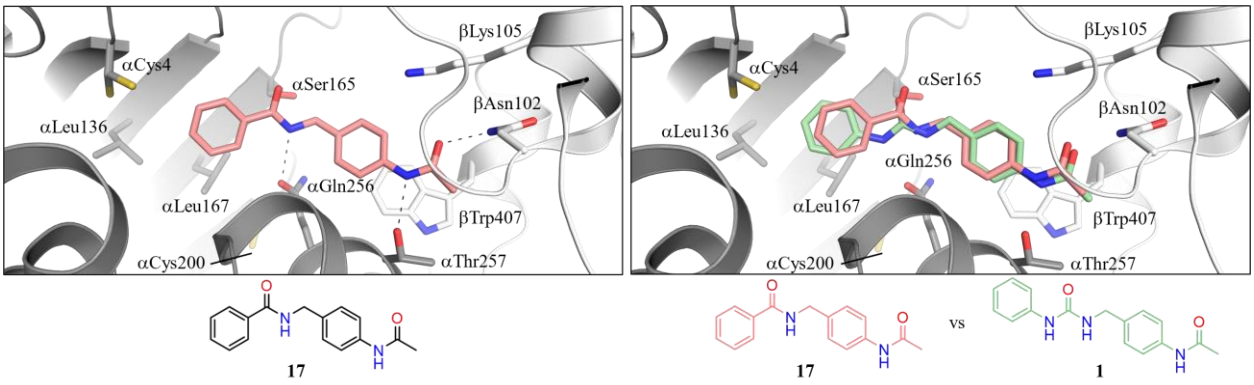


Figure III-41. Left: Close-up of the interaction between compound **17** and tubulin (gray) in the crystal structure. Right: Superimposition of the crystal structures of compound **1** and **17** when in complex with tubulin. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbons, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored in red, nitrogen atoms in navy blue and sulfur atoms in gold. Hydrogen bonds are shown as black dashed lines.

III.2-4.5.1 Acrylamide and Amide-Containing Todalam Site Potential Covalent Binders

Moreover, on the basis of the previous results obtained from X-ray crystallography, tubulin polymerization assays, and chemistry feasibility studies, the scaffold containing an acrylamide was also identified as promising for the substitution of the urea-containing molecules. Of the four hit molecules containing the acrylamide central linker (**9**, **10**, **11**, **12**), three were found to be bound in the crystal structure (**9**, **11**, **12**). Further *in vitro* biochemical analysis revealed that molecules **11** and **12** were identified to be similarly effective in reducing tubulin assembly as the reference compound todalam, indicating the potential as a relevant scaffold for further development.

Furthermore, in order to expand the potential for designing a diverse range of todalam site binders, we incorporated to the substructure search terminal alkyne, primary amine and acrylic acid-containing building blocks presenting new warheads (Fig. III-42) in addition to the ones included in the previous search (Fig. III-36). The warheads we screened for in this work were found by searching in the relevant literature. We looked for electrophilic functional groups that, in general, have a great affinity with the thiol group of a cysteine residue. In particular, in a paper by Dalton *et al.*[98] and in databases such as CovPDB[99] and CovalentInDB[100], we could identify which warheads had been used in previous works that had successfully targeted cysteine residues covalently.

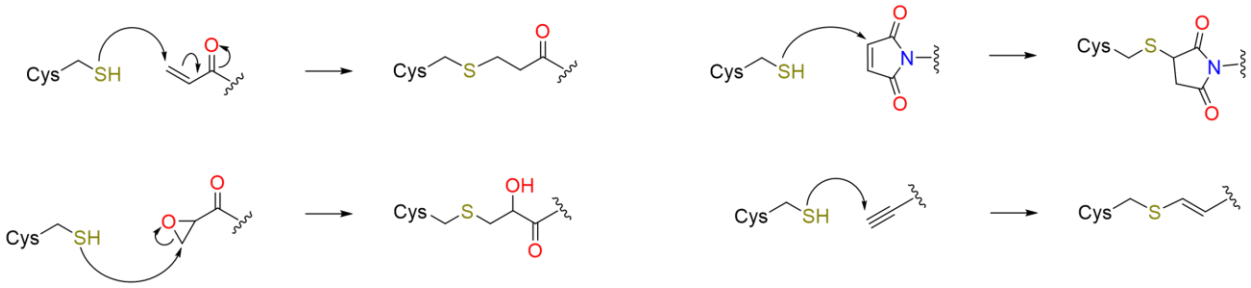


Figure III-42. Nucleophilic groups capable of covalently interacting with a cysteine: αβ-unsaturated ketone, maleimide, epoxide, and alkyne.

Efforts in the synthetic chemistry laboratory were focused on the synthesis of an acetaminophenyl (anchor point) attached to an acrylic acid. The strategy followed to perform the substructure search was to screen the compounds from the ZINC and the Enamine library by using the SMARTS queries shown in Fig. III-43. Additionally, a local search in SciFinder was carried out to try to find new outputs.

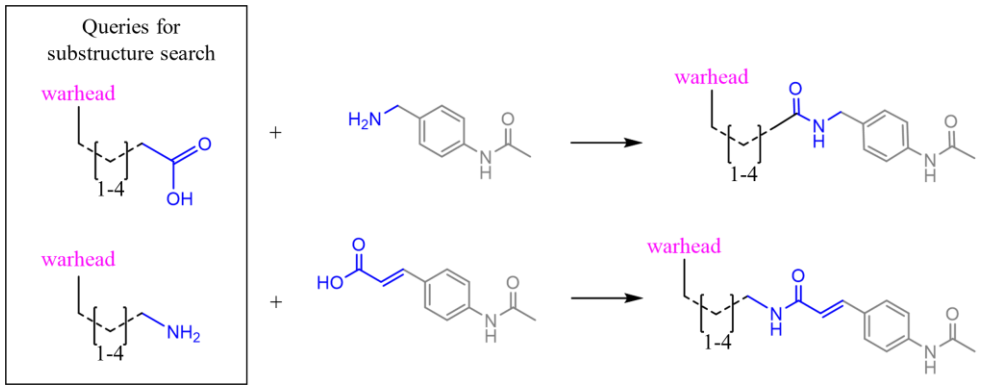
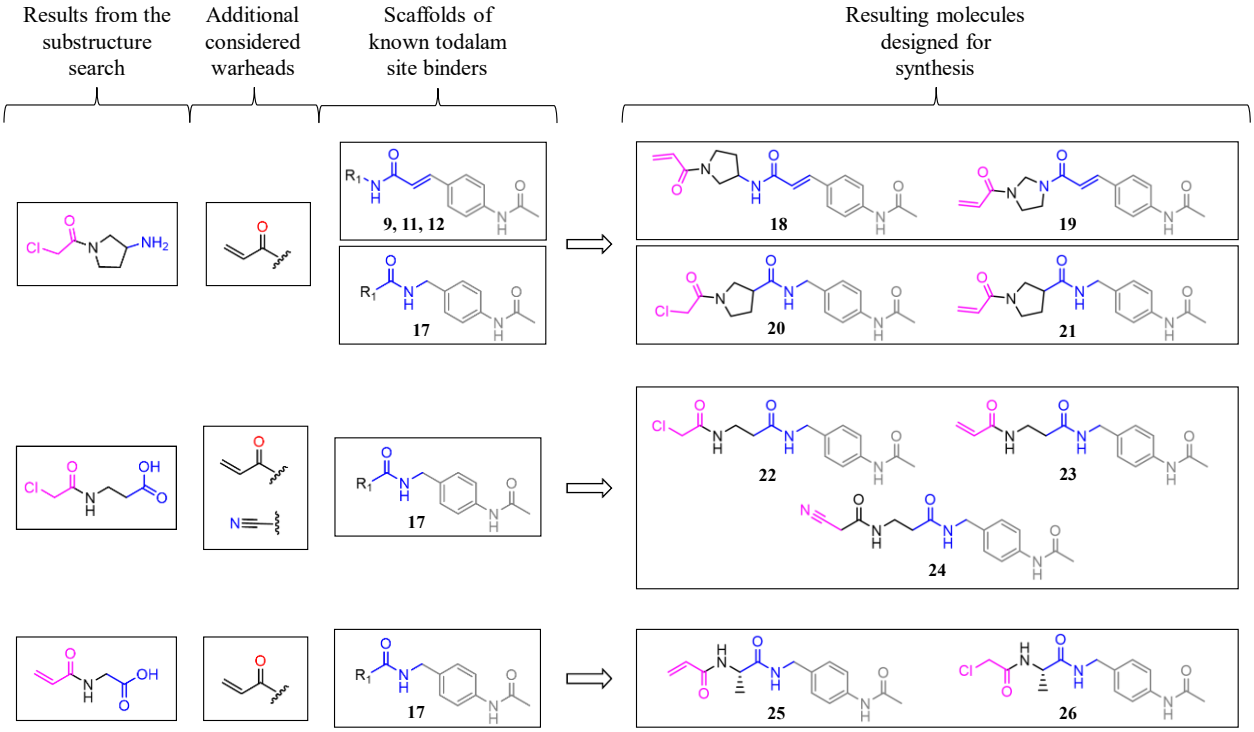


Figure III-43. Planned synthetic strategy for the synthesis of warhead-containing todalam site binders that present an amide and an acrylamide group as the central linker of the molecule. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink. On the left the SMARTS queries used to perform the substructure search are highlighted with a square.

The selection process of the hits was the same as the one carried out in the first building block search, but instead of generating urea-containing molecules with the building blocks, we created both amide and acrylamide-containing molecules depending on whether the terminal chemical group of the building block was an acrylic acid or a primary amine, respectively. Additionally, we added a last filtering step, which consisted of discarding those building blocks that cost more than 200 €/gram. The selected building blocks are shown on the left side of Fig. III-44, highlighted by a black square. Once again, on the basis of the computationally found building blocks, in collaboration with the chemistry team, we designed the molecules that could be synthesized in their laboratory. Therefore, we carried out a computational re-evaluation of the more synthetic-friendly 35 compounds by performing covalent and non-covalent docking, evaluating their docking score, binding mode, and established contacts within the todalam site. Of the 35 molecules, 10 of them were predicted to be suitable for our purpose and were accepted for synthesis. The strategy employed for the creation of the 10 chosen molecules is depicted in Fig. III-44.



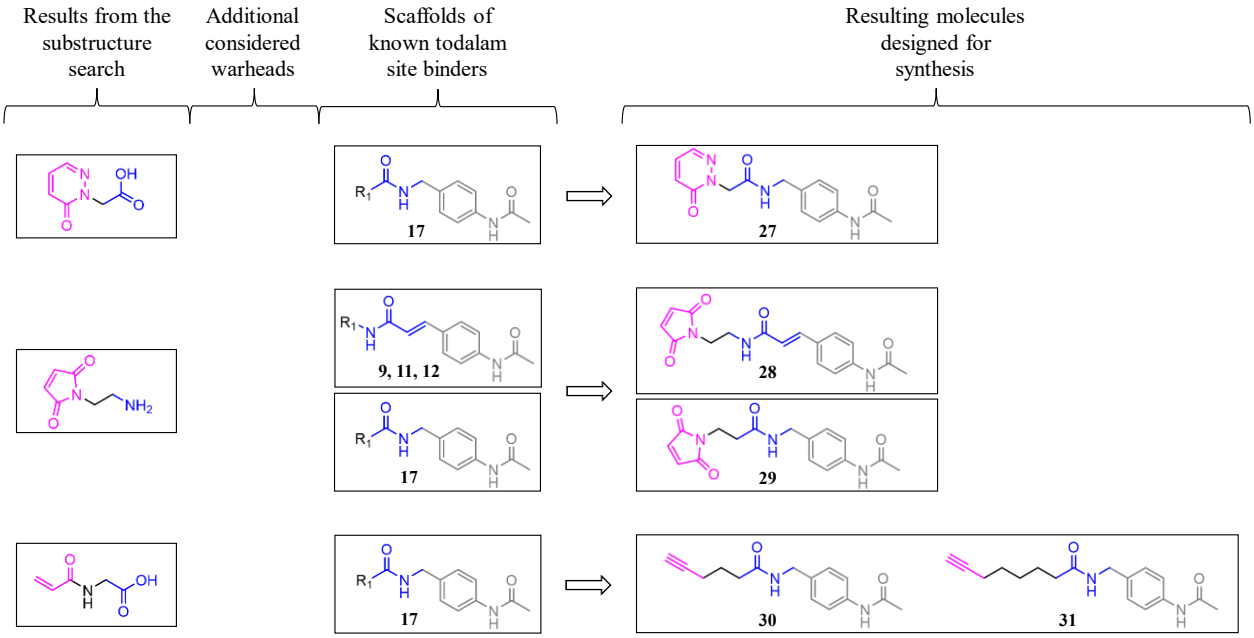
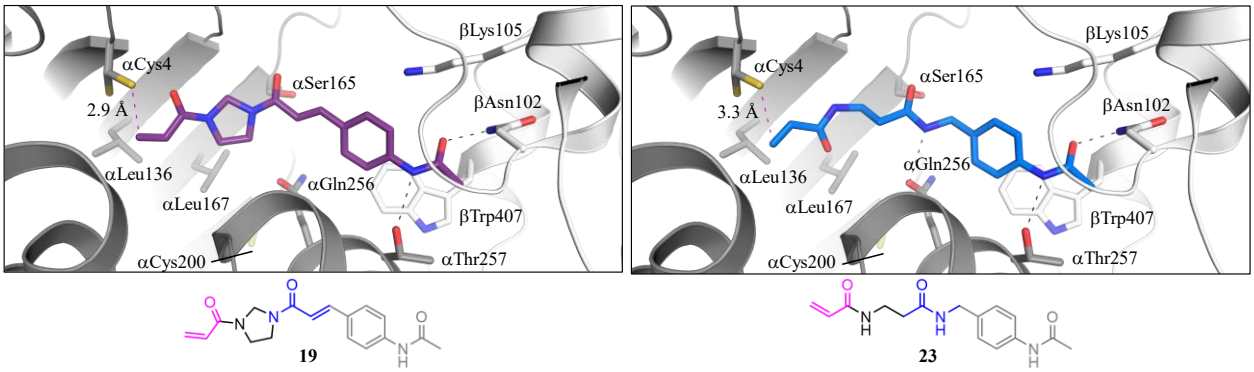


Figure III-44. Proposed warhead-containing todalam site binders designed in collaboration with the synthetic chemistry group. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink.

As an example, Fig. III-45 illustrates the results from the computational evaluation of some of these amide and acrylamide-containing new covalent binders. For simplicity, only a limited number of results are presented here, but the results from the other molecules were just as meaningful. Fig. III-45 shows the binding pose of the best conformer of compounds **19**, **23**, **25**, **27**, **29**, **30** and **31** obtained by S4MPLE. The non-covalent docking experiment identified a favorable binding pose for the potential covalent binders, with the distance between the warhead and the cysteine residue being within an acceptable range (~3 Å) for the reaction to happen. The binding mode was in line with our expectations. These results were consistent across multiple molecules tested, with similar results observed for the other molecules (**18-31**).



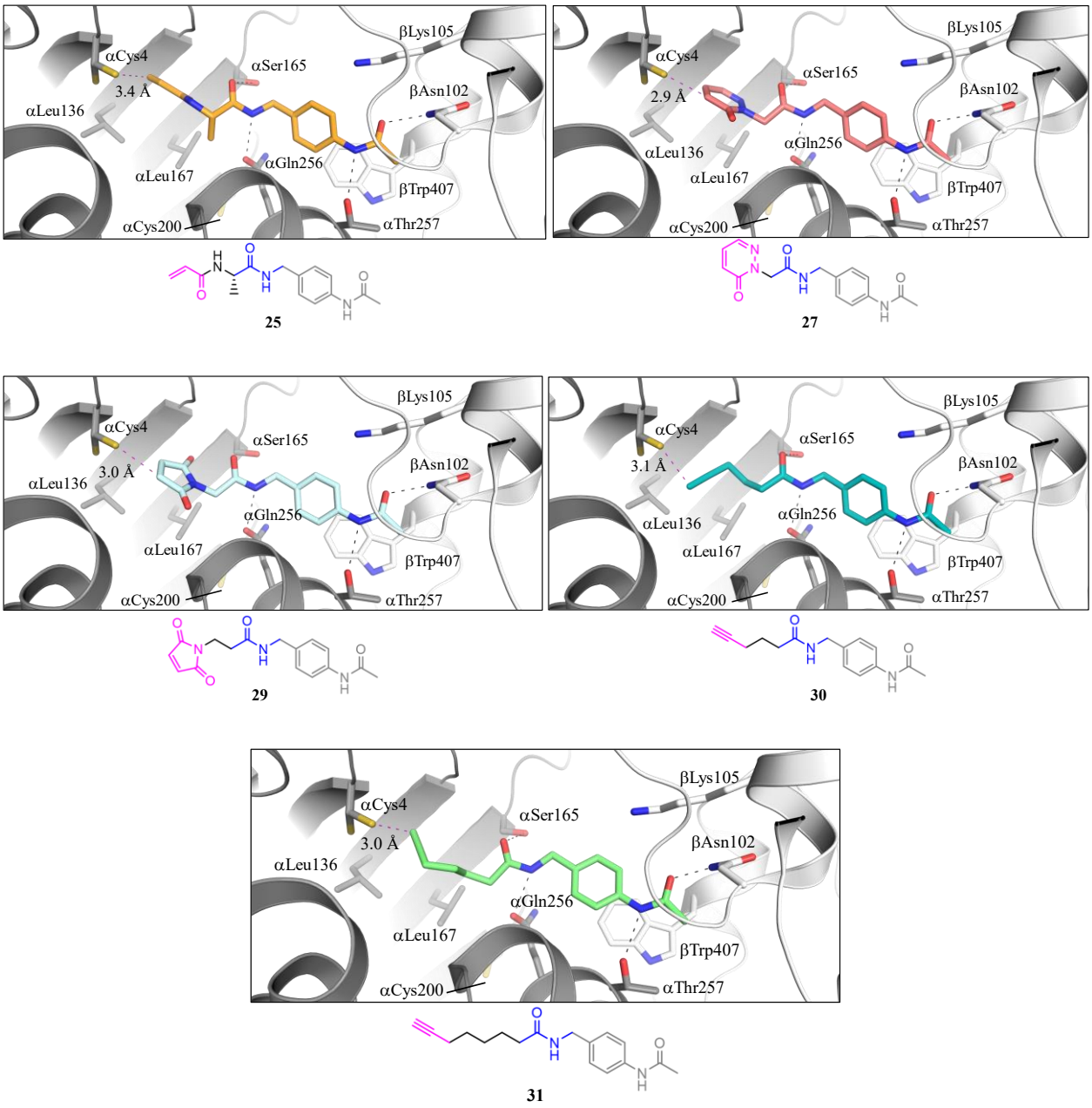


Figure III-45. Close-up of the interaction between compounds **19**, **23**, **25**, **27**, **29**, **30** and **31** and tubulin (gray). The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, chlorine atoms in green and fluorine atoms in cyan. Hydrogen bonds are displayed as black dashed lines. The distance between the reacting atom of each warhead and the thiol group of αCys4 is shown as a pink dashed line.

MD of these molecules was carried out by Anne-Catherine Abel at the University of Milan as a computational docking post-processing method that allowed the validation and refinement of the docking results and the analysis of the ligand-tubulin dynamics.

III.2-4.5.2. Triazole-containing Todalam Site Potential Covalent Binders

Furthermore, the synthetic chemistry team within the TubInTrain consortium has expertise performing 'click chemistry', which is a term used to describe a type of chemical reaction that is fast, efficient, and highly selective. The term was originally coined in 2001 to describe the copper-catalyzed azide-alkyne cycloaddition reaction (CuAAC).[101] This reaction allows the efficient and selective synthesis of 1,2,3-triazoles from alkyne and azide-containing compounds. In fact, the development of this reaction earned Sharpless and Meldal the Nobel Prize in Chemistry in 2022. Even though we observed that the triazole-containing hits were unable to form such specific interactions at the todalam site as those formed by the amide, acrylamide or urea-containing hits (Tab. III-2), the synthetic feasibility of the triazole-containing molecules made them attractive to us and were included in this second warhead-containing todalam site binders design campaign. In a physiological media the solvents would not affect the chemical structure of the triazole ring because of its stability in these conditions.

This substructure search campaign was also planned in collaboration with the chemistry group. The goal of the campaign was to identify suitable alkyne-containing building blocks to facilitate the synthesis of a new batch of triazole-containing compounds, as illustrated in Fig. III-46.

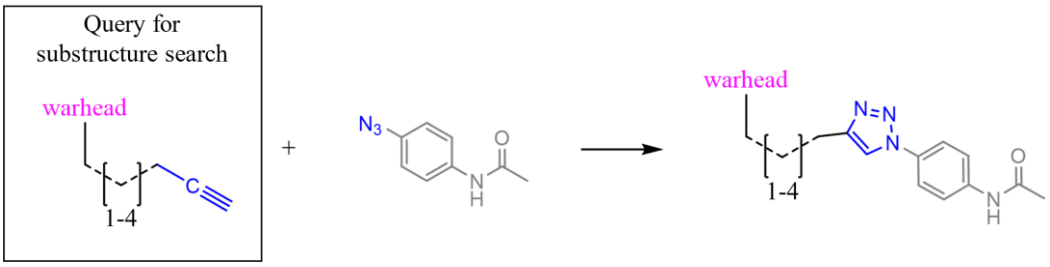


Figure III-46. Planned synthetic strategy for the synthesis of warhead-containing todalam site binders that present a triazole group in the central linker part of the molecule. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink. On the left is highlighted with a square the SMARTS query used to perform the substructure search.

In the present study, a substructure search campaign was carried out manually in SciFinder to identify economically accessible azide-containing building blocks. Seven interesting building blocks were identified and further used to generate compounds **32**-**38** shown in Fig. III-47.

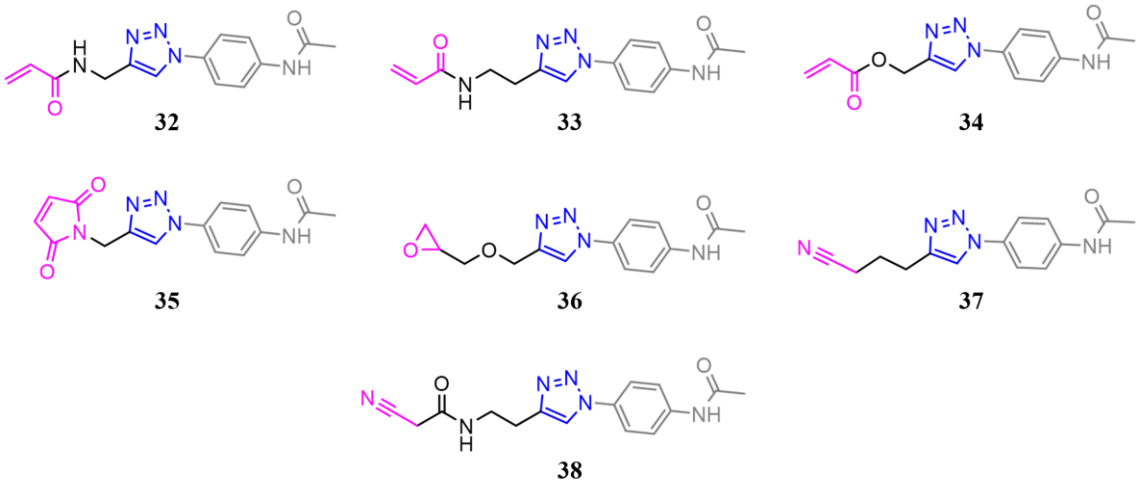


Figure III-47. Proposed warhead-containing todalam site binders that present a triazole ring as the central linker. The atoms involved in the formation of the anchor point are colored in light gray, the ones from the central linker in navy blue, and the warhead is represented in pink.

The designed molecules were docked both covalently and non-covalently. In the results of the non-covalent docking experiments, the distance between the warhead and the thiol group of the αCys4 was measured as exemplified in Fig. III-48.

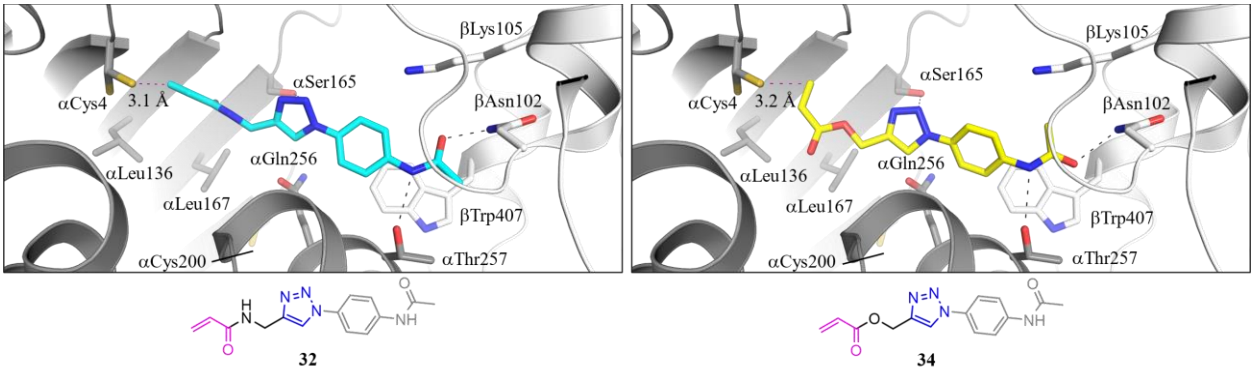


Figure III-48. Close-up of the interaction between compounds **32** and **34** and tubulin (gray). The α-tubulin and β-tubulin are represented in dark gray and light gray ribbon, respectively. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red and nitrogen atoms in navy blue. Hydrogen bonds are displayed as black dashed lines. The distance between the reacting atom of each warhead and the thiol group of αCys4 is shown as a pink dashed line.

III.2-4.6. X-Ray Crystallography Assays

To date, the chemistry group has synthesized five of the ten suggested compounds containing an amide as the central linker (molecules **22**, **25**, **26**, **30** and **31)** (Fig. III-44) which have already been tested by X-ray crystallography at PSI by Anne-Catherine Abel. All of these molecules were found to be bound to the todalam site, with molecule **31** exhibiting a continuous density towards the αCys4 thiol group, indicating a potential covalent bond formation (Fig. III-49). Further research is

necessary to confirm these findings such as refinement of the crystallographic findings and the performance of mass spectrometry. Nonetheless, the presented data provide valuable information on the optimal number of atoms to include in the molecules after the central linker to reach the thiol group of αCys4. Furthermore, the amide-based central linker was validated again since all of the molecules containing it were found to be bound to the todalam site.

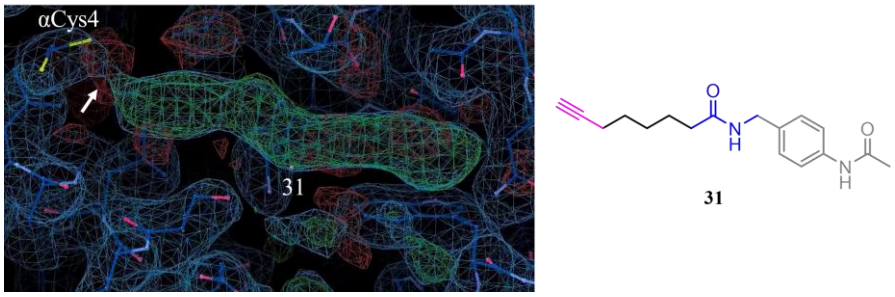


Figure III-49. Electron-density map of compounds **31** bound to the todalam site of tubulin in its corresponding T_2R-TTL complex. The SigmaA-weighted 2mFo - DFc (deep blue mesh) and mFo-DFc (green/red mesh) maps are contoured at +1.0 and +/-3.0, respectively.

The seven suggested triazole-containing molecules were synthesized and subsequently evaluated by X-ray crystallography. Of the seven molecules examined, only molecules **32**, **33** and **34** (Fig. III-47) were found to be associated with the todalam site, yet none of them exhibited covalent binding. It was previously acknowledged that triazole was not the most specific central linker among the ones identified. Nevertheless, even though the bound molecules do not appear to exhibit covalent binding, the eventual determination of their crystal structures will provide valuable structural information regarding potential chemical groups that could effectively bind to the todalam site.

The polymerization assays are yet to be performed in Madrid (Spain) at the CIB-CSIC facility.

III.2-5. Conclusions

In this study, we have established and validated a robust pipeline within the TubInTrain network, for the exploration of new scaffolds and molecules that bind to the novel tubulin site, known as the todalam site. The pipeline includes VS and molecular design, MD simulations to refine docking results, synthesis, X-ray crystallography, and *in vitro* assays. The application of computer-aided methods in the early stages of the pipeline has played a significant role in this study. Our VS of compounds utilizing structure-based approaches yielded numerous hits that were experimentally

validated. Substructure search methods have the advantage of computational efficiency and ability to handle large datasets, making them well-suited for the early stages of VS in this study, where the goal was to quickly filter out compounds that are not relevant to the task at hand. These initial results were then further filtered using docking experiments, which provide a more detailed understanding of the potential interactions between a given virtual hit and the target biomolecule.

X-ray crystallography data showed that the urea moiety was the most effective central linker in terms of establishing hydrogen bond interactions with the residues from the todalam binding site. However, due to the challenges in synthesizing isocyanates and the unavailability of commercial alternatives, we decided to focus on the amide, acrylamide, and triazole central linkers for further optimization into warhead-containing compounds. These central linkers were considered to be more synthetically feasible, thereby providing a more viable option for further development of todalam binders.

Our established pipeline enabled the rational design of 12 small molecules specifically targeting the todalam site, which were subsequently developed into warhead-containing compounds with a focus on binding to αCys4 within the todalam site. Of the 14 potential covalent binders designed, synthesized and tested, nine were found to be bound to the todalam site by X-ray crystallography experiments, while only one of them showed a continuous density suggesting possible covalent binding. Multiple design iterations, testing, and optimization of various compounds were carried out to find the most promising candidate, compound **31**. This compound allowed us to find an optimal length for a covalent binder targeting the todalam site, specifically the thiol group of the residue αCys4.

Our newly discovered todalam site binders open up a whole new realm of possibilities for future research, as they provide a foundation for the development of more potent and selective binders targeted toward the todalam site.

CHAPTER IV: Computer-Aided Maytansinol Functionalization to Develop Molecular Probes for Studying Microtubule Dynamics

Maytansine (Fig. IV-I, **1a**) is a natural product derived from the *Maytenus ovatus* plant and it is known for its potent cytotoxicity due to its ability to inhibit MT assembly upon binding with high affinity to β-tubulin (K_d = 6.8–14 nM).[102-104] Despite initial promising *in vitro* results, clinical trials using maytansine as a cancer treatment were unsuccessful due to poor efficacy and unacceptable toxicity.[105,106] However, based on the promising cytotoxicity shown by maytansine, scientists designed new maytansine derivatives which have been used successfully as antibody-drug conjugates (ADCs).[107-111] This success further prompted the development of new maytansine derivatives as cancer therapeutics, e.g., in the form of nanoparticles.[112,113]

Maytansinol (**1b**) is a precursor compound to maytansinoids and was obtained by isolation from the *Putterlickia verrucose* plant and its ester hydrolysis at the C3 position in maytansine. Maytansinol showed weaker inhibitory activity in tubulin polymerization than maytansine, indicating that the ester moiety at the C3 position of ansamitocins, such as maytansine and maytansinoids, plays an important role in their biological activity and cell permeability.[114,115] In addition, the carbonyl oxygen atom of the ester moiety has been found to form a strong intramolecular interaction with the hydroxyl at position C9, fixing the bioactive conformation.[116]

Maytansinol has been regarded as a valuable precursor for the preparation of both natural and new semisynthetic maytansinoids, which differ in ester side chain substituents.[117,118] The acylation reaction of maytansinol is a crucial step in the preparation of maytansinoid ADCs or nanoparticles, which, as aforementioned, constitute an emerging class of cancer therapeutics (Fig. IV-1).[117-121] Furthermore, considering that the maytansine binding site is one of the most recently identified and least explored site on tubulin, the design of new maytansinoids could also serve as molecular probes to better understand the structure-activity relationships of maytansinoids or to identify new ligands able to bind to the maytansine site.[67,122]

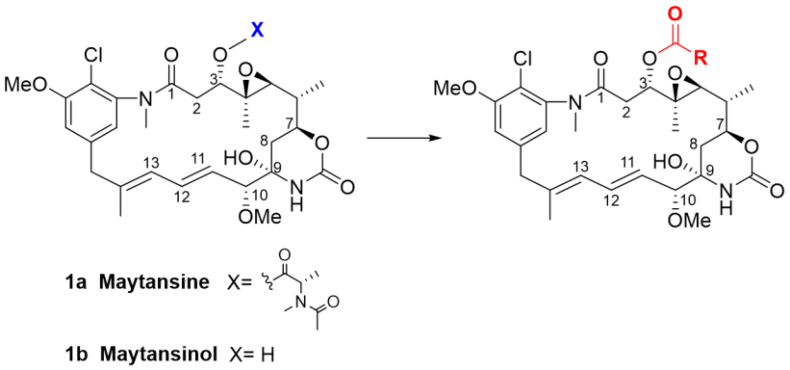


Figure IV-1. 2D molecular structure of maytansine (**1a**), maytansinol (**1b**) and the general acylation reaction of maytansinol.

In 2014, Prota *et al.* solved the high-resolution crystallographic structure of the complex formed by β-tubulin and maytansine (PDB ID: 4TV8).[38] This structure allows to understand the intermolecular interactions between maytansine and β-tubulin at the atomic level and the spatial arrangement of the complex, thus providing the basis for the rational design of maytansine derivatives (Fig. IV-2).

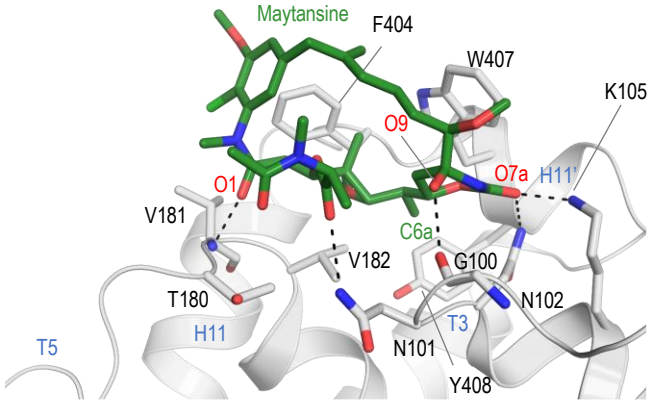


Figure IV-2. 3D representation of the complex formed by the association of maytansine (green) with tubulin (gray) extracted from crystallographic data.[38]

The potent biological activity of maytansinoids, their diverse potential applications, and their highly functionalized macrocyclic structures inspired us to investigate the design of derivatives of maytansinol. Pursuing our interest in the chemistry and biological activity of tubulin binders,[123-126] a set of new maytansinoids were obtained as a result of the maytansinol acylation reaction and other structural transformations not previously reported. The research was carried out in two stages: (1) The first stage focused on the design and synthesis of long-chain maytansinoids and maytansinoid conjugates with nucleotide mimetics. These maytansinoids were computationally studied and experimentally tested; (2) The second stage of the study consisted of using molecular

docking to predict the binding mode of seven short-chain maytansinoids when bound to the maytansine site in β-tubulin. These maytansinoids were byproducts of the reactions performed by the chemistry team during the synthesis of the maytansinoid conjugates. In addition, binding affinity assays, cytotoxicity tests, and X-ray crystallography were performed on these molecules. The results of this study are important both for the design and synthesis of new effective maytansinoids, for their use as molecular probes and potential therapeutic applications.

IV.1 Design and Modeling of Long Chain Maytansinoids and Maytansinoid Conjugates

IV.1-1. Introduction

The generation of maytansine-based molecular probes is challenging due to the complexity of the natural product scaffold. New studies on the chemistry of maytansine derivatives could help to guide the attachment of fluorophore tags or radionuclides.[67,127] Therefore, we decided to further investigate new C3 functionalized maytansinoids.

Maytansine C3 actually happens to be in close proximity (~12 Å) from the exchangeable nucleotide site (E-site) occupied by GDP/Mg^{2+} (Fig. IV-3). The first interesting factor that could be exploited to exert a novel effect on MTs is designing long-chain maytansinoids that can interact with the Mg^{2+}. Furthermore, the design of bivalent compounds containing maytansinol and nucleotide mimetics would a) increase the understanding of the ability of maytansinoids to accommodate at the binding site independently of the size of the substituents, and b) test the hypothesis that such linking is feasible, and see whether one could gain some synergy from bidentate binding, in spite of the pressure of high-concentration GDP acting against its replacement.

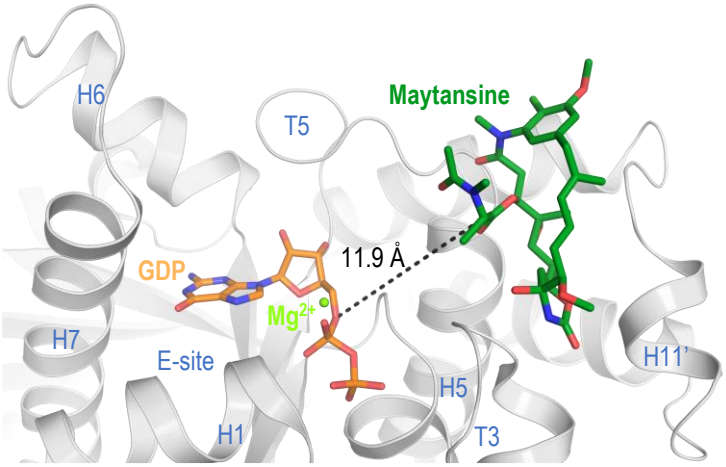


Figure IV-3. Section of the crystal structure with PDB ID 4TV8 in which maytansine (dark green), GDP (orange), Mg^{2+} (green sphere) and β-tubulin (gray) are present. The black dashed line represents the distance between the position C3 of maytansine and the 5'-OH of GDP which is ~12 Å. The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in light green.

Tubulin inhibition by nucleotide analogues is very challenging because the concentration of GTP is around 300 μM inside the cells.[128] However, we sought to investigate whether the presence of the maytansinol moiety in a bivalent compound could favor binding of the nucleotide portion. The maytansinol moiety would act as an anchor point that holds the modified nucleotide portion in close proximity to the E-site through a flexible linker, thus favoring the nucleotide exchange.

Consequently, we decided to prepare long-chain maytansinol derivatives and maytansinoid conjugates to target the nucleotide-bound Mg^{2+} or the nucleotide binding site, respectively (Fig. IV-4). Different linkers to magnesium chelators or nucleotide mimetics were selected by their ability to favorably interact with the protein while spanning the required distance. Furthermore, we were interested to investigate whether the new molecules would produce different effects on tubulin structure and dynamics.

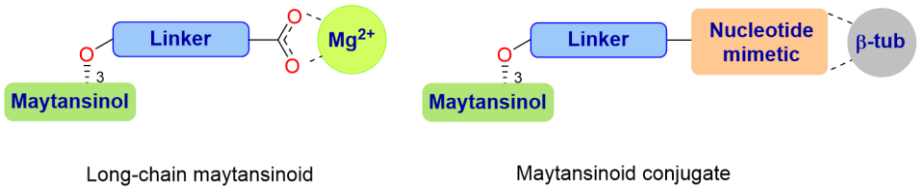


Figure IV-4. Schematic drawing of both the maytansinoids and the conjugates designed to target the nucleotide-bound Mg^{2+} or the exchangeable nucleotide site (E-site) of tubulin (β-tub).

In this study, we report the tolerance of the C3 position of maytansinol to the addition of bulky substituents by not being sterically hindered. These findings provide a solid foundation for further in-depth exploration of maytansinoids as molecular tools for investigating MT dynamics.

IV.1-2. Aim of the Work

The aim of this research was to discover by VS linkers that fit in at position C3 of maytansinol and: (1) coordinate the Mg^{2+} located in the E-site, (2) bridge the distance to the E-site while picking up favorable interactions with the protein.

IV.1-3. Computational Details

The docking tool S4MPLE[53,54] was used to model: a) the long-chain maytansinoids and maytansine conjugates containing Acyclovir *via* molecular growth, and b) the maytansine conjugates containing guanosine *via* molecular linkage.

The high-resolution crystallographic structure used for the docking studies was obtained from the PDB entry with ID 4TV8.[38]

IV.1-3.1. Fragment-Based Docking Protocol of S4MPLE

S4MPLE has been developed to provide versatile control of degrees of freedom, especially in docking scenarios where a portion of the ligand (the 'anchor' fragment(s)) can be kept fixed in place while selected degrees of freedom in the protein may be actively sampled. Anchors can be fragments in their X-ray-determined binding pockets, as in Fragment-Based Drug Design, which can be 'linked' or 'grown' using S4MPLE. Here, the anchors were: a) maytansinol in view of 'growth' by long chains designed to reach Mg^{2+} or long chains attached to the Acyclovir moiety to reach the nucleotide site, or b) maytansinol and guanosine in view of 'linking'. Flexibility of the protein was not allowed. The steps of this growth/linking protocol are similar to those previously reported,[53] with some minor amendments:

IV.1-3.1.1. Target Site Preparation

In S4MPLE, the binding site is defined by generating a cut-out zone around the ligands. Two different input target sites were generated.

For the docking of molecules **2**-**4**, we considered all the atoms of the target site present in residues within a radius of 10 Å around the maytansine molecule present in the structure 4TV8 including the cation Mg^{2+}. For the docking of molecules **5**-**8**, we considered all the atoms of the target site present within a radius of 10 Å around the maytansine and GDP molecules. In this case, the cation

Mg^{2+} was excluded. Docking input files for the target protein β-tubulin were generated using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.[76] The protein site file was stored in the 'ref.mol2' format in Sybyl MOL2 in a site directory, together with the list of fixed atoms (here, all protein atoms were considered fixed).

IV.1-3.1.2. Preparation of Anchors

Anchor fragments must be specified in SDF format, with the coordinates taken from the fragment pose in the protein site from the crystal structure. The anchor fragment must be fully hydrogenated in agreement with the assumed protonation status of putatively ionizable groups, with one notable exception: the hydrogen atoms yielding the free valency expected to be used to connect to the linkers must be deleted, while the concerned connecting heavy atoms must be marked using the atom mapper functionality of structure editors such as MarvinSketch.[75] In a 'growth' scenario, the only connecting heavy atom in the single anchor fragment must be mapped to '1'. For 'linking', both anchors must be given in the same SDF, with connecting heavy atoms mapped as '1' and '2', respectively. At least some of the atoms in both anchor fragments must be defined as fixed. A specific 'anchor' directory was provided containing the aforementioned SDF and a list of fixed atoms.

IV.1-3.1.3. Preparation of Linker Fragments

Mono (for 'growth') or bidentate fragments (for 'linking') must be provided in SMILES format, with respective connecting atoms mapped '1' or '1' and '2' such as to match the marks of associated anchor connectors. The mapping was performed using MarvinSketch.[75]

IV.1-3.1.4. Preparation of the Ligand Structures

The SMILES string of each linker fragment is first converted to 3D, with explicit hydrogens as expected at physiological pH, using the same ChemAxon API-based tool serving to preprocess regular ligands. The resulting fragment SDF is merged with the anchor, and the list of anchor fixed atoms is imported. Then, S4MPLE is invoked with the 'connector' directive, creating covalent bonds between marked atoms of the same map number. Note that, at this point, neither the required GAFF[57] atom types nor the partial charges are assigned. However, S4MPLE will proceed to a coarse fitting of the linker to the (fixed) anchor atoms, based on the 'default' force field parameters,

based on a few generations of evolutionary algorithms. This will coarsely preposition the linker in the reference frame of the anchor, all while roughly ensuring reasonable bond lengths and valence angle values (as far as possible) and avoiding anchor-linker atom clashes. The resulting ligand geometry is stored together with the inherited list of fixed anchor atoms. It is next subjected to GAFF force field typing, Gasteiger partial charge calculation and detection of saturated rings requiring explicit sampling, and then added to a tar archive as ready for docking.

IV.1-3.1.5. Docking Protocol

S4MPLE is used with the 'fitted'[53,54] parameterization of the AMBER[55,56]/GAFF[57] engine, including an implicit desolvation model, as previously reported. The docking protocol followed in this study is similar to that explained in section III.2-3.2.2.1. consisting of five different main steps that follow the linker insertion step already explained: (1) in-site linker sampling, (2) ligand relaxation, (3) RMSD assessment, (4) free-ligand sampling, and (5) ranking. In the case of the docking protocol followed to dock the long-chain maytansinoids containing a terminal carboxylic acid **2**-**4** and the maytansinoid conjugates **7** and **8** containing an Acyclovir moiety as nucleotide mimetic, only the maytansinol crystal structure was fixed as an anchor point. In the case of the docking protocol followed to dock the maytansinoid conjugates **5** and **6** containing a guanosine moiety as nucleotide mimetic, both the maytansinol and guanosine crystal binding poses were kept as (fixed) fragments (Fig. IV-5).

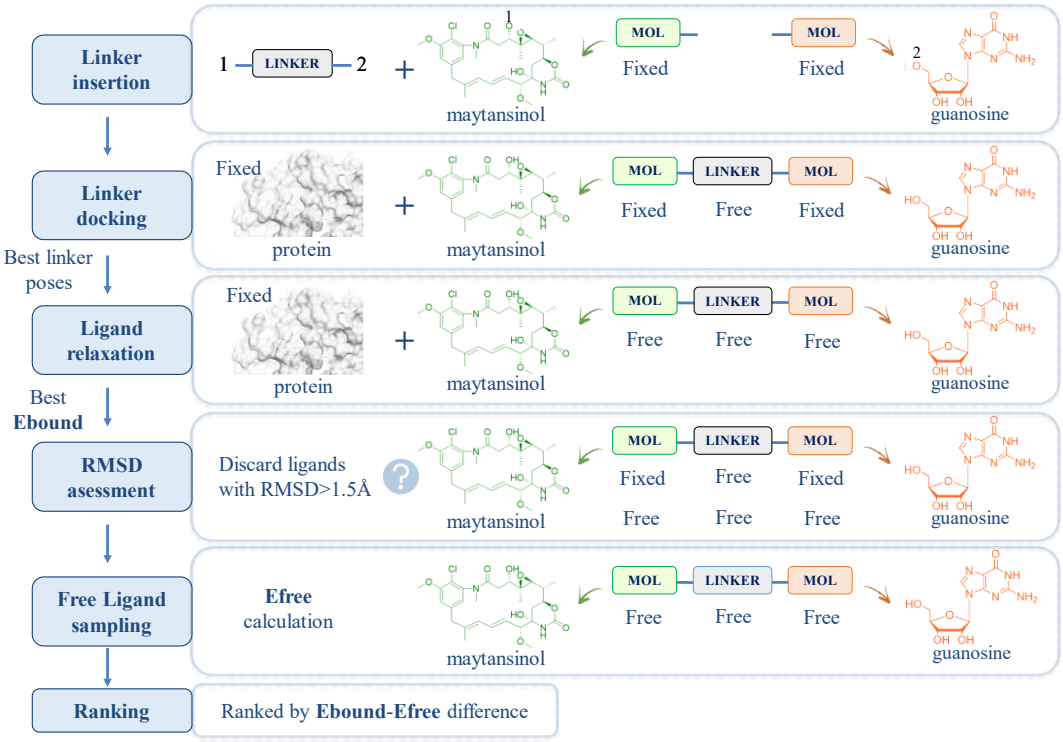


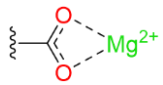
Figure IV-5. Docking protocol of S4MPLE.

IV.1-4. Results and Discussion

IV.1-4.1. Long-Chain Maytansinoids

The presence of a Mg^{2+} at the E-site of β-tubulin plays a key role in the coordination of GTP/GDP binding by stabilizing the negatively charged phosphate groups of the molecules, thus increasing the binding strength. The Mg^{2+} also acts as a catalytic agent in the hydrolysis of GTP to GDP, which is an essential step in the dynamic instability of MTs, allowing for polymerization and depolymerization processes. As explained, interactions involving metals may play an important role in biological systems such as MTs. Chemical groups that possess a lone electron pair can form either an ionic bond or a coordination bond with an alkaline earth metallic cation or a metallic cation, respectively. Common residues involved in these type of interactions include aspartate, glutamate, histidine and cysteine and the main (alkaline earth) metallic ions encountered are zinc, iron, calcium, magnesium, copper, and manganese.[129] Unlike transition metals, Mg^{2+} has a low propensity for coordinative covalent binding to nucleophilic entities and is more likely to interact with anionic ligands (Fig. IV-6). The distance of a coordination bond depends on the chemical

groups involved and particularly the metal.[130] In practice, a maximum empirical distance ranging from 2.6 Å to 2.8 Å is normally used to identify this type of interaction.



Double coordination

Figure IV-6. Example of ligand-metal ion coordination bonding.

Based on this information, we suggested that it could be possible for a ligand containing a carboxylate to establish this type of interaction with the Mg^{2+} present at the E-site while binding to the maytansine site in β-tubulin. However, the effectiveness of binding of the ligand to the Mg^{2+} would depend on how well it can compete with the GDP phosphates which partially compensate for its positive charge. It is difficult to accurately predict using FF methods whether a $-COO^-$ terminated linker of the appropriate length will establish an interaction with Mg^{2+}. This is because FF methods do not take into account the fine balance between factors such as electrostatics, coordinative bonding contributions, and the entropic loss that results from fixing the flexible linker in a position that reaches out to Mg^{2+}. Docking considers only the electrostatic interactions and a continuum desolvation contribution, which is likely to be inaccurate for scenarios involving metal ions. This is because the solvent around metal ions is usually structured and does not act like a dielectric, so the estimate of the contribution of the solvent is likely to be flawed. In this instance, only experimental proof can be deemed reliable.

For the discovery of effective linkers for magnesium chelators we conducted a comprehensive screening of a library of possible linkers with varying lengths and substitutions containing a carboxylate at the tip of the chain that could potentially interact with the Mg^{2+}. By computationally investigating the distance between the C3 of maytansinol and the Mg^{2+} using the visualization tool PyMol, we estimated that the length of the long-chain substituents should range between 9-13 carbon atoms. This would provide the chain with the necessary degrees of freedom to correctly accommodate and interact with Mg^{2+}. To ensure synthesis feasibility, a linker library of 10 different linkers of several lengths and substituents was designed in collaboration with the TubInTrain chemistry team. The screening process involved the evaluation of each linker based on its ability to interact with the Mg^{2+} ion. After careful analysis, we selected the top three linkers shown in Fig. IV-7.

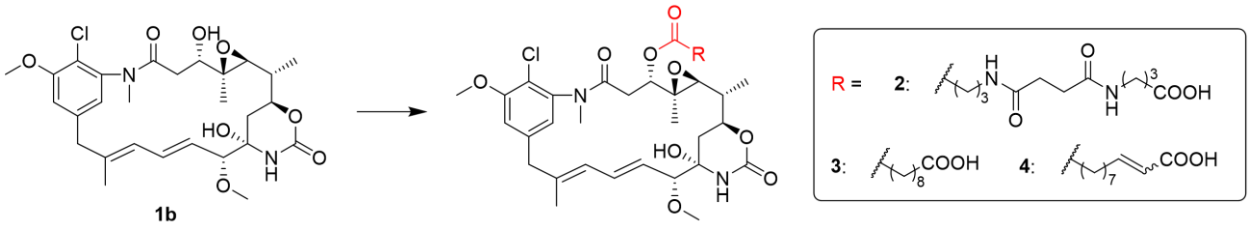


Figure IV-7. 2D molecular structure of maytansinol (**1b**) and the long-chain maytansinoids **2**-**4**.

The results of the docking studies indicated that these three long chains have the correct length and flexibility to reach the magnesium ion and interact with it while binding to the maytansine site maintaining the core ring in the correct conformation (Fig. IV-8). Furthermore, it was observed that the specific stereo configuration (E or Z isomerism) of molecule **4** is not crucial to form the correct interactions with the Mg^{2+}, as both isomers were able to establish the same interactions despite slight variations in the positioning of the linkers.

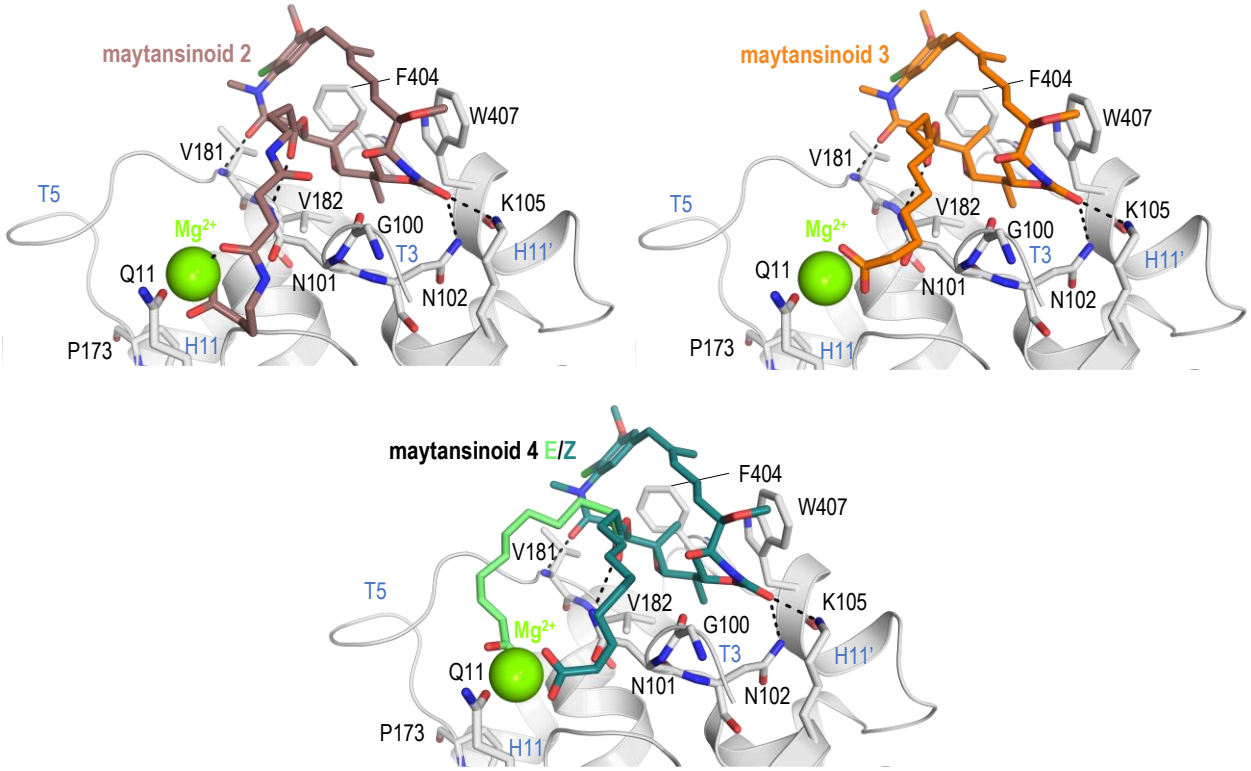


Figure IV-8. Docking results of the long-chain maytansinoids **2**, **3** and **4** when bound to β-tubulin in the presence of the Mg^{2+} (green sphere). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green. Hydrogen bonds are displayed as black dashed lines.

IV.1-4.2. Maytansinoid Conjugates

We first decided to conjugate the maytansinol moiety with guanosine mimetics and based the rational linker design on the distance of ~12 Å (Fig. IV-3). In the linker design stage, the main goal was to define the appropriate length and flexibility to secure the accommodation of maytansinol and nucleotide-like scaffolds at their respective β-tubulin binding sites.

We focused on linkers that could allow the conjugation of the two compounds of interest by means of click chemistry, because the chemistry team showed clear interest in synthesizing the maytansine conjugates by a CuAAC reaction between the fragment of an alkyne and an azide providing the corresponding triazole derivatives. For this, we generated a library of 54 possible linkers in which the position of the triazole group and the length of the carbon tails attached to it would vary. Furthermore, since the measured spatial distance between the two functional groups of interest was of ~12 Å, we considered linker lengths that ranged from 12-14 atoms. Therefore, the linker size was rationally chosen by determining that the sum of A and B should always be greater or equal to 5 and smaller or equal to 13 (Fig. IV-9).

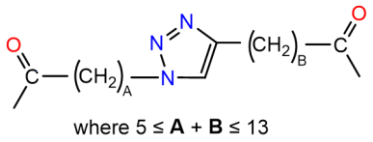


Figure IV-9. Linker design strategy: schematic version of the applied approach followed for the generation of 54 triazole-containing linkers of different lengths.

For the linker screening, the maytansinol and the guanosine molecules were treated as fragments and were fixed in their states seen in the crystal structure 4TV8. Then, we carried out linker screening and used the S4MPLE docking software to find the linker that best bridged the gap between the two fragments (maytansinol and guanosine). After the calculations, the RMSD values for all of the ligands were determined. The small difference observed between the lowest (0.543 Å) and the highest (0.750 Å) RMSD values suggested that all linkers were suitable. Thus, since none of the generated linkers put any strain on the warheads (maytansinol and guanosine), the linkers to be synthesized were chosen on the basis of their synthetic convenience. We also explored the possibility of designing a more rigid linker by adding a *p*-phenylene spacer in the side chain of maytansinol. As expected, in this case we observed that the degrees of freedom of the linker were reduced, thus better orienting the linker moiety towards the nucleotide pocket.

Therefore, based on this computationally predicted structural information study, the chemistry team decided to proceed with the synthesis of the maytansine conjugates **5** and **6** (Fig. IV-10).

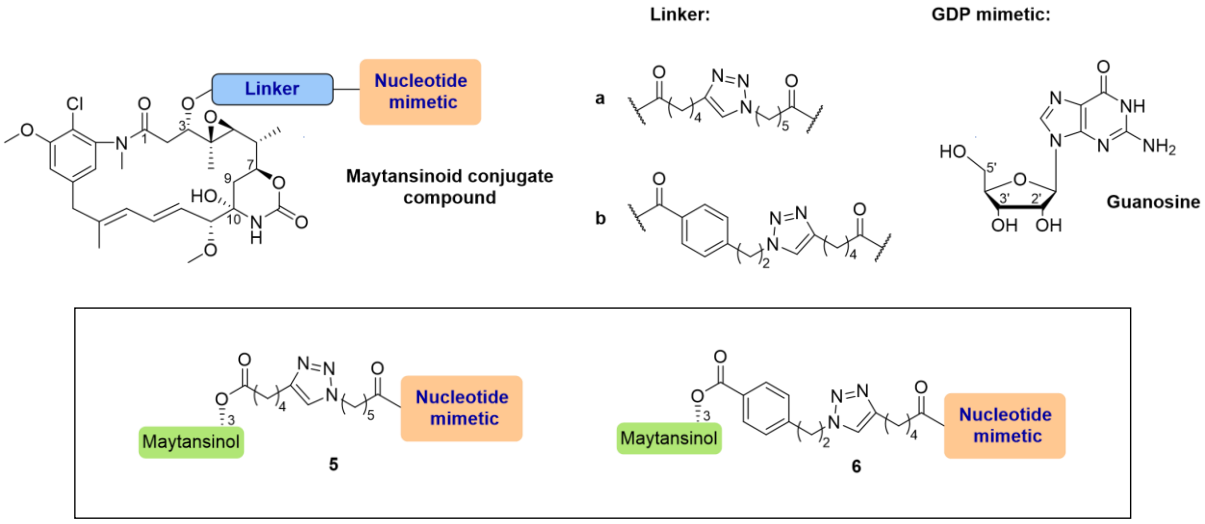
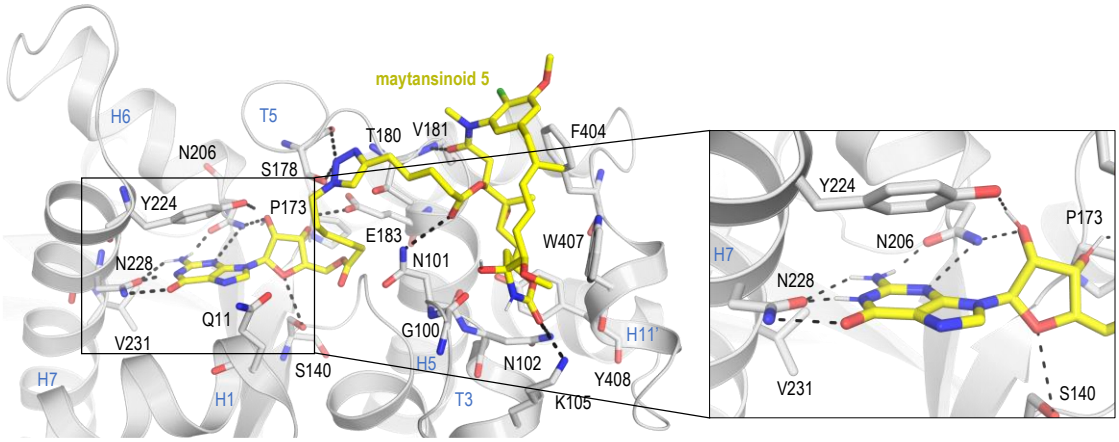


Figure IV-10. Structures of maytansinoid conjugate compounds designed to target maytansine and nucleotide binding sites.

We analyzed the validity of the linkers present in conjugates **5** and **6** with docking. The results showed that both linkers can flexibly connect the maytansinol and the guanosine entities in a suitable distance, which allows the two entities to bind by maintaining a proper geometry (Fig. IV-11).



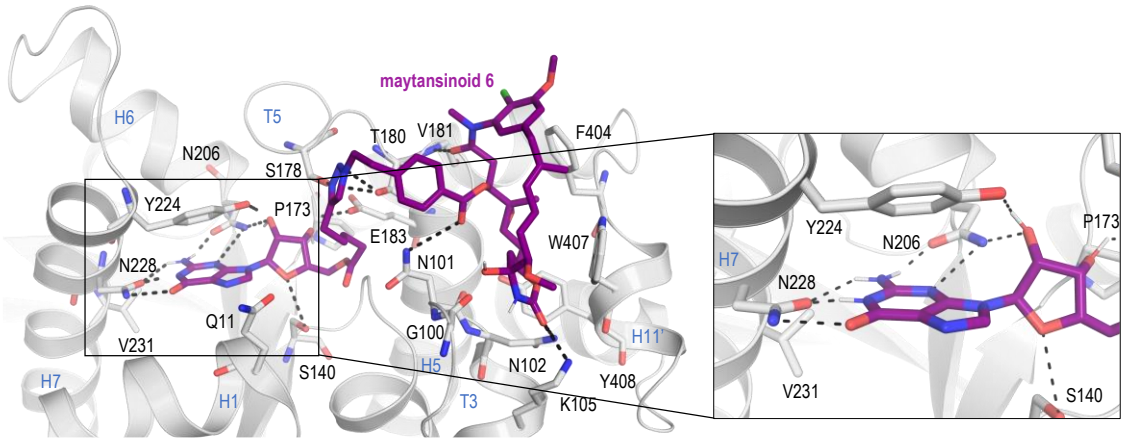


Figure IV-11. Docking results of maytansinoid-guanosine conjugates **5** (yellow) and **6** (violet). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green. Hydrogen bonds are displayed as black dashed lines.

Moreover, according to the chemistry team, the synthesis of the guanosine-containing bivalent compounds is a very challenging task. Among other hurdles, the protection and deprotection of the 2' and 3' hydroxyl groups of the guanosine ribose add two additional steps to the synthetic pathway enhancing its complexity. Therefore, our objective was to also design bivalent compounds capable of binding to the maytansine and GDP sites present in β-tubulin in which the guanosine entity would be simplified by a guanosine substitute. To design a guanosine substituent that could fit into the GDP binding site, we thoroughly analyzed the interactions established by GDP with tubulin. We could observe that guanine was the main moiety for binding of guanosine to β-tubulin. The base establishes very precise and specific interactions (hydrogen bonds and π-stacking) with the target protein β-tubulin (Fig. IV-11). Since guanine was identified as the hallmark of the guanosine molecule and the absence of the 2' and 3' hydroxyl groups from the guanosine ribose would simplify the synthetic steps of the compounds, we decided to substitute the guanosine molecule by a nucleic acid mimetic that would fulfill these requirements. In order to explore different nucleic acid mimetics we investigated the structure of several antiviral agents. This is how we selected the antiviral drug Acyclovir as a mimetic for GDP in our maytansinoid conjugates.

Based on this information, in collaboration with the chemistry team, we designed the maytansinoid conjugates **7** and **8** for synthesis containing Acyclovir (Fig. IV-12).

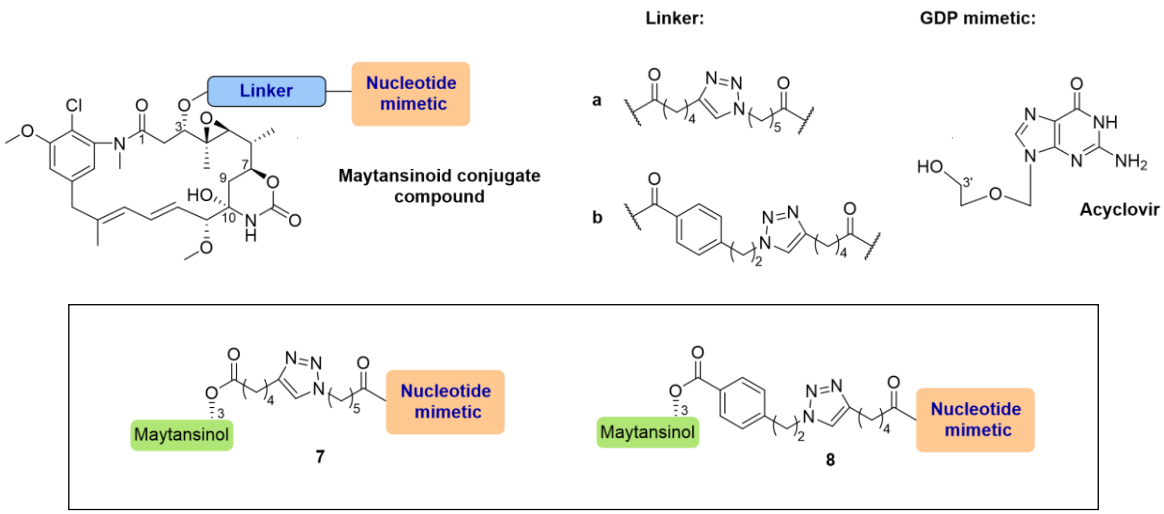
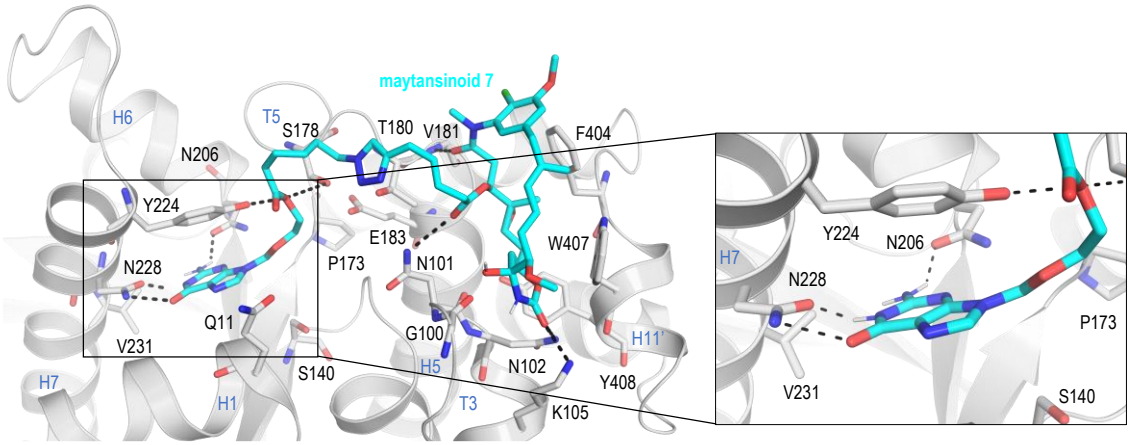


Figure IV-12. Structures of maytansinoid conjugate compounds designed to target maytansine and nucleotide binding sites.

In our docking experiments of the maytansinoid conjugates containing Acyclovir, only the crystallographic binding pose of maytansinol was fixed, while the linker and the attached Acyclovir moiety were allowed to explore the binding site in a flexible manner. This was necessary, since there is no X-ray data describing the preferred pose of Acyclovir in the E-site of tubulin. This more challenging setup allowed us to assess whether the sampling protocol is sufficient to permit the Acyclovir moiety sample and occupy the nucleotide binding site and establish interactions similar to those of guanosine within the site. The results showed that both maytansinoid **7** and **8** were able to reach the nucleotide binding site and establish hydrogen bond interactions through the base with the sidechain atoms of the amino acids Asn228 and Asn206 (Fig. IV-13) in a similar manner to guanosine (Fig. IV-11).



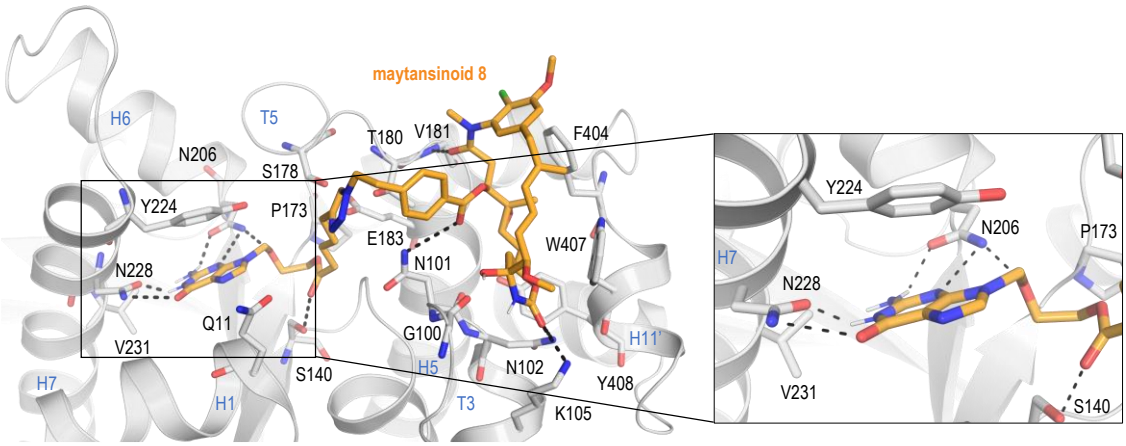


Figure IV-13. Docking results of maytansinoid-Acyclovir conjugates **7** (cyan) and **8** (orange). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green. Hydrogen bonds are displayed as black dashed lines.

Docking studies confirmed that the 5-membered ring of the ribose can, in principle, be substituted for binding to β-tubulin. In Fig. IV-14, we can observe in the superimposition of the best docking conformers of maytansinoid conjugates containing the same linker **5** and **7** and molecules **6** and **8** that the modes acquired by these pair of molecules are very similar even after deletion of the 5-member guanosine ring. This choice led to simplifications in the synthesis of maytansinoid conjugates.

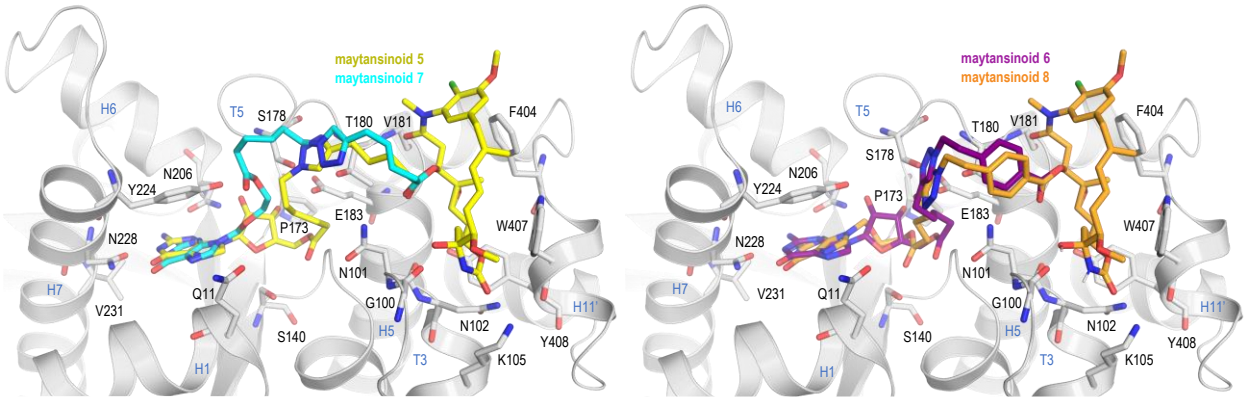


Figure IV-14. Left: Docking result of the maytansinoid-guanosine conjugate **5** (yellow) superimposed on the docking result of the maytansinoid-Acyclovir conjugate **7** (cyan) when bound to β-tubulin. Right: Docking result of the maytansinoid-guanosine conjugate **6** (violet) superimposed on the docking result of the maytansinoid-Acyclovir conjugate **8** (orange) when bound to β-tubulin. The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green.

IV.1-4.3. Biological Assays

IV.1-4.3.1. Binding Constant Determination

The binding affinity of compounds **4-8** to the maytansine site was assessed using a competition assay at CIB-CSIC, where their binding to tubulin was compared to a reference compound (Fc maytansine). In a competition assay, the tested compound and the reference compound are added to the target site and the strength of their interaction is determined by measuring the amount of reference compound, in this case maytansine, that remains bound to the target in the presence of the tested compound. This assay allowed to determine the K_d (dissociation constant) values of each tested maytansinoid. K_d represents the concentration of a ligand at which half of the binding sites on a protein are occupied, so a lower K_d means a stronger binding interaction. K_d and binding affinity are inversely related, meaning that as K_d decreases, the binding affinity increases. Therefore, with the displacement assay, the goal was to obtain the binding affinities of the compounds and compare them with the binding affinities of maytansine. The results showed that the K_d of compounds **4-8** to the maytansine site is well conserved, suggesting that bulky substituents have a minimal effect on the interaction between the maytansine core and tubulin (Tab. IV-1). Furthermore, the results imply that the nucleotide moiety has a limited contribution to the interaction between the compounds and tubulin.

Table IV-1. Dissociation constant of maytansinoid compounds. The data come from three independent measurements.

Compound	1a	4	5	6	7	8
K_d (nM)	11 ± 1	188 ± 22	83 ± 12	83.4 ± 7.8	137 ± 14	125 ± 14

IV.1-4.3.2. Cell Viability Assays

To understand the relationship between binding affinity and cytotoxicity and to assess the ability of the compounds to counteract multidrug resistance caused by membrane pumps, at CIB-CSIC in Madrid they also tested the cytotoxicity of the compounds in A549 cells and in the isogenic pair A2780/A2780AD (P-gp-overexpressing) cell lines (Tab. IV-2). A lower IC_{50} value means that the compound is more potent in killing cells, as it indicates a lower concentration of compound required to achieve a 50% decrease in cell survival.

Table IV-2. IC_{50} of maytansinoids in A549 and A2780/A2780AD cell lines. The data come from three independent measurements.

Compound	A549 IC_{50} (nM)	A2780 IC_{50} (nM)	A2780AD IC_{50} (nM)
1a	0.92 ± 0.07	0.26 ± 0.03	12 ± 0.4
4	4.9 ± 0.3	2.0 ± 0.4	400 ± 70
5	11600 ± 700	620 ± 50	21000 ± 1000
6	4800 ± 300	160 ± 10	16000 ± 2000
7	10.4 ± 0.5	4.4 ± 0.5	900 ± 100
8	49 ± 3	35 ± 4	3600 ± 500

The cytotoxicity results obtained from the tumoral cell lines are in agreement with the binding affinity data of compound **4**, which exhibits potent activity. On the contrary, compounds **5-8**, which contain a triazole moiety in their structures generated by a CuAAC reaction, experienced a substantial decrease in activity while retaining binding affinity, indicating that the consequent bulky substituents impede internalization of the compounds within the cells.

IV.1-4.4. X-Ray Crystallography

We aimed to investigate the effect of attaching long chains to the maytansine scaffold on the binding mode of the molecules by determining the structures of tubulin-maytansinoid complexes using X-ray crystallography. Anne-Catherine Abel at PSI performed the crystallographic analysis of compounds **2-8** following both a soaking and co-crystallization protocol. The analysis confirmed the binding of all compounds to the maytansine site, as evidenced by the presence of ligand-induced density differences at the site for all compounds, indicating that none of the introduced moieties hinders binding in the crystal system. However, only the central ring of the maytansine scaffold was well defined in the crystal structures, while the introduced linkers remained indistinct in any of the maytasinoid-tubulin complexes. This is consistent with the observed conservation of binding affinities.

The maytansine core ring of all evaluated molecules showed good superimposition with the parent compound maytansine, and no major modifications in the binding mode were observed. The insertion of long chains at the C3-O position did not influence the previously described interactions between the core ring of maytansinoids and tubulin, including the hydrogen bonding of C1 carbonyl with nitrogen in the main chain of Val181, C24 carbonyl with the amine groups of the Lys105 and Asn102 side chains, and C9-OH with carbonyl in the main chain of Gly100. This

indicates that the insertion of bulky groups at the C3-O position does not result in any changes in the binding mode of the maytansinoids.

The ligand density seen in the structures of the maytansinoids extended up to 2-3 carbon atoms beyond the well-defined ester group that connects the chain to the C3 carbon atom, suggesting that the added linker groups adopt multiple conformations. The absence of a well-defined density of the linker chain indicates that compounds **2**-**4** were not coordinating the Mg^{2+}, as coordination would lead to a fixed position of the linker. The exposed binding site toward the solvent and the flexibility of the long carbon chains likely prevent stable coordination of the ions. Additionally, no evidence of exchange of GDP by the nucleotide mimetics was found in either soaking or co-crystallization experiments, likely due to the high affinity of nucleotides for tubulin preventing replacement of the already bound nucleotide.

IV.1-5. Conclusion

Our research aimed to examine the tolerance of maytansinoids toward modifications that introduce bulky groups and their impact on tubulin binding and the possible effects these could have in tubulin or MTs. We utilized computational methods to assist the design of a series of maytansinoids that incorporated long carbon chains, carboxylic acids, or nucleotide mimetics.

The crystallography studies conducted in this work provided conclusive evidence that the attachment of long carbon chains to the C3-O position of maytansinol does not affect the binding mode of the core ring of the maytansinoids. Maytansinoids retained a fundamental spatial arrangement with respect to β-tubulin that was not altered by the different attached substituents. Therefore, we are confident that the addition of bulky substituents to the maytansine core should not affect its binding mode at the maytansine binding site. These results were supported by binding affinity and cytotoxicity assay data, demonstrating that the observed effects of both carboxylate substituted maytansinoids and nucleotide mimetic substituted maytansinoids were not attributed to metal ion coordination or nucleotide substitution, respectively.

The cell permeability was impacted by the introduction of larger groups, as indicated by the decreased cytotoxicity. Our research highlights the C3-O position as an ideal location for modifications to the maytansine scaffold, enabling the incorporation of long flexible chains or

even bulky groups. This sets the foundation for further exploration of the C3 position of maytansinoids by integrating functional modifications, such as specific fluorophores or chelators, which could serve as probes to study the dynamics of MTs *in vitro*. Furthermore, these modifications could include other types of chemical entities, paving the way for the development of maytansinol-based bivalent compounds capable of targeting specific proteins of interest (e.g., PROTACS).

IV.2 Characterization of Short-Chain Maytansinoids

IV.2-1. Introduction

The synthetic chemistry team encountered many challenges in the process of synthesizing the aforementioned maytansinoid conjugates. During the synthesis process for obtaining maytansinoid conjugates some byproducts were generated. Given the availability of these compounds, we were curious to study how the changes in their molecular structure affected their binding mode when in complex with tubulin and their biological activity.

IV.2-2. Aim of the Work

Characterization of the binding mode of the short-chain maytansinoids within the maytansine site using molecular docking.

IV.2-3. Computational Details

IV.2-3.1. Docking with AutoDock Vina

The high-resolution crystallographic structure used for the docking studies of the short-chain maytansinoids was obtained from the PDB entry with ID 4TV8.[34] Docking input files for the target protein β-tubulin and the ligand structures were prepared in UCSF Chimera 1.14.[131]

IV.2-3.1.1. Receptor Preparation for Docking with AutoDock Vina

The preparation of the structure of the receptor for the docking calculations was accomplished using the *dock prep* feature of UCSF Chimera. However, it is important to note that this tool does not automatically remove unwanted molecules, such as additional ligands or subunits. Therefore, these molecules were previously removed before using the *dock prep* feature, ultimately obtaining the β-tubulin monomeric structure in which maytansine was present bound in the crystal structure.

The *dock prep* feature of UCSF Chimera streamlines the process of preparing the receptor structure by performing multiple tasks, including removal of water molecules and ions, repair of truncated side chains, addition of hydrogens, assignment of partial charges, and saving the structure in the MOL2 file format. These tasks can be completed individually in UCSF Chimera, but *dock prep* offers a convenient and efficient way to complete all of them in one go. In this particular case, no

side chains were needed to be repaired. We generated protonation states appropriate for physiological pH by adding hydrogens to the protein structure. The aspartic acid and glutamic acid side chains were assumed to be negatively charged, while the arginine and lysine side chains were assumed to be positively charged. Charges for standard amino acids were obtained from the AMBER ff14SB force field.[86]

IV.2-3.1.2. Ligand Preparation for Docking with AutoDock Vina

The molecular structure of each derivative was designed using the *build structure* tool implemented in UCSF Chimera 1.14 by modifying the aforementioned crystal structure of the maytansine molecule. Therefore, the initial conformation of the macrocycle was derived from the maytansine crystal structure.

The structure of each ligand was prepared for docking by opening its PDB file in UCSF Chimera and first adding hydrogens to it using the *structure editing – addH* tool. Subsequently, the addition of partial charges to the ligand structure was performed using the UCSF Chimera *add charge* tool. This tool is a call to the antechamber program, which is a set of auxiliary programs for molecular mechanics studies. The partial charges of each ligand were calculated using the Gasteiger method.[78]

IV.2-3.1.3. Autodock Vina – Posing and Scoring Functions

The calculations were performed using AutoDock Vina[69] which was run in UCSF Chimera 1.14 for the analysis and visualization of the results. AutoDock Vina is a freely available docking software that, for the search for possible poses of the docked ligand with respect to the target protein, applies a stochastic search method. It consists of a GA with a local gradient optimization using an Iterated Local Search Optimizer (ILSO) similar to that by Abagyan *et al.*[132]

The ILSO algorithm begins with an initial solution and utilizes a combination of local search and global search techniques to explore the space of possible solutions. The local search technique uses small perturbations to the coordinates of the initial solution, while the global search technique uses random moves to explore a wider range of solutions. The goal of the ILS optimizer is to find the global minimum of the binding energy function, which represents the most stable binding mode of the ligand to the protein. The ILSO uses an acceptance criterion to decide whether to accept or

reject the new solutions found by local and global moves. The acceptance criterion is based on the energy change of the new solutions, compared to the current solution. Solutions with a lower energy change are accepted, whereas solutions with a higher energy change are rejected. The optimization process terminates when the maximum number of iterations is reached, or when the energy change is lower than a predefined threshold, indicating that the algorithm has reached a local minimum.

AutoDock Vina implements an empirical scoring function for the prediction of ligand-protein binding affinities.

IV.2-3.1.4. Docking Set-Up

In the setup of our docking calculations, AutoDock Vina considered all the atoms of the target site included in a cubic grid box with a grid spacing of 0.375 Å and a grid size of 22 Å for the geometry search within the docking process. The origin of the grid was placed at the center of the maytansine coordinates in 4TV8. Finally, the resulting docking models were classified by the value of binding-free energy ΔG^0 (kcal/mol) and the best solution with the lowest energy was selected for each maytansine derivative.

IV.2-3.1.5. Molecular Superimposition

After the X-ray crystallography structures of the previously computationally characterized short-chain maytansinoids were available, we performed a qualitative analysis to study the similarity of the binding mode acquired by our models of molecules **9a-c, 10**, **11a-c**, with respect to that observed in the crystal structures. To validate the outcomes of the computational modelling experiments, the best conformer of each maytansinoid was superimposed to its corresponding crystal structure. The structure comparison tool *Match Maker* in the software UCSF Chimera 1.14[131] was used for the superposition of molecular structures.

The structural renderings were obtained using PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.[76]

IV.2-4. Results and Discussion

Within the TubInTrain consortium, we were particularly interested in the investigation of novel maytansinoids to explore the biological effects these may have on MTs. For this reason, we decided to investigate further the byproducts obtained by the TubInTrain chemistry group while in the process of synthesizing the previously presented maytansinoid conjugates (Fig. IV-15).

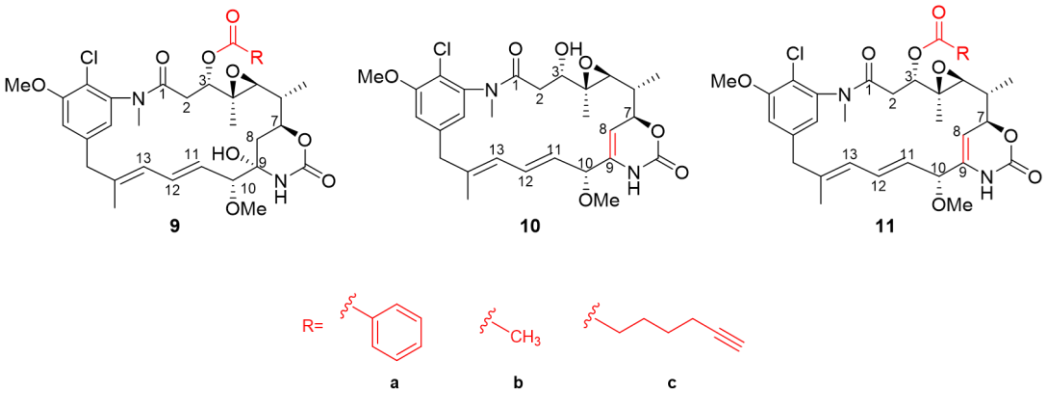


Figure IV-15. General structures of the short-chain maytansinoids.

IV.2-4.1. Molecular Modeling of Short-Chain Maytansinoids

Docking simulations were used to predict the binding mode of the molecules contained in this small library of short-chain maytansinoids within the maytansine site. The accuracy of the AutoDock Vina docking engine was evaluated by re-docking the crystallographic structure of maytansine bound to β-tubulin as a positive control measure. The geometry assigned by AutoDock Vina for maytansine overlapped with its crystallographic orientation (Fig. IV-16). The successful reproduction of the crystallographic findings confirmed the suitability of this docking software for this particular investigation. In this case, the input geometry of the ligand was given in PDB format and was that of the conformer extracted from the crystal structure 4TV8.

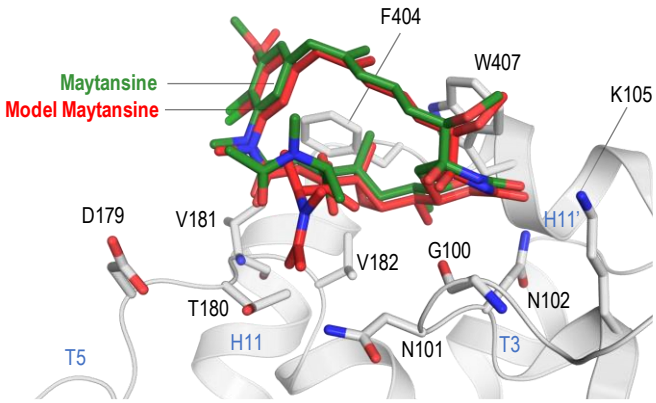


Figure IV-16. Superimposition of the predicted (red) and crystal (dark green) binding poses of maytansine. The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green.

AutoDock Vina used conformational sampling to determine the most accurate experimental pose of the maytansine molecule when bound to β-tubulin, however, the starting geometry already influenced the phase space explored by the molecule. Despite this, the results of our study benefited from this bias. Previous research on long-chain maytansinoids and maytansine conjugates showed that the binding mode of different maytansinoids was preserved even in the presence of bulky substituents. Hence, it was deemed appropriate to use AutoDock Vina as docking software, since by docking the different studied byproducts containing the biased ring conformation, we were able to save time and computational resources.

Therefore, we used AutoDock Vina to predict the spatial coordinates of the binding mode acquired by the proposed maytansinoids within the maytansine site, which is located in a shallow pocket on β-tubulin facing the inter-dimer interface.[38] As expected, in all cases the orientation of the maytansinol ring remained in the same spatial arrangement, acquiring a binding mode similar to that of the parent compound maytansine (Fig. IV-17). The introduction of bulky substituents at position C3 did not alter the predicted 3D arrangement of the core of the molecule. Therefore, we assumed that the binding of maytansinoids to β-tubulin is very tolerant to modifications of the C3 hydroxyl group and expected that the binding mode of the molecules resembles that of the parent compound.

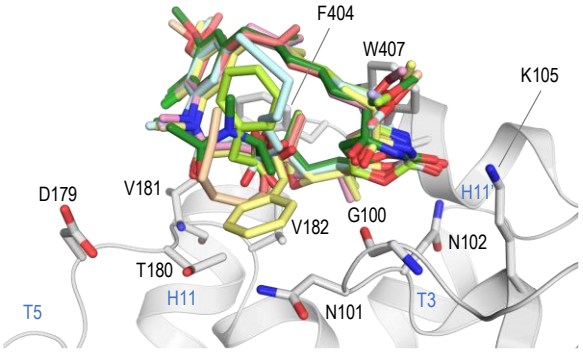


Figure IV-17. Superimposition of the binding pose of the best conformers obtained by AutoDock Vina of maytansinoids **9a-c**, **10** and **11a-c** and the crystal structure of maytansine (dark green). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored in red, nitrogen atoms in navy blue and chlorine atoms in green.

The present study aimed to investigate in more detail the binding mode of compounds **9a-c**, **10**, and **11a-c**, within the maytansine binding site. Our docking results show that all of the best conformers of the docked compounds form hydrogen bonds between the C1-O and the main chain nitrogen atom of Val181 and between the C24-O and the side chains of Lys105 and Asn102 (Fig. IV-18 and Fig. IV-19). Furthermore, we found that compounds **9a-c** establish a specific hydrogen bond with the main chain carbonyl group of Gly100 through their C9-OH group (Fig. IV-18).

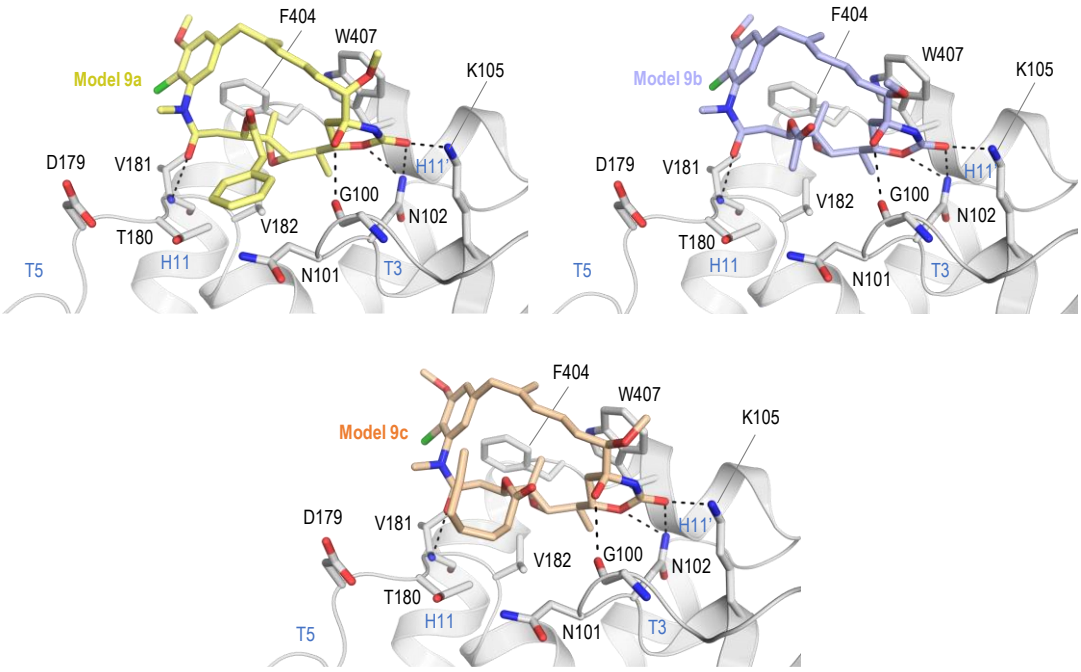


Figure IV-18. Predicted binding poses of maytansinoids **9a** (pale yellow), **9b** (violet), and **9c** (pale orange). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green. Hydrogen bonds are displayed as black dashed lines.

Structural analysis of maytansinoids **10**, **11a-c** further revealed that elimination of the C9-hydroxy group does not affect the binding mode of the ligands, since the heterocycle is anchored by two hydrogen bonds established between the C1-O and the nitrogen atom of the main chain of Val181 and C24-O and the nitrogen of the side chains of Asn102 and Lys105 (Fig. IV-7).

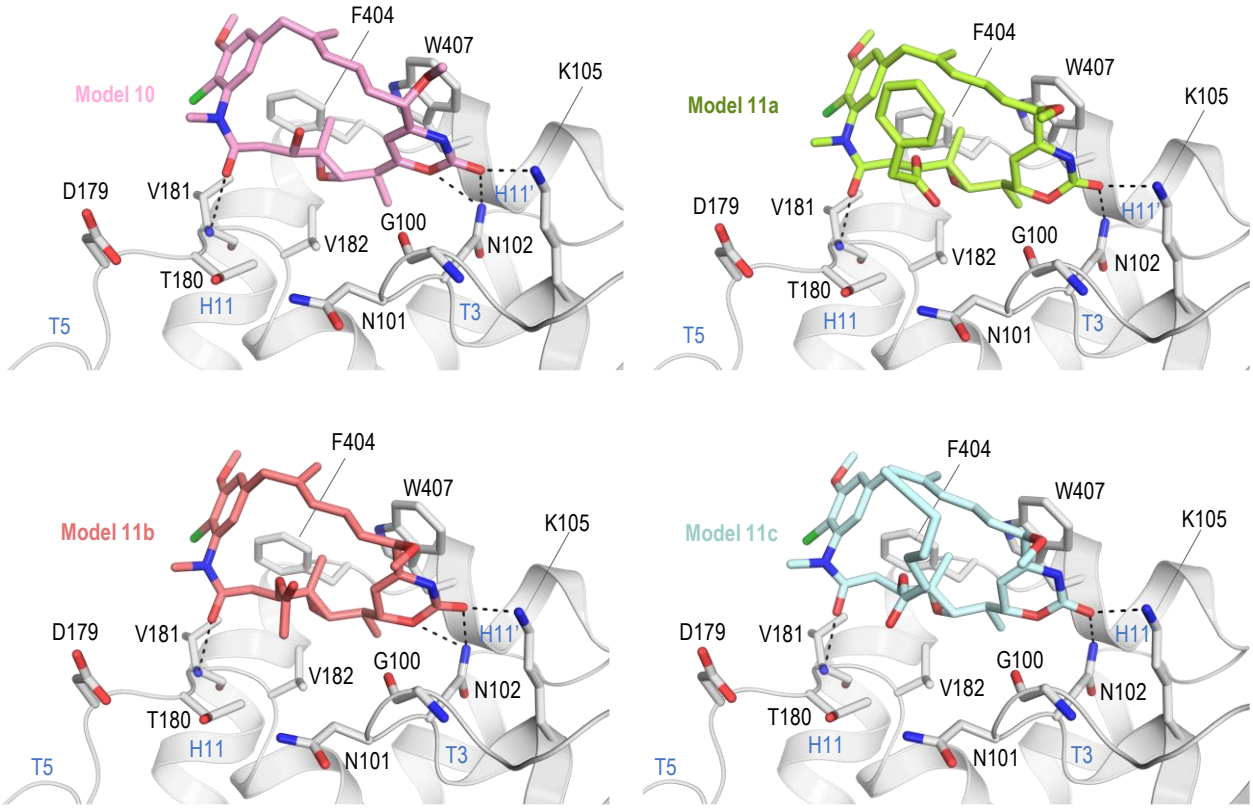


Figure IV-19. Predicted binding poses of maytansinoids **10** (pink), **11a** (light green), **11b** (pink-orange) and **11c** (cyan). The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green. Hydrogen bonds are displayed as black dashed lines.

In addition to the aforementioned findings, our computational investigation also revealed that in the studied maytansinoids, all of the modifications introduced at position C3 point toward the solvent and do not perturb the close environment of the maytansine site. This means that the modifications made to the C3 position do not have a significant impact on the immediate environment surrounding the maytansine site. This allows for greater flexibility in the design of new maytansinoids without disrupting the essential interactions within the binding site. Our results suggest that the binding mode of maytansinoids is primarily determined by the interactions between the functional groups C1-O and C24-O and specific amino acid residues within the binding site, rather than the modifications made to the C3 position.

IV.2-4.2. Biological Assays

IV.2-4.2.1. Binding Constant Determination

Encouraged by the positive results of our docking studies, these compounds were synthesized by Zlata Boiarska in Milan (Italy) and later submitted for biological evaluation to the research center CIB-CSIC in Madrid (Spain) that is part of the TubInTrain consortium. The goal was to evaluate the binding of these molecules to tubulin dimers, and for this they used a displacement assay. This experiment involves measuring the ability of the compounds to displace a fluorescently labeled version of the compound maytansine. The results showed that the three C3-modified maytansinoids **9a–c** displayed affinities toward tubulin comparable to the one of maytansine or even higher (Tab. IV-3).

Table IV-3. Dissociation constant of short-chain maytansinoids. The data come from three independent measurements.

Compound	1a	1b	9a	9b	9c	10	11a	11b	11c
K_d (nM)	14 ± 2	780 ± 40	51 ± 3	11 ± 1	20 ± 2	830 ± 20	1090 ± 40	860 ± 30	1800 ± 120

More specifically, the results revealed that compounds **9b** and **9c** had the strongest binding affinity to tubulin, with dissociation constants similar to those of maytansine (**1a**). This suggests that the replacement of N-acetyl-N-methyl-L-alanine in the maytansine molecular structure with small or large substituents does not greatly affect its binding affinity to tubulin. However, compound **9a**, which had a benzoate ester at position C3, had a lower binding affinity. Compounds **10**, **11a** and **11b**, which had no hydroxyl group at position C9, had binding affinities in the range of 1 μM, suggesting that C9-hydroxyl plays a key role in facilitating interaction with the tubulin site, as seen in the case of maytansinol (**1b**). Lastly, compound **11c**, which also lacked hydroxyl at position C9 and had modifications at C3, showed binding affinities in the submillimolar range.

IV.2-4.2.2. Cell Viability Assays

To investigate the potential of these compounds to overcome multidrug resistance mediated by membrane pumps, at the CIB-CSIC they also determined their cytotoxicity in A549 (small lung cell carcinoma) and in the isogenic pair A2780/A2780AD (P-gp-overexpressing) cell lines.

The results of this study revealed a correlation between the cytotoxicity of the compounds and their binding affinities to tubulin. Compounds **9a-c**, which had high binding affinities, also

exhibited potent cytotoxicity in the nano to subnanomolar range (Tab. IV-4). On the contrary, compounds with lower binding affinities in the submicromolar and micromolar ranges were, as expected, less cytotoxic.

Table IV-4. IC_{50} of maytansinoids in A549 and A2780/A2780AD cell lines. The data come from three independent measurements.

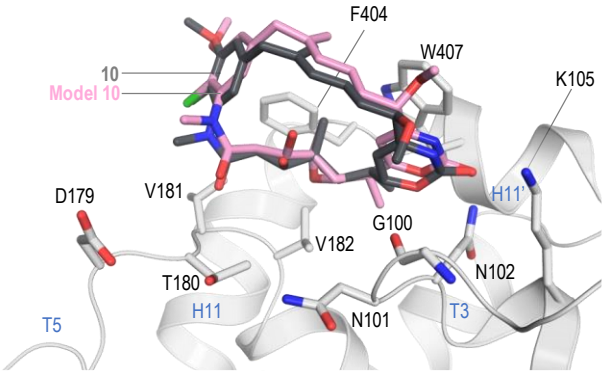
Compound	A549 IC_{50} (nM)	A2780 IC_{50} (nM)	A2780AD IC_{50} (nM)
1a	0.278 ± 0.04	0.31 ± 0.02	19.5 ± 1.14
1b	60 ± 3	23.78 ± 1.65	1459.94 ± 148.27
9a	0.24 ± 0.03	0.11 ± 0.01	11 ± 1
9b	1.2 ± 0.3	0.25 ± 0.01	20 ± 1
9c	0.07 ± 0.008	0.033 ± 0.003	4.7 ± 0.6
10	28 ± 2	74 ± 4	>19706
11a	>570	338.40 ± 4.65	2626.48 ± 186.56
11b	190 ± 40	130 ± 20	>6156
11c	320 ± 40	225 ± 15	>7956

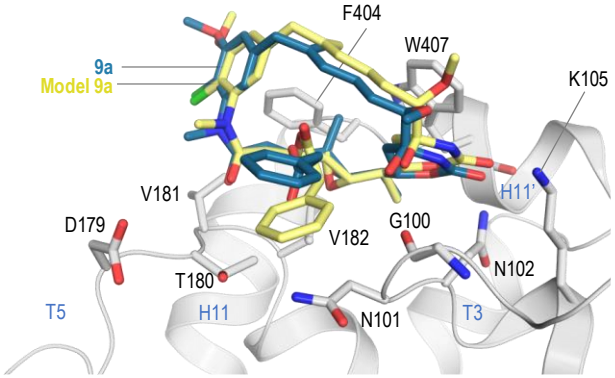
IV.2-4.3. X-Ray Crystallography

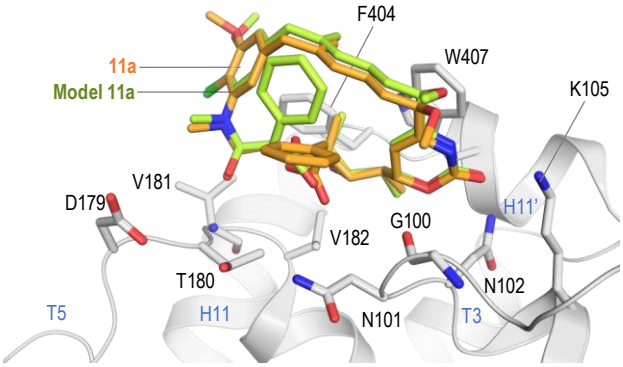
To gain a deeper understanding of the binding mode of the studied maytansinoids, Anne-Catherine Abel at PSI determined the structures of tubulin-compound complexes using X-ray crystallography following a soaking protocol. This allowed us to validate the already computationally predicted common binding mode of the selected maytansinoids and to confirm the accuracy of our docking results. They examined the binding of the tested ligands to the maytansine site of the β-tubulin chain in the T_2R-TTL complex.

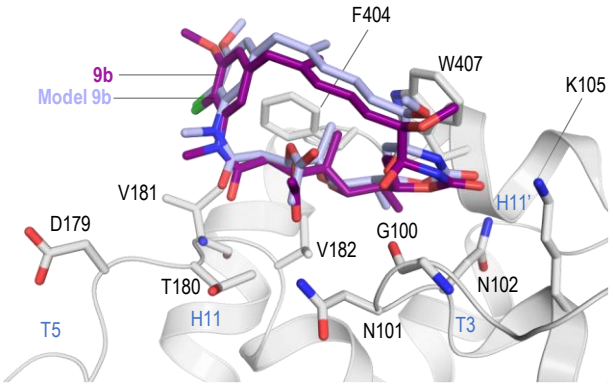
To assess the accuracy of the docking results, the predicted binding modes of the maytansinoids were compared with their solved crystal structures. This evaluation involved superimposing the docked models onto the corresponding crystal structures. The results showed that the binding poses of the docked compounds were highly consistent with those of the crystal structures, providing further confidence in the computational approach used (Fig. IV-20). The interactions observed in the X-ray structures were in close agreement with the predictions made in the molecular docking section. This consistency between the experimental and computational results provides strong support for the proposed binding mode of maytansinoids to the target protein.

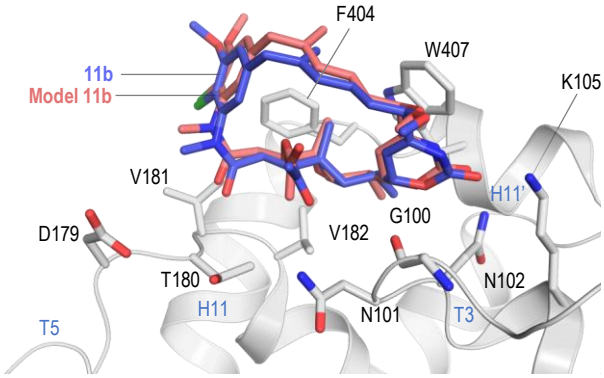
Furthermore, T_2R-TTL-maytansinoid complexes were found to align closely with the protein structure observed in the absence of a ligand (PDB ID 4I55), indicating that binding of maytansinoids does not alter the overall conformation of tubulin.











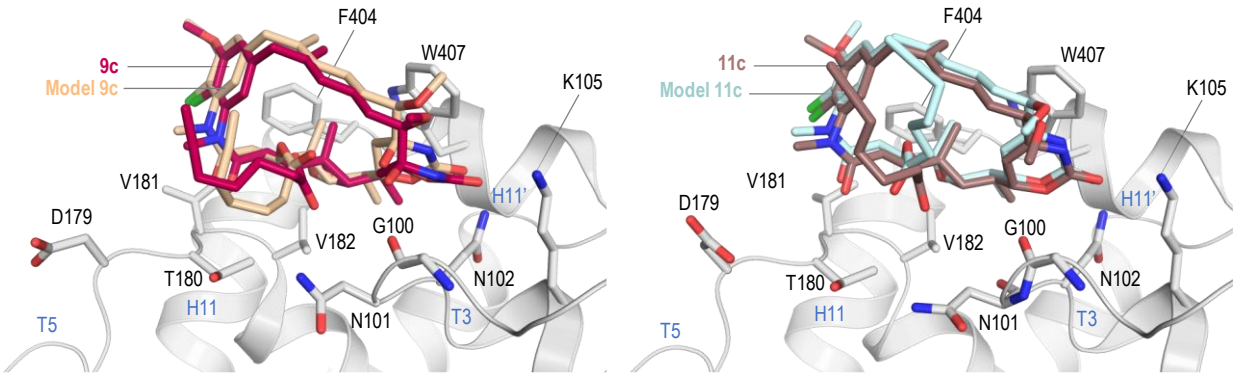


Figure IV-20. Superimposition of the crystallographic binding mode of compounds **9a-c**, **10** and **11a-c** and the best conformer predicted by AutoDock Vina of each of these compounds when bound to the maytansine binding site of β-tubulin. The β-tubulin is represented in light gray ribbon. The interacting residues and ligands are represented as sticks. Oxygen atoms are colored red, nitrogen atoms in navy blue, and chlorine atoms in green.

A quantitative analysis to validate the docking results of the studied byproducts was performed by calculating the RMSD values between the crystal conformer and the docking conformer of each compound (Tab. IV-5). This value represents the root-mean-square deviation between the atomic positions of the two conformers and provides an objective measure of the difference between the two structures.

Table IV-5. RMSD values calculated between the predicted and crystal binding poses of each compound.

Compound	**1**	**9a**	**9b**	**9c**	**10**	**11a**	**11b**	**11c**
RMSD (Å)	1.32	1.47	1.11	1.25	1.17	1.32	1.10	1.24

All RMSD values are <2 Å indicating that the docking method accurately predicted the binding mode of all the ligands, and the results can be considered reliable.

The loss of a hydrogen bond between the C9-OH group and the main chain carbonyl of Gly100 may account for the observed lower affinities and efficacies of these maytansinoids compared to their hydroxylated analogues, even if the structure of the ligand-protein complex is not apparently perturbed.

IV.2-5. Conclusion

In conclusion, within the TubInTrain consortium, we applied both computational and experimental methods to design and evaluate the binding of short-chain maytansinoids to tubulin.

The docking simulations predicted the binding mode of the seven short-chain maytansinoids, which were later successfully validated by X-ray crystallography. The results of these experiments revealed that the binding mode of the studied maytansinoids is primarily determined by hydrogen bond interactions between the C1-O and the nitrogen atom of the main chain of Val181 and C24-O and the side chains of Lys105 and Asn102 within the maytansine binding site. Additionally, it was found that the introduction of substituents at position C3 did not alter the predicted 3D arrangement of the core ring of the molecule and did not have a significant impact on the immediate environment surrounding the maytansine site.

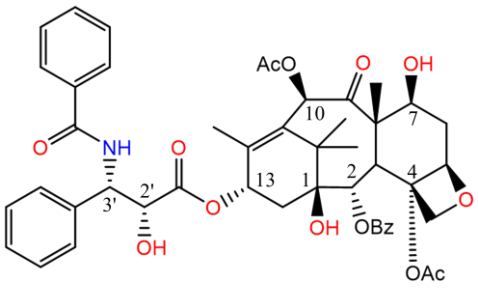
Furthermore, compounds **10** and **11a-c**, which were missing the C9-OH group, showed a significantly lower impact on cells and tubulin in all assays performed. However, compounds **9a-c** presented biological results similar to those of maytansine, suggesting that the attachment of larger groups at this position has no impact on the binding mode, binding affinity or cytotoxicity.

The X-ray structures reported here support the usefulness of structure-based molecular design studies on maytansine derivatives to exploit previously unexploited features of the maytansine binding site of β-tubulin and its surroundings.

CHAPTER V: Computer-Aided Development of New C7-Derivatives of Paclitaxel to Study the Effect on the Structural Signaling of Microtubules

V-1. Introduction

Paclitaxel (**1**), marketed as Taxol®, is a well-studied natural product derived from the Pacific yew tree (*Taxus brevifolia*) with anti-cancer properties. It works by disrupting tubulin dynamics and hindering the polymerization/depolymerization process of MTs, causing a shift in the equilibrium in favor of the MT polymerized form (Fig. V-1).



1

Figure V-1. 2D molecular structure of paclitaxel (**1**).

Studies on Structure-Activity Relationships (SAR) have revealed important details about the critical components of the taxane framework that influence MT stability and drug resistance.[133-135]

The C13 lateral region has significant relevance as demonstrated by the disparity in activity between paclitaxel (with an association constant of 10^7 M^{-1}) and its analog lacking the C13 chain, baccatin III (10^5 M^{-1}).[136] According to molecular modeling, the 2'-OH group accounts for 80% of the binding free energy of the C13 lateral region, emphasizing the important role of this site in taxoid analogues. This is further exemplified by the fact that 2'-deoxy-paclitaxel exhibits a 100-fold reduced affinity for MTs compared to paclitaxel and only slightly higher affinity than baccatin III.

Modifications at positions C7 and C10 in the molecular structure of paclitaxel have minimal impact on its binding to tubulin. When bulky substituents occupy C7 or C10 individually, there may be a slight decrease or enhancement in cytotoxicity within a range of one order of magnitude.

It has been suggested that modifications at these positions could potentially overcome P-gp-mediated drug resistance.[137] However, while some C7 analogs exhibited significantly reduced interaction with P-gp, none of the analogs demonstrated effective cytotoxicity against multidrug resistant cells (MDR).

The binding mode of paclitaxel when in complex with polymerized tubulin has been revealed in three structures: (1) a cryo-electron microscopy (Cryo-EM) reconstruction of a paclitaxel-stabilized MT (PDB code 5SYF, with a resolution of 3.5 Å),[33] (2) a crystallographic structure of a tubulin dimer stabilized with paclitaxel through zinc-induced sheets (PDB code 1JFF, with a resolution of 3.5 Å)[3] and (3) a Cryo-EM structure of a section of a paclitaxel-stabilized MT (PDB code 6WVL, with a resolution of 3.2 Å).[138] Analysis of these structures shows that while substituents C4, C7, and C10 are largely solvent exposed, most interactions occur specifically through the 3′-benzamide, the 2′-hydroxyl group, and the oxetane ring.

X-ray fiber diffraction experiments have shown that MTs experience lattice expansion in the presence of paclitaxel.[139] The experiments were carried out using MTs that were polymerized from GTP-tubulin, which were either left untreated or treated with paclitaxel during or after the polymerization process. The MTs treated with paclitaxel, both during and after polymerization, displayed an increase in dimer length as compared to the drug-free MTs, although the increase was slightly smaller in the case of MTs treated with paclitaxel after assembly. Likewise, MTs that were assembled from GDP-tubulin and treated with paclitaxel during polymerization also showed an extended lattice. Note that GDP-tubulin does not undergo polymerization in the absence of paclitaxel, and therefore the effect of paclitaxel could not be assessed under post-assembly conditions (Tab. V-1).

Table V-1. Results of X-ray fiber diffraction experiments performed in the absence and presence of paclitaxel.

MT samples	Drug free MTs	Paclitaxel-MTs Pre-assembly	Paclitaxel-MTs Pre-assembly (with GDP)	Paclitaxel-MTs Post-assembly
Avg. monomer length (nm)	4.06 ± 0.01	4.18 ± 0.01	4.18 ± 0.01	4.16 ± 0.01
Avg. dimer length (nm)	8.12 ± 0.02	8.35 ± 0.02	8.35 ± 0.02	8.32 ± 0.02
1nm band peak position (nm^{-1})	6.19 ± 0.01	6.02 ± 0.01	6.02 ± 0.01	6.03 ± 0.01

Recent research carried out at CIB-CSIC has demonstrated that the MT structure is heterogeneous and exhibits an expanded lattice region area at the tip.[140] This 'structural signal' most likely signifies the release of various types of MT-associated proteins (MAPs), such as regulators that control MT growth and motor proteins that track cargo along these intracellular highway fibers. Therefore, lattice expansion is a relevant structural signal in MTs that is recognized by many MAPs. Stabilizers, such as paclitaxel, that disperse this minority signal along the entire fiber, may have a significant impact on the interaction of MTs with other protein partners.

It is also plausible that the movement of motor proteins such as kinesin, which moves along β-subunits of tubulin in a single-step motion, could be impacted by variations in step length due to lattice expansion in the presence of stabilizers such as paclitaxel. By affecting axonal transport in neurons, this modification may be responsible for the peripheral neurotoxicity that MT-stabilizing drugs cause. Understanding this biological pathway would be a crucial step in the development of MT-targeting drugs for the treatment of neurodegenerative diseases.

Fluorescence spectroscopy is one of many methods that have been used in the past 20 years to investigate how paclitaxel (molecule **1**) interacts with tubulin and/or the MT filament. Fluorescein and difluorescein tagged taxanes flutax-1 (**2a**) and flutax-2 (**2b**) serve as examples of how different fluorescent probes derived from molecule **1** have been generated, frequently by inserting a fluorophore at the C7 position by suitable linkage (Fig. V-2).[141] The centrosome and spindle pole MTs have been identified as the primary targets for taxane-induced cell death, and these taxane-site probes enable the evaluation of thermodynamic and kinetic parameters for the binding process between taxane and tubulin.

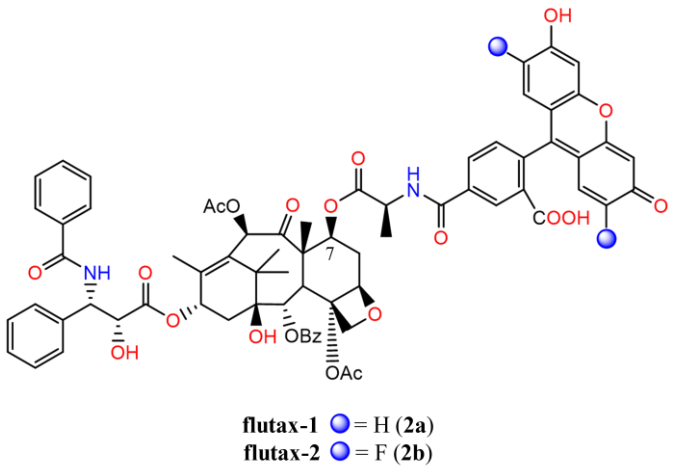


Figure V-2. 2D molecular structure of flutax-1 (**2a**) and flutax-2 (**2b**).

Recent X-ray fiber diffraction measurements on MTs in the presence of several MTAs were undertaken by the TubInTrain biology team in Madrid at CIB-CSIC (data not yet published). Flutax-2, which has the same structure as paclitaxel with an additional bulky substituent at position C7, promotes MT lattice compaction, while paclitaxel promotes MT lattice expansion (Tab. V-2). This was an intriguing result, as until now, it was believed that alteration of position C7 in paclitaxel analogues had little impact on their interaction with MTs.

Table V-2. Results of X-ray fiber diffraction experiments performed in the absence and presence of different MTAs.

MT samples	GDP – MT	Paclitaxel – GDP	Laulimalide – GDP	Peloruside A – GDP	Pelophen B – GDP	Flutax 2 – GDP
MT radius (nm)	11.42 ± 0.10	10.87 ± 0.10	12.21 ± 0.61	10.84 ± 0.30	11.74 ± 0.10	13.17 ± 0.66
Avg. pf number	12.91 ± 0.10	12.37 ± 0.10	13.96 ± 0.70	11.80 ± 0.45	12.96 ± 0.10	14.77 ± 0.89
Inter-pf distances (nm)	5.50 ± 0.03	5.45 ± 0.01	5.46 ± 0.27	5.71 ± 0.22	5.64 ± 0.03	5.56 ± 0.33
Avg. monomer length (nm)	4.06 ± 0.01	4.18 ± 0.010	4.05 ± 0.01	4.06 ± 0.010	4.06 ± 0.010	4.05 ± 0.01
1nm band peak position (nm^{-1})	6.19 ± 0.01	6.02 ± 0.01	6.20 ± 0.01	6.19 ± 0.01	6.19 ± 0.01	6.20 ± 0.01

Additionally, flutax-2 appears to have the ability to compact GMPCPP MTs (Tab. V-3), which have an intrinsically expanded lattice:

Table V-3. Results of X-ray fiber diffraction experiments performed in the absence and presence of paclitaxel and flutax-2.

MT samples	GMPCPP – MT	Paclitaxel – GMPCPP	Flutax-2 – GMPCPP
MT radius (nm)	11.63 ± 0.10	11.07 ± 0.55	13.35 ± 0.36
Avg. pf number	13.29 ± 0.08	12.49 ± 0.63	14.87 ± 0.51
Inter-pf distances (nm)	5.45 ± 0.03	5.51 ± 0.28	5.59 ± 0.22
Avg. monomer length (nm)	4.18 ± 0.01	4.18 ± 0.01	4.05 ± 0.01
1nm band peak position (nm^{-1})	6.02 ± 0.01	6.02 ± 0.01	6.20 ± 0.01

The discovery that flutax-2 does not induce lattice expansion in MTs as paclitaxel led us, the TubInTrain team, to conduct further research. Our strategy was to synthesize a series of derivatives of paclitaxel containing bulky substituents in position C7, with the aim of understanding the specific steric requirements that influence the inversion of the expansion behavior in the MT lattice. The team intended to use different bulk aromatic substituents, such as indole, naphthylamine, and adenine, with the help of an appropriate spacer. In flutax-2, the spacer is an alanine which contains a chiral center,[141] but our aim was to simplify the spacer used to design the new paclitaxel derivatives. These derivatives would be synthesized and further evaluated

utilizing X-ray fiber diffraction studies, which do not require the new substituents to be fluorescent. This provides more freedom in the design of new paclitaxel derivatives.

V-2. Aim of the Work

Exploration of potential flutax-2 analogues, by varying the spacer/bulky substituent combinations in search of compounds affecting lattice expansion of MTs.

V-3. Computational Details

The underlying mechanisms behind the unexpected impact of flutax-2 on MTs remain unclear. However, we hypothesize that the interactions of the linker (spacer and substituent) with the protein is the cause. Based on this rationale, we proposed that other combinations of spacer and substituent that elicit a similar interaction pattern may have a similar global impact on MTs. For this reason, we selected spacer and substituent combinations not on the basis of their computed interaction energy with the protein, but rather based on their similarity to the computed binding model of flutax-2.

V-3.1. Fragment-Based Docking Protocol of S4MPLE

S4MPLE[53,54] is a computational tool that has been developed to provide versatile control of degrees of freedom in molecular docking simulations. This tool is particularly useful in scenarios where a portion of the ligand, referred to as the 'anchor' fragment, needs to be kept fixed in place while selected degrees of freedom in the protein are actively sampled, as in this case. The anchor fragments can be obtained from experimentally solved structures and are commonly used in Fragment-Based Drug Design (FBDD) to identify promising starting points for the discovery of new interesting molecules.

In this study, the 'growth' protocol of S4MPLE was utilized to 'grow' the structure of paclitaxel as the anchor fragment. The growth process involved the addition of linkers composed of a spacer and a bulky substituent to the C7-OH position of the structure of paclitaxel (Fig. V-3). The use of this 'growth' protocol was guided by previously reported methods,[53] with some minor modifications made to optimize the process for the specific scenario being studied.

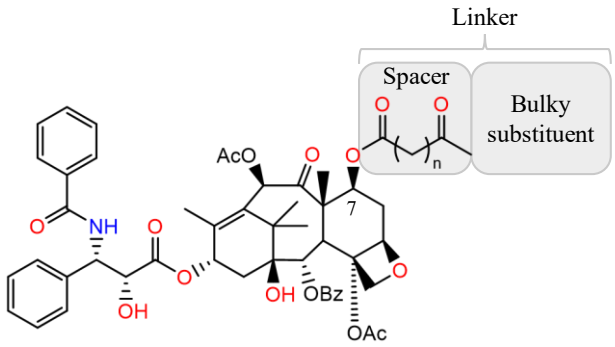


Figure V-3. 2D general molecular structure of the designed paclitaxel derivatives.

V-3.1.1. Target Site Preparation

S4MPLE uses a cut-out zone approach to define the binding site for molecular docking simulations. In this study, one target site was generated for the docking of the designed paclitaxel derivatives. The target site was defined as all atoms present in residues within a radius of 15 Å around the paclitaxel molecule present in the 6WVL electron microscopy structure, including some residues of the adjacent β-tubulin monomer. This was done to account for the potential flexibility of the linkers in the designed paclitaxel derivatives and to ensure that the sampling of the linker would have enough space within the site.

The docking input files for the target protein were generated using the software PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC,[76] with the protein site file stored in the Sybyl MOL2 format in a 'site' directory, together with a list of fixed atoms. All protein atoms were considered fixed during the simulation, since the focus was on the study of the effects of the addition of the different linkers to the anchor fragment (paclitaxel) on its binding mode within the protein site.

V-3.1.2. Preparation of Anchors

The specification of anchor fragments in S4MPLE is an important aspect of molecular docking simulations. The anchor fragment was provided in SDF format and the coordinates were taken from the fragment pose (paclitaxel) in the 6WVL electron microscopy structure. The anchor fragment was fully hydrogenated in accordance with the assumed protonation states of potentially ionizable groups. However, the hydrogen atom that would form the free valency to be used to connect to the linkers was deleted, and the concerned connecting heavy atoms were marked using

the atom mapper functionality of the structure editor MarvinSketch.[75] In this 'growth' scenario, the only connecting heavy atom in the single anchor fragment was mapped to '1'.

To facilitate the use of anchor fragments in S4MPLE, a specific 'anchor' directory was provided containing the SDF file and a list of fixed atoms. In this study, all atoms of the anchor fragment were considered fixed, ensuring that the focus was on the linker part that promoted the growth of the anchor fragment and the impact on its binding mode within the protein site.

V-3.1.3. Linker Generation

Linker generation was based on the spacers and bulky substituents suggested by the chemistry team, and it was performed using the ChemAxon Reactor[142] software. This tool was used for the rapid and efficient preparation of 50 linkers to attach to the C7 position of the molecular structure of paclitaxel. The ChemAxon Reactor allows virtual modeling of chemical reactions using defined reaction schemas, enabling the integration of organic chemistry knowledge and virtual enumeration of a large number of reactants and products, and predicting reaction outcomes. The application is available as a desktop application. In ChemAxon Reactor, the following steps were taken to model the acylation chemical reactions needed to form amides/esters containing linkers from carboxylic acids, primary amines, and alcohols:

1. **Reaction selection**: We drew two acylation reaction equations in Reactor (Fig. V-4). Linkers were generated by coupling each acid with each amine/alcohol.

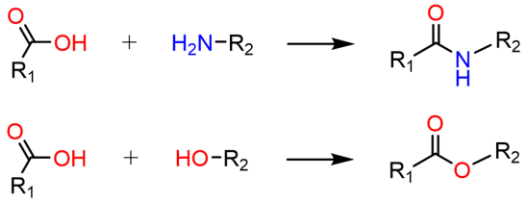


Figure V-4. General synthetic pathway for the synthesis of amide and ester bonds.

2. **Specification of the reactant molecules**: We prepared three different ChemDraw files, one containing five spacers (acids), one including six bulky substituents with a terminal primary amine and a third one presenting four bulky substituents with a terminal alcohol (Fig. V-5). These files were imported to Reactor to be set up as the input molecules that were going to be used as reactants of the acylation reactions.

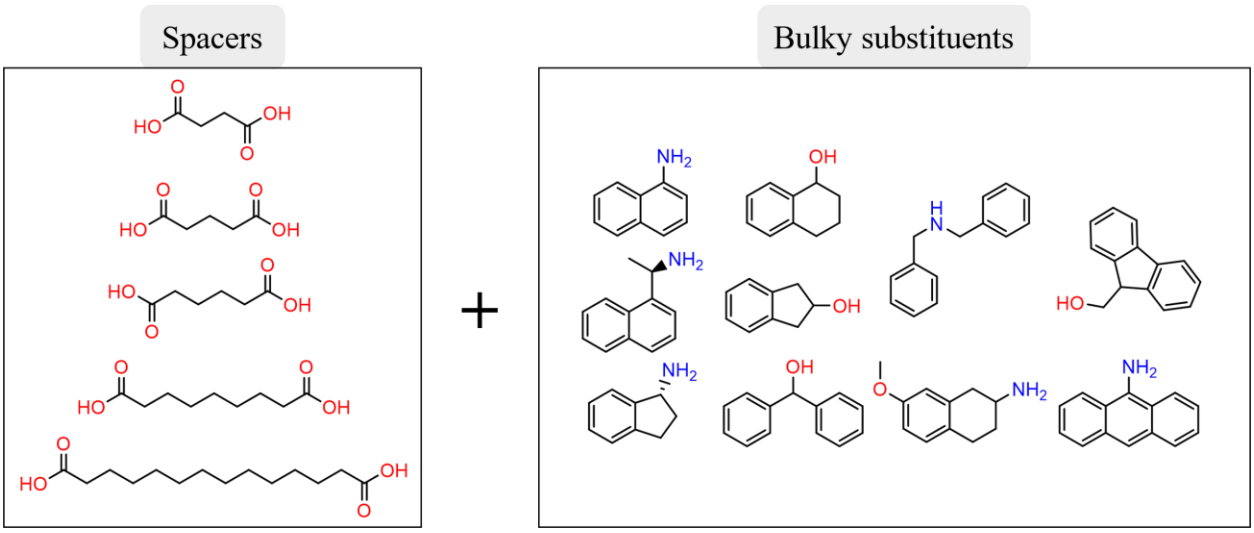


Figure V-5. 2D structure of the spacers (left) and bulky substituents (right) combined for the generation of 50 linkers.

3. **Reactor setup**: ChemAxon Reactor allows for the simulation of multiple chemical reactions in a single run, generating multiple possible products through the use of a reactants library. For this, we set up the running mode to 'combinatorial' and the output type to 'product'.
4. **Analysis of the results**: We visually evaluated the results of the simulations to determine if the resulting products were the expected amides/esters. Furthermore, we retrieved the data for further computational experiments. The integration of Reactor with other ChemAxon tools, such as the Marvin Suite, facilitated downstream analysis and visualization of the results.
5. **Linker preparation for docking**: To ensure that these linkers (monodentate fragments) were compatible with the anchor connector, they were first converted to the SMILES format and then mapped to '1' using MarvinSketch to match the marks of the associated anchor connectors.

V-3.1.4. Docking Protocol

S4MPLE is used with the 'fitted'[53,54] parameterization of the AMBER[55,56]/GAFF[57] engine, including an implicit desolvation model, as previously reported. The docking protocol followed in this study is similar to that explained in section III.2-3.2.2.1. consisting of five different main steps that follow the linker insertion step already explained (Fig. V-6): (1) linker insertion, (2) in-site linker sampling, (3) ligand relaxation, (4) RMSD assessment, (5) free-ligand sampling, and (6) ranking.

First of all, flutax-2 was docked using the electron microscopy binding mode of paclitaxel in the structure with PDB ID 6WVL as an anchor fragment. The best conformer was selected based on the docking score, with the conformer having the lowest score being selected.

Then, the other 50 paclitaxel derivatives were docked. Docking fingerprints were used to rank the docking results of these 50 paclitaxel derivatives based on the resulting flutax-2 docking binding mode. These contact fingerprints were monitoring, for selected protein atoms, the nearest ligand atoms placed in their vicinity. In this way, contact fingerprints are useful for identifying molecules with similar spatial arrangements. This allowed the comparison of the binding mode of flutax-2 to all the docked ligands from the generated virtual library. This information was then used to sort the docking results on the basis of their similarity to the binding mode of flutax-2. This method was a convenient and efficient way to prioritize and analyze the virtual screening results in this particular case.

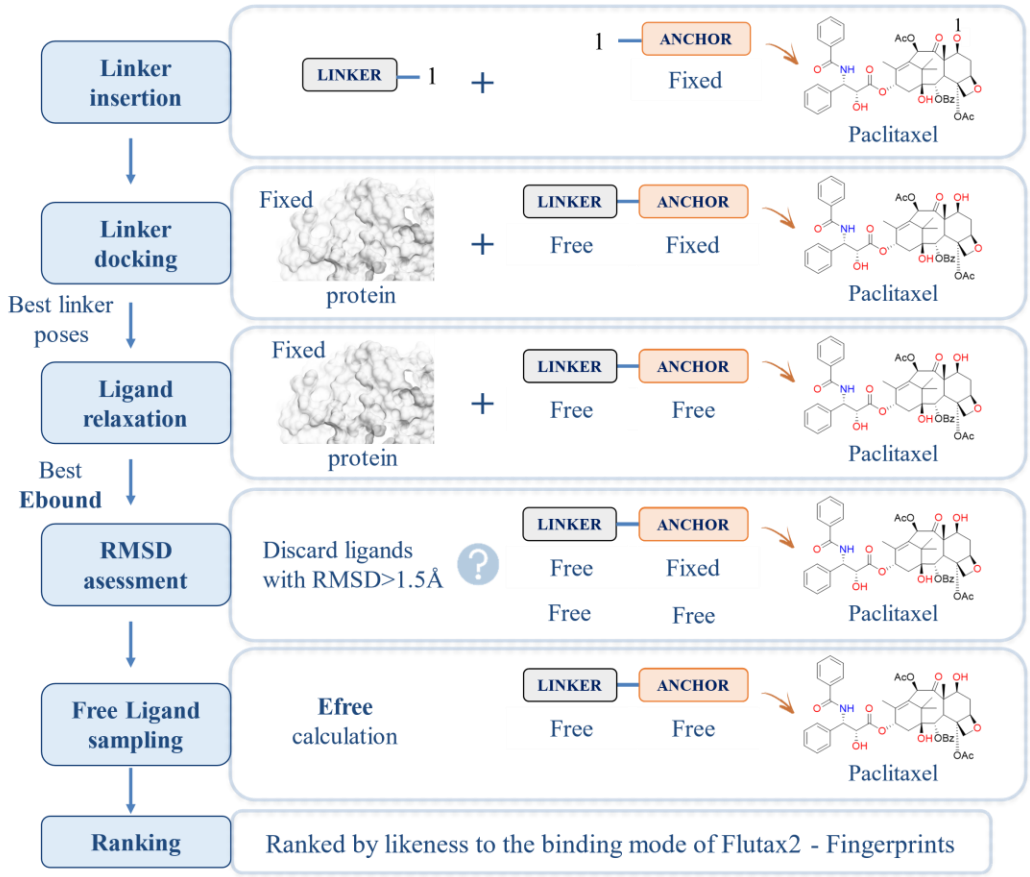


Figure V-6. Covalent docking protocol followed by S4MPLE.

V-4. Results and Discussion

As mentioned in the introduction, in the presence of flutax-2 the biology team from the TubInTrain consortium have observed an atypical behavior in the lattice expansion process in MTs through X-ray fiber diffraction experiments. To further investigate this interesting result, we proposed synthesizing a series of derivatives of paclitaxel presenting bulky substituents in position C7, in order to gain more insights into the steric requirement that produces the abnormal behavior of the MT lattice.

V-4.1 Molecular Modeling of New C7-Derivatives of Paclitaxel

In this study, the goal was to synthesize paclitaxel derivatives that mimic the binding mode of flutax-2 to tubulin. To achieve this goal, the chemistry team proposed five different spacers and 10 different bulky substituents, including six amines and four alcohols that were available in their laboratory (Fig. V-5). We combined these molecular scaffolds, obtaining a total number of 50 *in silico* generated linkers to attach at the C7-OH position of paclitaxel to design new derivatives.

The 50 paclitaxel derivatives were then subjected to molecular docking studies to evaluate their binding pose in the taxane binding pocket of tubulin. It is important to note that the aim of this study was not to find a high-affinity binder, but rather a paclitaxel derivative that adopts a similar binding pose to flutax-2 when bound to tubulin. This is because we wanted to identify a compound likely to produce a similar effect as flutax-2 on MTs, as the binding mode and conformation of the molecule in the tubulin binding pocket should be key determinants of its effect in MTs.

Up to date, the solved experimental structure of flutax-2 in complex with tubulin/MTs is not available. According to the structural biologists of the TubInTrain team, this structure is very difficult to obtain due to several challenges they have to face when working with flutax-2. Therefore, in this study, the binding mode of flutax-2 when bound to the taxane site of tubulin was evaluated using molecular docking (Fig. V-7). The docking result showed a hydrogen bond interaction between flutax-2 and βArg278. However, an unburied hydrogen bond in solvent is not considered significant.

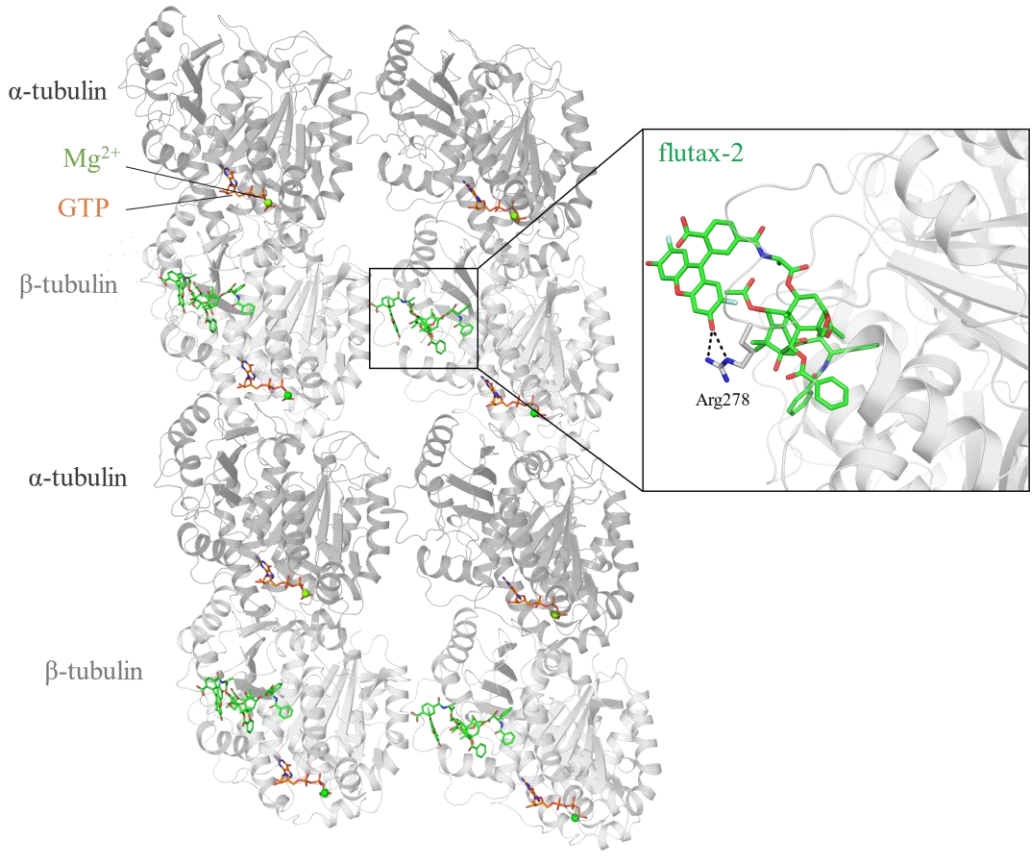


Figure V-7. Section of two adjacent protofilaments containing two tetramers of α,β-tubulin. The docking conformer of flutax-2 (green) is highlighted in a black square and its interactions with the binding site residue βArg278 are displayed as dashed black lines. The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, and fluorine atoms in cyan. The α-tubulin and β-tubulin are represented in dark gray and light gray ribbons, respectively.

The binding mode predicted for the best conformer of each of the 50 proposed paclitaxel derivatives was then compared to that predicted for flutax-2 as illustrated in Fig. V-8. Some of the linkers (made up of a certain spacer and bulky substituent) were found to bind to the same site as the flutax-2 linker, but establish hydrophobic interactions only (molecules **3**-**6**). These linkers are not functionalized, and therefore it is impossible to reproduce the double hydrogen bond established between flutax-2 and βArg278. In the results, we also included one of the 50 predicted paclitaxel derivatives that was predicted to bind very differently to flutax-2 based on docking fingerprints (molecule **7**), to demonstrate the diversity of the binding modes among the docked derivatives.

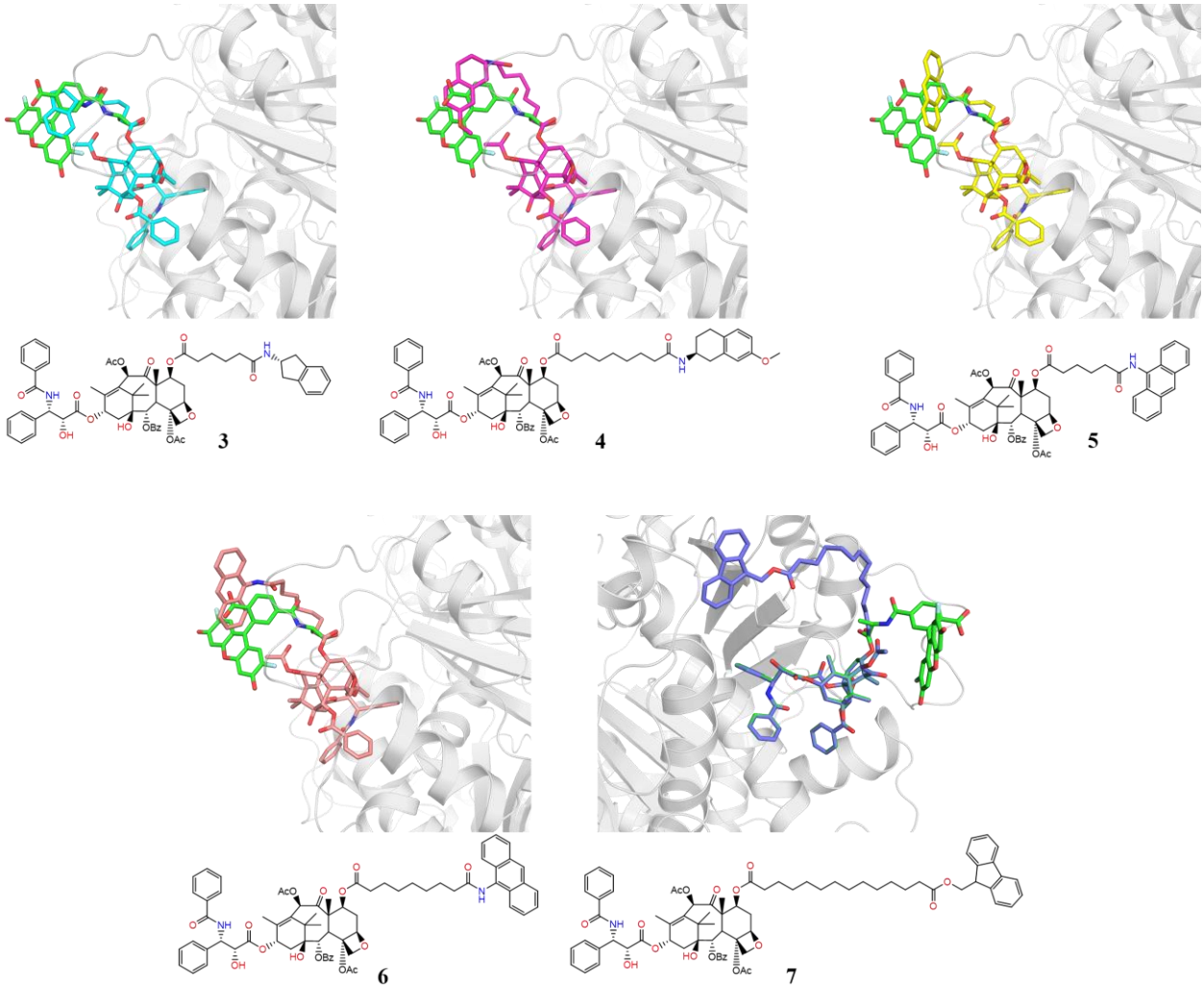


Figure V-8. Superimposition of the selected hits from the virtual library of new paclitaxel derivatives with the docking conformer of flutax-2 (green). The interacting residues and ligands are represented as sticks. Oxygen atoms are in red, nitrogen atoms in navy blue, and fluorine atoms in cyan. β-tubulin is represented by a light gray ribbon.

The results suggest that some of the computationally tested paclitaxel derivatives may have potential for further investigation as new MTAs, but further experimental studies are needed to fully understand their binding mode. Currently, only molecule **4** of the four best candidates proposed has been synthesized and tested experimentally.

V-4.2. Experimental Evaluation

The experimental tests of compound **4** were carried out at the research center CIB-CSIC in Madrid (Spain) by Francesca Bonato, a Ph.D. student of the TubInTrain program.

V-4.2.1. Biochemical Polymerization Assays

In order to explore how compound **4** affects the dynamics of tubulin assembly *in vitro,* biologists performed duplicated polymerization assays in aqueous PEDTA buffer. This buffer contained 10 mM NaPi, 1 mM EDTA, and 1 mM GTP, and had a pH of 7. Additionally, they included 6 mM $MgCl_2$ in the experimental conditions. These conditions were chosen because tubulin is not inclined to polymerize under these circumstances, unless polymerization enhancers such as paclitaxel are present. The results from the tubulin polymerization assays shown in Fig. V-9 indicate that the C-7 modified taxane **4** is a weaker enhancer of the MT assembly compared to both paclitaxel and flutax-2.

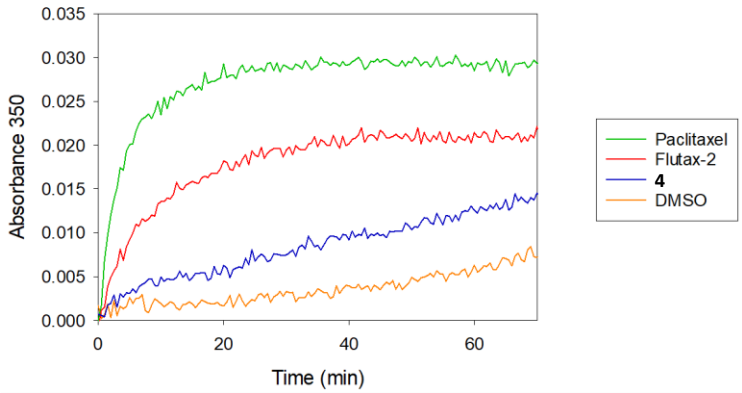


Figure V-9. The graph displays the impact of selected compounds on the inhibition of tubulin assembly, with all experiments being duplicated across two independent experiments. The time courses of the assembly of 25 μM tubulin were recorded in the presence of either the vehicle (DMSO) depicted by the orange lines, or paclitaxel shown in green lines, flutax-2 shown in red lines, or compound **4** depicted by blue lines.

V-4.2.2. Binding Constant Determination

To establish a correlation between the inhibition of tubulin assembly and the binding affinities of the compounds to the taxane site, their affinities were determined by means of competition against flutax-2 (Tab. V-4). Compound **4** was significantly less potent than the control, docetaxel, which is an analog of paclitaxel. These results were consistent with those from the polymerization assays. These findings suggest that the C7-decoration has a considerable impact on the interaction of the paclitaxel core of the derivatives with tubulin, which was not previously reported in the literature.

Table V-4. Binding affinities of taxane compounds. Data are the mean ± SEM values of three independent experiments with duplicates in each one.

Compound		Docetaxel	4
26°C	K_b [M^{-1}]	1.313 ± 0.008 x10^7	1.3 ± 0.1 x10^5
	K_d [M]	7.62 ± 0.04 x10^{-8}	7.9 ± 0.6 x10^{-6}
35°C	K_b [M^{-1}]	8.9 ± 0.01 x10^6	7.8 ± 0.8 x10^4
	K_d [M]	1.2 ± 0.2 x10^{-7}	1.3 ± 0.1 x10^{-5}
42°C	K_b [M^{-1}]	8 ± 1 x10^6	6.5 ± 0.9 x10^4
	K_d [M]	1.4 ± 0.2 x10^{-7}	1.7 ± 0.2 x10^{-5}

V-4.2.3. Cell Viability Assays

Subsequently, the toxicity of the compounds was assessed in five distinct cell lines: human non-small cell lung cancer A549, human cervical carcinomas HeLa and HeLaβIII (multidrug-resistant overexpressing βIII tubulin isotype), and human cervical carcinomas Kb3.1 and KbVb (multidrug-resistant overexpressing P-glycoprotein).

The results shown in Tab. V-5 indicate that the IC_{50} values of compound **4** are higher than those of the parental compound paclitaxel by at least one order of magnitude.

Table V-5. IC_{50} of taxanes in A549, HeLa, HeLaβIII, Kb3.1 and KbVb cell lines. Data are the mean ± SEM values of three independent experiments with duplicates in each one.

Compound	IC_{50} [nM]				
	A549	HeLa	HeLaβIII	Kb3.1	KbVb
Paclitaxel	3.1 ± 0.7	1.2 ± 0.4	28.7 ± 0.7	2.4 ± 0.2	2700 ± 600
4	220 ± 20	100 ± 20	720 ± 80	150 ± 20	19000 ± 1000

V-4.2.4. X-Ray Fiber Diffraction Experiments

The structural characteristics of MTs in different nucleotide-bound states were analyzed in the presence of certain compounds using Shear-flow X-ray fiber diffraction. As shown in the introduction, this method allows for precise determination of the average diameter, inter-protofilament distances, and axial interdimeric spacing of a MT population *in vitro*.[143]

To investigate two distinct lattice conformations, namely compacted and expanded, the MTs that were polymerized were initially examined in the presence of either GTP (resulting in GDP-MTs due to GTPase activity) or GMPCPP (yielding GMPCPP-MTs).[144] Currently, GMPCPP-MTs are considered to be the most accurate representation of the MT lattice bound to GTP. GMPCPP, a

non-hydrolyzable nucleotide, binds to the nucleotide E-site, leading to rapid tubulin polymerization and stabilization of MTs. The stabilization of MTs is associated with the induction of an axial lattice expansion.[29,145]

Tab. V-6 shows that when compound **4** was present, the average monomer length detected in GDP MTs was 4.05 nm, indicating the presence of a compact axial lattice, as in the case of flutax-2 in these conditions (Tab. V-2). However, in the case of MTs that were polymerized in the presence of compound **4** and GMPCPP, the average monomer length was 4.18 nm, suggesting the existence of an expanded axial lattice, as expected in the presence of this particular non-hydrolyzable nucleotide. Thus, we confirmed that compound **4** is able to stabilize the MT without inducing major changes in its structure, as it did not cause an expansion of the GDP-lattice or a compaction of the GMPCPP-lattice.

Table V-6. Fiber diffraction analysis of MTs in various nucleotide-bound states. Data are Mean ± StdErr.

MT samples	GDP – MT	GMPCPP – MT	4 – GDP	4 – GMPCPP
MT radius (nm)	11.4 ± 0.4	11.6 ± 0.3	11.6 ± 0.3	11.8 ± 0.3
Avg. pf number	13.1 ± 0.5	13.3 ± 0.4	13.3 ± 0.4	13.6 ± 0.5
Inter-pf distances (nm)	5.4 ± 0.1	5.4 ± 0.2	5.4 ± 0.6	5.4 ± 0.1
Avg. monomer length (nm)	4.05 ± 0.01	4.18 ± 0.01	4.05 ± 0.01	4.17 ± 0.01
1nm band peak position (nm^{-1})	6.20 ± 0.01	6.02 ± 0.01	6.20 ± 0.01	6.20 ± 0.01

V-5. Conclusion

This study provides insights into the predicted binding mode of flutax-2 when bound to β-tubulin. Furthermore, based on the results of this investigation, four new paclitaxel derivatives were designed and selected from a virtual library of 50, taking into account their chemical feasibility, the availability of reagents, and their probability of exhibiting a binding mode similar to flutax-2. These molecules will provide an opportunity to further explore the impact of bulky substituents at the C7-OH position in paclitaxel and determine whether there are steric requirements that result in the inversion of MT expansion. One of these four derivatives (compound **4**) has already been successfully synthesized and is undergoing further biochemical characterization. The results from the various already performed biological experiments suggest that the C-7 modification in compound **4** had an impact on the interaction of the paclitaxel core of the derivatives with tubulin.

In particular, the tubulin polymerization and binding affinity assays showed that compound **4** was a weaker enhancer of MT assembly that both paclitaxel and flutax-2 and had lower binding affinity than docetaxel, a paclitaxel analog. Additionally, the cell viability assays indicated that compound **4** had higher IC_{50} values than paclitaxel. Finally, the X-ray fiber diffraction experiments demonstrated that compound **4** had a more neutral structural effect on MTs *in vitro*, unlike paclitaxel and flutax-2. These findings suggest that the bulky substituent at the C7 position in compound **4** may cause steric hindrance that affects the interaction of the compound with tubulin and leads to a weaker effect on MT assembly and cell viability, in comparison to paclitaxel and flutax-2.

These findings represent a step forward in the study of new paclitaxel C7 derivatives and their effect on MT dynamics which could be potentially exploited to design new molecular probes able to target the taxane site.

General Conclusion and Future Perspectives

This Ph.D. thesis entitled "*Computer-Aided Molecular Design and Modeling of Tubulin Targeting Agents*" has explored the use of computational chemistry techniques in molecular design, with a focus on developing tubulin targeting agents. The studies presented in this thesis demonstrate how VS of commercially available and in-house designed chemical libraries *via* substructure search and docking strategies, as well as molecular dynamics simulations can be used to identify potential binders to specific protein sites and analyze protein-ligand biomolecular interactions. The application of computational chemistry techniques to tubulin dimers and oligomers has proven to be a valuable tool for identifying potential hit molecules and optimizing their structure and properties. The successful application of these computational strategies in the early stages of the molecular discovery pipeline has significantly reduced the time and cost of small molecule development and improved the hit success rate.

In Chapter III, we explored the possibility of bypassing the complex and expensive synthesis of structurally complex MTAs such as pironetin, potentially useful for tubulin immobilization, by studying new, simpler tubulin ligands that target the todalam site. The screening campaign successfully identified a new set of promising chemically diverse todalam site binders using computational methods, which were later experimentally validated. Of the 15 proposed molecules, 12 were observed to bind to the todalam site, resulting in an 80% hit rate, allowing to discover five different scaffolds able to target this site. The results set the basis for the further development of the identified scaffolds into covalent binders using computer-aided approaches. Further research enabled the design of 14 potential covalent binders targeting the todalam site, with nine found to be bound to the site by X-ray crystallography experiments. One of the nine compounds was identified as the most promising candidate, providing an optimal length for a covalent binder targeting the thiol group of the residue αCys4, presenting a continuous density towards αCys4. The newly discovered todalam site binders provide a foundation for the development of more potent and selective binders, opening up possibilities for future research.

The research in Chapter IV successfully examined the tolerance of maytansinoids towards modifications that introduce bulky groups and their impact on their binding to tubulin. Computational methods were used to design a series of maytansinoids that incorporated long carbon chains, carboxylic acids, or nucleotide mimetics. The studies provided conclusive evidence

that the attachment of bulky substituents to the C3-O position of maytansinol does not affect the binding mode of the core ring of the maytansinoids at the maytansine binding site in tubulin. The research highlights the potential of these modified maytansinoids in future research for their use as potential molecular probes to study MT dynamics.

The research presented in Chapter V provides insights into the predicted binding mode of flutax-2 when bound to β-tubulin, and four new paclitaxel derivatives were designed and selected from a virtual library of 50 compounds based on their chemical feasibility, the availability of reagents, and their probability of exhibiting a binding mode similar to flutax-2. One of these derivatives has already been synthesized and is undergoing further biochemical characterization. These molecules will provide an opportunity to explore the impact of bulky substituents at the C7-OH position in paclitaxel and determine steric requirements that result in the inversion of MT lattice expansion. The research highlights the potential for further investigation into paclitaxel derivatives and their impact on MT lattice parameters.

The computational structure-based approaches used in the different studies for the exploration of new scaffolds and hit compound derivatives that bind to the novel tubulin site, known as the todalam site, as well as the maytansine site and the better-known taxane-binding site, yielded numerous hits that were experimentally validated. Herein, we used up-to-date tools such as docking algorithms, fragment-growing strategies, accurate force fields, and molecular dynamics simulations to identify potential ligands and analyze protein-ligand interactions. The study encompassed multiple design iterations, testing, and optimization of various compounds, which ultimately provided novel insights into different binding sites on the two tubulin subunits that make up MTs, offering promising avenues for further investigation.

The identification of new potential compounds for tubulin immobilization and the design of molecular probes for studying MT dynamics could lead to the development of new drugs that target MTs and improve the treatment of diseases such as cancer and neurodegeneration. Future research could focus on optimizing the identified compounds and testing their efficacy *in vivo*, as well as developing new molecules that target the studied tubulin-binding sites.

This Ph.D. thesis also showcased successful integration of different areas of expertise from the multidisciplinary team of TubInTrain towards designing chemical scaffolds with target-oriented and tailor-made biochemical properties. Computational chemistry played a significant role in this

achievement by providing a foundation for chemical compound design and optimization, and by guiding organic synthesis efforts, which were then followed by X-ray crystallography, biochemical, and cellular experiments. The interdisciplinary approach adopted by the TubInTrain team sets a new standard for future MT research, highlighting the importance of collaboration between experts in computational chemistry, organic chemistry, biochemistry, and structural biology. The thesis demonstrates how interdisciplinary collaboration can overcome hurdles more easily when forces are combined, ultimately leading to successful molecular design strategies.

Bibliography

[1] P. G. McKean, S. Vaughan, K. Gull, *J. Cell Sci.* 2001, *114* (15), 2723–2733.
[2] K. H. Downing, E. Nogales, *Curr. Opin. Cell Biol.* 1998, *10* (1), 16–22.
[3] J. Löwe, H. Li, K. H. Downing, E. Nogales, *J. Mol. Biol.* 2001, *313* (5), 1045–1057.
[4] C. Dumontet, M. A. Jordan, *Nat. Rev. Drug Discov.* 2010, *9* (10), 790–803.
[5] A. Soliman, L. Bakota, R. Brandt, *Curr. Neuropharmacol.* 2022, *20* (4), 782–798.
[6] L. Pellegrini, A. Wetzel, S. Grannó, G. Heaton, K. Harvey, *Cell. Mol. Life Sci.* 2016, *74* (3), 409–434.
[7] M. O. Steinmetz, A. E. Prota, *Trends Cell Biol.* 2018, *28* (10), 776–792.
[8] T. Mitchison, M. Kirschner, *Nature* 1984, *312* (5991), 237–242.
[9] J. M. Kollman, A. Merdes, L. Mourey, D. A. Agard, *Nat. Rev. Mol. Cell Biol.* 2011, *12* (11), 709–721.
[10] N. Teixidó-Travesa, J. Roig, J. Lüders, *J. Cell Sci.* 2012, *125* (19), 4445-4456.
[11] P. Liu, M. Würtz, E. Zupa, S. Pfeffer, E. Schiebel, *Curr. Opin. Cell Biol.* 2021, 68, 124–131.
[12] G. J. Brouhard, L. M. Rice, *Nat. Rev. Mol. Cell Biol.* 2018, *19* (7), 451–463.
[13] J. Roostalu, C. Thomas, N. I. Cade, S. Kunzelmann, I. A. Taylor, T. Surrey, *eLife* 2020, 9, e51992.
[14] M. A. Jordan, L. Wilson, *Nat. Rev. Cancer* 2004, *4* (4), 253–265.
[15] L. Wordeman, J. J. Vicente, *Cancers* 2021, *13* (22), 5650.
[16] L. Lafanechère, *Front. Pharmacol.* 2022, *13*, 823571.
[17] E. C. Amoroso, *Nature* 1935, *135* (3407), 266–267.
[18] G. G. Borisy, E. W. Taylor, *J. Cell Biol.* 1967, *34* (2), 525–533.
[19] G. G. Borisy, E. W. Taylor, *J. Cell Biol.* 1967, *34* (2), 535–548.
[20] S. Matthew, Q.-Y. Chen, R. Ratnayake, C. S. Fermaintt, D. Lucena-Agell, F. Bonato, A. E. Prota, S. T. Lim, X. Wang, J. F. Díaz, A. L. Risinger, V. J. Paul, M. Á. Oliva, H. Luesch, *P. Natl. A. Sci.* 2021, *118* (9), e2021847118.
[21] T. Mühlethaler, L. Milanos, J. A. Ortega, T. B. Blum, D. Gioia, B. Roy, A. E. Prota, A. Cavalli, M. O. Steinmetz, *Angew. Chem. Int. Ed.* 2022, *61* (25), e202204052.
[22] K.-S. Chan, C.-G. Koh, H.-Y. Li, *Cell Death Dis.* 2012, *3* (10), e411.
[23] T. Mühlethaler, D. Gioia, A. E. Prota, M. E. Sharpe, A. Cavalli, M. O. Steinmetz, *Angew. Chem. Int. Ed.* 2021, *60* (24), 13331–13342.
[24] J. H. Gill, K. L. Rockley, C. De Santis, A. K. Mohamed, *Pharmacol. Therapeut.* 2019, *202*, 18–31.
[25] B. Bhattacharyya, S. Kapoor, D. Panda, *Methods in Cell Biol.* 2010, 301–329.
[26] T. Schvartz, N. Aloush, I. Goliand, I. Segal, D. Nachmias, E. Arbely, N. Elia, *Mol. Biol. Cell* 2017, *28* (21), 2747–2756.
[27] R. Gerasimaitė, J. Bucevičius, K. A. Kiszka, S. Schnorrenberg, G. Kostiuk, T. Koenen, G. Lukinavičius, *ACS Chem. Biol.* 2021, *16* (11), 2130–2136.
[28] C. Elie-Caille, F. Severin, J. Helenius, J. Howard, D. J. Muller, A. A. Hyman, *Curr. Biol.* 2007, *17* (20), 1765–1770.
[29] G. M. Alushin, G. C. Lander, E. H. Kellogg, R. Zhang, D. Baker, E. Nogales, *Cell* 2014, *157* (5), 1117–1129.
[30] A. E. Prota, K. Bargsten, D. Zurwerra, J. J. Field, J. F. Díaz, K.-H. Altmann, M. O. Steinmetz, *Science* 2013, *339* (6119), 587–590.
[31] R. B. G. Ravelli, B. Gigant, P. A. Curmi, I. Jourdain, S. Lachkar, A. Sobel, M. Knossow, *Nature* 2004, *428* (6979), 198–202.
[32] A. Dorléans, B. Gigant, R. B. G. Ravelli, P. Mailliet, V. Mikol, M. Knossow, *P. Natl. A. Sci.* 2009, *106* (33), 13775–13779.
[33] E. H. Kellogg, N. M. A. Hejab, S. Howes, P. Northcote, J. H. Miller, J. F. Díaz, K. H. Downing, E. Nogales, *J. Mol. Biol.* 2017, *429* (5), 633–646.
[34] A. E. Prota, K. Bargsten, P. T. Northcote, M. Marsh, K.-H. Altmann, J. H. Miller, J. F. Díaz, M. O. Steinmetz, *Angew. Chem. Int. Ed.* 2014, *53* (6), 1621–1625.

[35] B. Gigant, C. Wang, R. B. G. Ravelli, F. Roussi, M. O. Steinmetz, P. A. Curmi, A. Sobel, M. Knossow, *Nature* 2005, *435* (7041), 519–522.
[36] A. Cormier, M. Marchand, R. B. G. Ravelli, M. Knossow, B. Gigant, *EMBO rep.* 2008, *9* (11), 1101–1106.
[37] A. Maderna, M. Doroski, C. Subramanyam, A. Porte, C. A. Leverett, B. C. Vetelino, Z. Chen, H. Risley, K. Parris, J. Pandit, A. H. Varghese, S. Shanker, C. Song, S. C. K. Sukuru, K. A. Farley, M. M. Wagenaar, M. J. Shapiro, S. Musto, M.-H. Lam, F. Loganzo, C. J. O'Donnell, *J. Med. Chem.* 2014, *57* (24), 10527–10543.
[38] A. E. Prota, K. Bargsten, J. F. Diaz, M. Marsh, C. Cuevas, M. Liniger, C. Neuhaus, J. M. Andreu, K.-H. Altmann, M. O. A Steinmetz, *P. Natl. A. Sci.* 2014, *111* (38), 13817–13821.
[39] A. E. Prota, J. Setter, A. B. Waight, K. Bargsten, J. Murga, J. F. Diaz, M. O. Steinmetz, *J. Mol. Biol.* 2016, *428* (15), 2981–2988.
[40] J. Yang, Y. Wang, T. Wang, J. Jiang, C. H. Botting, H. Liu, Q. Chen, J. Yang, J. H. Naismith, X. Zhu, L. Chen, *Nat. Commun.* 2016, *7* (1).
[41] H. Pérez-Peña, A.-C. Abel, M. Shevelev, A. E. Prota, S. Pieraccini, D. Horvath, *Biomolecules* 2023, *13* (2), 285.
[42] S. Kumar, S. Kumar, *In Silico Drug Design* 2019, 161–189.
[43] Enamine, Ltd. https://enamine.net/ (accessed Feb 27, 2023).
[44] J. J. Irwin, K. G. Tang, J. ZIBYoung, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, R. A. Sayle, *J. Chem. Inf. Model.* 2020, *60* (12), 6065–6073.
[45] N. Kochev, V. Monev, I. Bangov, *Chemoinformatics* 2003, 291–318.
[46] G. F. Mangiatordi, D. Trisciuzzi, D. Alberga, N. Denora, R. M. Iacobazzi, D. Gadaleta, M. Catto, O. Nicolotti, *European J. Med. Chem.* 2017, *139*, 792–803.
[47] Q. Guo, H. Zhang, Y. Deng, S. Zhai, Z. Jiang, D. Zhu, L. Wang, *European J. Med. Chem.* 2020, *196*, 112328.
[48] Y. Zhou, B. Di, M.-M. Niu, *Molecules* 2019, *24* (17), 3181.
[49] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, *Proteins* 2002, *47*, 409–443.
[50] V. B. Sulimov, D. C. Kutov, A. V. Sulimov, *Curr. Med. Chem.* 2019, *26*, 7555–7580.
[51] V. T. Sabe, T. Ntombela, L. A. Jhamba, G. E. M. Maguire, T. Govender, T. Naicker, H. G. Kruger, *Eur. J. Med. Chem.* 2021, *224*, 113705.
[52] E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, A. G. Taranto, *Front. Chem.* 2020, *8*, 343.
[53] L. Hoffer, J.-P. Renaud, D. Horvath, *J. Chem. Inf. Model.* 2013, *53* (4), 836–851.
[54] L. Hoffer, C. Chira, G. Marcou, A. Varnek, D. Horvath, *Molecules* 2015, *20*, 8997–9028.
[55] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* 1995, *117* (19), 5179–5197.
[56] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct. Funct. Bioinform.* 2006, *65* (3), 712–725.
[57] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* 2004, *25* (9), 1157–1174.
[58] D. Horvath, *J. Med. Chem.* 1997, *40* (15), 2412–2423.
[59] D. Horvath, D. van Belle, G. Lippens, S. J. Wodak, *J. Chem. Phys.* 1996, *104* (17), 6679–6695.
[60] M. K. Gilson, B. Honig, *J. Comput. Aid. Mol. Des.* 1991, *5* (1), 5-20.
[61] A. R. Leach, 3rd ed., *Pearson Education*, Harlow, England, 2001.
[62] M. Karplus, *Accounts Chem Res.* 2002, *35* (6), 321–323.
[63] L. Euler, Opera Omnia, ser. 1, vol. 1, *Impensis Academiae Imperialis Scientiarum*, Petropoli, 1824.
[64] L. Verlet, *Phys. Rev*. 1967, *159* (1), 98–103.
[65] R. W. Hockney, J. W. Eastwood, *crc Press*, Boca Raton, FL, 1988.
[66] A. Castro-Alvarez, O. Pineda, J. Vilarrasa, *ACS Omega* 2018, *3* (2), 1770–1782.
[67] G. Menchon, A. E. Prota, D. Lucena-Agell, P. Bucher, R. Jansen, H. Irschik, R. Müller, I. Paterson, J. F. Díaz, K.-H. Altmann, M. O. Steinmetz, *Nat. Commun.* 2018, *9* (1), 5382.

[68] Malvern Panalytical. (2021). WAVE: Wave-guided interferometry for label-free biomolecular interaction analysis. Retrieved from https://www.malvernpanalytical.com/en/products/product-range/WAVE
[69] O. Trott, A. J. Olson, *J. Comput. Chem.* 2010, *31*, 455-461.
[70] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, *J. Chem. Inf. Model.* 2021, *61* (8), 3891–3898.
[71] Virtual Screening Web Server. http://infochim.u-strasbg.fr/webserv/VSEngine.html, December 2020.
[72] ChemAxon, Standardizer, C, version 5.12; ChemAxon, Ltd: Budapest, Hungary, 2012.
[73] O. Korb, T. Stützle, T. E. Exner, *J. Chem. Inf. Model.* 2009, *49* (1), 84–96.
[74] ChemAxon. JChem, Version 20.8.3, ChemAxon, Ltd: Budapest, Hungary 2020.
[75] MarvinSketch version 21.3, calculation module developed by *ChemAxon*, http://www.chemaxon.com, 2021.
[76] PyMOL | pymol.org http://pymol.org (accessed Feb 27, 2023).
[77] prepLig4Docking https://chematlas.chimie.unistra.fr/WebTools/prepLig4Docking.php (accessed Feb 27, 2023).
[78] J. Gasteiger, M. A Marsili, *Tetrahedron Lett.* 1978, *19* (34), 3181–3184.
[79] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* 2015, *1–2*, 19–25.
[80] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl, *Lect. Notes Comput. Sc.* 2015, 3–27.
[81] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham et al., *Amber* 2021, University of California, San Francisco.
[82] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* 2006, *25* (2), 247–260.
[83] Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria et al., *Gaussian, Inc., Wallingford CT*, 2016.
[84] A. W. Sousa da Silva, W. F. Vranken, *BMC Research Notes* 2012, *5* (1), 633.
[85] K. L. Meagher, L. T. Redman, H. A. Carlson, *J. Comput. Chem.* 2003, *24* (9), 1016–1025.
[86] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* 2015, *11* (8), 3696–3713.
[87] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 1983, *79* (2), 926–935.
[88] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* 1993, *98* (12), 10089–10092.
[89] VF2 Algorithm – NetworkX 3.0 documentation
https://networkx.org/documentation/stable/reference/algorithms/isomorphism.vf2.html (accessed Feb 27, 2023)
[90] A. Jüttner, P. Madarasi, *Discrete Applied Mathematics* 2018, 242, 69–81.
[91] RDKit: Open-source cheminformatics. https://www.rdkit.org
[92] SciFinder [database]. Columbus (OH): *American Chemical Society*; (accessed 2023 Feb 23).
[93] prepFrg4Docking https://chematlas.chimie.unistra.fr/WebTools/prepFrg4Docking.php (accessed Feb 27, 2023).
[94] F. Huang, X. Han, X. Xiao, J. Zhou, *Molecules* 2022, *27* (22), 7728.
[95] Pfizer. 2021. Pfizer's Novel COVID-19 Oral Antiviral Treatment Candidate Reduced Risk of Hospitalization or Death by 89% Compared to Placebo in Interim Analysis of Phase 2/3 EPIC-HR Study [Press release]. https://www.pfizer.com/news/press-release/press-release-detail/pfizers-novel-covid-19-oral-antiviral-treatment-candidate (accessed Feb 23, 2023).
[96] P. Ábrányi-Balogh, L. Petri, T. Imre, P. Szijj, A. Scarpino, M. Hrast, A. Mitrović, U. P. Fonovič, K. Németh, H. Barreteau, D. I. Roper, K. Horváti, G. G. Ferenczy, J. Kos, J. Ilaš, S. Gobec, G. M. Keserű, *Eur. J. Med. Chem.* 2018, *160*, 94–107.
[97] F. Huang, X. Han, X. Xiao, J. Zhou, *Molecules* 2022, *27* (22), 7728.
[98] S. E. Dalton, S. Campos, *ChemBioChem* 2020, *21* (8), 1080–1100.
[99] M. Gao, A. F. A. Moumbock, A. Qaseem, Q. Xu, S. Günther, *Nucleic Acids Res.* 2021, *50* (D1), D445–D450.
[100] H. Du, J. Gao, G. Weng, J. Ding, X. Chai, J. Pang, Y. Kang, D. Li, D. Cao, T. Hou, *Nucleic Acids Res.* 2020, *49* (D1), D1122–D1129.
[101] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* 2001, *40* (11), 2004–2021.
[102] S. M. Kupchan, Y. Komoda, W. A. Court, G. J. Thomas, R. M. Smith, A. Karim, C. J. Gilmore, R. C. Haitiwanger, R. F. Bryan, *J. Am. Chem. Soc.* 1972, *94* (4), 1354-1356.

[103] S. Remillard, L. I. Rebhun, G. A. Howie, S. M. Kupchan, *Science* 1975, *189*, 1002–1005.
[104] M. Lopus, E. Oroudjev, L. Wilson, S. Wilhelm, W. Widdison, R. Chari, M. A. Jordan, *Mol. Cancer Ther.* 2010, *9*, 2689–2699.
[105] B. F. Issell, S. T. Crooke, *Cancer Treat. Rev.* 1978, *5*, 199–207.
[106] J. M. Cassady, K. K. Chan, H. G. Floss, E. Leistner, *Chem. Pharma. Bull.* 2004, *52*, 1–26.
[107] L. Amiri-Kordestani, G. M. Blumenthal, Q. C. Xu, L. Zhang, S. W. Tang, L. Ha, W. C. Weinberg, B. Chi, R. Candau-Chacon, P. Hughes, A. M. Russell, S. P. Miksinski, X. H. Chen, W. D. McGuinn, T. Palmby, S. J. Schrieber, Q. Liu, J. Wang, P. Song, N. Mehrotra, L. Skarupa, K. Clouse, A. Al-Hakim, R. Sridhara, A. Ibrahim, R. Justice, R. Pazdur, P. Cortazar, *Clin. Cancer Res.* 2014, *20*, 4436–4441.
[108] S. Wedam, L. Fashoyin-Aje, X. Gao, E. Bloomquist, S. Tang, R. Sridhara, K. B. Goldberg, B. L. King-Kallimanis, M. R. Theoret, A. Ibrahim, L. Amiri-Kordestani, R. Pazdur, J. A. Beaver, *Clin. Cancer Res.* 2020, *26*, 4180–4185.
[109] S. Garcia-Alonso, A. Ocana, A. Pandiella, *Trends Cancer* 2020, *6*, 130–146.
[110] J. I. Geller, J. G. Pressey, M. A. Smith, R. A. Kudgus, M. Cajaiba, J. M. Reid, D. Hall, D. A. Barkauskas, S. D. Voss, S. Y. Cho, S. L. Berg, J. S. Dome, E. Fox, B. J. Weigel, *Cancer* 2020, *126*, 5303–5310.
[111] P. Zhao, Y. Zhang, W. Li, C. Jeanty, G. Xiang, Y. Dong, *Acta Pharm. Sin. B* 2020, *10*, 1589–1600.
[112] S. J. M. Hale, R. D. Perrins, C. E. GarcÍa, A. Pace, U. Peral, K. R. Patel, A. Robinson, P. Williams, Y. Ding, G. Saito, M. A. Rodriguez, I. Perera, A. Barrientos, K. Conlon, S. Damment, J. Porter, T. Coulter, *Bioconjugate Chem.* 2019, 30, 703–713.
[113] J. Porter, Y. Ding, S. J. M. Hale, R. D. Perrins, A. Robinson, M. P. Mazanetz, Y. Wu, Y. Ma, K. Conlon, T. Coulter, *Bioorganic Med. Chem. Lett.* 2020, 30, 127634.
[114] S. Ikeyama, M. Takeuchi, *Biochem. Pharmacol.* 1981, *30*, 2421–2425.
[115] T. W. Yu, H. G. Floss, in *Anticancer Agents from Nat. Prod.* (Eds: G. M. Cragg, D. J. Newman), 2nd ed., Taylor & Francis/CRC Press, Boca Raton, 2012, p.407.
[116] W. Li, M. Huang, Y. Li, A. Xia, L. Tan, Z. Zhang, Y. Wang, J. Yang, *Biochem. Bioph. Res. Co.* 2021, *566*, 197–203.
[117] A. Kawai, H. Akimoto, Y. Kozai, K. Ootsu, S. Tanida, N. Hashimoto, H. Nomura, *Chem. Pharma. Bull.* 1984, *32*, 3341–3351.
[118] W. C. Widdison, S. D. Wilhelm, E. E. Cavanagh, K. R. Whiteman, B. A. Leece, Y. Kovtun, V. S. Goldmacher, H. Xie, R. M. Steeves, R. J. Lutz, R. Zhao, L. Wang, W. A. Blattler, R. V. J. Chari, *J. Med. Chem.* 2006, *49*, 4392–4408.
[119] T. Nittoli, M. P. Kelly, F. Delfino, J. Rudge, A. Kunz, T. Markotan, J. Spink, Z. Chen, J. Shan, E. Navarro, M. Tait, K. Provoncha, J. Giurleo, F. Zhao, X. Jiang, D. Hylton, S. Makonnen, C. Hickey, J. R. Kirshner, G. Thurston, N. Papadopoulos, *Bioorg. Med. Chem.* 2018, *26*, 2271–2279.
[120] W. C. Widdison, J. F. Ponte, J. A. Coccia, L. Lanieri, Y. Setiady, L. Dong, A. Skaletskaya, E. E. Hong, R. Wu, Q. Qiu, R. Singh, P. Salomon, N. Fishkin, L. Harris, E. K. Maloney, Y. Kovtun, K. Veale, S. D. Wilhelm, C. A. Audette, J. A. Costoplus, R. V. J. Chari, *Bioconjugate Chem.* 2015, *26*, 2261–2278.
[121] J. A. Costoplus, K. H. Veale, Q. Qiu, J. F. Ponte, L. Lanieri, Y. Setiady, L. Dong, A. Skaletskaya, L. M. Bartle, P. Salomon, R. Wu, E. K. Maloney, Y. V. Kovtun, O. Ab, K. Lai, R. V. J. Chari, W. C. Widdison, *ACS Med. Chem. Lett.* 2019, *10*, 1393–1399.
[122] A. E. Prota, K. Bargsten, J. F. Diaz, M. Marsh, C. Cuevas, M. Liniger, C. Neuhaus, J. M. Andreu, K. H. Altmann, M. O. Steinmetz, *P. Natl. A. Sci.* 2014, *111*, 13817–13821.
[123] Z. Boiarska, D. Passarella, *Drug Discov. Today* 2020, *26*, 604–615.
[124] P. D. Bonandi, E. Foschi, F. Marucci, C. Dapiaggi, F. Sironi, M. Pieraccini, S. Christodoulou, M. S. de Asis Balaguer, F. Diaz, J. F. Zidar, *ChemPlusChem* 2019, *84*, 98–102.
[125] G. Cappelletti, D. Cartelli, M. Christodoulou, D. Passarella, *Curr. Pharm. Design* 2017, *23*, 784–808.
[126] D. P. C. Marucci, M. Christodoulou, S. Pieraccini, M. Sironi, F. Dapiaggi, D. Cartelli, A. Calogero, G. Cappelletti, C. Vilanova, S. Gazzola, G. Broggini, *Eur. J. Org. Chem.* 2016, 2029–2036.

[127] P. Marzullo, Z. Boiarska, H. Pérez-Peña, A. C. Abel, B. Alvarez-Bernad, D. Lucena-Agell, F. Vasile, M. Sironi, K. H. Altmann, A. E. Prota, J. F. Diaz, S. Pieraccini, D. Passarella, *Chem. Eur. J.* 2022, *28*, e202103520.
[128] T. W. Traut, *Mol. Cell. Biochem.* 1994, *140*, 1–22.
[129] M. M. Harding, *Acta Crystallogr. D Biol. Crystallogr.* 2001, *57* (3), 401–411.
[130] M. M. Harding, *Acta Crystallogr. D Biol. Crystallogr.* 2006, *62* (6), 678–682.
[131] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* 2004, *25*, 1605–1612.
[132] R. Abagyan, M. Totrov, D. Kuznetsov, *J. Comput. Chem.* 1994, *15* (5), 488–506.
[133] R. Matesanz, I. Barasoain, C.-G. Yang, L. Wang, X. Li, C. de Inés, C. Coderch, F. Gago, J. J. Barbero, J. M. Andreu, W.-S. Fang, J. F. Díaz, *Chem. Biol.* 2008, *15* (6), 573–585.
[134] X. Xiao, J. Wu, C. Trigili, H. Chen, J. W. K. Chu, Y. Zhao, P. Lu, L. Sheng, Y. Li, F.J. Sharom, I. Barasoain, J. F. Diaz, W. Fang, *Bioorg. Med. Chem. Lett.* 2011, *21* (16), 4852–4856.
[135] C. C. Rohena, S. L. Mooberry, *Nat. Prod. Rep.* 2014, *31* (3), 335–355.
[136] J. M. Andreu, I. Barasoain, *Biochemistry* 2001, *40* (40), 11975–11984.
[137] R. Yusuf, Z. Duan, D. Lamendola, R. Penson, M. Seiden, *Curr. Cancer Drug Tar.* 2003, *3* (1), 1–19.
[138] G. E. Debs, M. Cha, X. Liu, A. R. Huehn, C. V. Sindelar, *P. Natl. A. Sci.* 2020, *117* (29), 16976–16984.
[139] A. Rai, T. Liu, S. Glauser, E. A. Katrukha, J. Estévez-Gallego, R. Rodríguez-García, W.-S. Fang, J. F. Díaz, M. O. Steinmetz, K.-H. Altmann, L. C. Kapitein, C. A. Moores, A. Akhmanova, *Nat. Mater.* 2019, *19* (3), 355–365.
[140] J. Estévez-Gallego, F. Josa-Prado, S. Ku, R. M. Buey, F. A. Balaguer, A. E. Prota, D. Lucena-Agell, C. Kamma-Lorger, T. Yagi, H. Iwamoto, L. Duchesne, I. Barasoain, M. O. Steinmetz, D. Chrétien, S. Kamimura, J. F. Díaz, M. A. Oliva, *eLife* 2020, *9*, 1–26.
[141] A. A. Souto, A. U. Acuña, J. M. Andreu, I. Barasoain, M. Abal, F. Amat-Guerri, *Angew. Chem. Int. Ed.* 1996, *34* (2324), 2710–2712.
[142] ChemAxon Reactor version 21.3, calculation module developed by *ChemAxon*, http://www.chemaxon.com, 2021
[143] S. Kamimura, Y. Fujita, Y. Wada, T. Yagi, H. Iwamoto, *Cytoskeleton* 2016, *73* (3), 131–144.
[144] J. Estévez-Gallego, F. Josa-Prado, S. Ku, R. M. Buey, F. A. Balaguer, A. E. Prota, D. Lucena-Agell, C. Kamma-Lorger, T. Yagi, H. Iwamoto, L. Duchesne, I. Barasoain, M. O. Steinmetz, D. Chrétien, S. Kamimura, J. F. Díaz, M. A. Oliva, *eLife* 2020, *9*, 1–26.
[145] R. Zhang, G. M. Alushin, A. Brown, E. Nogales, *Cell* 2015, *162*, 849–859.