Etude de la structure et la dynamique de Mcl-1 : application en cancérologie

Mohammed Benabderrahmane

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité CHIMIE

Préparée au sein de l'Université de Caen Normandie

Etude de la structure et la dynamique de Mcl-1 : application en cancérologie

Présentée et soutenue par
Mohammed BENABDERRAHMANE

Thèse soutenue publiquement le 11/12/2020 devant le jury composé de		
M. PASCAL BONNET	Professeur des universités, Université d'Orléans	Rapporteur du jury
M. LAURENT CHALOIN	Directeur de recherche, Institut de Recherche en Infectiologie	Rapporteur du jury
M. GAUTIER MOROY	Maître de conférences HDR, Université Paris 7 Paris Diderot	Membre du jury
Mme JANA DE OLIVEIRA SANTOS	Professeur des universités, Université Caen Normandie	Directeur de thèse
M. RONAN BUREAU	Professeur des universités, Université Caen Normandie	Co-directeur de thèse
Mme ANNE-SOPHIE VOISIN-CHIRET	Professeur des universités, Université Caen Normandie	Président du jury

Thèse dirigée par JANA DE OLIVEIRA SANTOS et RONAN BUREAU, Centre d'études et de recherches sur le médicament de normandie (Caen)

UNIVERSITÉ CAEN NORMANDIE

ED 508 NC Normande de Chimie

À Madame. VOISIN-CHIRET ANNE SOPHIE,

Professeure à l'Université de Caen Normandie

À Monsieur. BONNET PASCAL,

Professeur à l'Université d'Orléans

À Monsieur. CHALOIN LAURENT,

Directeur de recherche à l'Institut de Recherche en Infectiologie de Montpellier

À Monsieur. MOROY GAUTIER,

Maître de conférences à l'Université Paris Diderot

À Madame. SOPKOVA-DE OLIVEIRA SANTOS JANA,

Professeure à l'Université de Caen Normandie

À Monsieur. BUREAU RONAN,

Professeur à l'Université de Caen Normandie

Qui me font l'honneur de lire ce manuscrit,
Et de juger ces travaux,
En témoignage de ma profonde reconnaissance,

Remerciements

Que soit remercié ici, le Professeur Patrick Dallemagne, directeur du Centre d'Etudes et de Recherche sur le Médicament de Normandie, de m'avoir accueilli au sein de l'unité pour ce projet doctoral.

Que soit remerciée ici, la Prof. Sopkova-De Oliveira Santos Jana pour avoir dirigé cette thèse. Merci pour toute ta patience, ta disponibilité, ton savoir, ta bonne humeur, partagés tout au long de ces trois années. Merci pour m'avoir donné la chance d'évoluer tant sur le plan scientifique qu'humain.

Que soit remercié le Prof. Ronan Bureau pour avoir co-dirigé cette thèse. Merci pour ta disponibilité, pour toutes les discussions partagées. Merci également, pour tout ce que tu fais pour créer un environnement scientifique propice à la créativité au niveau de la plateforme de chémo-informatique.

Jana, Ronan. Merci infiniment pour la liberté scientifique et personnelle accordée pendant mon passage au CERMN. Elle est essentielle à mon existence.

Que soient remerciés, le Dr. Massimiliano Bonomi.(Institut Pasteur), le Prof. Carlo Camilloni (Univ Milan), et la Société Schrödinger pour avoir rendu mon séjour à Lugano possible pendant l'école d'été (CECAM-2019). Merci pour toute la science et la joie partagées pendant ce séjour.

Que soient remerciés, tous les membres du CERMN, collègues, amis que j'ai pu rencontrer. Merci pour l'ambiance joviale et les moments partagés.

Que soient remerciés, mes parents et tous les membres de ma famille pour leur soutien indéfectible tout au long de ces années. (Tack så mycket !).

Table des matières

Résumé

Ce travail ayant pour objet l'étude de la structure et la dynamique de Mcl-1, une protéine anti-apoptotique d'intérêt en cancérologie, est scindé en trois parties.
Une première étude concerne une caractérisation du mode d'interaction du Pyridoclax (un inhibiteur BH3-mimétique) avec Mcl-1 par des approches expérimentales (RMN) et théoriques (simulations de dynamique moléculaire). Une deuxième partie est consacrée à l'étude et à la caractérisation des espaces conformationnels de Mcl-1 et son mode d'inhibition allostérique. Le troisième volet de ce travail, traite d'une approche d'analyse, basée sur des simulations de métadynamique avec comme application la détection du répertoire des poches cryptiques de Mcl-1.

CHAPITRE 1.

INTRODUCTION

1. Introduction générale

1.1. Le cancer dans l'histoire et données actuelles

Le cancer est actuellement considéré comme une pathologie majeure réduisant l'espérance de vie au 21ème siècle. Toutefois, le cancer n'est pas une maladie récente. En effet, le terme cancer est dérivé du terme grec "Karkinos", un terme qui fut utilisé par Hippocrate pour décrire les carcinomes. Mais Hippocrate ne fut pas le premier à avoir noté cette pathologie. Les premières évidences d'un cancer remontent à des manuscrits de l'Egypte ancienne (Papyrus D'Edwin Smith et Georges Ebers).[1,2]

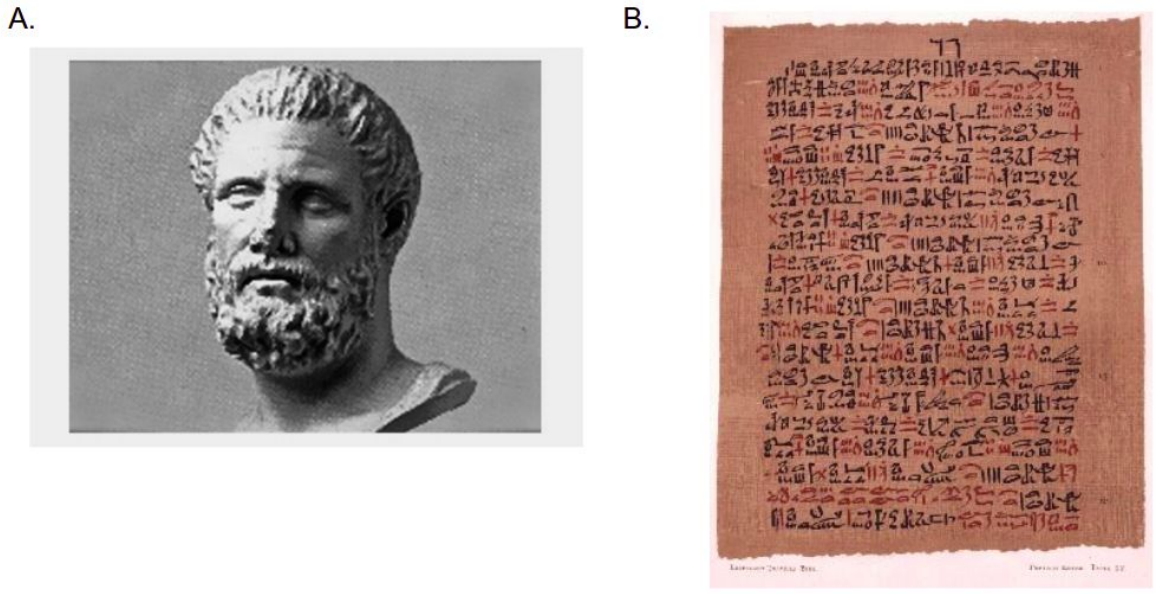


Figure 1. Le cancer dans l'histoire. A. Buste d'Hippocrate B. Exemple d'une page du Papyrus d'Ebers.

Les données de l'organisation mondiale de la santé pour l'année 2015, considèrent le cancer comme étant la première ou la deuxième cause de mort prématurée (avant 70 ans) dans le monde (**Figure 2**).

L'incidence du cancer et son taux de mortalité augmentent. En 2018, le même rapport (GLOBOCAN 2018)[3], estime une incidence de 18,1 millions de cas avec une mortalité approchant les 9,5 millions de décès dans le monde.

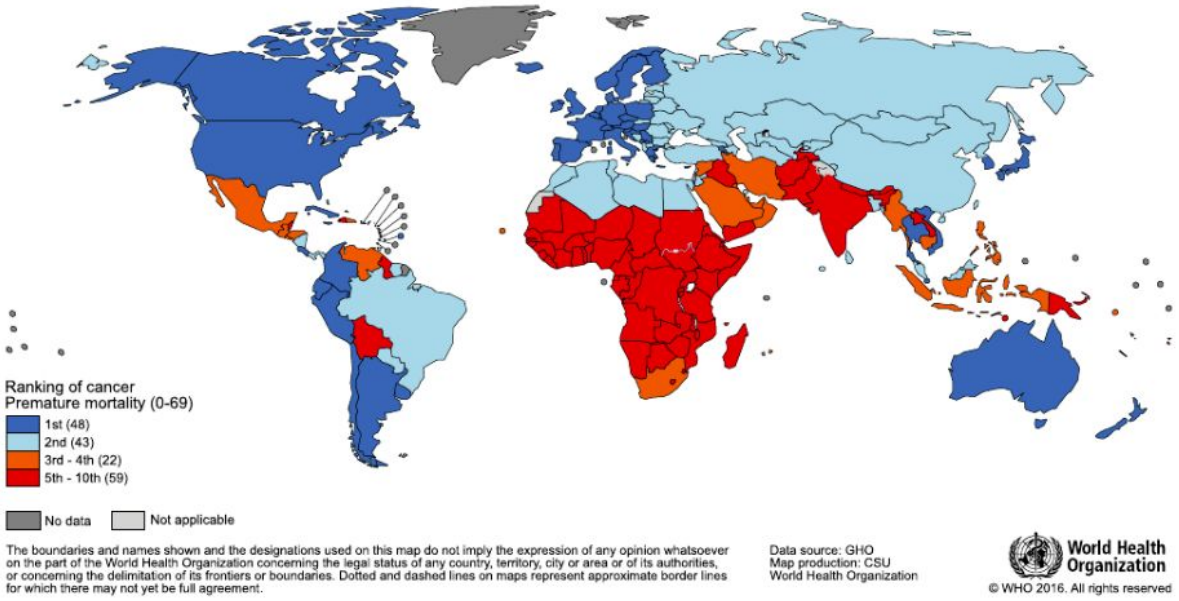


Figure 2. Le cancer comme cause majeure de décès prématuré dans le monde (source Globocan 2018)

La distribution des incidences/décès en 2018 est présentée dans la **Figure 3.** Il est estimé que la moitié des cas/décès est recensé en Asie (taux approchant les 57% de mortalité). Chez les hommes (**Figure 3.B**), l'incidence est estimée à 9,5 millions de cas [49 % en Asie, 23% EU, 20 % US/Canada et 5% Afrique] avec une mortalité de 5,4 millions décès. Chez les femmes (**Figure 3.C**), une incidence de 8,6 millions de cas contre 4,2 millions de décès a été relevée. Au vue de ces données actuelles, les cancers touchent davantage les hommes que les femmes.

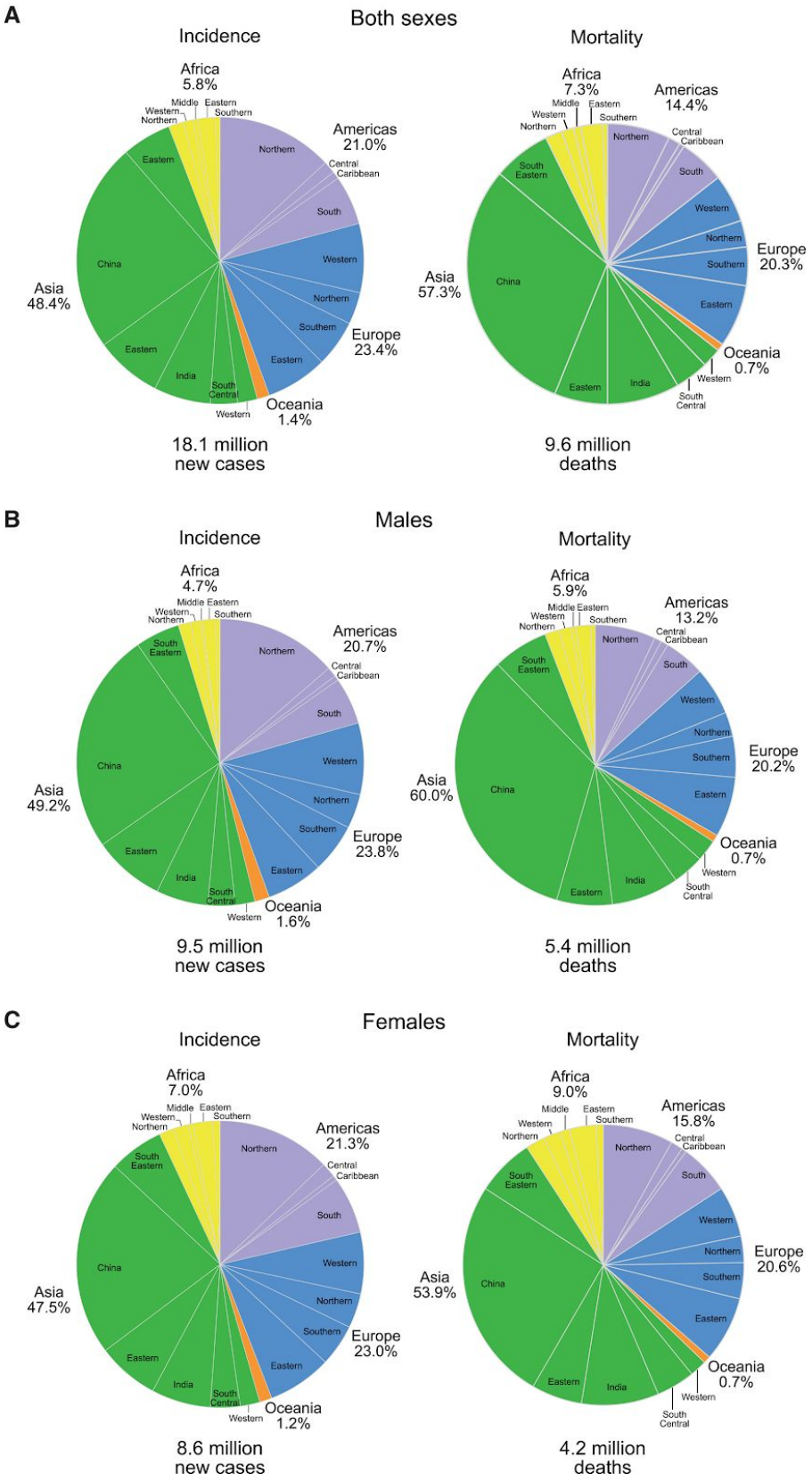


Figure 3. Distribution de l'incidence/décès dus aux cancers en 2018. (source: GLOBOCAN 2018)

S'agissant des types de cancer les plus fréquents, le cancer du sein est le cancer le plus fréquent dans le monde, suivi du cancer de la prostate puis des poumons. Ce dernier est le plus virulent et est responsable de la majeure partie des décès enregistrés. (**Figure 4**).

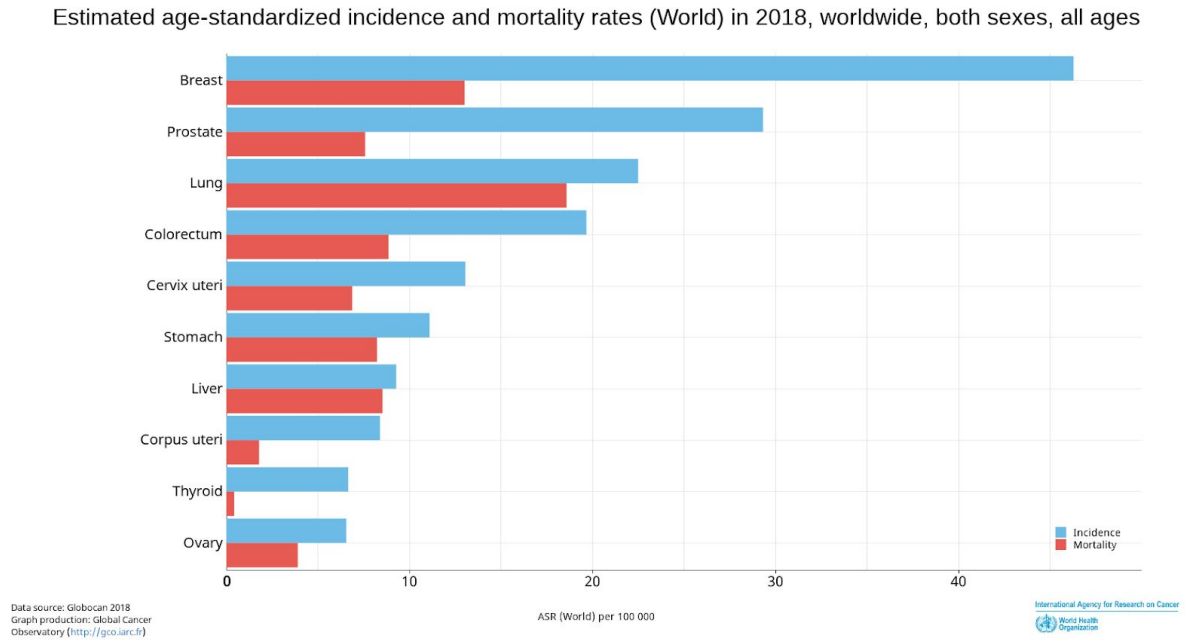


Figure 4. Incidence et mortalité des dix cancers majeurs dans le monde en 2018. (GLOBOCAN 2018)

1.2. Carcinogenèse, apoptose et famille des protéines BCL-2

Le maintien de tout tissu biologique nécessite un respect d'un équilibre entre la vie et la mort des cellules qui le constituent. Tout déséquilibre dans l'homéostasie cellulaire, provoque une panoplie de pathologies qui mettent en péril la survie de l'organisme.

1.2.1. Apoptose et carcinogenèse

En 1972, Kerr et al.,[4] ont associé le terme "Apoptosis" [du grec "apo": se séparer et "ptosis": qui tombe; une référence à la chute des feuilles d'un arbre] à un ensemble de caractéristiques liées à une mort de cellules différente de celles précédemment connues. En effet, ils ont observé une condensation du noyau cellulaire et une fragmentation cellulaire suivie par une phagocytose de ces fragments par les cellules avoisinantes. En 1976,[5] Sulston a décrit une mort cellulaire similaire chez les vers ronds (C.elegans), avec une synchronisation entre le type de cellules entrant un cycle apoptotique et le cycle du développement cellulaire. Par la suite, en 1983,[6] Horvitz a identifié deux mutations sur deux gènes: ced-3 et ced-4 qui ont perturbé cette mort cellulaire programmée. Ceci fut la première suggestion d'une mort cellulaire contrôlée génétiquement. Ces travaux ont été récompensés par le prix Nobel de médecine et physiologie 2002.[7]

Le cancer peut donc être vu comme le résultat d'une succession de modifications génétiques durant lesquelles une cellule saine est transformée en une cellule cancéreuse.
La carcinogenèse est donc le processus de transformation d'une cellule saine en cellule cancéreuse en échappant à l'apoptose (mort cellulaire programmée).

1.2.1.1. Mécanismes de l'apoptose

La compréhension des mécanismes moléculaires contrôlant l'apoptose est donc une étape nécessaire pour contrer la carcinogenèse.
Il existe trois voies de signalisations moléculaires qui régulent l'apoptose: une voie extrinsèque (ou voie du récepteur de la mort cellulaire); une voie intrinsèque (voie mitochondriale) et enfin une troisième voie, peu connue, qu'est la voie du réticulum endoplasmique (RE).[8] La **Figure 5** résume les deux voies principales d'activation de l'apoptose.

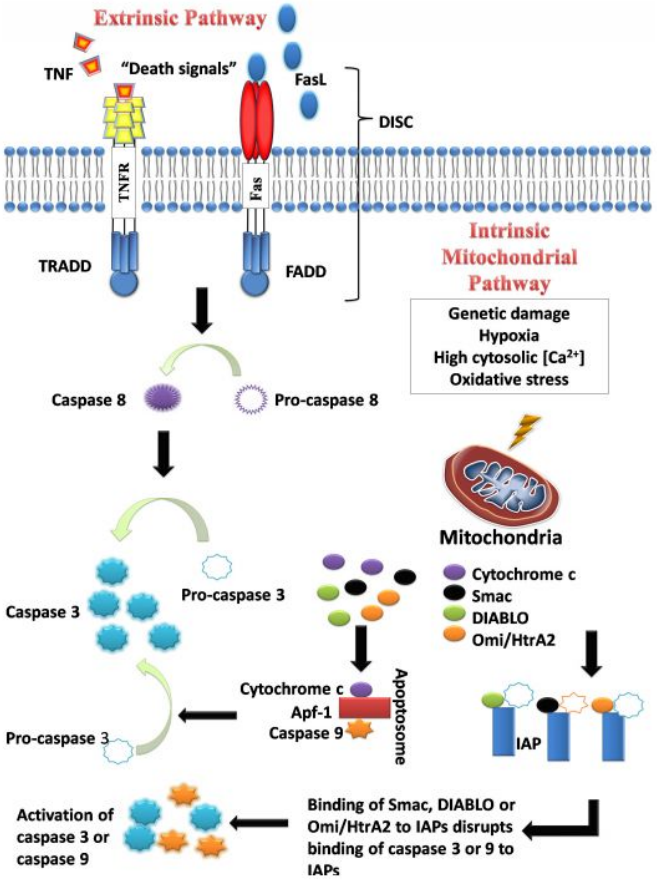


Figure 5. Mécanismes d'activation de l'apoptose. Deux voies principales: la voie des récepteurs de la mort (voie extrinsèque) et voie mitochondriale (voie intrinsèque). D'après Wong et al.,[9]

La voie extrinsèque est initiée lorsque les ligands de la mort se fixent sur leur récepteur. Plusieurs récepteurs de la mort ont été décrits, mais les mieux connus sont les récepteurs TNF de type 1 (TNFR1) et la protéine FAS et leurs ligands respectifs TNF et FasL.[10]

La fixation des ligands sur les récepteurs membranaires entraîne le recrutement des protéines intracellulaires adaptatrices (*FADD: Fas-associated Domain et TRADD: TNF-receptor associated death domain*). L'ensemble récepteur-ligand-protéine adaptatrice forme un complexe dénoté DISC (Death-inducing signaling pathway).[8] Ce complexe initialise par la suite l'assemblage et l'activation de la pro-caspase 8. La forme activée de la caspase 8 initie l'apoptose en clivant d'autres caspases effectrices.

La voie d'activation intrinsèque est une voie intracellulaire. Plusieurs facteurs peuvent initier cette voie: mutations, hypoxie, forte concentration en calcium, stress oxydatif. Indépendamment du facteur initiateur, cette voie est le résultat de la perméabilisation accrue de la membrane externe mitochondriale et la libération des cytochromes-c dans le cytoplasme.[11]

La libération du cytochrome-c au niveau du cytoplasme permet l'activation de la caspase-3 par la formation de l'apoptosome: un complexe formé du cytochrome-c, de l'Apaf-1 et de la caspase-9.[12] D'autres molécules libérées à partir de la mitochondrie telles que SMAC (Second Mitochondria-derived Activator of Caspase) ou DIABLO (Direct Binding protein with Low Pi), permettent l'activation des caspases-3 en se liant à l'IAP (Inhibition of Apoptosis Protein) ce qui induit l'inhibition de la formation du complexe IAP-caspase-3.[13]

La phase d'exécution de l'apoptose est lancée lorsque les caspases-3 sont activées (point de convergence des deux voies d'activation de l'apoptose). L'activation de la cascade des caspases induit des modifications marquantes sur les protéines réparatrices de l'ADN, des protéines du cytosquelette, ce qui contribue au phénotype cellulaire observé lors de l'apoptose.

1.2.2. Famille des protéines Bcl-2

Dans la section précédente, l'apoptose a été décrite comme étant un processus impliquant deux voies de signalisations: une voie externe et une voie mitochondriale interne. Le contrôle de la voie intrinsèque est l'affaire des protéines de la famille Bcl-2.
Cette famille de protéines prend son nom, de l'oncogène situé au point de translocation des chromosomes 14-18 décrit pour la première fois sur un lymphome folliculaire B humain (BCL: B Cell Lymphoma).[14]
Les protéines de la famille Bcl-2 partagent au moins un domaine d'homologie conservé BH (BH: Bcl-2 Homology). Ce domaine d'homologie permet de diviser la famille en 3 groupes **(Figure 6)**.

a. Les protéines anti-apoptotiques multi-domaines, comprennent les protéines Bcl-2, Bcl-xL, Mcl-1, A1, Bcl-w.

b. Les protéines pro-apoptotiques à un seul domaine, le domaine BH3 (les BH3-only): BIM, BID, PUMA, BIK, BMF, HRK, NOXA, BAD.
c. Les protéines pro-apoptotiques multi domaines, comprennent les protéines BAX, BAK et BOK.

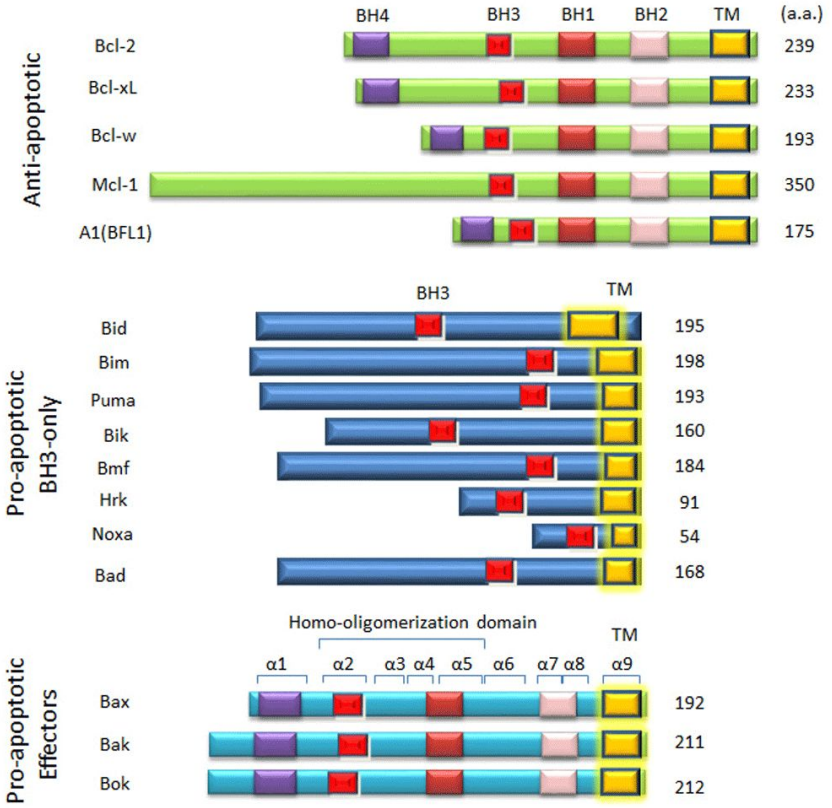


Figure 6. Représentation schématique des membres de la famille Bcl-2 (d'après Luo et al.)[15]

Les protéines pro-apoptotiques à multi domaines BAX et BAK sont responsables de la perméabilisation de la membrane externe mitochondriale (MOMP).

Les protéines BAX et BAK se trouvent sous forme inactive, au niveau du cytosole pour BAX et au niveau de la membrane mitochondriale pour BAK.[16] Les formes actives de BAK et BAX permettent leur homo-oligomérisation au sein de la membrane, formant ainsi des pores laissant échapper les cytochromes-c.[17,18]

Les membres BH3-only et les anti-apoptotiques régulent la perméabilisation de la membrane en jouant sur l'activation des protéines effectrices BAX et BAK.

Du fait du rôle central que jouent les protéines effectrices BAX et BAK dans la perméabilisation de la membrane externe, leurs mécanismes d'activation ont toujours été un terrain de débat pour la communauté scientifique. Traditionnellement, deux modèles d'activation ont été décrits. Un modèle d'activation directe et un modèle d'activation

indirecte. Récemment, un troisième modèle a été proposé par O'Neill et al.[19] Ces trois modèles sont résumés dans la **Figure 7.**

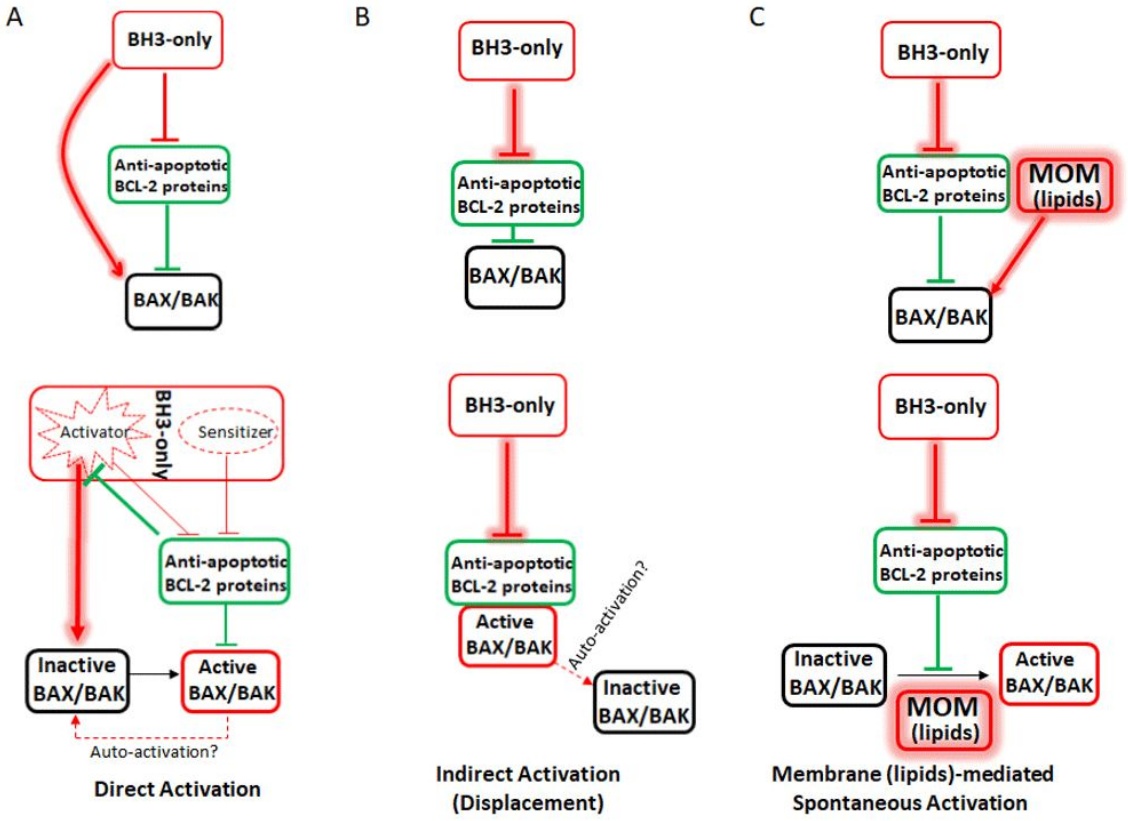


Figure 7. Les trois modèles d'activation des protéines effectrices BAX/BAK (d'après Luo et al.)[15]

Le modèle d'activation directe, classifie les BH3-only comme membres activateurs (BID, BIM, PUMA) et membres sensibilisateurs (BAD).
Lorsque les BH3-only activateurs ne sont pas séquestrés par les membres anti-apoptotiques, ils interagissent directement avec BAX/BAK et les activent.
Les BH3-only sensibilisateurs ne peuvent pas activer directement BAX/BAK mais peuvent interagir avec les membres anti-apoptotiques.
Pour initier l'apoptose, les membres BH3-only sensibilisateurs interagissent avec les anti-apoptotiques libérant ainsi les BH3-only activateurs, leur permettant ainsi, d'activer BAX/BAK.

Le mode d'activation indirecte est basé sur deux hypothèses.
La première hypothèse, stipule que BAX existe intrinsèquement sous forme active, mais séquestrée par les protéines anti-apoptotiques. Durant l'apoptose, les BH3-only interagissent avec les membres anti-apoptotiques permettant ainsi la libération de BAX activée.
La deuxième hypothèse est basée sur le fait que BAX mitochondriale pourrait recruter BAX cytosolique (inactive) vers la mitochondrie (l'auto-activation).

Le dernier mode d'activation récemment décrit, propose quant à lui, un modèle d'activation spontanée médiée par les lipides de la membrane externe mitochondriale. Ce modèle porte

l'hypothèse d'une liaison entre les membres BH3-only avec les membres anti-apoptotiques. Cela permettrait aux lipides de jouer le rôle d'activateur direct de BAX (on suppose une translocation par simple diffusion vers la membrane). Le domaine transmembranaire (TM) de BAX (hélice α9) est un élément central dans ce modèle.

Les différents modèles d'activation des protéines BAX/BAK effectrices, permettent d'illustrer la promiscuité de l'interactome au sein de la famille Bcl-2.

On se pose la question par la suite quant au rôle des membres anti-apoptotiques sur le contrôle de BAX/BAK ?
Il est admis, que BAX existe majoritairement sous forme monomérique inactive au cytosol. Selon le troisième modèle d'activation, BAX cherche incessamment une transition vers la membrane via son domaine TM. Les protéines anti-apoptotiques contribuent en déplaçant la population BAX mitochondriale vers le cytosol.[20]
Hormis le rôle du domaine transmembranaire de Bcl-x_L qui pourrait faciliter cette translocation,[20] le mécanisme moléculaire précis reste une question ouverte.

Le rôle des protéines de la famille Bcl-2 n'est plus à démontrer dans le contrôle de l'apoptose et par conséquent dans le contrôle de la carcinogenèse.

Dans le cadre de cette thèse, nous nous intéressons à un membre anti-apoptotique de cette famille: la protéine Mcl-1.

1.3. La protéine anti-apoptotique Mcl-1

Mcl-1 (Myeloid Cell Leukemia-1) est un membre anti-apoptotique de la famille des protéines Bcl-2. Son gène a été initialement décrit au cours de la différentiation de cellules myéloïdes leucémiques chez l'homme [21] et fut le deuxième à être découvert après la protéine Bcl-2.

1.3.1. Pourquoi Mcl-1 ?

Mcl-1, tout comme les autres membres de la famille Bcl-2, est largement exprimée dans les tissus biologiques. Elle retient toutefois, des caractéristiques et rôles physiologiques qui lui sont propres. En effet, plusieurs cellules en dépendent pour leur survie et leur développement. Cette protéine est essentielle à titre d'exemple dans le processus de l'embryogenèse [22] et sa suppression est létale, à l'encontre de la protéine Bcl-2.[23] Mcl-1 est également primordiale pour le développement des cellules lymphocytaires B et T,[24] et au développement des neurones.[25] Elle est également indispensable pour la régulation des macrophages [26] et à la survie des cellules souches hématopoïétiques.[27]
Ces fonctions et implications dans des processus physiopathologiques la rendent une cible d'étude d'intérêt.

1.3.2. Aspects structuraux de Mcl-1

La multitude de fonctions assurées par Mcl-1 doit être le fruit d'une structure particulière au sein de la famille des protéines Bcl-2. En effet, au sein des membres anti-apoptotiques, Mcl-1 est la seule à contenir 3 domaines BH (BH1, BH2 et BH3) (voir **Figure 8**).

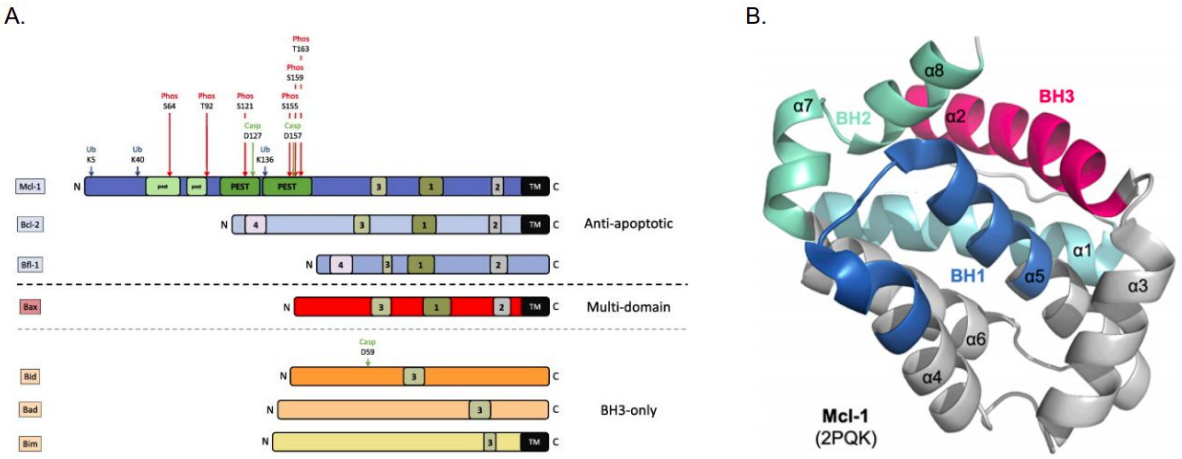


Figure 8. Organisation et aspects structuraux de Mcl-1. **A.** Organisation en domaines fonctionnels BH et régions PEST; **B.** Structure tridimensionnelle de Mcl-1.

De plus, Mcl-1 a une séquence plus longue si comparée aux autres protéines de la famille Bcl-2 (350 aa contre 233 aa de Bcl-xL par exemple). Cette différence est liée essentiellement à la présence d'un domaine non-homologue aux protéines Bcl-2 dans sa partie N-terminale (résidus 1-170). Ce domaine N-terminal est peu ordonné et est très enrichi en séquences PEST (proline (P); acide glutamique (E); serine (S); thréonine (T)).[21]
L'enrichissement en motif PEST est le signe des protéines labiles (taux de turnover très rapide). Ce domaine N-terminal est le lieu des modifications post-traductionnelles (phosphorylation de la Ser121 par exemple ou l'ubiquitinylation sur la Lys40).

La partie C-terminale de la protéine contient le domaine transmembranaire (TM), essentiel à son insertion au sein de la membrane externe mitochondriale. Ce domaine n'est pas résolu expérimentalement, tout comme la partie N-terminale.

La région centrale de la protéine, résolue expérimentalement (résidus 170-320), abrite une cavité hydrophobe, située entre les hélices α3, α4 et α2 (voir **Figure 9**).

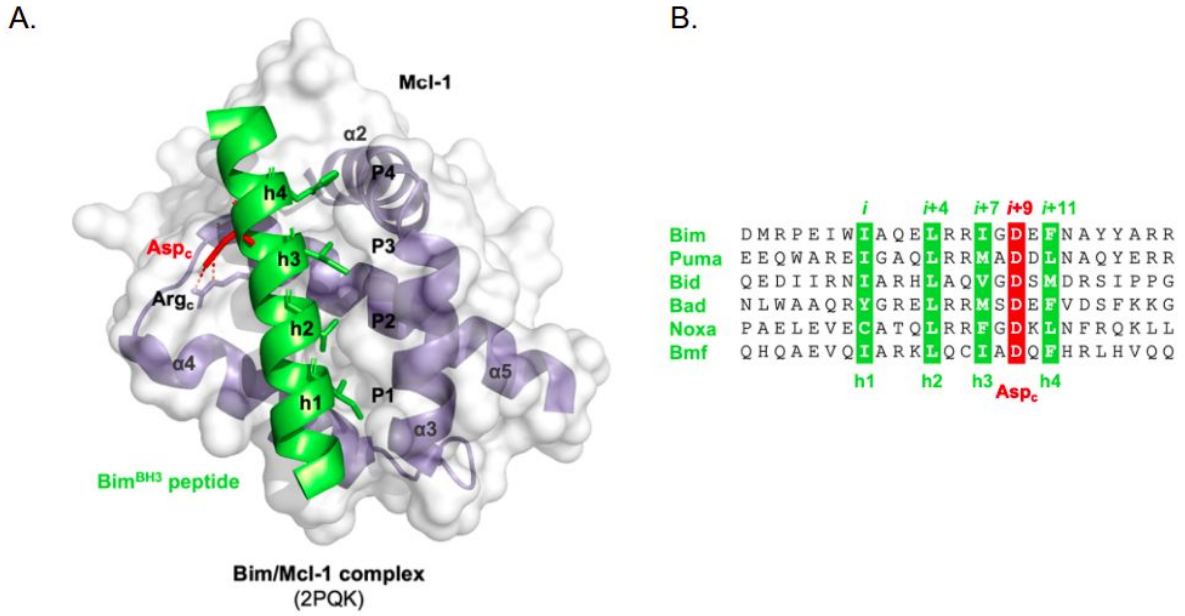


Figure 9. Cavité hydrophobe, hotspots de la protéine Mcl-1. **A.** Peptide BIM-BH3 séquestré dans la cavité hydrophobe de Mcl-1. **B.** Hotspots et interactions conservées (d'après Denis et al.).[28]

Cette cavité hydrophobe étendue, représente le site d'interaction principal de la protéine. C'est le lieu permettant la séquestration des protéines BH3-only par exemple. L'interaction se fait par le biais d'interactions protéine-protéine. Au sein de cette cavité, 4 résidus conservés en position *i, i+4, i+7 et i+11* des protéines BH3-only interagissent avec 4 poches orthostériques principales (nommées P1, P2, P3 et P4).
Il est également à noter, une interaction via pont salin conservée entre un acide aspartique en position *i+9* des BH3-only et une arginine conservée parmi toutes les protéines anti-apoptotiques .

Ces deux composantes: interactions hydrophobes de surface et formation d'un pont salin avec l'Arg 263 de Mcl-1 représentent le mode d'interaction protéine-protéine et protéine-ligand le plus fréquent pour ce site orthostérique.[28]

Récemment, deux autres sites d'interactions permettant la modulation de la fonction de Mcl-1 ont été proposés. Le premier site allostérique (Cys286) (**Figure 10**) a été proposé par Lee et al.[29] Le deuxième site est situé au sein de la cavité principale de Mcl-1. Il s'agit du triplet (Q221 R222 et N223). Zhang et al.,[30] ont démontré que ce motif QRN joue le rôle d'un switch conformationnel.

Une conformation hélicoïdale de ce motif favorise l'ubiquitination in vitro de Mcl-1.

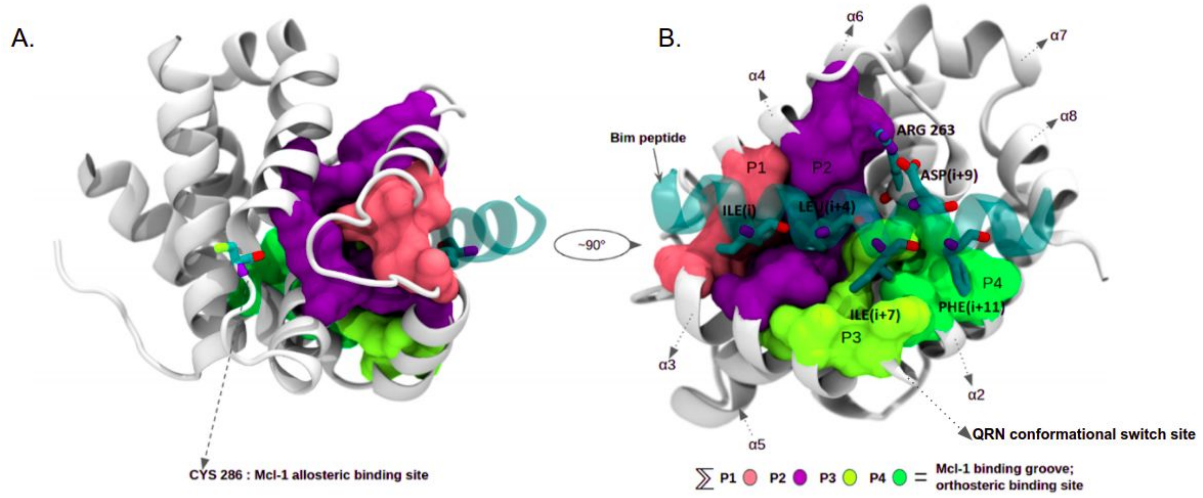


Figure 10. Sites fonctionnels de la protéine anti-apoptotique Mcl-1. **A.** Cys286, site de modulation allostérique; **B.** Surface et cavités hydrophobes formant le site d'interaction orthostérique.

1.3.3. Outils thérapeutiques récents ciblant la protéine Mcl-1

Le gène de Mcl-1 est parmi les gènes les plus surexprimés dans plusieurs types de cancer.[31] (**Figure 11**). Elle est particulièrement surexprimée dans le cancer de l'ovaire, du sein, et les lymphomes. Cette surexpression est une stratégie utilisée par les cellules cancéreuses pour échapper à l'apoptose.

Mcl-1 est également responsable de la résistance aux traitements anticancéreux visant d'autres protéines anti-apoptotiques telles que Bcl-2 ou Bcl-x_L.

Ces paramètres rendent Mcl-1 une cible thérapeutique de choix dans la lutte contre le cancer. Une des stratégies utilisées pour développer des inhibiteurs de Mcl-1, est de mimer le réseau d'interactions protéine-protéine entre les peptides BH3-only et la large cavité hydrophobe qui les séquestre.

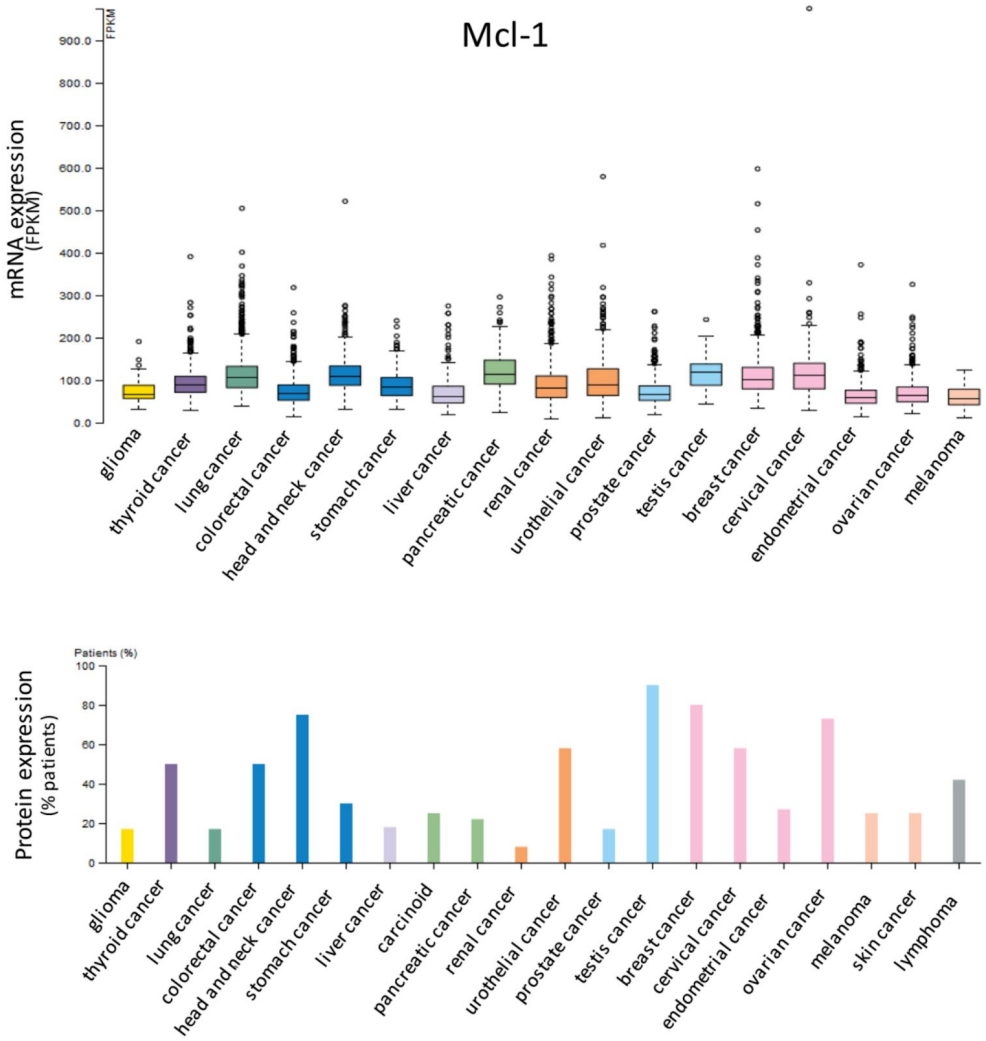


Figure 11. Profil d'expression (mRNA et protéique) de Mcl-1 sur différents cancers. (source Human Protein Atlas database-d'après D'Aguanno et al.)[32]

Cette approche a donné naissance aux inhibiteurs BH3-mimétiques.[33] La **Figure 12** résume schématiquement les principaux inhibiteurs BH3-mimétiques visant les protéines anti-apoptotiques.

La première génération des inhibiteurs BH3-mimétiques avait une sélectivité limitée envers les protéines anti-apoptotiques. L'exemple est le pan-inhibiteur obatoclax (GX15-070), capable d'interagir avec toutes les protéines anti-apoptotiques.[34] Un deuxième exemple est le composé AT-101, dérivé du gossypol, qui interagit avec Bcl-2, Bcl-x_L et Mcl-1.[35,36] Ce composé a été évalué en phase clinique I et II sur des cancers hématologiques et solides. Mais, du fait de leur toxicité, le développement de ces deux inhibiteurs a été arrêté.[32]

L'ABT-737 a été l'un des premiers BH3-mimétiques proposés.[37] Cet inhibiteur de Bcl-xL, Bcl-2 et Bcl-w a été évalué sur des cellules de leucémie chronique (CLL),[38] mais également sur des tumeurs solides.[37]

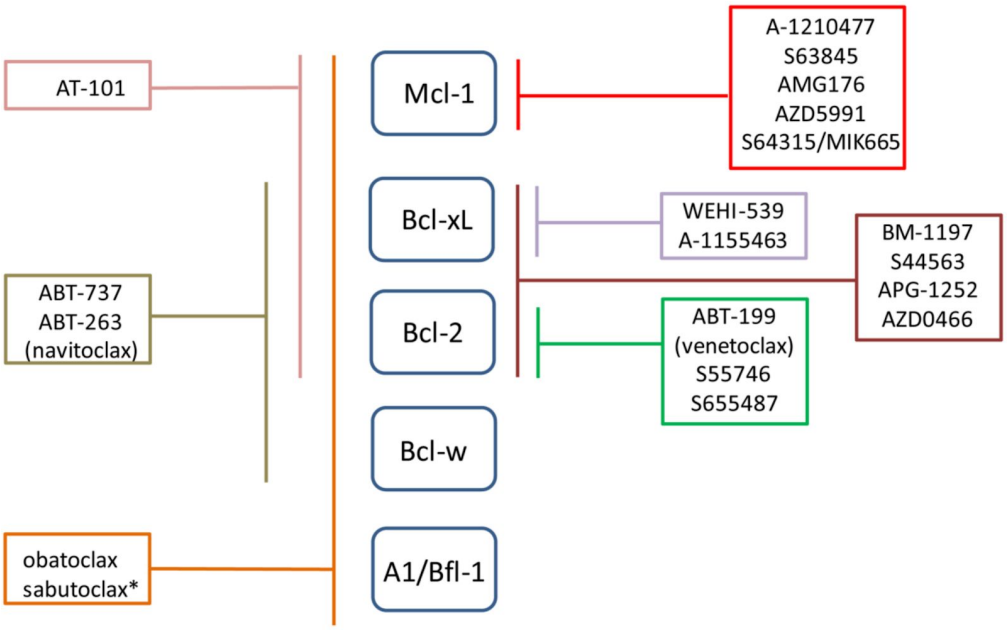


Figure 12. Représentation schématique des principaux inhibiteurs BH3-mimétiques des protéines anti-apoptotiques. D'après D'aguanno et al.,[32]

Du fait de sa faible biodisponibilité orale, l'ABT-737 a été limité, et le Navitoclax, son homologue biodisponible oralement a été proposé. Le Navitoclax a montré une efficacité thérapeutique *in-vivo* sur des modèles de leucémie.[39] Toutefois, le Navitoclax a également provoqué une thrombocytopénie médiée notamment par l'inhibition de Bcl-x_L.[40]

Cette toxicité du Navitoclax a été récemment levée par le développement du DT-2216, un PROTAC (Proteolysis-targeting Chimera), qui permet de diriger Bcl-x_L à interagir avec la VHL (Von Hippel-Lindau) E3 Ligase, causant ainsi sa dégradation. Étant donné que la VHL est peu exprimée dans les cellules plaquettaires, la toxicité est réduite tout en gardant un bon potentiel thérapeutique.[41]

S'agissant des inhibiteurs sélectifs pour Mcl-1, plusieurs sont actuellement disponibles. L'inhibiteur UMI-77,[42] fut l'un des premiers, et il a montré une efficacité *in vivo* sur des lignées pancréatiques (K_i = 490 nM pour Mcl-1).

En 2015, notre laboratoire (CERMN) en collaboration avec l'unité BioTICLA, a proposé un BH3-mimétique (le Pyridoclax) qui a montré une sélectivité envers Mcl-1 en combinaison avec des stratégies anti-Bcl-x_L sur des lignées tumorales ovariennes.[43]

En 2016, les équipes de *Servier et Vernalis* ont proposé le composé S63845, inhibiteur sélectif sur Mcl-1 (K_i = 1,2 nM) possédant une activité *in vivo*. Ce composé active l'apoptose BAX/BAK dépendante *in vivo* et *in vitro* sur des lignées hématologiques.[44]

Un composé dérivé du S-63845, le S-64315/MIK-665 est sous évaluation clinique en phase I chez des patients atteints de lymphomes, AML, et en thérapie combinée avec le Venetoclax. Cette molécule est administrée par perfusion.[45]

L'inhibiteur AMG-176 a été proposé par *AMGEN*. Cette molécule est très sélective pour Mcl-1 (K_i < 1 nM). Elle exprime une efficacité à restaurer l'apoptose sur des lignées cellulaires hématologiques. L'AMG-176 est sous évaluation clinique en monothérapie chez des patients AML ou en thérapie combinée avec le Venetoclax (NCT03797261). Il est important de noter toutefois, que des données sur la toxicité cardiaque [46] de cette série (notamment l'AMG-397, NCT03465540) ont été relevées.

Le laboratoire *Astrazeneca* a également proposé un inhibiteur avec une affinité en dessous du nanomolaire (K_i = 0,2 nM) pour Mcl-1: le composé AZD-5991.[47] De façon similaire au composé de Servier, l'AZD-5991 restaure l'apoptose de façon BAX/BAK dépendante (voie intrinsèque).

La **Figure 13** montre des exemples de structures chimiques d'inhibiteurs sélectifs pour Mcl-1.

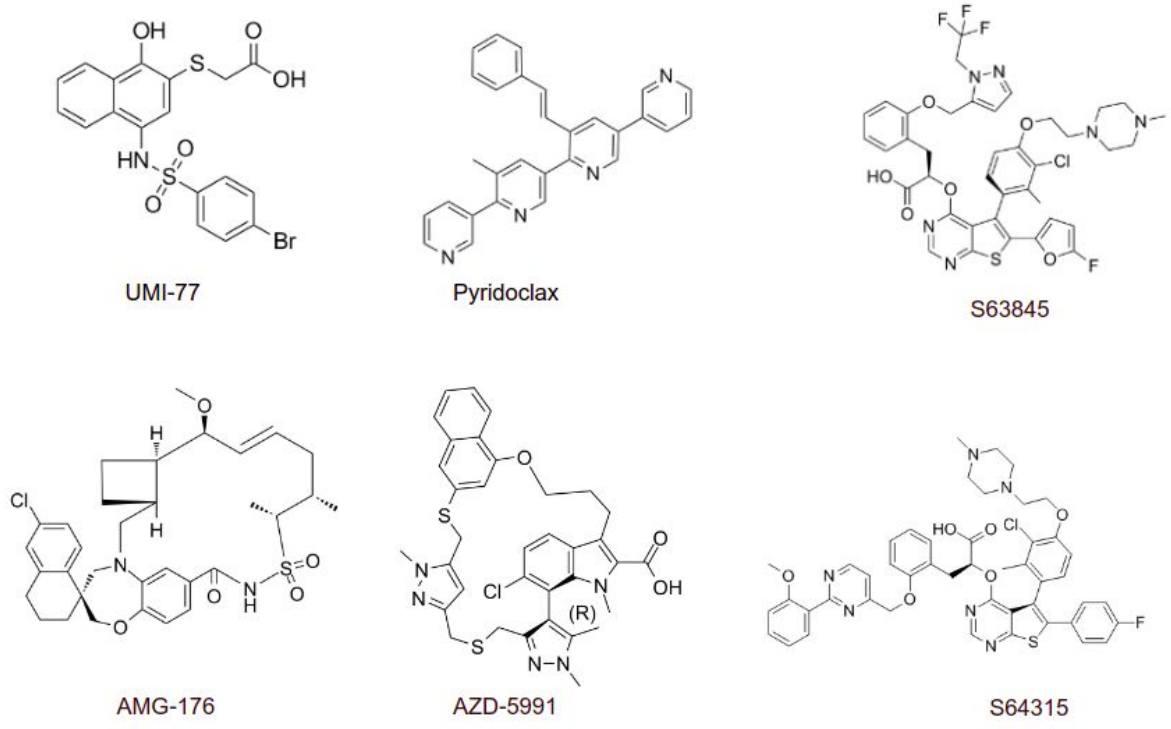


Figure 13. Structures des principaux inhibiteurs BH3-mimétiques et inhibiteurs sélectifs Mcl-1.

2. Aspects méthodologiques

Dans le cadre des travaux de cette thèse, nous faisons appel aux méthodes de simulation de dynamique moléculaire (DM). Dans cette section les aspects théoriques relatifs à ces simulations seront résumés.

2.1. Simulation de dynamique moléculaire

Dans les dernières décennies, beaucoup de progrès dans la détermination des structures tridimensionnelles des protéines ont été apportés par des méthodes expérimentales, en particulier, la cristallographie aux rayons-X,[48] la résonance magnétique nucléaire [49] ou récemment la cryo-microscopie électronique.[50]

Aller au-delà de l'image statique des structures reste toutefois un défi. Même si un grand nombre de techniques a été proposé telles que la relaxation RMN, la spectroscopie de fluorescence ou la cristallographie aux rayons-X à résolution temporelle,[51] aucune permet l'accès à la flexibilité intrinsèque des protéines sur des échelles fonctionnelles.
Malgré les récents progrès, les méthodes expérimentales ayant une résolution spatio-temporelle entre la nano/microseconde et le nanomètre sont peu disponibles pour une analyse de routine. Pour cela, l'information concernant les états conformationnels de la protéine *in vivo* reste obscure. En particulier, les processus de transition entre deux états conformationnels, nécessaires à la compréhension de la fonction d'une protéine, sont en général mal décrits.

Dans ce contexte, les simulations de dynamique moléculaire (DM) représentent une approche intéressante pour obtenir une information sur la dynamique des protéines à résolution atomique et à l'échelle de la micro/milliseconde.

Depuis la première DM d'une protéine, il y a environ 40 ans,[52] la simulation de la dynamique moléculaire est aujourd'hui considérée comme la méthode de choix pour étudier la dynamique des biomolécules.

2.1.1. Principes et approximations

La théorie sous-jacent les simulations de dynamique moléculaire reste simple.
Pour un système biomoléculaire à **N** particules (atomes), la résolution de l'équation de Schrödinger temps-dépendante:

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \hat{H}\Psi(r,t) \qquad (Eq.1)$$

pour la fonction d'onde à **N** particules Ψ**(r,t)** est très prohibitive sinon impossible.
Plusieurs approximations sont donc nécessaires pour simuler une biomolécule solvatée, à l'échelle de la microseconde.

Approx 1. La première approximation est relative à la position du noyau et des électrons. Du fait que la masse du noyau est largement supérieure à celle des électrons, et par conséquent la grande vélocité des électrons par rapport au noyau, les électrons sont assumés suivre le mouvement des noyaux. De ce fait, dans le cadre de l'approximation de Born-Oppenheimer, seulement la position des noyaux est à considérer, avec les degrés de liberté électroniques influençant la dynamique du noyau traités sous la forme d'une énergie potentielle **V(r)**.

Approx 2. La deuxième approximation essentielle utilisée dans le cadre des simulations de DM, est la description classique du mouvement des noyaux par la seconde équation de Newton:

$$m_i \frac{d^2 r_i}{dt^2} = -\nabla_i V(r_1, r_2, ..., r_n) \qquad (Eq.2)$$

où $\mathbf{m_i}$ est la masse du noyau **i** et $\mathbf{r_i}$ est la position du noyau **i**, l'opérateur nabla ∇ représente le gradient du potentiel.

Avec la description classique du mouvement des noyaux, l'équation de Schrödinger pour les degrés de libertés électroniques doit être résolue pour obtenir l'énergie potentielle **V(r)**.
Cependant, du fait du nombre élevé d' électrons, une troisième approximation est nécessaire.

Approx 3. Un champ de force semi-empirique est introduit pour approximer **V(r)**. Un champ de force est un ensemble de termes additifs et fonctionnels permettant de modéliser les interactions liantes et non liantes d'une molécule. Une forme générale d'un champ de force est donnée par l'équation 3:

$$V(r) = V_{liaisons} + V_{angles} + V_{diedre} + V_{coul} + V_{LJ} =$$
$$\sum_{liaison} k_i (l_i - l_{i,0})^2 + \sum_{angles} k_i^{\Theta} (\Theta_i - \Theta_{i,0})^2 + \sum_{diedre} V_n (1 + cos(n\phi - \delta)) + \sum_{coul} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon_r r_{ij}} +$$
$$\sum_{LJ} 4\epsilon_{ij} \left[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6 \right] \qquad (Eq.3)$$

Les termes simples sont de type harmonique (ex. $\mathbf{V_{liaison}}$, $\mathbf{V_{angles}}$) ou motivés par une loi physique (ex. potentiel de Coulomb $\mathbf{V_{coul}}$, potentiel de Lennard-Jones $\mathbf{V_{L,J}}$). Ces termes sont définis par leur forme fonctionnelle et un nombre de paramètres (ex. un rayon atomique pour les interactions de Van der Waals).

Tous les paramètres sont déterminés en utilisant des calculs quantiques *ab-initio,* ou par comparaison des mesures expérimentales structurales et thermodynamiques avec des

moyennes d'ensembles de DM de petites molécules. Entre les différents champs de force, le nombre des termes énergétiques et leurs paramètres peuvent varier.

Étant donné la description des protéines comme points massiques (avec une position $\mathbf{r_i}$ et une vitesse $\mathbf{v_i}$) évoluant dans un potentiel classique sous une force $\mathbf{F_i}$, une simulation de dynamique moléculaire intègre l'équation du mouvement sous forme discrétisée, avec un pas d'intégration $\boldsymbol{\Delta t}$ à l'échelle de la femtoseconde (~ 1 ou 2 fs). Ce pas correspond à la plus élémentaire fréquence vibratoire du système.

L'intégration se fait suivant un schéma numérique. Un exemple est donné par l'algorithme et l'intégrateur *Leap-frog* [53]:

$$\left\{ v_i(t+\frac{\Delta t}{2}) = v_i(t-\frac{\Delta t}{2}) + \frac{F_i(t)}{m_i}\Delta t \quad ; \quad r_i(t+\Delta t) = r_i(t) + v_i(t+\frac{\Delta t}{2})\Delta t \right\} \qquad (Eq.4)$$

Mis à part les interactions avec les membranes plasmiques et d'autres molécules (ADN, co-facteurs, autres protéines), l'eau constitue l'environnement principal des protéines. Pour une simulation d'un système modèle, se rapprochant le plus possible des conditions *in vivo*, les molécules d'eau et les ions, dans des concentrations physiologiques, sont ajoutées afin de solvater le système protéique.

Ayant une "boîte de simulation" contenant le soluté et le solvant, des artefacts peuvent apparaître. Pour éviter ce type d'effets, les conditions périodiques sont souvent appliquées. De ce fait, le système étudié n'a pas de surface de contact (on parle d'image ou de continuum). Cependant, cela pourrait induire d'autres effets si le système moléculaire interagit avec sa propre image (à cause des interactions électrostatiques longue-distance par exemple).

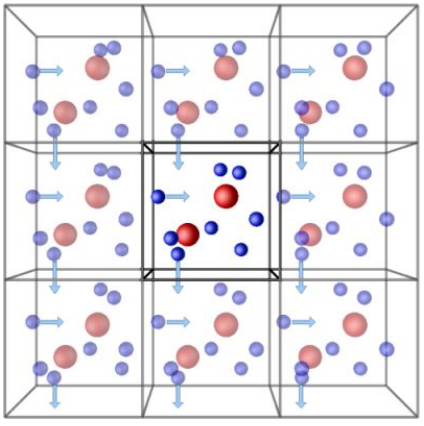


Figure 14. Schéma illustrant la notion de boîte de simulation cubique et conditions périodiques.

En pratique, ces effets sont minimisés en augmentant la taille de la boîte de simulation. Plusieurs choix de cellules unitaires sont possibles: ”cubique”, “octaédrique” … etc. Ces choix de la forme permettent un compromis entre la forme de la protéine, le nombre de molécules d’eau à utiliser et par conséquent l’échantillonnage nécessaire à réaliser.

Étant donné que le solvant (eau ou autre) induit des modifications sur la structure et la dynamique des protéines, les modèles d’eau construits doivent rendre une description précise de ses propriétés. À part les modèles d’eau implicites, où l’eau est représentée sous forme d’un continuum, plusieurs modèles d’eau explicites ont été présentés.[54]

Ces modèles diffèrent sur le nombre de particules utilisées pour représenter les molécules d’eau et les charges statiques qui leur sont attribuées, permettant de rendre compte de la polarisation dans la majorité des champs de force.
Du fait du caractère statique de ces charges durant les simulations de DM, les effets de polarisation explicite ne sont pas considérés. Récemment, plusieurs modèles d’eau polarisables ont été proposés.[54]

Par la résolution de l’équation du mouvement, l’énergie totale du système est conservée, ce qui conduit à un ensemble microcanonique NVE (N: nombre d’atomes constant; V: volume constant et E:énergie constante). Toutefois, les systèmes biologiques simulés échangent constamment de l’énergie avec leur milieu. De plus, une pression usuelle d’un bar est présente.

Pour prendre en compte ces caractéristiques, des algorithmes sont introduits afin de coupler le système à un thermostat et un barostat permettant de simuler un ensemble canonique NPT (N: nombre d’atomes; P: pression constante et T:température constante).

2.1.2. Limitations des simulations de dynamique moléculaire

En dépit des avantages que permettent les simulations de dynamique moléculaire pour l’étude à l’échelle atomique du mouvement des protéines, des limitations techniques et méthodologiques persistent.

Les limitations méthodologiques sont liées à la description classique des atomes et l’approximation de leurs interactions par des termes d’énergie simples, au lieu de l’équation de Schrödinger. Cela implique que les réactions chimiques (formation et déformation des liaisons chimiques) ne soient pas décrites dans le cadre d’une DM simple. D’autres méthodes permettant un traitement au niveau électronique de la protéine existent: “calculs QM/MM”.[55]
Par contre, ces méthodes restent encore très coûteuses en temps de calcul et le traitement QM est généralement réservé à un nombre réduit de résidus.
La deuxième source de limitations est la disponibilité de moyens et d’architectures de calculs permettant de couvrir le nombre d’itérations nécessaires à simuler des mouvements

fonctionnels. Même si les liaisons atomiques des hydrogènes sous contraintes permettent d'éliminer les mouvements à très haute fréquence, le pas d'intégration ne peut être supérieur à 4 fs. De ce fait, une nanoseconde (10^{-9}s) nécessite environ 250000 itérations pour évaluer la force et propager les positions des atomes.

Malgré l'apport des nouveaux progrès algorithmiques et des nouvelles architectures de calcul (notamment GPU), la simulation des mouvements et des échelles temporelles au niveau desquelles les processus biologiques s'effectuent reste inachevable de façon routinière. De là, il est facile de constater que les simulations de DM classiques seules, ne peuvent résoudre le problème d'échantillonnage de l'espace des configurations.

Pour pallier ces limitations, plusieurs méthodes ont été proposées.
L'une des solutions consiste à réduire le nombre d'atomes. Les molécules d'eau par exemple, peuvent être traitées de façon implicite. Une deuxième possibilité considère une diminution du nombre d'atomes par regroupement, on parle alors de modèles en gros-grains.[56] Dans ces modèles, les atomes sont regroupés ensemble (exemple: 4 molécules d'eau peuvent être regroupées en un "pseudo-atome"). Ces regroupements d'atomes ont deux effets:

- Le premier est la réduction de la taille du système à considérer.
- Le deuxième concerne le pas d'intégration, qui peut être augmenté en fonction des mouvements les plus rapides du système.

Le regroupement d'atomes ne se limite pas aux molécules d'eau et plusieurs modèles pour les lipides et protéines existent et permettent une forte réduction du temps de calcul.[56]

L'efficacité d'échantillonnage (par lissage de la surface d'énergie, voir **Figure 15**), qui résulte de la réduction du nombre d'atomes, a par ailleurs une conséquence sur la précision du modèle physique (description des interactions), comparé au modèle tout-atome. Cela réduit en effet, le domaine d'applicabilité de cette approche, notamment pour l'étude des interactions protéine-ligand où une description la plus fine est souvent nécessaire.

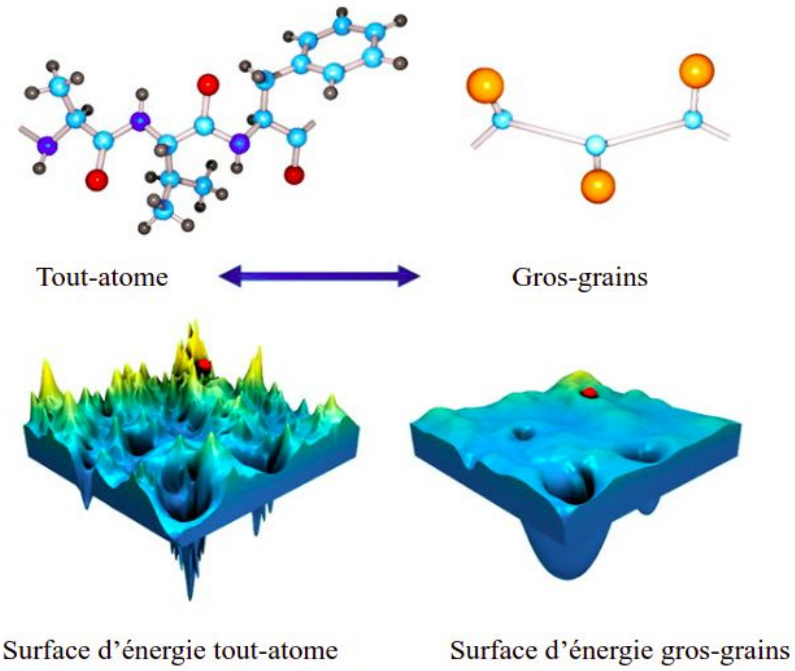


Figure 15. Effet du passage entre une représentation tout-atome d'un système moléculaire (peptide) vers une représentation en gros-grain. La réduction du nombre d'atomes permet de "lisser" la surface d'énergie, permettant ainsi une accélération de l'échantillonnage de l'espace conformationnel. Adapté d'après Kmiecik et al.[56]

2.2. Méthodes d'échantillonnage amélioré

L'objectif premier d'une simulation de DM est l'estimation d'observables macroscopiques (pouvant être mesurés au laboratoire par exemple) à partir des états microscopiques (décrits par la position et la vitesse des atomes) du système étudié.
Le lien entre ces deux niveaux de description se fait par le biais de la mécanique statistique.

2.2.1. Surface d'énergie potentielle (PES) et surface d'énergie libre (FES)

On s'intéresse à un système à **N** atomes (**N** de 1 à 10^6 atomes) dans un espace des phases 2dN-dimensionnel (**N** de 1 à 3).

Une dynamique moléculaire propage le mouvement du système étudié en intégrant l'équation de Newton du mouvement à température **T** (ou la température inverse $\boldsymbol{\beta = 1/k_BT}$, $\mathbf{k_B}$ est la constante de Boltzmann) sous un champ de force classique permettant de rendre compte du potentiel interatomique.

On s'intéresse également aux mouvements occurrents à des échelles temporelles largement supérieures aux fréquences vibrationnelles les plus élémentaires du système (> 1 ou 2 fs^{-1}).
Cela implique que les phénomènes étudiés soient considérés comme des processus ou événements rares.

Deux concepts liés découlent des événements rares: la séparabilité de l'échelle temporelle des mouvements et la markovianité de la dynamique.

a. Séparabilité des échelles temporelles.
Cela sous-entend que la dynamique du système évolue sous des régimes différents, impliquant la présence de puits d'énergie profonds permettant de les distinguer. En d'autres termes, le temps de décorrélation d'une variable du système ($\boldsymbol{\tau}_b$) dans un puits d'énergie est beaucoup plus rapide que celui permettant la sortie du puits ($\boldsymbol{\tau}_e$) ($\boldsymbol{\tau}_b << \boldsymbol{\tau}_e$) (Voir **Figure 16**)

b. Markovianité du processus.
Un processus est dit Markovien ou ayant la propriété de Markov, si l'état du système à un temps (t+1) ne dépend que de son état à l'instant (t) et non pas des états précédents:

$$P(X_t = x_t | X_0 = x_0, ..., X_{t-1} = x_{t-1}) = P(X_t = x_t | X_{t-1} = x_{t-1}) \qquad (Eq.5)$$

Cela implique pour un système moléculaire sortant d'un puits d'énergie profond, une "perte du chemin permettant l'accès au puits". Cette propriété est raisonnable, puisqu'un simple changement dans les conditions initiales (par une perturbation du thermostat par exemple) engendre une trajectoire différente sur la surface d'énergie.

Donc, une dynamique moléculaire est un processus de Markov avec une composante déterministe (champ de force et l'énergie potentiel U(r)) et une composante stochastique (émanant des fluctuations du thermostat par exemple).[57]

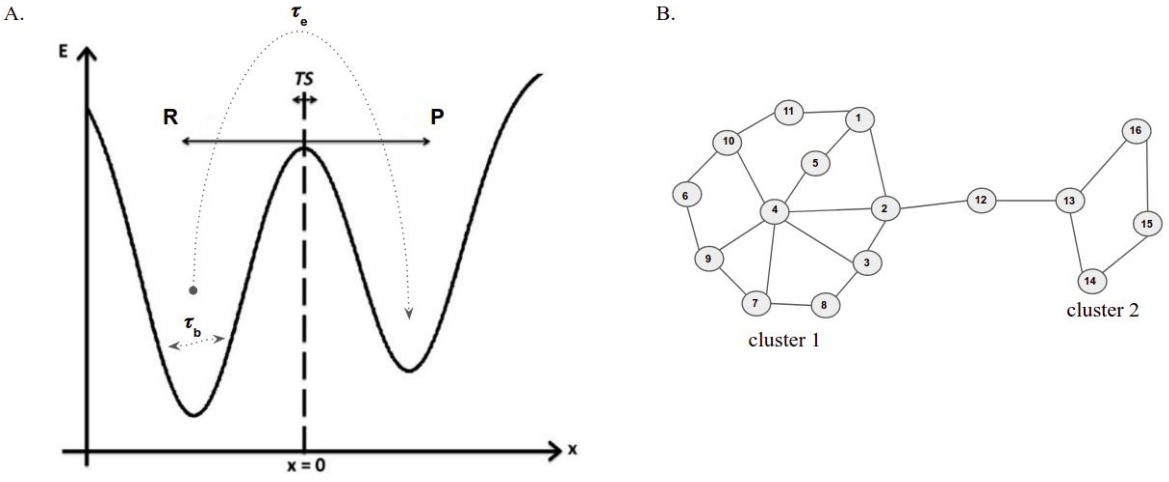


Figure 16. Séparation des échelles temporelles et markovianité d'un processus.
A. Schéma d'un profil d'énergie unidimensionnel. (x) représente une variable réactionnelle et (y) l'énergie. (R): réactif; (P): produits; (TS): état de transition; La profondeur des puits est largement supérieure à K_BT ou la fluctuation d'énergie associée à x. De ce fait, le passage du bassin 1 (R) vers le bassin 2 (P) devient un événement rare et les états de transitions sont très peu visités (séparation des échelles temporelles).
B. Analogie avec une marche aléatoire sur un graphe. Pour un processus Markovien, la transition du cluster 1 (état 2) vers le cluster 2 ne dépend que de l'état de la matrice de Markov à l'état 2 et non des états précédents.

Une analogie intéressante pour illustrer ces deux propriétés, est la marche aléatoire dans un graphe. En effet, un marcheur aléatoire passe beaucoup plus de temps (donc une énergie plus faible) à explorer les états accessibles dans le cluster 1 (intra-cluster) que de temps passé dans les états de transitions inter-clusters (énergie plus élevée). Ceci illustre la propriété de séparabilité des échelles temporelles. La markovianité indique que la transition cluster 1 -> cluster 2 ne dépend que de l'état du système (l'état de la matrice des transitions ou matrice de Markov) à l'état 2 et non des états précédents (dans le cluster 1) (voir **Figure 16**)

En évoquant les deux propriétés précédentes, la notion de paysage d'énergie peut être introduite.

La généralisation de cette notion, à l'espace des coordonnées atomiques **R** à partir de l'espace des configurations (**3N** dimensionnel) possibles, et en considérant l'énergie potentielle **U(R)** du système, nous permet de caractériser et de définir la notion de surface d'énergie potentielle (PES).

Plusieurs méthodes sont disponibles dans la littérature pour caractériser une surface d'énergie potentielle (PES).

Un exemple de ces approches est la méthode de l'hyperdynamique.[58,59] Dans cette approche, l'exploration de l'espace des phases est accélérée en surélevant les bassins d'énergie sans modifier les états de transitions.
Une variante de cette méthode, est connue sous le nom générique de "dynamique moléculaire accélérée",[60] dans laquelle le biais additionnel (ou "boost"): $\Delta V(R) = \Delta V\{U(R)\}$, où **U(R)** est l'énergie potentielle du système, permet au système d'échapper la trappe cinétique.

Ces méthodes sont traditionnellement appliquées aux systèmes à taille réduite, du fait de leur coût de calcul élevé. En effet, une surface d'énergie potentielle contient un nombre élevé d'états de transition ("saddle points"), on parle alors de rugosité du paysage d'énergie, un concept mis en avant par Frauenfelder pour les protéines.[61,62]

Une stratégie alternative consiste à caractériser une surface d'énergie libre (FES: "Free Energy Surface"), en fonction d'un ensemble de variables collectives (ou coordonnées réactionnelles) $\mathbf{s=\{s_1,...s_k\}}$ où $\mathbf{k << N}$.
Cet ensemble de variables collectives est donc un sous-ensemble de l'espace des phases. Son choix est souvent dicté par le processus ou le mécanisme à étudier.

En terme de l'énergie potentielle du système **U(R)**, l'énergie libre **F(s)** peut être définie par:

$$F(s) \quad = \quad -\beta \ln \int dR\delta(s - s(R))e^{-\beta U(R)} \qquad (Eq.6)$$

Cette définition diffère donc de la définition traditionnelle de l'énergie de Helmholtz [63] par le terme **δ(s-s(R))**. Ce terme permet la sélection d'une région dans l'espace des phases associée à une valeur **s** de la variable collective.

Plusieurs méthodes basées sur la construction d'une surface d'énergie libre (FES) en utilisant une ou plusieurs variables collectives ont émergé. Une revue faisant le point sur ces méthodes a été proposée par Allison et al.[64]

En résumé, deux méthodes phares sont représentatives de ces approches: la méthode d'échantillonnage "parapluie" ou (Umbrella Sampling: US) et la Métadynamique (MetaD).
Ces méthodes sont communément appelées: dynamiques moléculaires biaisées, dans le sens où un biais **V(s)** est utilisé dans l'espace des variables collectives **(s)** pour encourager le système à explorer davantage l'espace des configurations.

L'Umbrella Sampling (US) a été proposée par Torrie et Valleau en 1977.[65] Dans cette approche, un biais $\mathbf{V_{umb}(s)}$ sous forme d'un potentiel harmonique sur une variable collective **(s)**, est ajouté au système.

$$V_{umb}(s) \quad = \quad \frac{1}{2}K(s - s_0)^2 \qquad (Eq.7)$$

Où **s** est une variable collective; $\mathbf{s_0}$ est la valeur de la variable collective pour laquelle le biais est ajouté et **K** la constante de force appliquée.

Cette méthode nécessite un ensemble de simulations (une simulation pour chaque valeur $\mathbf{s_0}$) appelées "fenêtres" (voir **Figure 17**) nécessaires pour couvrir **(s)**. Chaque simulation permet une estimation biaisée de l'énergie libre. Une estimation non-biaisée est obtenue par

$$F(s) = \frac{-1}{\beta} lnN(s) - V_{umb}(s) \qquad (Eq.8)$$

où **N(s)** est un histogramme des états visités; $\mathbf{V_{umb}(s)}$ est le potentiel appliqué; $\boldsymbol{\beta = 1/k_BT}$ avec $\mathbf{k_B}$ constante de Boltzmann et **T** la température en Kelvin.

L'ensemble des profils d'énergie libre non-biaisée, obtenus pour chaque valeur s_0 de s est combiné par une analyse d'histogramme WHAM ("Weighted Histogram Analysis Method")[66] afin de reconstruire le profil d'énergie libre complet sur **(s)**.

L'application de l'US, nécessite donc une connaissance préalable de l'état initial et final du système. De plus, la définition et la segmentation d'une variable collective reliant les deux états n'est pas intuitive, et un chevauchement des profils sur **s** doit être vérifié.

2.2.2. Métadynamique

La métadynamique (MetaD) a été proposée par Laio et Parrinello en 2002.[67] Elle reprend l'essence de l'Umbrella Sampling, à savoir déposer un biais sur un espace de variables collectives pour favoriser l'exploration de la surface d'énergie libre.
Mais à l'encontre de l'US, une seule simulation est réalisée (variante classique de la MetaD) et le biais est non pas statique, mais dynamique.

Dans une simulation de MetaD, un biais répulsif sous une forme de gaussienne est 'versé' de façon "histoire-dépendante" dans l'espace des variables collectives.

Dans sa variante classique, le potentiel $\mathbf{U_G(s,t)}$ est de la forme:

$$U_{G(s,t)} = \quad w \sum_{t'=\tau,2\tau,\ldots,<t} e^{\frac{-(s(R,t) \quad - \quad s(R,t'))^2}{2\sigma^2}} \qquad (Eq.9)$$

Avec: ***(w)*** la hauteur de la gaussienne; **(σ)** sa largeur; **(*τ*)** est le pas de déposition; **s(R,t)** est la valeur de la variable collective au temps **t** pour les coordonnés **(R)**.

Une fois la convergence de la MetaD atteinte, la surface d'énergie libre est "noyée" par le potentiel déposé tout au long de la trajectoire. Une estimation non-biaisée de l'énergie libre sur la variable collective **(s)** est donnée lorsque *t* tend vers l'infini par la limite:

$$\lim_{t\to\infty} U_G(s,t) \quad \sim \quad -F(s) \qquad (Eq.10)$$

La variante classique de la MetaD peut mener à des erreurs une fois que les variables collectives deviennent diffusives. En effet, après l'aplatissement de la surface d'énergie libre à la fin de la simulation, l'algorithme classique de la MetaD continue de déposer un biais tout au long de la variable collective (voir **Figure 17**).
Afin de remédier à ce problème, une variante "bien tempérée" WT-MetaD (well-tempered MetaD) a été proposée par Barducci et al.[68]

Dans la WT-MetaD, le taux de déposition (c.à.d la hauteur de la gaussienne par unité de temps) est diminuée de façon adaptative là où le biais est accumulé.

La hauteur de la gaussienne est recalculée à chaque pas selon l'équation:

$$w = w_0 e^{-U(s,t).(k_B \Delta T)^{-1}} \qquad (Eq.11)$$

où **Δt** est un paramètre permettant d'adapter la vitesse à laquelle la hauteur **w** est réévaluée par le facteur-biais ("Bias Factor") **B = T+Δt/T**. La valeur standard du facteur-biais pour des simulations de MetaD de protéines est entre 10 et 20.

L'échantillonnage de l'espace de la variable collective est ainsi effectué à une température élevée ou régulée par le facteur-biais (d'où le nom de cette variante):

$$U(s, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(s) + constante \qquad (Eq.12)$$

Lorsque **Δt** approche de 0, le régime d'une dynamique conventionnelle est retrouvé, et lorsque **Δt** tend à ∞, la variante classique de la MetaD est obtenue.

L'effet de la diminution des hauteurs des gaussiennes, permet une exploration et une reconstruction plus précise de la surface d'énergie libre. Cependant, cela a un coût: lorsque le taux de déposition est réduit, le temps de simulation nécessaire pour la convergence se trouve prolongé.

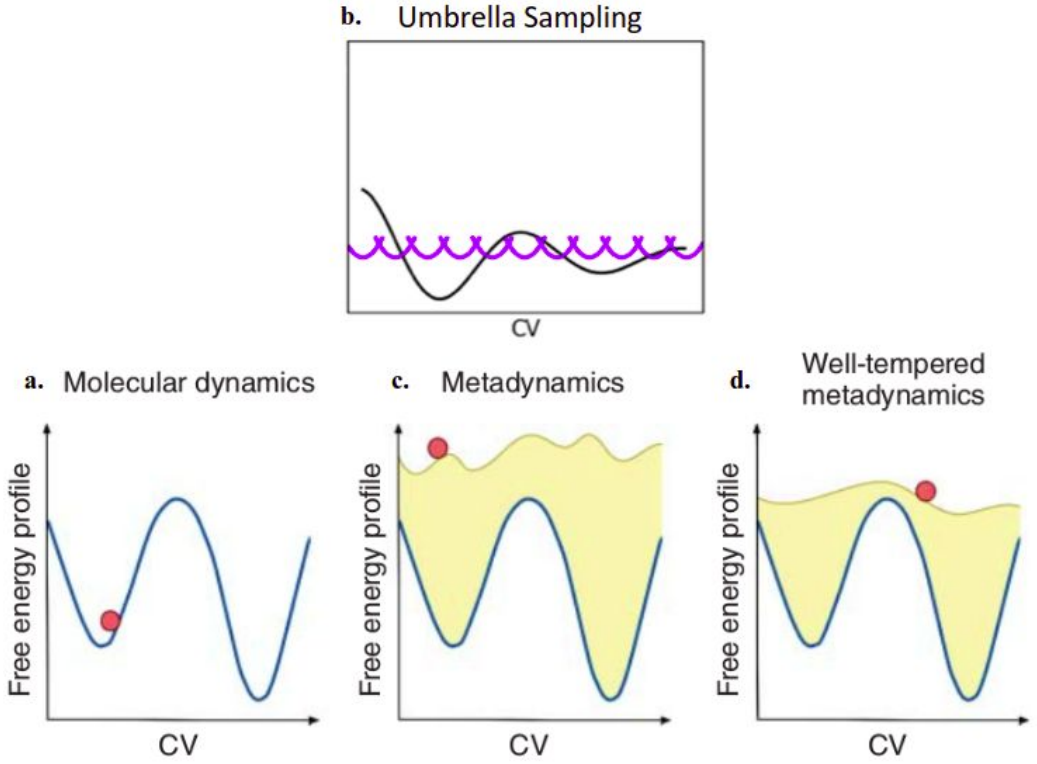


Figure 17. Principales méthodes d'échantillonnage amélioré pour les simulations de DM.
a. Dans une simulation de DM non biaisée, le système explore les conformations disponibles dans un puits d'énergie sans pouvoir explorer d'autres.
b. Pour la méthode de l'Umbrella Samling (US), un biais harmonique est construit tout au long des segmentations (fenêtres) sur la variable collective (CV).
c. Métadynamique conventionnelle (MetaD): un biais sous forme d'une gaussienne est déposé tout au long de la variable collective, permettant une estimation du profil d'énergie.
d. Métadynamique bien tempérée (WT-MetaD). Le biais déposé pendant la simulation décroît selon un schéma déterminé. Cela permet de réduire l'erreur sur le profil d'énergie libre.

2.2.3. Choix des variables collectives et réduction de dimensionnalité.

La convergence des surfaces d'énergie libre (FES) obtenues et le choix des variantes de la MetaD à utiliser est dicté par le processus étudié, les ressources de calcul disponibles mais surtout de la possibilité à identifier un ensemble de variables collectives à même de distinguer les différents états conformationnels de la protéine.

En pratique, la première étape pour choisir une/des variables collectives (CVs) est l'analyse des données expérimentales disponibles. Pour les protéines, les ensembles structuraux résultants des différentes études par RMN et diffraction aux rayons X, peuvent être utilisés pour cette tâche.[69] Nous donnons un exemple d'une extraction de variables collectives à travers des ensembles structuraux de Mcl-1 au chapitre 3.

Dans le cas où les processus étudiés sont assez complexes, notamment à cause de leur dimensionnalité inhérente, d'autres types de procédures sont nécessaires afin d'estimer les CVs. En effet, l'utilisation des approches d'apprentissage-machine (supervisées ou non) et de réduction de dimensionnalité, peuvent être d'utilité pour ce type de problèmes.

L'Analyse en Composantes Principales (ACP) est l'une des méthodes non-supervisées de réduction de dimensionnalité.[70] On appelle communément la dynamique extraite par réduction de dimensionnalité sur une trajectoire de dynamique moléculaire: dynamique essentielle.

L'espace de la dynamique essentielle résulte d'une projection des coordonnées cartésiennes ("snapshots" de la trajectoire), sur un espace orthogonal formé par les vecteurs propres. Ces vecteurs propres sont obtenus par une diagonalisation d'une matrice de covariance **C.**

$$C = \left\langle \quad (R - \langle R \rangle)(R - \langle R \rangle)^T \quad \right\rangle \qquad (Eq.13)$$

où : **R** est un vecteur des coordonnées (atomes C_α par exemple) des N résidus de la protéine et **<R>** est la moyenne sur la trajectoire.

Les composantes principales sont obtenues par une décomposition de **C** en valeurs propres et vecteurs propres:

$$C = \sum_{i=1}^{3N} \sigma_i p_i p_i^T \qquad (Eq.14)$$

où: $\mathbf{p_i}$ est le ième vecteur propre et $\boldsymbol{\sigma_i}$ sa valeur propre associée.

Ainsi, la première composante principale (PC1 ou $\mathbf{p_1}$) et sa variance $\boldsymbol{\sigma_1}$ associée, définissent le premier mode de basse fréquence de la dynamique de la protéine.

Les premiers résultats démontrant l'utilité et l'intérêt des modes à basse fréquence pour les protéines proviennent de l'analyse des modes normaux (NMA) de la trypsine pancréatique.[71]

L'analyse par modes normaux assume toutefois l'harmonicité de la dynamique autour d'un puits d'énergie (les modes normaux sont calculés par diagonalisation de la matrice Hessienne [1] d'une seule structure). L'analyse par modes normaux reste toutefois limitée, face aux mouvements anharmoniques, qui sont souvent d'intérêt, et liés à la diffusion des populations conformationnelles de la protéine entre différents puits d'énergie. De ce fait, la dynamique essentielle est ici privilégiée.

Nous avons mis à profit la dynamique essentielle, pour l'étude de la dynamique conformationnelle de Mcl-1. Des exemples d'application sont donnés au chapitre 3 et 4.

3. Organisation du manuscrit et objectifs

Ce manuscrit est organisé en quatre chapitres. Le présent chapitre est une introduction qui présente une vue d'ensemble sur le contexte de l'étude et les aspects méthodologiques utilisés tout au long de cette thèse.

Le deuxième chapitre concerne une étude sur le mode d'interaction du Pyridoclax (un inhibiteur BH3-mimétique) avec la protéine Mcl-1.

Dans le troisième chapitre, la dynamique et l'espace conformationnel de Mcl-1 sont étudiés et un mécanisme d'inhibition allostérique de Mcl-1 est proposé.

Le quatrième chapitre, traite d'une approche de détection de poches cryptiques sur Mcl-1 par dynamique moléculaire biaisée.

Enfin, une conclusion générale permet de synthétiser les résultats obtenus.

[1] Dérivée seconde de l'énergie potentielle V: $(\partial^2 V)(\partial x_i \partial x_j)$

4. Références

(1) Faguet, G. B. A Brief History of Cancer: Age-Old Milestones Underlying Our Current Knowledge Database. *Int. J. Cancer* **2015**, *136* (9), 2022–2036.
(2) *Papyros Ebers: Das Hermetische Buch über Die Arzeneimittel Der Alten Ägypter in Hieratischer Schrift*; Ebers, G., Stern, L. C., Eds.; W. Engelmann: Leipzig, 1875.
(3) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2018**, *68* (6), 394–424.
(4) Kerr, J. F.; Wyllie, A. H.; Currie, A. R. Apoptosis: A Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics. *Br. J. Cancer* **1972**, *26* (4), 239–257.
(5) Sulston, J. E. Post-Embryonic Development in the Ventral Cord of Caenorhabditis Elegans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1976**, *275* (938), 287–297.
(6) Horvitz, H. R.; Sternberg, P. W.; Greenwald, I. S.; Fixsen, W.; Ellis, H. M. Mutations That Affect Neural Cell Lineages and Cell Fates during the Development of the Nematode Caenorhabditis Elegans. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *48 Pt 2*, 453–463.
(7) Letai, A. Apoptosis and Cancer. *Annu. Rev. Cancer Biol.* **2017**, *1* (1), 275–294.
(8) O'Brien, M. A.; Kirby, R. Apoptosis: A Review of pro-Apoptotic and Anti-Apoptotic Pathways and Dysregulation in Disease. *J. Vet. Emerg. Crit. Care* **2008**, *18* (6), 572–585.
(9) Wong, R. S. Y. Apoptosis in Cancer: From Pathogenesis to Treatment. *Journal of Experimental & Clinical Cancer Research*. 2011. https://doi.org/10.1186/1756-9966-30-87.
(10) Hengartner, M. O. Apoptosis: Corralling the Corpses. *Cell* **2001**, *104* (3), 325–328.
(11) Danial, N. N.; Korsmeyer, S. J. Cell Death: Critical Control Points. *Cell* **2004**, *116* (2), 205–219.
(12) Kroemer, G.; Galluzzi, L.; Brenner, C. Mitochondrial Membrane Permeabilization in Cell Death. *Physiol. Rev.* **2007**, *87* (1), 99–163.
(13) LaCasse, E. C.; Mahoney, D. J.; Cheung, H. H.; Plenchette, S.; Baird, S.; Korneluk, R. G. IAP-Targeted Therapies for Cancer. *Oncogene* **2008**, *27* (48), 6252–6275.
(14) Tsujimoto, Y.; Finger, L. R.; Yunis, J.; Nowell, P. C.; Croce, C. M. Cloning of the Chromosome Breakpoint of Neoplastic B Cells with the t(14;18) Chromosome Translocation. *Science* **1984**, *226* (4678), 1097–1099.
(15) Luo, X.; O'Neill, K. L.; Huang, K. The Third Model of Bax/Bak Activation: A Bcl-2 Family Feud Finally Resolved? *F1000Res.* **2020**, *9*. https://doi.org/10.12688/f1000research.25607.1.
(16) Suzuki, M.; Youle, R. J.; Tjandra, N. Structure of Bax: Coregulation of Dimer Formation and Intracellular Localization. *Cell* **2000**, *103* (4), 645–654.
(17) Dewson, G.; Kratina, T.; Sim, H. W.; Puthalakath, H.; Adams, J. M.; Colman, P. M.; Kluck, R. M. To Trigger Apoptosis, Bak Exposes Its BH3 Domain and Homodimerizes via BH3:groove Interactions. *Mol. Cell* **2008**, *30* (3), 369–380.
(18) George, N. M.; Evans, J. J. D.; Luo, X. A Three-Helix Homo-Oligomerization Domain Containing BH3 and BH1 Is Responsible for the Apoptotic Activity of Bax. *Genes Dev.* **2007**, *21* (15), 1937–1948.
(19) O'Neill, K. L.; Huang, K.; Zhang, J.; Chen, Y.; Luo, X. Inactivation of Prosurvival Bcl-2

Proteins Activates Bax/Bak through the Outer Mitochondrial Membrane. *Genes Dev.* **2016**, *30* (8), 973–988.
(20) Edlich, F.; Banerjee, S.; Suzuki, M.; Cleland, M. M.; Arnoult, D.; Wang, C.; Neutzner, A.; Tjandra, N.; Youle, R. J. Bcl-xL Retrotranslocates Bax from the Mitochondria into the Cytosol. *Cell* **2011**, *145* (1), 104–116.
(21) Kozopas, K. M.; Yang, T.; Buchan, H. L.; Zhou, P.; Craig, R. W. MCL1, a Gene Expressed in Programmed Myeloid Cell Differentiation, Has Sequence Similarity to BCL2. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (8), 3516–3520.
(22) Rinkenberger, J. L.; Horning, S.; Klocke, B.; Roth, K.; Korsmeyer, S. J. Mcl-1 Deficiency Results in Peri-Implantation Embryonic Lethality. *Genes Dev.* **2000**, *14* (1), 23–27.
(23) Veis, D. J.; Sorenson, C. M.; Shutter, J. R.; Korsmeyer, S. J. Bcl-2-Deficient Mice Demonstrate Fulminant Lymphoid Apoptosis, Polycystic Kidneys, and Hypopigmented Hair. *Cell* **1993**, *75* (2), 229–240.
(24) Opferman, J. T.; Letai, A.; Beard, C.; Sorcinelli, M. D.; Ong, C. C.; Korsmeyer, S. J. Development and Maintenance of B and T Lymphocytes Requires Antiapoptotic MCL-1. *Nature* **2003**, *426* (6967), 671–676.
(25) Arbour, N.; Vanderluit, J. L.; Le Grand, J. N.; Jahani-Asl, A.; Ruzhynsky, V. A.; Cheung, E. C. C.; Kelly, M. A.; MacKenzie, A. E.; Park, D. S.; Opferman, J. T.; Slack, R. S. Mcl-1 Is a Key Regulator of Apoptosis during CNS Development and after DNA Damage. *J. Neurosci.* **2008**, *28* (24), 6068–6078.
(26) Steimer, D. A.; Boyd, K.; Takeuchi, O.; Fisher, J. K.; Zambetti, G. P.; Opferman, J. T. Selective Roles for Antiapoptotic MCL-1 during Granulocyte Development and Macrophage Effector Function. *Blood* **2009**, *113* (12), 2805–2815.
(27) Opferman, J. T. Obligate Role of Anti-Apoptotic MCL-1 in the Survival of Hematopoietic Stem Cells. *Science*. 2005, pp 1101–1104. https://doi.org/10.1126/science.1106114.
(28) Denis, C.; Sopková-de Oliveira Santos, J.; Bureau, R.; Voisin-Chiret, A. S. Hot-Spots of Mcl-1 Protein. *J. Med. Chem.* **2020**, *63* (3), 928–943.
(29) Lee, S.; Wales, T. E.; Escudero, S.; Cohen, D. T.; Luccarelli, J.; Gallagher, C. G.; Cohen, N. A.; Huhn, A. J.; Bird, G. H.; Engen, J. R.; Walensky, L. D. Allosteric Inhibition of Antiapoptotic MCL-1. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 600–607.
(30) Song, T.; Wang, Z.; Ji, F.; Feng, Y.; Fan, Y.; Chai, G.; Li, X.; Li, Z.; Zhang, Z. Deactivation of Mcl-1 by Dual-Function Small-Molecule Inhibitors Targeting the Bcl-2 Homology 3 Domain and Facilitating Mcl-1 Ubiquitination. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (46), 14250–14256.
(31) Le Gouill, S.; Podar, K.; Harousseau, J.-L.; Anderson, K. C. Mcl-1 Regulation and Its Role in Multiple Myeloma. *Cell Cycle* **2004**, *3* (10), 1259–1262.
(32) D'Aguanno, S.; Del Bufalo, D. Inhibition of Anti-Apoptotic Bcl-2 Proteins in Preclinical and Clinical Studies: Current Overview in Cancer. *Cells* **2020**, *9* (5). https://doi.org/10.3390/cells9051287.
(33) Merino, D.; Kelly, G. L.; Lessene, G.; Wei, A. H.; Roberts, A. W.; Strasser, A. BH3-Mimetic Drugs: Blazing the Trail for New Cancer Medicines. *Cancer Cell* **2018**, *34* (6), 879–891.
(34) Nguyen, M.; Marcellus, R. C.; Roulston, A.; Watson, M.; Serfass, L.; Murthy Madiraju, S. R.; Goulet, D.; Viallet, J.; Belec, L.; Billot, X.; Acoca, S.; Purisima, E.; Wiegmans, A.; Cluse, L.; Johnstone, R. W.; Beauparlant, P.; Shore, G. C. Small Molecule Obatoclax (GX15-070) Antagonizes MCL-1 and Overcomes MCL-1-Mediated Resistance to

Apoptosis. *Proceedings of the National Academy of Sciences*. 2007, pp 19512–19517. https://doi.org/10.1073/pnas.0709443104.

(35) Oliver, C. L.; Bauer, J. A.; Wolter, K. G.; Ubell, M. L.; Narayan, A.; O'Connell, K. M.; Fisher, S. G.; Wang, S.; Wu, X.; Ji, M.; Carey, T. E.; Bradford, C. R. In Vitro Effects of the BH3 Mimetic, (-)-Gossypol, on Head and Neck Squamous Cell Carcinoma Cells. *Clin. Cancer Res.* **2004**, *10* (22), 7757–7763.

(36) Paulus, A.; Chitta, K.; Akhtar, S.; Personett, D.; Miller, K. C.; Thompson, K. J.; Carr, J.; Kumar, S.; Roy, V.; Ansell, S. M.; Mikhael, J. R.; Dispenzieri, A.; Reeder, C. B.; Rivera, C. E.; Foran, J.; Chanan-Khan, A. AT-101 Downregulates BCL2 and MCL1 and Potentiates the Cytotoxic Effects of Lenalidomide and Dexamethasone in Preclinical Models of Multiple Myeloma and Waldenström Macroglobulinaemia. *British Journal of Haematology*. 2014, pp 352–365. https://doi.org/10.1111/bjh.12633.

(37) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. An Inhibitor of Bcl-2 Family Proteins Induces Regression of Solid Tumours. *Nature* **2005**, *435* (7042), 677–681.

(38) Del Gaizo Moore, V.; Brown, J. R.; Certo, M.; Love, T. M.; Novina, C. D.; Letai, A. Chronic Lymphocytic Leukemia Requires BCL2 to Sequester Prodeath BIM, Explaining Sensitivity to BCL2 Antagonist ABT-737. *J. Clin. Invest.* **2007**, *117* (1), 112–121.

(39) Suryani, S.; Carol, H.; Chonghaile, T. N.; Frismantas, V.; Sarmah, C.; High, L.; Bornhauser, B.; Cowley, M. J.; Szymanska, B.; Evans, K.; Boehm, I.; Tonna, E.; Jones, L.; Manesh, D. M.; Kurmasheva, R. T.; Billups, C.; Kaplan, W.; Letai, A.; Bourquin, J.-P.; Houghton, P. J.; Smith, M. A.; Lock, R. B. Cell and Molecular Determinants of in Vivo Efficacy of the BH3 Mimetic ABT-263 against Pediatric Acute Lymphoblastic Leukemia Xenografts. *Clin. Cancer Res.* **2014**, *20* (17), 4520–4531.

(40) Schoenwaelder, S. M.; Jarman, K. E.; Gardiner, E. E.; Hua, M.; Qiao, J.; White, M. J.; Josefsson, E. C.; Alwis, I.; Ono, A.; Willcox, A.; Andrews, R. K.; Mason, K. D.; Salem, H. H.; Huang, D. C. S.; Kile, B. T.; Roberts, A. W.; Jackson, S. P. Bcl-xL-Inhibitory BH3 Mimetics Can Induce a Transient Thrombocytopathy That Undermines the Hemostatic Function of Platelets. *Blood* **2011**, *118* (6), 1663–1674.

(41) Khan, S.; Zhang, X.; Lv, D.; Zhang, Q.; He, Y.; Zhang, P.; Liu, X.; Thummuri, D.; Yuan, Y.; Wiegand, J. S.; Pei, J.; Zhang, W.; Sharma, A.; McCurdy, C. R.; Kuruvilla, V. M.; Baran, N.; Ferrando, A. A.; Kim, Y.-M.; Rogojina, A.; Houghton, P. J.; Huang, G.; Hromas, R.; Konopleva, M.; Zheng, G.; Zhou, D. A Selective BCL-XL PROTAC Degrader Achieves Safe and Potent Antitumor Activity. *Nature Medicine*. 2019, pp 1938–1947. https://doi.org/10.1038/s41591-019-0668-z.

(42) Abulwerdi, F.; Liao, C.; Liu, M.; Azmi, A. S.; Aboukameel, A.; Mady, A. S. A.; Gulappa, T.; Cierpicki, T.; Owens, S.; Zhang, T.; Sun, D.; Stuckey, J. A.; Mohammad, R. M.; Nikolovska-Coleska, Z. A Novel Small-Molecule Inhibitor of Mcl-1 Blocks Pancreatic Cancer Growth In Vitro and In Vivo. *Molecular Cancer Therapeutics*. 2014, pp 565–575. https://doi.org/10.1158/1535-7163.mct-12-0767.

(43) Gloaguen, C.; Voisin-Chiret, A. S.; Sopkova-de Oliveira Santos, J.; Fogha, J.; Gautier, F.; De Giorgi, M.; Burzicki, G.; Perato, S.; Pétigny-Lechartier, C.; Simonin-Le Jeune, K.; Brotin, E.; Goux, D.; N'Diaye, M.; Lambert, B.; Louis, M.-H.; Ligat, L.; Lopez, F.; Juin, P.; Bureau, R.; Rault, S.; Poulain, L. First Evidence That Oligopyridines, α-Helix

Foldamers, Inhibit Mcl-1 and Sensitize Ovarian Carcinoma Cells to Bcl-xL-Targeting Strategies. *J. Med. Chem.* **2015**, *58* (4), 1644–1668.
(44) Kotschy, A.; Szlavik, Z.; Murray, J.; Davidson, J.; Maragno, A. L.; Le Toumelin-Braizat, G.; Chanrion, M.; Kelly, G. L.; Gong, J.-N.; Moujalled, D. M.; Bruno, A.; Csekei, M.; Paczal, A.; Szabo, Z. B.; Sipos, S.; Radics, G.; Proszenyak, A.; Balint, B.; Ondi, L.; Blasko, G.; Robertson, A.; Surgenor, A.; Dokurno, P.; Chen, I.; Matassova, N.; Smith, J.; Pedder, C.; Graham, C.; Studeny, A.; Lysiak-Auvity, G.; Girard, A.-M.; Gravé, F.; Segal, D.; Riffkin, C. D.; Pomilio, G.; Galbraith, L. C. A.; Aubrey, B. J.; Brennan, M. S.; Herold, M. J.; Chang, C.; Guasconi, G.; Cauquil, N.; Melchiore, F.; Guigal-Stephan, N.; Lockhart, B.; Colland, F.; Hickman, J. A.; Roberts, A. W.; Huang, D. C. S.; Wei, A. H.; Strasser, A.; Lessene, G.; Geneste, O. The MCL1 Inhibitor S63845 Is Tolerable and Effective in Diverse Cancer Models. *Nature* **2016**, *538* (7626), 477–482.
(45) Pervushin, N. V.; Senichkin, V. V.; Zhivotovsky, B.; Kopeina, G. S. Mcl-1 as a "Barrier" in Cancer Treatment: Can We Target It Now? *Int. Rev. Cell Mol. Biol.* **2020**, *351*, 23–55.
(46) Wang, X.; Bathina, M.; Lynch, J.; Koss, B.; Calabrese, C.; Frase, S.; Schuetz, J. D.; Rehg, J. E.; Opferman, J. T. Deletion of MCL-1 Causes Lethal Cardiac Failure and Mitochondrial Dysfunction. *Genes Dev.* **2013**, *27* (12), 1351–1364.
(47) Tron, A. E.; Belmonte, M. A.; Adam, A.; Aquila, B. M.; Boise, L. H.; Chiarparin, E.; Cidado, J.; Embrey, K. J.; Gangl, E.; Gibbons, F. D.; Gregory, G. P.; Hargreaves, D.; Hendricks, J. A.; Johannes, J. W.; Johnstone, R. W.; Kazmirski, S. L.; Kettle, J. G.; Lamb, M. L.; Matulis, S. M.; Nooka, A. K.; Packer, M. J.; Peng, B.; Rawlins, P. B.; Robbins, D. W.; Schuller, A. G.; Su, N.; Yang, W.; Ye, Q.; Zheng, X.; Secrist, J. P.; Clark, E. A.; Wilson, D. M.; Fawell, S. E.; Hird, A. W. Discovery of Mcl-1-Specific Inhibitor AZD5991 and Preclinical Activity in Multiple Myeloma and Acute Myeloid Leukemia. *Nat. Commun.* **2018**, *9* (1), 5341.
(48) Yonath, A. X-Ray Crystallography at the Heart of Life Science. *Curr. Opin. Struct. Biol.* **2011**, *21* (5), 622–626.
(49) Bax, A.; Clore, G. M. Protein NMR: Boundless Opportunities. *J. Magn. Reson.* **2019**, *306*, 187–191.
(50) Murata, K.; Wolf, M. Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules. *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862* (2), 324–334.
(51) Šrajer, V.; Schmidt, M. Watching Proteins Function with Time-Resolved X-Ray Crystallography. *J. Phys. D Appl. Phys.* **2017**, *50* (37). https://doi.org/10.1088/1361-6463/aa7d32.
(52) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585–590.
(53) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1* (3), 173–185.
(54) Onufriev, A. V.; Izadi, S. Water Models for Biomolecular Simulations. *WIREs Comput Mol Sci* **2018**, *8* (2), e1347.
(55) Steinbrecher, T.; Elstner, M. QM and QM/MM Simulations of Proteins. *Methods Mol. Biol.* **2013**, *924*, 91–124.
(56) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936.

(57) Noé, F. Machine Learning for Molecular Dynamics on Long Timescales. *arXiv [physics.chem-ph]*, 2018.
(58) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78* (20), 3908–3911.
(59) Voter, A. F. A Method for Accelerating the Molecular Dynamics Simulation of Infrequent Events. *J. Chem. Phys.* **1997**, *106* (11), 4665–4677.
(60) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929.
(61) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254* (5038), 1598–1603.
(62) Frauenfelder, H.; Leeson, D. T. The Energy Landscape in Non-Biological and Biological Molecules. *Nat. Struct. Biol.* **1998**, *5* (9), 757–759.
(63) Chandler, D. *Introduction to Modern Statistical Mechanics*; ui.adsabs.harvard.edu, 1987.
(64) Allison, J. R. Computational Methods for Exploring Protein Conformations. *Biochem. Soc. Trans.* **2020**, *48* (4), 1707–1724.
(65) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.
(66) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16* (11), 1339–1350.
(67) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.
(68) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
(69) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the Free-Energy Landscapes of Proteins by NMR-Guided Metadynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (17), 6817–6822.
(70) Sittel, F.; Jain, A.; Stock, G. Principal Component Analysis of Molecular Dynamics: On the Use of Cartesian vs. Internal Coordinates. *J. Chem. Phys.* **2014**, *141* (1), 014111.
(71) Brooks, B.; Karplus, M. Harmonic Dynamics of Proteins: Normal Modes and Fluctuations in Bovine Pancreatic Trypsin Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80* (21), 6571–6575.

CHAPITRE 2.

Mode d’interaction entre le Pyridoclax et Mcl-1

1. Motivation

Mcl-1 (Myeloid Cell Leukemia-1) est une protéine anti-apoptotique, membre de la famille des protéines Bcl-2. Cette protéine se retrouve surexprimée dans un ensemble de cancers.

Le pyridoclax est un inhibiteur BH3-mimétique, qui interagit avec Mcl-1 et restaure l'apoptose dans des lignées de tumeurs ovariennes résistantes à des traitements existants.

Dans cette étude, des données expérimentales sont complémentées par des simulations de dynamique moléculaire afin de proposer un mode d'interaction entre le Pyridoclax et Mcl-1.

2. Article 1

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
2020, VOL. 38, NO. 14, 4162–4178
https://doi.org/10.1080/07391102.2019.1680434

Binding mode of Pyridoclax to myeloid cell leukemia-1 (Mcl-1) revealed by nuclear magnetic resonance spectroscopy, docking and molecular dynamics approaches

A. Bourafai-Aziez[a,b], M. Sebban[a], M. Benabderrahmane[b], B. Marekha[b], C. Denis[b], H. Paysant[c,d], L. B. Weiswald[c,d], L. Carlier[e], R. Bureau[b], G. Coadou[a], D. Ravault[e], A. S. Voisin-Chiret[b], J. Sopková-de Oliveira Santos[b] and H. Oulyadi[a]

[a]CNRS Laboratoire COBRA (UMR 6014 & FR 3038), Normandie Université, UNIROUEN, INSA de Rouen, Rouen, France; [b]Normandie Université, UniCaen, CERMN, F-14000 Caen, France; [c]Normandie Université, UNICAEN, Inserm U1086 ANTICIPE « Interdisciplinary Research Unit for Cancer Prevention and Treatment », Biologie et Thérapies Innovantes des Cancers de l'ovaire (BioTICLA), Caen, France; [d]Centre de Lutte Contre le Cancer F. Baclesse, Unicancer, Caen, France; [e]Laboratoire Des Biomolécules, LBM, Sorbonne Université, École Normale Supérieure, PSL University, CNRS, Paris, France

Communicated by Ramaswamy H. Sarma

ABSTRACT

Myeloid cell leukemia-1 (Mcl-1) is an anti-apoptotic member of the Bcl-2 family proteins. Its amplification is one of the most frequent genetic aberrations found in human cancers. Pyridoclax, a promising BH3 mimetic inhibitor, interacts directly with Mcl-1 and induces massive apoptosis at a concentration of 15 µM in combination with anti-Bcl-x_L strategies in chemo-resistant ovarian cancer cell lines. In this study, a combined experimental and theoretical approach was used to investigate the binding mode of Pyridoclax to Mcl-1. The representative poses generated from dynamics simulations compared with NMR data revealed: (i) Pyridoclax bound to P1 and P2 pockets of Mcl-1 BH3 binding groove through its styryl and methyl groups establishing mainly hydrophobic contacts, (ii) one of the ending pyridines interacts through electrostatic interaction with K234 side chain, a negatively charged residue present only in this position in Mcl-1.

ARTICLE HISTORY
Received 9 July 2019
Accepted 26 September 2019

KEYWORDS
Pyridoclax; Mcl-1; protein-ligand interaction; NMR spectroscopy; docking; molecular dynamics

Abbreviations: ATCC: American Type Culture Collection; Bcl-2: B-cell lymphoma 2; Bcl-xL: B-cell lymphoma-extra large; BH: Bcl-2 Homology; CCPNMR: Collaborative Computational Project for NMR; CCSP: Combined Chemical Shift Perturbations; CGENFF: CHARMM GENeral Force Field; CHARMM: Chemistry At Harvard Molecular Mechanics; DAPI: 4′,6′-diamidino-2-phenylindole; DMSO-d6: Deuterated Dimethylsulfoxide; DTT: Dithiothreitol; EDTA: Ethylene-Diamine-Tetra-Acetic acid; GOLD: Genetic Optimization for Ligand Docking; HEPES: 4-(2-HydroxyEthyl)-1-PiperazineEthaneSulfonic acid; HSQC: Heteronuclear Single Quantum Correlation; IMAC: Immobilized Metal Ion Affinity Chromatography; IPTG: Isopropyl-Beta-D-Thiogalactoside; KD: Dissociation Constant; LP: Linear Prediction; MBP: Maltose-Binding Protein; Mcl-1: Myeloid cell leukemia-1; NMR: Nuclear Magnetic Resonance; NOE: Nuclear Overhauser Effect; NPT: Number, Pressure, and Temperature; NVT: Number, Volume, and Temperature; PARP: Poly- Adenosine-DiPhosphate-Ribose-Polymérase; PBS: Phosphate-Buffer Saline; PDB: Protein Data Bank; PME: Particle Mesh Ewald; PMSF: PhenylMethaneSulfonyl Fluoride; PVDF: PolyVinylidene DiFluoride; RMSD: Root-Mean-Square Deviation; RPMI: Roswell Park Memorial Institute; RT: Room Temperature; SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis; SOFAST: band-Selective Optimized Flip Angle Short Transient; HMQC: Heteronuclear Multiple Quantum Coherence; TBS: Tris Buffered Saline; TITAN: TITration ANalysis; TPPI: Time Proportional Phase Incrementation

Introduction

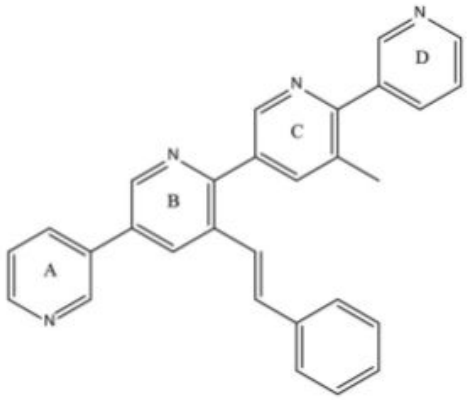
Apoptosis, an evolutionarily conserved process of programmed cell death, was first defined in the early 1970s (Kerr, Wyllie, & Currie, 1972) and plays a pivotal role in the regulation of cell proliferation. Avoidance of apoptosis is critical for the development and sustained expansion of tumors and underpins resistance to diverse anti-cancer treatments (Hanahan & Weinberg, 2011).

The regulation of apoptosis among others depends on the Bcl-2 family proteins, which contain one or more Bcl-2 homology (BH) sequence motifs. Based on their functions these proteins can be divided into three classes (Gelinas & White, 2005; Shore & Warr, 2008):

1. Anti-apoptotic (i.e., pro-survival) multi-domain proteins, such as Mcl-1, Bcl-2, Bcl-x_L, Bcl-w, and Bfl-1/A1;

CONTACT M. Sebban muriel.sebban@univ-rouen.fr Normandie Université, UNIROUEN, INSA de Rouen, CNRS Laboratoire COBRA (UMR 6014 & FR 3038), 76000 Rouen, France; J. Sopková-de Oliveira Santos jana.sopkova@unicaen.fr Normandie Université, UNICAEN, CERMN, FR CNRS 3038 INC3M, SF 4206 ICORE bd Becquerel, F-14000 Caen, France.

Supplemental data for this article is available online at https://doi.org/10.1080/07391102.2019.1680434.



Scheme 1. Structure of Pyridoclax.

2. Pro-apoptotic multi-domain proteins such as Bax, Bak and Bok, which exhibit a similar architecture as multi-domain pro-apoptotic proteins;
3. Pro-apoptotic proteins comprising only a single BH sequence motif (BH3), such as Bid, Bad, Bim, Puma, Noxa, Hrk, Bmf and Nbk/Bik ('BH3-only' proteins).

The Bcl-2 family members interact with each other *via* their BH3 domain. These protein-protein interactions control commitment to apoptosis (Petros et al., 2000).

In many cancers, the balance between the pro- and anti-apoptotic Bcl-2 family members is tipped towards survival as a consequence of genetic or epigenetic changes (Delbridge, Grabow, Strasser, & Vaux, 2016); Mcl-1 is among the most frequently amplified genes in human cancers (Beroukhim et al., 2010). Therefore, it is considered as a high priority target for research on anti-cancer therapy. Although development of small-molecule inhibitors directed against Mcl-1 has proved challenging, several Mcl-1 inhibitors have been reported recently (Abid, Sonawane, Contreras, Rana, & Natarajan, 2017; Abulwerdi, Liao, Liu, et al., 2014; Abulwerdi, Liao, Mady, et al., 2014; Ashkenazi, Fairbrother, Leverson, & Souers, 2017; Beekman & Howell, 2016; Belmar & Fesik, 2015; Caenepeel et al., 2018; Kotschy et al., 2016; Petros et al., 2014; Tron et al., 2018; Wan, Dai, Tang, & Fang, 2018).

In this context, our teams designed and synthesized small compounds based on a pyridyl scaffold, called oligopyridines, which target Mcl-1. These foldamers can adopt a well-adapted spatial conformation to mimic desired side chains on the amphipathic helix of the BH3-only proteins, thus inhibiting the overexpressed Mcl-1 protein in ovarian cancer. Previously, we demonstrated that the leading first-generation oligopyridine, Pyridoclax (Scheme 1), interacts directly with Mcl-1, releases the pro-apoptotic partners Bim and Bak, and induces massive apoptosis at a concentration of 15 µM in combination with anti-Bcl-x_L strategies in chemo-resistant ovarian cancer cell lines (Gloaguen et al., 2015).

Characterizing the structural aspects of Pyridoclax binding to Mcl-1 is an important step to understand its interaction mode that should enable the development of new drugs. Nuclear Magnetic Resonance (NMR) spectroscopy has evolved into a powerful tool for characterizing protein-ligand interactions in solution under near physiological conditions. It is now frequently used to assess the affinity and specificity of interactions, to identify binding epitopes on proteins and ligands, and to characterize structural rearrangements induced by ligand binding (Cala, Guillière, & Krimm, 2014; Meyer & Peters, 2003; Orts & Gossert, 2018). NMR data can also be exploited to validate the ligand poses generated by molecular modeling approaches.

In this work, a combined experimental and theoretical study, based on multidimensional NMR spectroscopy and molecular modeling, was carried out to characterize the binding modes of Pyridoclax to Mcl-1 and to propose a representative model of the complex Mcl-1:Pyridoclax with the ultimate aim of rationalizing the design of new Mcl-1 inhibitors.

Material and methods

Biological assays

Cell culture

The human ovarian carcinoma cell lines SKOV-3 was purchased from ATCC (Manassas, VA, USA) and grown in RPMI (Roswell Park Memorial Institute) 1640 medium supplemented with 2 mM GlutamaxTM, 25 mM HEPES, 10% fetal calf serum, and 33 mM sodium bicarbonate (Fisher Scientific, Illkirch, France).

Treatments

Cells were seeded in 25 cm^2 flasks. After 24 h, exponentially growing cells were exposed to Pyridoclax as a single agent or in combination with ABT-737 for 24 h at the indicated concentrations.

Morphological characterization of apoptotic cells by nuclear staining with DAPI

After treatment, both detached and adherent cells were pooled after trypsinization, applied to a polylysine-coated glass slide by cytocentrifugation and fixed with a solution of ethanol/chloroform/acetic acid (6:3:1). The preparations were then incubated for 15 min at room temperature with 1 µg/ml DAPI (4',6'-diamidino-2-phenylindole) solution (Boehringer Mannheim-Roche, Mannheim, Germany), washed in distillated water, mounted under a coverslip in Mowiol (Calbiochem) and analyzed under a fluorescence microscope (BX51, Olympus, Rungis, France).

Western blot analysis of PARP (poly-ADP-ribose-polymérase-1) and caspase-3 cleavage

Cells were rinsed with ice-cold PBS, suspended in a lysis buffer [RIPA: NaCl 150 mM, Tris (pH–8) 50 mM, Triton X100 1%, PMSF 4 mM, EDTA 5 mM, NaF 10 mM, NaPPi 10 mM, Na3OV4 1 mM, aprotinin 0.5 µL/mL and 4.6 mL ultra-pure water] and incubated on ice for 30 min. Lysates were collected after centrifugation (13,200 xg, 10 min, 4 °C) and protein concentrations were determined using the Bradford assay (Bio-Rad, Hercules, USA).

25 µg of protein were separated by SDS–PAGE on a 4%–12% gradient polyacrylamide gel (Invitrogen, Cergy-Pontoise, France) and transferred to Hybond-PVDF membranes (Amersham, Orsay, France). After blocking non-specific binding sites for 1 h at RT by 5% (w/v) non-fat dry milk in TBS with 0.1% (v/v) Tween 20 (T-TBS), the membranes were incubated overnight at 4 °C with the PARP and caspase-3 rabbit monoclonal antibody (Cell Signaling Technology, Ozyme, Saint-Quentin-en-Yvelines, France) or Actin mouse monoclonal antibody (EMD, Millipore, France). Membranes were then washed with T-TBS and incubated for 1 h with the appropriate horseradish peroxidase-conjugated anti-rabbit (Cell Signaling Technology, France) or anti-mouse (Amersham, Orsay, France) secondary antibodies. The revelation was done using a luminescent Image Analyzer (GE Healthcare, Orsay, France).

Recombinant expression and purification of mcl-1 for NMR studies

Soluble Mcl-1 (172-327) was produced as a maltose-binding protein (MBP) fusion protein in *E. coli*, as proposed by Friberg et *al.* (Friberg et al., 2013). The gene encoding the region 172-327 of human Mcl-1 (Uniprot accession number Q07820) was synthesized and cloned into a modified pET-15b plasmid (Genscript). The resulting expression vector encodes a 60 kDa fusion protein that comprises an amino-terminal (His)6-tag (MBP), a PreScission protease cleavage site, and the Mcl-1 fragment (18 kDa). The pET-15b plasmid was transformed into an E. coli BL21-CodonPlus (DE3)-RIL expression strain (Agilent).

The recombinant protein was prepared from cells grown to an OD600 of 0.8 at 37 °C in a bacterial culture supplemented by 100 µg/mL ampicillin and 35 µg/mL chloramphenicol. Cells were cultured in either unlabeled LB-rich medium or uniformly labeled ^{15}N or $^{15}N/^{13}C$ minimum media. Protein expression was induced by the addition of 0.25 mM IPTG, and the culture was further grown for 16 h at 20 °C. The cells were pelleted by centrifugation (3000 xg for 25 min), resuspended in a 50 mM Tris-HCl buffer (pH 7.8) containing 500 mM NaCl and 1 mM DTT, and lysed by sonication. The soluble fraction containing the Mcl-1 fragment was separated from cell debris by centrifugation (30000 xg for 40 min at 4 °C) and loaded on a Ni-NTA column (GE Healthcare). The fusion protein was purified using IMAC standard protocols. Eluted fractions were pooled and dialyzed against PBS buffer, and recombinant (His)6-PreScission was added to cleave the MBP fusion partner. A second step of purification was performed with the Ni-NTA column to remove (His)6-PreScission and the cleaved (His)6-MBP partner. A third step of purification was finally carried out using size exclusion chromatography on a Superdex 75 16/60 column (GE Healthcare). The produced fragment (172-327) of Mcl-1 protein will be called Mcl-1 in the following paper.

NMR experiments

Pyridoclax was synthesized following a previously published protocol (Gloaguen et al., 2015).

Samples preparation for NMR studies

Different samples were prepared depending on the experiment to be performed.

To investigate the conformational properties of Pyridoclax, a sample containing 18 mM Pyridoclax solubilized in DMSO-d_6 was prepared.

For Mcl-1 backbone assignment, 600 µL of uniformly ^{15}N-^{13}C labeled Mcl-1 (250 µM) was prepared in PBS buffer (20 mM, 130 mM NaCl, 2 mM DTT, pH 7) containing 5% D_2O (*Sample A*). For titration experiments, Pyridoclax was dissolved in pure DMSO-d_6 to obtain a 5 mM stock solution. DMSO-d_6 was used because of the poor solubility of Pyridoclax in PBS buffer. This co-solvent was chosen in reference to the literature (Friberg et al., 2013; Song et al., 2016). Next, two samples were prepared to carry out the titration experiments: *Sample B* contained 570 µL of a solution of uniformly ^{15}N labeled Mcl-1 (final concentration: 50 µM) prepared in PBS buffer and 30 µL DMSO-d_6 (5%, v/v); *Sample C* contained 570 µL of a solution of uniformly ^{15}N labeled Mcl-1 (final concentration: 50 µM) prepared in PBS buffer with 30 µL of Pyridoclax stock solution prepared in DMSO-d_6 (final concentration of Pyridoclax: 250 µM). The titration series was incremented in steps of 0.5 molar equivalents of the ligand. To do this, adequate volumes of each sample (*B and C*) were mixed to obtain the following ratios L/P_0 (0/1 – 0.5/1 – 1/1 – 1.5/1 – 2/1 – 2.5/1 – 3/1 – 3.5/1 – 4/1 – 4.5/1 – 5/1).

To highlight the effect of the interaction on Mcl-1 side-chains, two samples with ^{13}C-^{15}N labeled Mcl-1 were prepared: *Sample D* contained 570 µL of ^{13}C-^{15}N labeled Mcl-1 (final concentration, 190 µM) was prepared in PBS buffer with 30 µL DMSO-d_6 (5%, v/v); *Sample E* contained 570 µL of uniformly ^{13}C-^{15}N labeled Mcl-1 (final concentration: 190 µM) was prepared in PBS buffer with 30 µL of Pyridoclax stock solution prepared in DMSO-d_6 (final concentration of Pyridoclax: 250 µM).

NMR measurements

NMR spectra were recorded at 298 K (for Mcl-1 and the complex) and at 313 K (for Pyridoclax in DMSO-d_6) on a Bruker Avance III 600 MHz spectrometer equipped with a triple resonance cryoprobe {^1H, ^{13}C, ^{15}N} including shielded z-gradients.

To perform the conformational study of Pyridoclax, a 2 D ^1H-^1H NOESY experiment was acquired with 512 indirect points, 2048 direct points (spectral width of 12 ppm in both dimensions), 8 scans, a recycle time of 3 s, and a mixing time of 1 s.

The backbone assignment of Mcl-1 resonances was carried out in reference to a previous work by Liu et *al.* (Liu et al., 2014) using a minimal set of triple resonance experiments: (HNCA, HN(CO)CA (Grzesiek & Bax, 1992) and CBCA(CO)NH (Grzesiek & Bax, 1993)). HN(CO)CA and CBCA(CO)NH spectra were collected as 2048 (F3, ^1H), 80 (F2, ^{15}N), and 128 (F1, ^{13}C) complex data points with 24 transients per F3 and F2 increments and a recycle delay of 1 sec. The spectral widths were 14 ppm in F3 (carrier frequency at 4,70 ppm), 32 ppm in F2 (carrier frequency at 119,5 ppm) and 32, 75 ppm, respectively, in F1 (carrier frequency at 54 and 43 ppm, respectively). The same parameters were used for the recording of the HNCA experiment except

the complex data points in F2 (^{15}N, 48), the spectra width in F2 (^{15}N, 40 ppm, carrier frequency at 51 ppm) and the number of scans (16 scans). Prior to Fourier transformation, the data were zero filled in F2 (256 complex points for HN(CO)CA and CBCA(CO)NH and 128 for HNCA) and F1 (512 complex points for HN(CO)CA and CBCA(CO)NH and 256 for HNCA). For HN(CO)CA and CBCA(CO)NH experiments, linear prediction (LP) was also applied in the indirect dimensions (F2, F1) (Forward LP on a complex data mode, Number of LP Coefficients, and Number of points for LP were set to 32). The time domain data were apodized by Sine Squared function (F1, F2 and F3) for the three 3 D experiments. Obtained spectra were analyzed using Sparky for signal assignment (Lee, Tonelli, & Markley, 2015).

The binding mode of Pyridoclax was characterized by recording 2D ^{1}H-^{15}N HSQC experiments (Mori, Abeygunawardana, Johnson, & Vanzijl, 1995) for each ratio L/P_0 described above.

2D ^{1}H-^{15}N HSQC spectra were collected with 128 indirect points, 1152 direct points using spectral width of 40 ppm (F2, ^{15}N) and 12 ppm (F1, ^{1}H), respectively, and 64 scans. A one-second relaxation delay was used, and quadrature detection in the indirect dimension was obtained with states-time proportional phase incrementation (TPPI) phase cycling (Marion & Wüthrich, 1983). Before Fourier transformation, linear predictions (LP) were applied in the indirect dimension (Forward LP on a complex data mode, Number of LP Coefficients, and Number of points for LP were set to 64 and 32, respectively). The time domain data were zero filled (size of the real spectrum was 2048x512 points) and apodized by Sine Squared function. Data were acquired and processed using Topspin 3.5 software (Bruker Biospin). Protein spectra were analyzed using Sparky for signal assignment (Lee et al., 2015), CCPNMR (Skinner et al., 2016) and TITAN (Waudby, Ramos, Cabrita, & Christodoulou, 2016) for the estimation of K_D and k_{off}.

For a given ^{1}H-^{15}N HSQC cross peak, proton and nitrogen chemical shift perturbations induced by Pyridoclax binding was defined as the combined difference between the corresponding chemical shifts in the bound and free states: the combined chemical shift perturbations (CCSP) were calculated from the ^{1}H and ^{15}N shift coordinates of the cross peaks using equation (1):

$$\mathrm{CCSP(ppm)} = \left[\Delta\delta H^2 + (0.2\Delta\delta N)^2\right]^{1/2} \quad (1)$$

where $\Delta\delta$H and $\Delta\delta$N are the observed chemical shifts differences along the proton and nitrogen dimensions, respectively, with respect to the Mcl-1 free ^{1}H-^{15}N HSQC spectrum.

The dissociation constant (K_D) was then obtained by monitoring the chemical shift changes of the backbone amide as a function of ligand concentration using a binding model with 1:1 stoichiometry (Williamson, 2013) in CCPNMR software (Skinner et al., 2016) (equation 2).

$$\Delta\delta_{obs} = \frac{\Delta\delta_{max}}{2}\left[1 + x + \frac{K_D}{P_0} - \sqrt{\left(1 + x + \frac{K_D}{P_0}\right)^2 - 4x}\right] \quad (2)$$

where $\Delta\delta_{obs}$ is the change in the observed shift from the free state and fitted $\Delta\delta_{max}$ is the maximum change in chemical shifts for the given peak, x is the molar L/P_0 ratio and P_0 is the total concentration of the labeled protein.

^{1}H-^{15}N HSQC spectra were also analyzed by the TITAN program (Waudby et al., 2016) using a two-state binding mode. Mcl-1 and Pyridoclax concentrations and NMR acquisition parameters were provided as inputs for the data fitting. Ten peaks with significant CCSPs were chosen to be fitted for each titration experiment. Error analysis was performed by the bootstrap re-sampling module in the program with 100 re-sampled spectra.

The effect of the interaction on Mcl-1 methyl groups was studied by recording 2D ^{1}H-^{13}C SOFAST-HMQC experiments (Schanda & Brutscher, 2005) using Samples D and E. 2D ^{1}H-^{13}C SOFAST-HMQC spectra were recorded with 256 indirect points, 772 direct points and 128 scans. A 0.2 s relaxation delay was used. The band selective ^{1}H excitation was centered at 0 ppm covering a bandwidth of 4 ppm. A quadrature detection in the indirect dimension was obtained with states-time proportional phase incrementation (TPPI) phase cycling (Marion & Wüthrich, 1983). The time domain data were zero filled (size of the real spectrum was 2048x512 points) and apodized by Sine Squared function. Data were acquired and processed using Topspin 3.5 software (Bruker Biospin).

Molecular modeling

NMR Pyridoclax solution structure

Pyridoclax solution structure was generated by dynamics simulations using CHARMM software with the all-atom force field parameters generated by CGENFF, (Vanommeslaeghe et al., 2009; Yu, He, Vanommeslaeghe, & MacKerell, 2012) during which measured NOE (distance) restraints were applied with a force constant of 25 kcal/mol. Starting from the energy minimized Pyridoclax X-ray structure (CCDC n°996607 - Gloaguen et al., 2015), an NVT dynamics simulation of 50 ns was carried out for each derivative at 300 K, with a time step of 1 fs. Van der Waals interactions were truncated using a shift function with a cut-off distance of 12 Å and electrostatic ones using a force switching function, between 8 and 12 Å. The dynamics were preceded by heating (during 6 ps) and equilibration (50 ps) step. During the production phase (50 ns), conformations were saved every 10 ps and energy minimized to a root-mean-square gradient of less than 0.001 kcal/(mol·Å^2). The 10 lowest-energy conformations obtained for Pyridoclax were used in the subsequent analysis.

Mcl-1 starting model

During the preliminary analysis of various Mcl-1 complexes co-crystallized with peptide ligands or with synthetic ligands, we have observed that the wide opening of the main hydrophobic pockets of the BH3 binding groove, especially P2, depended on the fixed ligands. For example, in the structure 5FC4 (Pelz et al., 2016) or 6B4U (Bruncko et al., 2015), the hydrophobic pocket P2 was opened while in the structure 5UUM (Jenson, Ryan, Grant, Letai, & Keating, 2017) it was

closed. Furthermore, various mutations in Mcl-1 sequence and flexible loop deletion were carried out to facilitate the Mcl-1 co-crystallization with synthetic ligands. For these reasons, a representative model of Mcl-1 with a relatively open pocket P2 was built for this study. The starting human Mcl-1 model (UniProtID: Q07820) was built by the @Tome2 server (Pons & Labesse, 2009) using four X-ray Mcl-1 structures (PDB ID: 5FC4 (Pelz et al., 2016), 4WMS (Clifton et al., 2015), 4ZBI (Burke et al., 2015) and 3WIX (Tanaka et al., 2013)) as templates. A missing loop was built and refined using the Modeler engine from the @Tome2 server backend pipeline.

Docking

Docking of Pyridoclax into the Mcl-1 groove highlighted by NMR was carried out using the GOLD program (v5.5) with the default parameters (Jones, Willett, & Glen, 1995; Jones, Willett, Glen, Leach, & Taylor, 1997). This program applies a genetic algorithm to explore conformational spaces and ligand binding modes. To evaluate the proposed ligand poses the ChemPLP fitness function was applied in the docking studies. The binding site in Mcl-1 was defined by the residues lying in a 14 Å sphere around the Cζ atom of Phe270 in the Mcl-1 model. The two conformers of Pyridoclax from its solved X-ray structure and its NMR representative conformation were used in the docking study (Gloaguen et al., 2015).

Molecular dynamics simulations

All simulations were carried out using NAMD 2.12 (Phillips et al., 2005) with the all-atom CHARMM 36 m force field for protein (Best et al., 2012; Huang et al., 2017) and CGENFF for Pyridoclax (Vanommeslaeghe et al., 2009; Yu et al., 2012). The parametrization of the torsion angles between the Pyridoclax pyridine rings was modified according to our previous published work (Sopkova-de Oliveira Santos et al., 2012) (SI - Figure S1). To simulate the aqueous solvent environment, each system was surrounded by a rectangular box of TIP3P water molecules (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) and 0.15 M of KCl was added to the system using CHARMmGUIsolvator (Jo, Kim, Iyer, & Im, 2008). The chosen box size ensured, for each complex, that the simulated complex was at a minimum distance of 10 Å from the edge. Periodic boundary conditions were applied to the systems using the IMAGE algorithm. Van der Waals interactions were truncated using a force switching function between 10 and 12 Å and the Particle Mesh Ewald (PME) (Darden, York, & Pedersen, 1993) was used to calculate long-range electrostatic interactions. The SHAKE algorithm was applied to restrain all bonds involving hydrogen atoms (Ryckaert, Ciccotti, & Berendsen, 1977). The vacuum dielectric constant was used during all calculations. Firstly, the systems underwent energy minimization in 10000 steps. Next, the minimized systems were heated to 303.15 K and the dynamics were temperature-equilibrated during 50 ps via heating reassignment under NVT conditions. Finally, the systems ran freely for 100 ns under NPT conditions. Langevin dynamics with a damping coefficient of 1 ps^{-1} were used to maintain the system temperature and the Nosé-Hover Langevin piston method was used to control the pressure at 1 atm. Production trajectories were saved every 2 ps and subsequent analyzes were performed using the CHARMM program version c40b2 (Brooks et al., 2009).

In order to analyze and resume the information within our trajectories, a clustering strategy was carried out using gromos algorithm (Daura et al., 1999) implemented within the Gromacs 5.1.2 *cluster* tool (Berendsen, van der Spoel, & van Drunen, 1995). The algorithm counted the number of neighbors using a cut-off, took the structure with largest number of neighbors with all its neighbors as the cluster and eliminated it from the pool of clusters. This scheme was repeated for the remaining structures in the pool. The trajectories were fitted on the backbone and the RMSD was computed for the protein-ligand index group. A set of cut-offs was tested for each trajectory and the most suitable cut-offs were chosen: 2.2 Å for pose 1 and 6 Å for pose 2.

Result and discussion

Definition of the hydrophobic pockets corresponding to the binding site on the Mcl-1 surface

The solved 3 D structures of Mcl-1/BH3 complexes showed that the BH3 domain, corresponding to an amphipathic helix, lay in a hydrophobic groove formed by the BH1, BH2, and BH3 motifs of Mcl-1 (Jenson et al., 2017). More precisely, the four conserved hydrophobic residues at positions i, $i+4$, $i+7$, and $i+11$ on the contact face of the BH3 helix are projected into the hydrophobic cavities of the Mcl-1 binding groove (Figure 1A). The receiving cavities of these four hydrophobic residues have been named P1(i), P2 ($i+4$), P3 ($i+7$) and P4 ($i+11$), respectively, but the precise delimitation of these pockets varies according to authors and therefore remains ambiguous. The lack of consensus on the definition of the pockets led us to adopt our own definition based on the following consideration:

Five structures of the human Mcl-1 protein in a complex with the BH3 domain of different pro-apoptotic proteins, available in the PDB data bank (SI - Table 1) were considered (Berman et al., 2000). These structures were mainly solved by X-ray diffraction techniques (Czabotar et al., 2011; Fire, Gulla, Grant, & Keating, 2010; Liu et al., 2010; Miles et al., 2016; Stewart, Fire, Keating, & Walensky, 2010). Using PyMol software, the analysis of the contact interface between Mcl-1 and the BH3 partners was carried out to identify the residues of Mcl-1 close to the key BH3 hydrophobic residues (i, $i+4$, $i+7$ and $i+11$) (Figure 1A). This analysis showed that the residues forming the Mcl-1 hydrophobic pockets could be defined as presented in Figure 1B.

Interaction between Pyridoclax and Mcl-1 revealed by NMR

NMR spectroscopy is a powerful technique to map the interacting site of a protein upon complexation with its partner. ^{1}H-^{15}N HSQC spectrum is a "fingerprint" of the protein because each peak corresponds to an amide function (NH) of each residue.

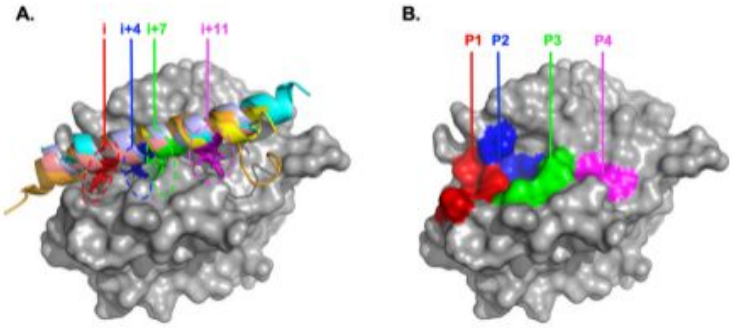


Figure 1. A. hMcl-1/BH3-domains complexes (Mcl-1 PDB ID-3MK8: yellow; hBax PDB ID-3PK1: cyan; hBid PDB ID-5C3F: light blue; hBid PDB ID-2KBW: bright orange and Bim PDB ID-2PQK: pink) pointing key residues for each BH3 peptide (*i*, *i* + 4, *i* + 7 and *i* + 11) (SI - Table 1). B. Determination of Mcl-1 hydrophobic pockets by analyzing the contact area between Mcl-1 and the key residues of BH3 domains (**P1**: L235, K234, V249, M231; **P2**: M250, F270, L267, V253; **P3**: A227, F228, H224, T266; **P4**: V216, V220, V265).

Table 1. The calculated mean interaction energy between Pyridoclax and Mcl-1.

	Mean interaction energy all trajectory [kcal/mol]	Mean interaction energy last 30 ns [kcal/mol]
Pose 1	−31.1 ± 5.4	−31.0 ± 5.0
Pose 2	−29.5 ± 3.7	−29.8 ± 2.5

The chemical shifts in both the nitrogen and proton dimensions are sensitive to the chemical environment of both nuclei.

An NMR study was carried out to confirm the binding of Pyridoclax to the Mcl-1 BH3 binding groove and to validate the predicted binding site (Gloaguen et al., 2015). For that purpose, we first expressed and purified uniformly labeled ^{15}N- and $^{15}N/^{13}C$ protein samples. The backbone assignment of Mcl-1 region 172-327 in the apo form was based on the previously reported assignments of Liu *et al.* (Liu et al., 2014), which were confirmed by the analysis of a series of three-dimensional triple resonance experiments. The backbone assignment reached 96% for $^{1}H_N$ and $^{15}N_H$, 96% for $^{13}C\alpha$, and 92% for $^{13}C\beta$ resonances. Evidence for direct binding of Pyridoclax to Mcl-1 was subsequently provided by 2D ^{1}H-^{15}N HSQC experiments of ^{15}N-Mcl-1 in the presence of stoichiometric amounts of Pyridoclax (SI - Figure S2).

The addition of Pyridoclax induced selective backbone amide chemical shifts changes (residues depicted in Figure 2A). Shifting of correlations in the ^{1}H-^{15}N HSQC of the complex compared to the ^{1}H-^{15}N HSQC of the apo protein reflect some changes in the environment of the corresponding residues and may be related either to a direct interaction between Pyridoclax and residues at the Mcl-1 interface or to conformational changes in the protein induced by the binding to Pyridoclax.

Combined chemical shift perturbations (CCSP) were calculated using equation (1). A plot of CCSPs values for each Mcl-1 residue was depicted in Figure 2A. The significance threshold for backbone amide chemical shift changes was calculated by the standard deviation, in accordance with standard methods (Marintchev, Frueh, & Wagner, 2007).

The chemical shifts changes upon the titration are relatively small, with an average chemical shift perturbation value for Mcl-1 residues not exceeding 0.08 ppm (Figure 2A), which is in accordance with CCSP values reported in the literature for other potential inhibitors of Mcl-1 (Abulwerdi et al., 2014; Cohen et al., 2012;). Of note, Pyridoclax has five aromatic rings expected to lead to ring-current effects prone to enhance CCSP values. The significant dynamics of Pyridoclax in the binding groove of Mcl-1 probably average these effects.

To define the interaction interface between the two partners, highly perturbed (CCSP > 2 standard deviation) and moderately perturbed (1 standard deviation < CCSP < 2 standard deviation) residues were mapped on the surface of Mcl-1 (Figure 2B and C). These residues formed a large contact zone centered mainly on pockets P2 (Met250, Phe270, Leu267 and Val253) and P1 (Lys234, Val249 and Met231), suggesting that the interaction site of Pyridoclax is centered on these two hydrophobic pockets.

Significant perturbations were also observed for residues located outside P1 and P2 pockets, in particular in the P4 pocket that is formed by the C-terminal part of helix α2. These additional perturbations may be explained by a direct interaction with Pyridoclax or by conformational fluctuation of the protein during the interaction with Pyridoclax, probably including movements of the protein's skeleton and reorientation of its side chains, especially for residues located in helix α2 (Arg215, Val216 (P4), Gly219, Val220 (P4)) and helix α4 (Val243, Ser247, His252). In this respect, Sankararamakrishnan *et al.* showed that helix α2 loses its helical structure during dynamics simulations of Mcl-1:BH3 peptide complexes due to the presence of three glycine's on this helix, and the absence of stabilizing interactions between helix α2 and the other helices (Modi & Sankararamakrishnan, 2017). Furthermore, in a study of the Mcl-1 complex formed with substituted anthraquinones, Wang et *al.* revealed hydrophobic interactions between the residue Val253 and one isopropyl group of their ligand (2,3-dihydroxy-6-((4-isopropylphenyl)thio)anthracene-9,10-dione); the ligand triggered the rupture of the *i*, *i* + 4 hydrogen bonds between the Val243-Ser247 and Leu246-Met250, leading to the deformation of the α4 helix. This helix limits the Mcl-1 BH3 binding groove and it is known to allow adaptation of the P2 interaction pocket to an optimal conformation for receiving the ligand (Wang et al., 2016).

Other highly perturbed residues located outside the interaction site include Thr295 in helix α6, Lys302 and Asp313, at

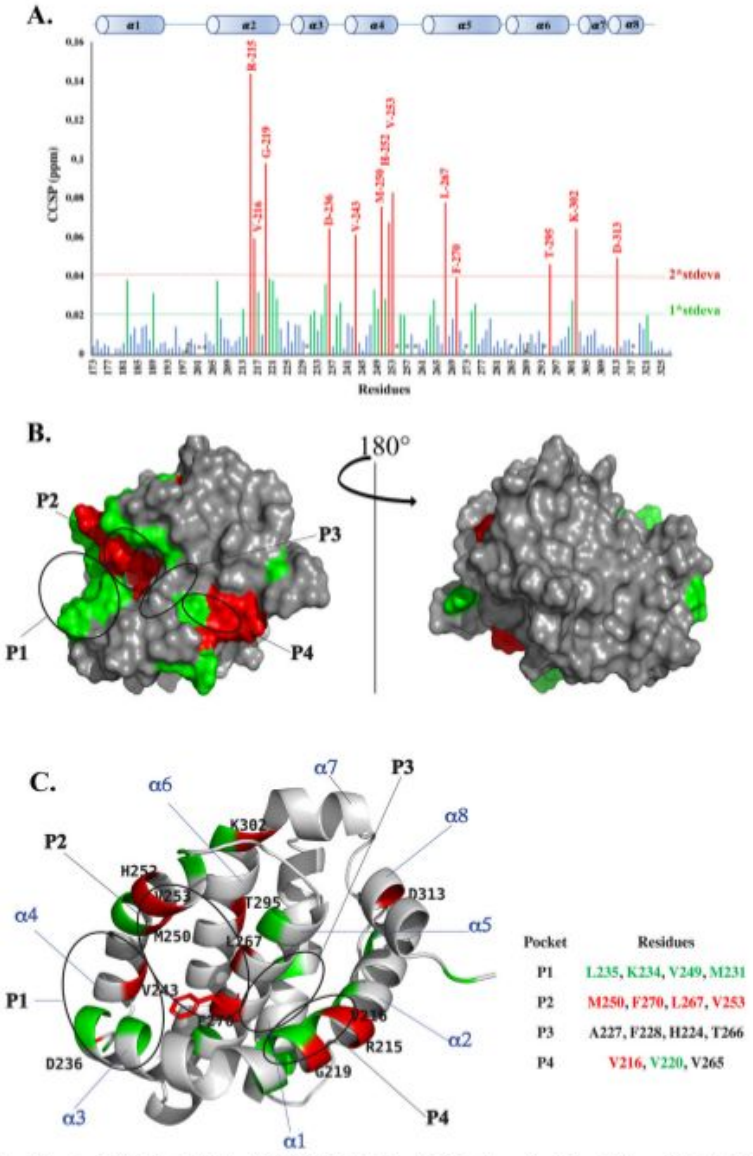


Figure 2. A. Measured Combined Chemical Shifts Perturbation (CCSP) of the Mcl-1 amide functions after the addition of Pyridoclax (ratio PyridoclaxMcl-1 = 1:1). (Pro) presence of Proline. (*) Undetermined. B. The most perturbed residues on the surface of the Mcl-1 protein, and C. the Cartoon representation of the Mcl-1 structure with annotation of most perturbed residues (residues in red).

the α6 and α8 helix ends, respectively, and Asp236 located in the loop between helices α3 and α4. These charged residues are probably involved in electrostatic interactions and their disruption is likely explained by an indirect effect of the interaction (Krzeminski, Fuentes, Boelens, & Bonvin, 2009).

Effect of the interaction on methyl groups

Methyl groups have been shown to be useful probes for the study of drug binding by NMR (Wiesner & Sprangers, 2015). Therefore, 2 D ^{1}H-^{13}C SOFAST HMQC spectra for Mcl-1 were recorded in the absence and the presence of Pyridoclax. The signal assignment of the ^{1}H-^{13}C SOFAST HMQC peaks was obtained by direct transfer from the previously reported assignment of Mcl-1 region 172-327 (BMRB 19654; Liu et al., 2014). Addition of Pyridoclax induced signal broadening and chemical shift perturbations of several aliphatic correlations including the methyl groups of Met231 and Met250, the β CH_2 of Leu213, and the γ CH_3 of Val253 (SI - Figure 3). The amide protons of these residues have been previously identified as hotspots using ^{1}H-^{15}N HSQC experiments.

Mapping these residues on the structure of Mcl-1 (Figure 3A) showed that the distance between the methyl groups of the Val253 and Met231 residues was 2.7 Å,

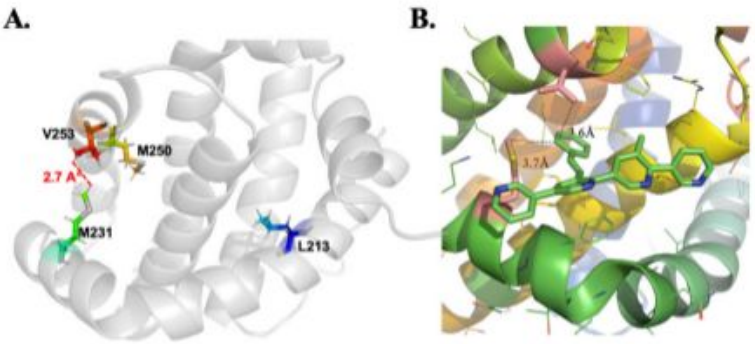


Figure 3. A. Mapping of residues on Mcl-1 for which the chemical shifts of the methyl groups were modified by the addition of Pyridoclax. B. Representation of the Pyridoclax in Mcl-1 binding site (structure of the complex deduced from NMR and molecular modeling).

signifying the occurrence of a hydrophobic interaction between the side chains of these two residues; therefore addition of Pyridoclax could disrupt this hydrophobic interaction and thus affect the chemical shifts of these methyl groups (Figure 3B). Met250 is located in the deep part of the P2 hydrophobic pocket, which formed the interaction site of Pyridoclax with Mcl-1 protein. Furthermore, Leu213 is located far from the hydrophobic pockets P1 and P2 at α2 helix. Its disruption was probably related to an indirect effect of the interaction. In addition, ring-current effects, mentioned above, may also be responsible for the perturbation of these residues.

Overall, our NMR study based on chemical shift perturbations indicates that Pyridoclax binds preferentially the P1 and P2 hydrophobic pockets of Mcl-1. However, the low solubility of Pyridoclax under our experimental conditions has prevented access to the distance constraints between Pyridoclax and Mcl-1 protons. For the same reason, no epitope mapping could be done.

K_D estimation

One of the most important steps in characterizing a protein-ligand complex is measuring its dissociation constant (K_D). For this purpose, chemical shift perturbations induced in the ^{1}H-^{15}N HSQC spectra of Mcl-1 by adding Pyridoclax were monitored. The shifted peaks showed linear concentration-dependent chemical shift changes (Figure 4A, B & C), which indicated a fast exchange regime on the NMR chemical shift time scale (Williamson, 2013). Subsequently, combined chemical shifts differences of backbone amides were displayed versus the ratio of Pyridoclax concentration relative to protein for selected isolated residues showing chemical shift perturbations (Figure 4D). Apparent K_D for each perturbed residue was estimated in the micromolar range, with an average value of 15 μM.

2D line-shape analyzes, carried out using the TITAN software, allowed the estimation of the kinetic (k_{off}) and thermodynamic (K_D) constants. Interestingly, it revealed a fast k_{off} value ($208 \pm 8\,s^{-1}$) and a large residence time τ of 4.8 ms ($\tau = 1/k_{off}$). These values are consistent with the moderate affinity of Pyridoclax to the Mcl-1 protein in the low micromolar range.

However, the binding affinity measured by NMR highly differs from that previously measured by the SPR method (Gloaguen et al., 2015) which indicated a K_D value of 25 nM. This inconsistency is likely due to the moderate solubility and low stability of Pyridoclax in aqueous solution at micromolar concentration. It may also be due to the non-specific surface absorption of proteins in SPR assays (Masson et al., 2006) or the different experimental conditions used for NMR and SPR measurements.

Biological assay

Ovarian cancer SKOV3 cell line was used to study the biological effect of Pyridoclax *in vitro*. As previously demonstrated, SKOV3 cells survival depends on both Bcl-x_L and Mcl-1 (Brotin et al., 2010), and since their concomitant inhibition leads to massive apoptosis in these cells, this model can be used for the screening of new pharmacological inhibitors of these two targets. Indeed, the detection of apoptosis induction in SKOV3 cells exposed to the evaluated drug after inhibition of Bcl-x_L by the BH3 mimetic molecule ABT-737 suggests that the molecule can act as a Mcl-1 inhibitor.

Whereas neither 15 μM Pyridoclax nor 5 μM ABT-737 exerts any apoptotic effect as single agents on SKOV3 cells, their combination led to a massive apoptotic cell death induction. Indeed, in response to this combination, nearly all the cells are shrunk and detached (Figure 5A), many nuclear condensations and fragmentations are observed after DAPI staining of the nuclei (Figure 5A), and caspase 3 and PARP cleavage are observed by western blot (Figure 5B). A similar effect is also observed with the combination of 10 μM Pyridoclax/5 μM ABT-737 but to a lesser extent. These results suggest that Pyridoclax efficiently inhibits Mcl-1 in SKOV3 cells in a concentration-dependent manner. Similar results have been obtained in another ovarian cancer cell line (IGROV1-R10, data not shown). This revealed that Pyridoclax induced massive apoptosis at a concentration of 15 μM in combination with Bcl-x_L inhibitor (ABT-737) in chemo-resistant ovarian cancer cell lines. These results are consistent with the estimated K_D value of ~15μM.

Docking of Pyridoclax to the Mcl-1 binding groove, selection of the representative poses

Before carrying out the docking of Pyridoclax to Mcl-1, the solved X-ray Pyridoclax structures (CCDC n°996607) were

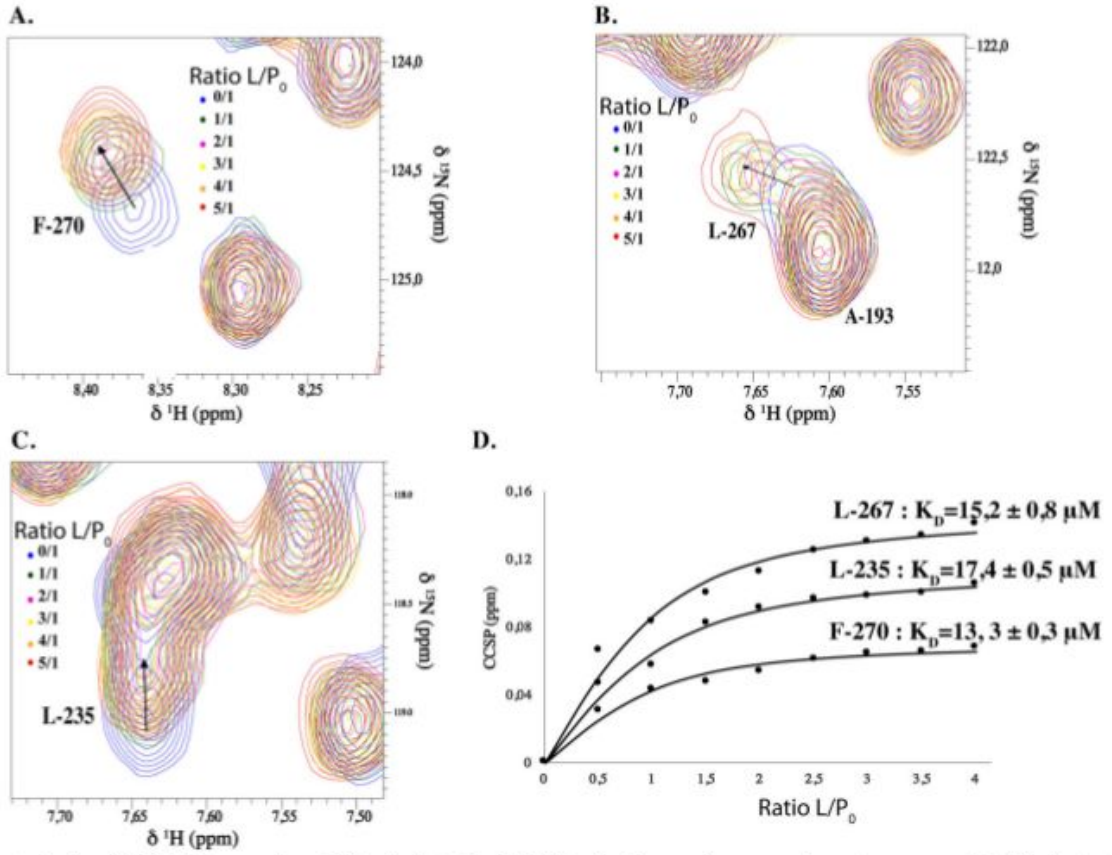


Figure 4. Overlay of ^{1}H-^{15}N HSQC spectra (expanded views) obtained for ^{15}N-Mcl-1 in the absence and presence of increasing amounts of Pyridoclax (various L/P_0 ratio). Chemical shift modifications of residues F270 (A), L267 (B), and L235 (C) are indicated by black arrows. D. Titration curves for L235, L267 and F270 and K_D estimation of each residue using Equation (2).

compared to the conformation of Pyridoclax in solution, determined from NMR experiments. Molecular dynamics simulations were carried out with the application of $^{1}H-^{1}H$ distance constraints determined from the NMR data of Pyridoclax in solution (SI – Table S2). Starting from Pyridoclax X-ray structure, these simulations led to Pyridoclax structure in solution. Ten lowest energy conformers (as shown in SI - Table S3A) were retained for subsequent analysis. All ten conformers were very close and the rings had turned in the same way. Generated NMR structure is close to the X-ray one (see Figure 6), but some differences exist. The methyl substituent is oriented toward the styryl in the NMR structure in contrast to the X-ray one which is due to the flip of the C and D ring block but the absolute preferential angles between rings are still conserved (SI-Table S3B). Our previously published studies (Perato et al., 2013; Sopkova-de Oliveira Santos et al., 2012) have shown that the adjacent rings turn as well to the left as to the right that there is not a preferential direction. The A-ring orientation differs also between X-ray and NMR structure, but in the X-ray structure A-ring nitrogen atom is engaged in weak intermolecular contacts among the neighboring molecules in the crystal.

Then, to gain further insight into the possible binding modes of Pyridoclax to the Mcl-1 binding groove highlighted from NMR, docking studies using the GOLD program applying the ChemPLP scoring function were carried out on our Mcl-1 model using the two Pyridoclax conformers extracted from its crystallographic structure and the representative NMR structure (Figure 6). GOLD is an automated ligand docking program that uses a genetic algorithm. GOLD's evolutionary algorithm modifies the position, orientation and conformation of a ligand to fit into one or more low energy states of the protein active site.

As the number of solved Mcl-1 structures has increased during the last 5 years and as first Mcl-1 structures co-crystalized with synthetic ligands were published, we were able to observe that the wide opening of hydrophobic pockets depended on the fixed ligands. Especially, the P2 pocket is more open and deeper in presence of the synthetic ligands compared to the peptide ones. Therefore, we built and used for this docking a representative Mcl-1 3D model, without mutation and without deletion, with the P2 pocket open. Even if the Mcl-1 model used during this docking was different compared to the previously published one (Gloaguen et al., 2015), based on structures co-crystalized with peptide ligands, the generated poses for Pyridoclax were globally similar. Two principal poses, close to ones observed in the previous docking (Gloaguen et al., 2015),

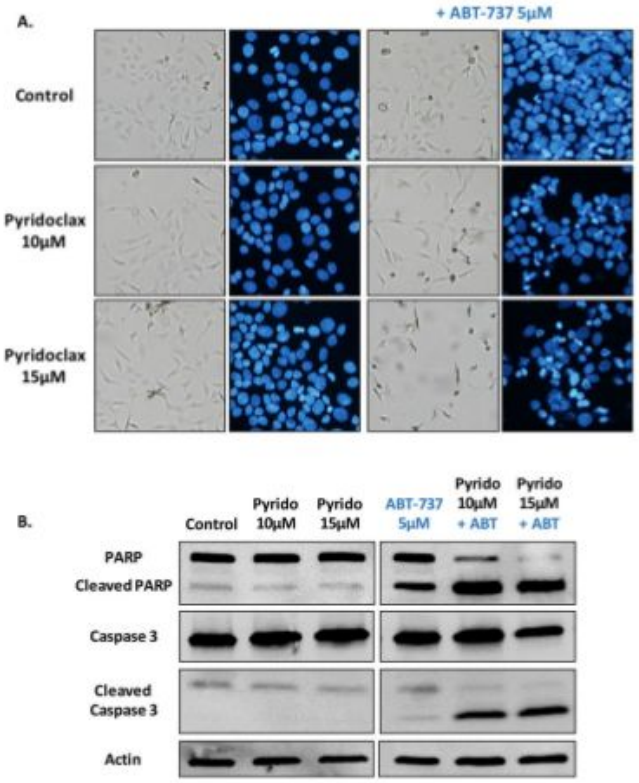


Figure 5. Apoptotic effect of Pyridoclax, alone or associated with ABT-737, on SKOV3 ovarian cancer cell lines. A. Nuclear morphology studied after DAPI staining. B. Caspase 3 and PARP cleavage observed by western blot.

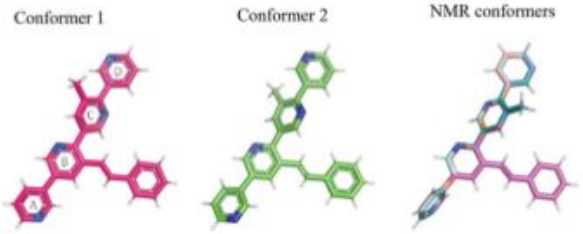


Figure 6. Comparison of the two conformers of Pyridoclax determined by X-ray crystallography (CCDC n°996607) with the 10 lowest energy NMR conformers calculated in this study (NMR structures are overlaid over non-hydrogen atoms).

emerged from this new docking study. The first pose with Pyridoclax positioned 180° relative to the second pose (see Figures 7A and B). In addition to these two poses, a new third-minority occupied pose appeared (see Figure 7C).

In the first two poses, the Pyridoclax styryl group was always anchored in the same Mcl-1 pocket, pocket P2 (it always stacked against Mcl-1 Phe270), even if the orientations of Pyridoclax in the binding cavity changed (Figure 7A and B). Moreover, one of the Pyridoclax ending pyridines was always close to Arg263 of Mcl-1. The methyl group pointed to different pockets in the various ligand poses. In the third pose, contrary to the previous two, P2 pocket was occupied by one terminal pyridine and not by the styryl ring. The pyridine ring went deeper into P2 pocket compared to the styryl ring in the previous two poses. In this third pose, the styryl group was situated on the protein surface, accessible to a solvent (see Figure 7C). In these three poses, the main point of attachment was the P2 pocket, as supported by the NMR analysis, and no interaction was observed between Pyridoclax and Mcl-1 P1 and P4 pockets.

MD simulation on the three representative poses from docking

To evaluate the relevance of these three poses, dynamics simulations of 100 ns were carried out on each one. In a previously published theoretical analysis (Perato et al., 2013; Sopkova-de Oliveira Santos et al., 2012) on the preferential orientation between variously substituted pyridine rings we showed that they did not follow the same rule as phenyl ones (see SI – Figure S1). The preferential orientation between the successive phenyl rings is either coplanar or perpendicular, while for pyridine rings it depends on the nitrogen atom position in the ring as well as the substituent position. In addition, experimentally observed preferential angles in Pyridoclax solid structure as well as in solution one were in agreement with our theoretical prediction. For this reason, we have modified the dihedral angle parametrization for successive pyridine in the Pyridoclax scaffold (see SI-Figure S1) for our molecular dynamics' simulations.

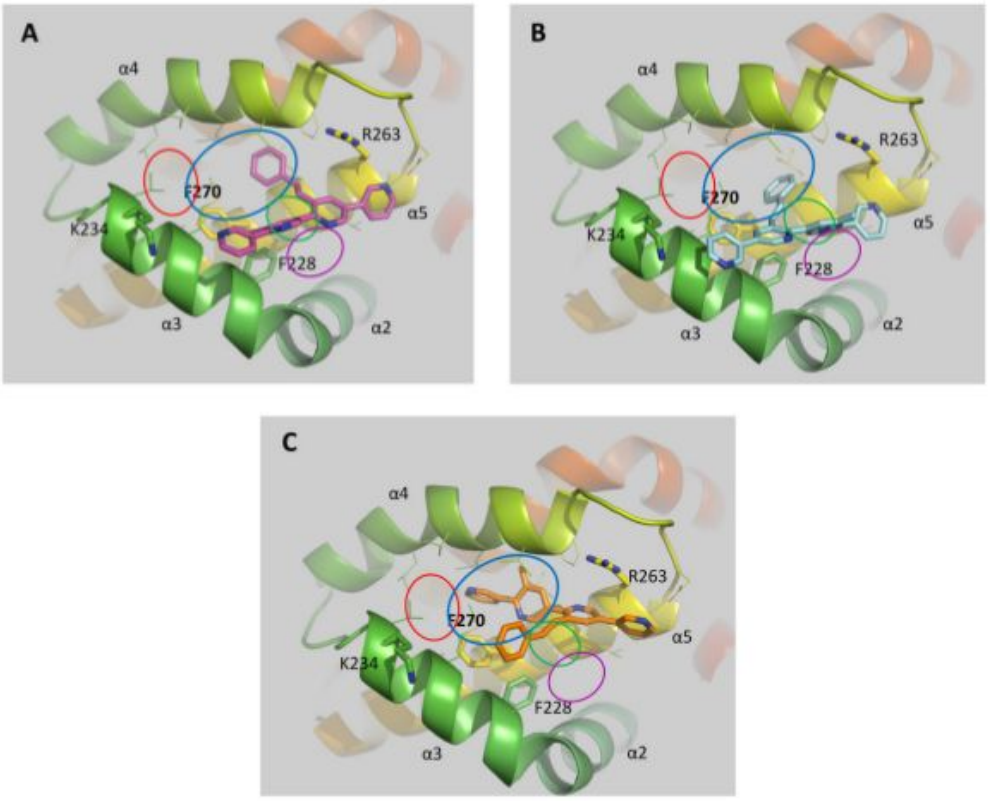


Figure 7. The representative presentation of three poses generated from docking studies using the GOLD program: (A) pose 1, (B) pose 2 and (C) pose 3. The protein is presented as a ribbon, and selected side chains and Pyridoclax as sticks. Mcl-1 binding pockets are indicated by open circles with P1 in red, P2 in blue, P3 in green, and P4 in purple.

The backbone root-mean-square deviation (RMSD) calculated on the Mcl-1 backbone with respect to the initial complex along the trajectories did not exceed 2.5 Å in all three poses which demonstrated the stability of the complexes during our simulations (SI - Figure S4).

The calculated backbone fluctuations during the simulations confirmed that the helical parts of the Mcl-1 protein were globally preserved and stable during the simulation. Fluctuation values in the helical parts did not exceed 1.0 Å (SI - Figure S5) with the exception of the α4 helix for which they increased to 1.6 Å. The Pyridoclax binding groove is delimited by three α-helices (α3, α4, and α5). The α4 helix was the most mobile of the three; it moved to adapt the Mcl-1 binding cavity for the incoming ligand.

In the non-helical parts, the greatest Mcl-1 flexibilities were observed in the N and C-terminal parts, and in the loop between the α1 and α2 helices (SI - Figure S5). This long flexible loop was usually shortened in the crystal structures with synthetic ligands to facilitate the crystal growth.

With the goal of analyzing Pyridoclax behavior in the Mcl-1 binding groove, evolution of Pyridoclax position RMSD with respect to the initial docking RMSD for each cluster was calculated (Figure 8). The visualization of dynamics for poses 1 and 2 showed that Pyridoclax remained in the binding groove throughout the whole 100 ns of the simulation and the Pyridoclax RMSD evolution along the simulations pointed out that the Pyridoclax position became stable after 70 ns of the simulation for pose 1 and after 40 ns for pose 2. The Pyridoclax RMSD in pose 1 and 2 dynamics stabilized around the same value, 9 Å. However, for pose 3 the Pyridoclax RMSD reached 12 Å at the end of the simulation and the trajectory visualization showed that this corresponds to the Pyridoclax leaving from the Mcl-1 binding site. Therefore, this pose was eliminated.

The interaction energy between Pyridoclax and Mcl-1 along the whole trajectory was also analyzed for each pose (Figure 9). According to the interaction energy criterion (Table 1), the most favorable appeared to be pose 1, but the energetic differences among the two poses are minimal. The interaction energy of pose 1 was about 2 kcal/mol more favorable with respect to pose 2 if the whole trajectory was considered, and only about 1 kcal/mol more favorable in the last stable part of the simulation (last 30 ns; Table 1).

To predict various possible modes of Pyridoclax binding in the Mcl-1 groove, clustering on the whole trajectory of pose 1 and pose 2 was carried out. Clustering on pose 1 and on pose 2 trajectories revealed six clusters in both cases (SI - Figure S6). In both clusters, the first three clusters were the

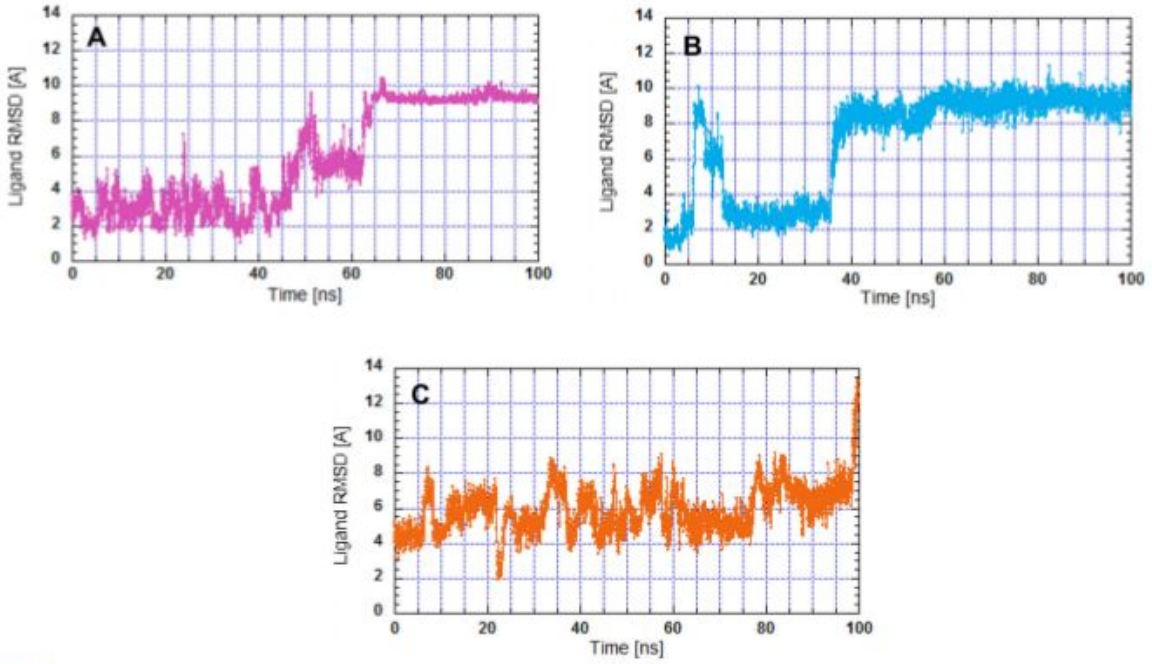


Figure 8. Pyridoclax root mean square deviations along the Pyridoclax/Mcl-1 simulated trajectory with respect to the initial structure: (A) pose 1, (B) pose 2, and (C) pose 3.

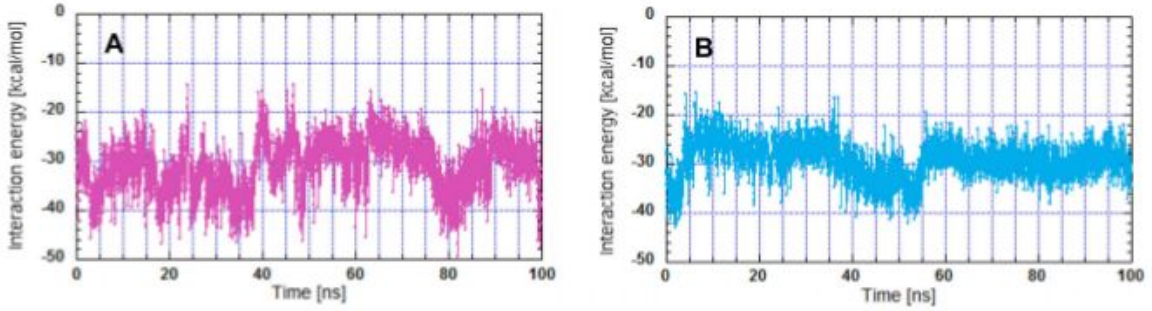


Figure 9. Evolution of interaction energy between Pyridoclax and Mcl-1 along the simulated trajectory of (A) pose 1and (B) pose 2.

ones mostly populated and therefore mainly representative. In the case of pose 1 dynamics, Pyridoclax remained in the same binding groove along whole trajectory, but its position changed. In the first cluster, corresponding to the beginning of the trajectory, Pyridoclax remained close to its initial docking position, *i.e.* the position with D pyridine in proximity of the Arg263 side chain. In the middle of the trajectory (third cluster), Pyridoclax started to move and became perpendicular to its initial position. Then, it rotated by 180° compared to its initial position (cluster 2, Figure 10) and D pyridine became close now to Lys234. This switch of Pyridoclax was accompanied by a distortion of the α4 helix and α3 helix displacement outside the BH3 binding groove, detected in fluctuation analysis (SI - Figure S5). This observation confirmed that perturbation of the residues held by these helices and located far from pockets P1 and P2 identified in the NMR study was related to an indirect effect.

Contrarily, clustering on pose 2 revealed two different Pyridoclax binding modes along the dynamics. The Pyridoclax position in the second cluster, at the start of its trajectory, was close to the initial docking pose 2, contrary to the Pyridoclax position in cluster1. In cluster 1 and cluster 3, corresponding to the stable ending part of the pose 2 simulation, Pyridoclax left pocket P3 and moved along the α3 helix occupying principally pockets P1 and P2 (Figure 10). This change of the Pyridoclax position was accompanied by the α4 helix moving as detected in the fluctuation analysis (SI - Figure S5).

The Pyridoclax interaction energy per Mcl-1 residue study highlighted the different Mcl-1 residues involved in the Pyridoclax binding. Comparison of the binding histograms calculated for representative pose of each cluster (Figures 10 and 11) and the residues perturbed by Pyridoclax binding, observed during the NMR study, led to the conclusion that

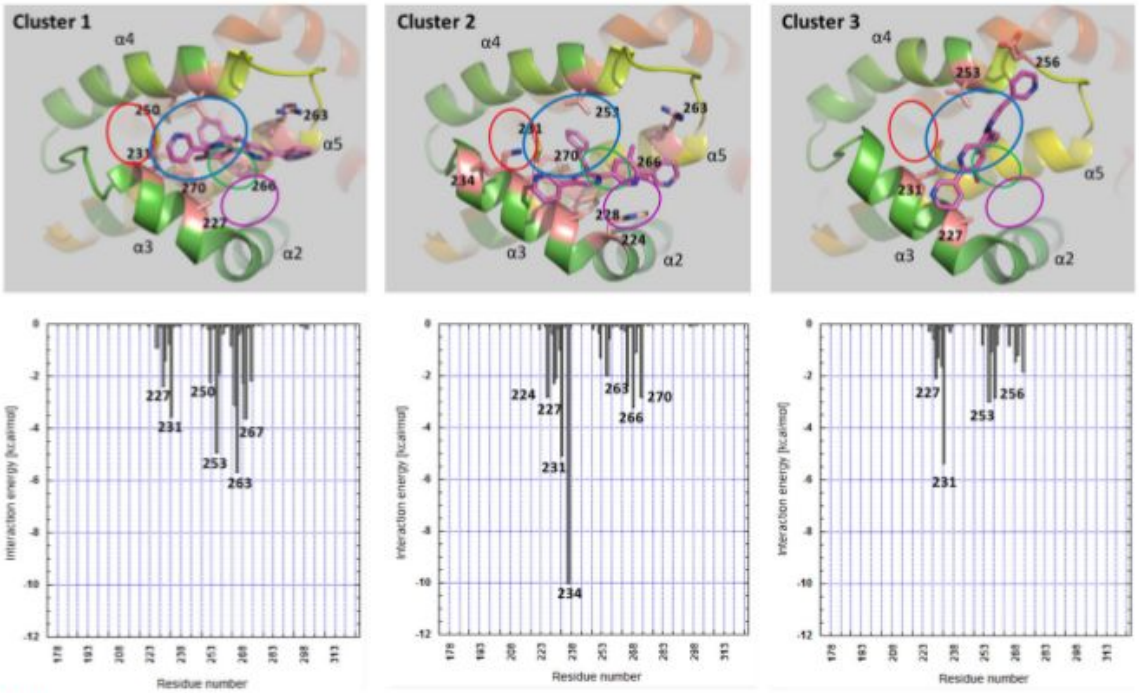


Figure 10. (Top panel) representations of the first three cluster centroids from the pose 1 trajectory. The protein is presented as ribbon, and the selected side chains (Interaction energy > 2 kcal/mol) as pink sticks. (Bottom panel) Histograms of the Pyridoclax interaction energy per Mcl-1 residue for each cluster centroid. Mcl-1 binding pockets are indicated by open circles with P1 in red, P2 in blue, P3 in green, and P4 in purple.

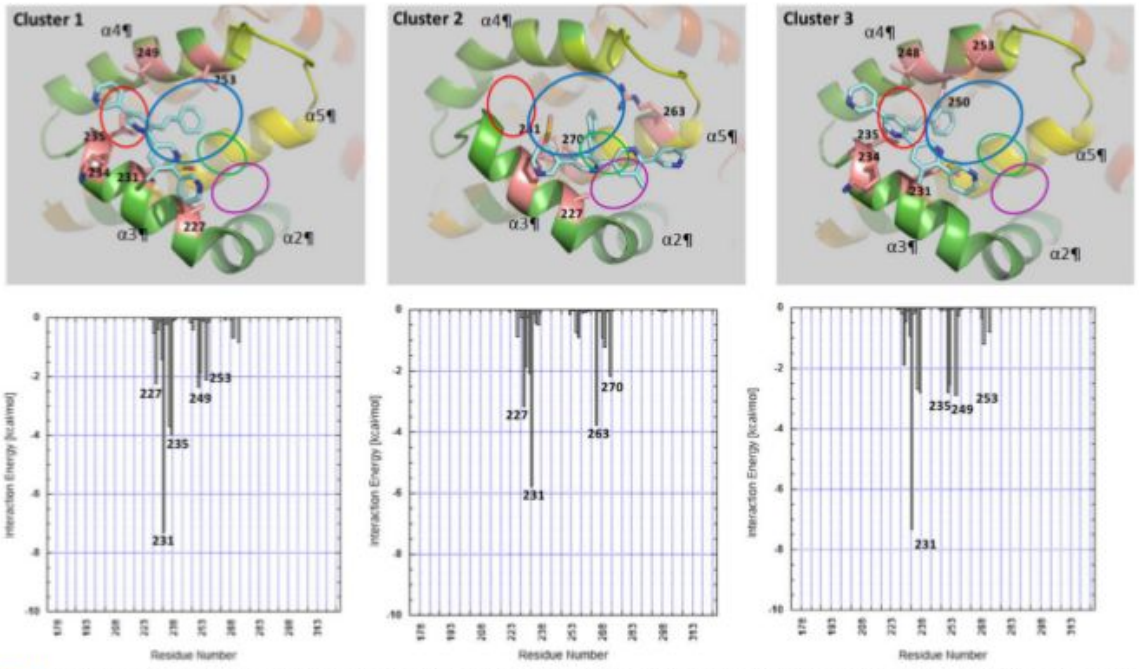


Figure 11. (Top panel) Representations of the first three cluster centroids from the pose 2 trajectories. The protein is presented as a ribbon, and selected side chains (Interaction energy > 2 kcal/mol) as pink sticks. (Bottom panel) Histograms of the Pyridoclax interaction energy per Mcl-1 residue for each cluster centroid. Mcl-1 binding pockets are indicated by open circles with P1 in red, P2 in blue, P3 in green, and P4 in purple.

Table 2. The Mcl-1 residues interacting with Pyridoclax with energy higher than 1kcal/mol for pose 1 cluster 2 and pose 2 cluster 1. Red: Highly disturbed residues, green: moderately perturbed residues.

Pose1 cluster 2		Pose 2 cluster 1	
Residues	Interaction energy [kcal.mol^{-1}]	Residues	Interaction energy [kcal.mol^{-1}]
Lys234	−10.02	Met231	−7.30
Met231	−5.08	Leu235	−3.97
Thr266	−3.24	Lys234	−3.69
His224	−2.82	Val249	−2.36
Phe270	−2.82	Ala227	−2.26
Arg263	−2.39	Val253	−2.11
Ala227	−2.31	Met250	−1.86
Phe228	−2.09	Gly230	−1.41
Val253	−2.01	–	–
Met250	−1.31	–	–
Leu267	−1.08	–	–

the preferred binding modes of Pyridoclax were either cluster 1 of pose 2 or cluster 2 of pose 1. To evaluate these two poses, residues with interaction energy higher than 1 kcal/mol (Table 2) were analyzed. In pose 1 cluster 2, Pyridoclax interacted strongly with Val253, Met250, Phe270, Leu267 (see Table 2), four of 13 residues detected as highly perturbed in NMR study (shown as red residues in Figure 2) during the Pyridoclax binding, while in pose 2 cluster 1 it interacted only with two highly perturbed residues (Val253 and Met250). The energetic contribution of these four highly perturbed residues in pose 1 cluster 2 was greater than in pose 2 cluster 1: -7.22 kcal/mol compared to -3.87 kcal/mol, respectively. Furthermore, three moderately perturbed residues for pose 1 cluster 2 (Lys234, Met231, Arg263; shown as green residues in Figure 2) and four moderately perturbed residues for pose2 cluster1 (Met231, Leu235, Lys234 and Val249) were also revealed. In conclusion, in contrast to the initial docking poses, the molecular dynamics study revealed also the importance of the P1 pocket in Pyridoclax binding. However, Pyridoclax was not shown to interact with Arg215, Val216, and Gly219 of pocket P4 in any of the clusters obtained by docking or molecular dynamics, while the NMR chemical shifts of these residues were significantly affected upon ligand binding. Visual analysis showed that Arg215 and Gly219 were situated on the Mcl-1 surface outside P4 pocket and Mcl-1 binding groove. Nevertheless, in several published NMR studies (Abulwerdi et al., 2014; Chen et al., 2016; Zhang et al., 2013), these residues were defined as perturbed ones without the establishment of any direct interaction with a ligand.

Interestingly, in pose 1 cluster 2 an electrostatic contact was established between the Lys234 side chain nitrogen and pyridine nitrogen. The structure activity study on the Pyridoclax family showed that replacement of ending pyridines by phenols abolished the compound pro-apoptotic activity. The presence at least of one ending pyridine was necessary for the activity. In pose 2 cluster 1 no electrostatic interaction with ending pyridine nitrogen was observed. Once again, this result favored pose 1 cluster 2. Therefore, NMR and molecular dynamics studies are in accordance and in favor of pose 1 cluster 2.

Conclusion

NMR titration experiments allowed us to gain insight into the binding interface of Mcl-1 by mapping residues with large CCSPs on the protein structure. They showed that Pyridoclax bound to the BH3 binding groove and more specifically to the P2 pocket of Mcl-1.

The docking and molecular dynamics simulations were compared with NMR data to predict more precisely the Pyridoclax binding mode. Simulation results showed that Pyridoclax styryl entered pocket P2; the methyl substituent was at the edge of pockets P2 and P3 and one of the ending pyridine interacts through electrostatic interaction with Lys234 side chain, a positively charged residue not conserved in Bcl-x_L and Bcl-2. This binding mode was consistent with previously established SAR study on the Pyridoclax family, which showed that the methyl and styryl hydrophobic motifs situated on the bipyridine core are necessary for the Pyridoclax pro-apoptotic activity as well as the presence of at least one terminal pyridine.

Therefore, study of the interaction between the Mcl-1 protein and Pyridoclax by NMR and molecular modeling allowed the proposal of a representative model of the formed complex.

Acknowledgements

We are also grateful to Dr. Laurent POULAIN, director of BioTICLA Unit (Inserm U1199) who provided expertise in bioassays that greatly assisted the research.

Author contributions

AB and MS performed NMR studies; DR and LC have carried out the production and purification of the labeled Mcl-1$^{172-327}$ domain; AV and CD synthesized Pyridoclax; HP and LW have carried out bioassays; BM, MB, RB, GC and JSO performed docking and dynamics simulations; AB, MS, AV, JSO and HO conceived the studies. AB, JSO, HO, AV and MS wrote the papers with corrections from all co-authors. All authors have given approval to the final version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work has been partially supported by Normandie Université (NU), the Région Normandie, the Centre National de la Recherche Scientifique (CNRS), Université de Rouen Normandie (URN), INSA Rouen Normandie, Labex SynOrg (ANR-11-LABX-0029), Innovation Chimie Carnot (I2C). Part of this work was performed using computing resources of CRIANN (Normandy, France) as well as the European Community (FEDER) for the molecular modeling software.

ORCID

A. Bourafai-Aziez http://orcid.org/0000-0001-5974-7641
M. Sebban http://orcid.org/0000-0002-7907-1092
C. Denis http://orcid.org/0000-0002-0417-1443

L. Carlier http://orcid.org/0000-0002-9036-6561
A. S. Voisin-Chiret http://orcid.org/0000-0001-5564-2244
J. Sopková-de Oliveira Santos http://orcid.org/0000-0002-4829-8120
H. Oulyadi http://orcid.org/0000-0001-5813-1482

References

Abid, M., Sonawane, Y. A., Contreras, J. I., Rana, S., & Natarajan, A. (2017). Recent advances in cancer drug development: Targeting induced myeloid cell leukemia-1 (Mcl-1) differentiation protein. *Current medicinal chemistry, 24*(40), 4488–4514. doi:10.2174/0929867324666170912092659

Abulwerdi, F., Liao, C., Liu, M., Azmi, A. S., Aboukameel, A., Mady, A. S. A., … Nikolovska-Coleska, Z. (2014). A novel small-molecule inhibitor of mcl-1 blocks pancreatic cancer growth in vitro and in vivo. *Molecular Cancer Therapeutics, 13*(3), 565–575. doi:10.1158/1535-7163.MCT-12-0767

Abulwerdi, F. A., Liao, C., Mady, A. S., Gavin, J., Shen, C., Cierpicki, T., … Nikolovska-Coleska, Z. (2014). 3-substituted- *N* -(4-hydroxynaphthalen-1-yl)arylsulfonamides as a novel class of selective Mcl-1 inhibitors: Structure-based design, synthesis, SAR, and biological evaluation. *Journal of Medicinal Chemistry, 57*(10), 4111–4133. doi:10.1021/jm500010b

Ashkenazi, A., Fairbrother, W. J., Leverson, J. D., & Souers, A. J. (2017). From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nature Reviews Drug Discovery, 16*(4), 273–284. doi:10.1038/nrd.2016.253

Beekman, A. M., & Howell, L. A. (2016). Small-molecule and peptide inhibitors of the pro-survival protein Mcl-1. *ChemMedChem, 11*(8), 802–813. doi:10.1002/cmdc.201500497

Belmar, J., & Fesik, S. W. (2015). Small molecule Mcl-1 inhibitors for the treatment of cancer. *Pharmacology & Therapeutics, 145*, 76–84. doi:10.1016/j.pharmthera.2014.08.003

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications, 91*(1–3), 43–56. doi:10.1016/0010-4655(95)00042-E

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research, 28*(1), 235–242. doi:10.1093/nar/28.1.235

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., … & Mc Henry, K. T. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283), 899. doi:10.1038/nature08822

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation, 8*(9), 3257–3273. doi:10.1021/ct300400x

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., … Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry, 30*(10), 1545–1614. doi:10.1002/jcc.21287

Brotin, E., Meryet-Figuière, M., Simonin, K., Duval, R. E., Villedieu, M., Leroy-Dudal, J., … Poulain, L. (2010). Bcl-x $_L$ and MCL-1 constitute pertinent targets in ovarian carcinoma and their concomitant inhibition is sufficient to induce apoptosis. *International Journal of Cancer, 126*, 885–895. doi:10.1002/ijc.24787

Bruncko, M., Wang, L., Sheppard, G. S., Phillips, D. C., Tahir, S. K., Xue, J., … Souers, A. J. (2015). Structure-guided design of a series of MCL-1 inhibitors with high affinity and selectivity. *Journal of Medicinal Chemistry, 58*(5), 2180–2194. doi:10.1021/jm501258m

Burke, J. P., Bian, Z., Shaw, S., Zhao, B., Goodwin, C. M., Belmar, J., … Fesik, S. W. (2015). Discovery of tricyclic indoles that potently inhibit Mcl-1 using fragment-based methods and structure-based design. *Journal of Medicinal Chemistry, 58*(9), 3794–3805. doi:10.1021/jm501984f

Caenepeel, S., Brown, S. P., Belmontes, B., Moody, G., Keegan, K. S., Chui, D., … Hughes, P. E. (2018). AMG 176, a selective MCL1 inhibitor, is effective in hematological cancer models alone and in combination with established therapies. *Cancer Discovery, 8*, 1–16. doi:10.1158/2159-8290.CD-18-0387

Cala, O., Guillière, F., & Krimm, I. (2014). NMR-based analysis of protein–ligand interactions. *Analytical and Bioanalytical Chemistry, 406*(4), 943–956. doi:10.1007/s00216-013-6931-0

Chen, L., Wilder, P. T., Drennen, B., Tran, J., Roth, B. M., Chesko, K., … Fletcher, S. (2016). Structure-based design of 3-carboxy-substituted 1,2,3,4-tetrahydroquinolines as inhibitors of myeloid cell leukemia-1 (Mcl-1). *Organic & Biomolecular Chemistry, 14*(24), 5505–5510. doi:10.1039/C5OB02063H

Clifton, M. C., Dranow, D. M., Leed, A., Fulroth, B., Fairman, J. W., Abendroth, J., … Serrano-Wu, M. H. (2015). A maltose-binding protein fusion construct yields a robust crystallography platform for MCL1. *PLoS One, 10*(4), e0125010. doi:10.1371/journal.pone.0125010

Cohen, N. A., Stewart, M. L., Gavathiotis, E., Tepper, J. L., Bruekner, S. R., Koss, B., … Walensky, L. D. (2012). A competitive stapled peptide screen identifies a selective small molecule that overcomes MCL-1-dependent leukemia cell survival. *Chemistry & Biology, 19*(9), 1175–1186. doi:10.1016/j.chembiol.2012.07.018

Czabotar, P. E., Lee, E. F., Thompson, G. V., Wardak, A. Z., Fairlie, W. D., & Colman, P. M. (2011). Mutation to bax beyond the BH3 domain disrupts interactions with pro-survival proteins and promotes apoptosis. *Journal of Biological Chemistry, 286*(9), 7123–7131. doi:10.2210/pdb3pk1/pdb

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh ewald: An *N* ·log (*N*) method for Ewald sums in large systems. *Journal of Chemical Physics, 98*(12), 10089–10092. doi:10.1063/1.464397

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., & Mark, A. E. (1999). Peptide folding: When simulation meets experiment. *Angewandte Chemie International Edition, 38*(1–2), 236–240. doi:10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M

Delbridge, A. R. D., Grabow, S., Strasser, A., & Vaux, D. L. (2016). Thirty years of BCL-2: Translating cell death discoveries into novel cancer therapies. *Nature Reviews Cancer, 16*(2), 99–109. doi:10.1038/nrc.2015.17

Fire, E., Gulla, S. V., Grant, R. A., & Keating, A. E. (2010). Mcl-1-bim complexes accommodate surprising point mutations via minor structural changes. *Protein Science, 19*(3), 507–519. doi:10.2210/pdb2pqk/pdb

Friberg, A., Vigil, D., Zhao, B., Daniels, R. N., Burke, J. P., Garcia-Barrantes, P. M., … Fesik, S. W. (2013). Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *Journal of Medicinal Chemistry, 56*(1), 15–30. doi:10.1021/jm301448p

Gelinas, C., & White, E. (2005). BH3-only proteins in control: Specificity regulates MCL-1 and BAK-mediated apoptosis. *Genes & Development, 19*(11), 1263–1268. doi:10.1101/gad.1326205

Gloaguen, C., Voisin-Chiret, A. S., Sopkova-de Oliveira Santos, J., Fogha, J., Gautier, F., De Giorgi, M., … Poulain, L. (2015). First evidence that oligopyridines, α-helix foldamers, inhibit Mcl-1 and sensitize ovarian carcinoma cells to Bcl-xL-targeting strategies. *Journal of Medicinal Chemistry, 58*(4), 1644–1668. doi:10.1021/jm500672y

Grzesiek, S., & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetism Reasoning, 96*(2), 432–440. doi:10.1016/0022-2364(92)90099-S

Grzesiek, S., & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *Journal of Biomolecular NMR, 3*(2), 185–204. doi:10.1007/bf00178261

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646–674. doi:10.1016/j.cell.2011.02.013

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., … MacKerell, A. D. (2017). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods, 14*(1), 71–73. doi:10.1038/nmeth.4067

Jenson, J. M., Ryan, J. A., Grant, R. A., Letai, A., & Keating, A. E. (2017). Epistatic mutations in PUMA BH3 drive an alternate binding mode to

potently and selectively inhibit anti-apoptotic Bfl-1. *eLife*, *6*, e25541. doi:10.7554/eLife.25541

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*(11), 1859–1865. doi:10.1002/jcc.20945

Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, *245*(1), 43–53 doi:10.1016/S0022-2836(95)80037-9

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, *79*(2), 926–935. doi:10.1063/1.445869

Kerr, J. F., Wyllie, A. H., & Currie, A. R. (1972). Apoptosis: A basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of Cancer*, *26*(4), 239. doi:10.1038/bjc.1972.33

Kotschy, A., Szlavik, Z., Murray, J., Davidson, J., Maragno, A. L., Le Toumelin-Braizat, G., … Geneste, O. (2016). The MCL1 inhibitor S63845 is tolerable and effective in diverse cancer models. *Nature*, *538*(7626), 477–482. doi:10.1038/nature19830

Krzeminski, M., Fuentes, G., Boelens, R., & Bonvin, A. M. J. J. (2009). MINOES: A new approach to select a representative ensemble of structures in NMR studies of (partially) unfolded states. Application to Delta25-PYP. *Proteins: Structure, Function, and Bioinformatics*, *74*(4), 895–904. doi:10.1002/prot.22197

Lee, W., Tonelli, M., & Markley, J. L. (2015). NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, *31*(8), 1325–1327. doi:10.1093/bioinformatics/btu830

Liu, Q., Moldoveanu, T., Sprules, T., Matta-Camacho, E., Mansur-Azzam, N., & Gehring, K. (2010). Apoptotic regulation by MCL-1 through heterodimerization. *Journal of Biological Chemistry*, *285*(25), 19615–19624. doi:10.1074/jbc.M110.105452

Liu, G., Poppe, L., Aoki, K., Yamane, H., Lewis, J., & Szyperski, T. (2014). High-quality NMR structure of human anti-apoptotic protein domain MCL-1(171-327) for cancer drug design. *PLoS One*, *9*(5), e96521. doi:10.1371/journal.pone.0096521

Marintchev, A., Frueh, D., & Wagner, G. (2007). NMR methods for studying protein–protein interactions involved in translation initiation. In *Methods in enzymology* (pp. 283–331). Elsevier BV. EDs Abelson J.N.and Simon, M. I., Division of Biology, California Institute f Technology, Pasadena, California. doi:10.1016/S0076-6879(07)30012-8

Marion, D., & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1H-1H spin-spin coupling constants in proteins. *Biochemical and Biophysical Research Communications*, *113*(3), 967–974. doi:10.1016/0006-291X(83)91093-8

Masson, J.-F., Battaglia, T. M., Cramer, J., Beaudoin, S., Sierks, M., & Booksh, K. S. (2006). Reduction of nonspecific protein binding on surface Plasmon resonance biosensors. *Analytical and Bioanalytical Chemistry*, *386*(7–8), 1951–1959 doi:10.1007/s00216-006-0834-2

Meyer, B., & Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition*, *42*(8), 864–890. doi:10.1002/anie.200390233

Miles, J. A., Yeo, D. J., Rowell, P., Rodriguez-Marin, S., Pask, C. M., Warriner, S. L., … Wilson, A. J. (2016). Hydrocarbon constrained peptides – understanding preorganisation and binding affinity. *Chemical Science*, *7*(6), 3694–3702. doi:10.2210/pdb5c3f/pdb

Modi, V., & Sankararamakrishnan, R. (2017). Binding affinity of pro-apoptotic BH3 peptides for the anti-apoptotic Mcl-1 and A1 proteins: Molecular dynamics simulations of Mcl-1 and A1 in complex with six different BH3 peptides. *Journal of Molecular Graphics and Modelling*, *73*, 115–128. doi:10.1016/j.jmgm.2016.12.006

Mori, S., Abeygunawardana, C., Johnson, M. O., & Vanzijl, P. C. M. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *Journal of Magnetic Resonance*, *108*(1), 94–98. doi:10.1006/jmrb.1995.1109

Orts, J., & Gossert, A. D. (2018). Structure determination of protein-ligand complexes by NMR in solution. *Methods*, *138–139*, 3–25. doi:10.1016/j.ymeth.2018.01.019

Pelz, N. F., Bian, Z., Zhao, B., Shaw, S., Tarr, J. C., Belmar, J., … Fesik, S. W. (2016). Discovery of 2-Indole-acylsulfonamide myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods. *Journal of Medicinal Chemistry*, *59*(5), 2054–2066. doi:10.1021/acs.jmedchem.5b01660

Perato, S., Fogha, J., Sebban, M., Voisin-Chiret, A. S., Sopkova-de Oliveira Santos, J., Oulyadi, H., & Rault, S. (2013). Conformation control of abiotic α-helical foldamers. *Journal of Chemical Information and Modeling*, *53*(10), 2671–2680. doi:10.1021/ci400365y

Petros, A. M., Nettesheim, D. G., Wang, Y., Olejniczak, E. T., Meadows, R. P., Mack, J., … Thompson, C. B. (2000). Rationale for Bcl-X_L/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Science*, *9*(12), 2528–2534. doi:10.1110/ps.9.12.2528

Petros, A. M., Swann, S. L., Song, D., Swinger, K., Park, C., Zhang, H., … Sun, C. (2014). Fragment-based discovery of potent inhibitors of the anti-apoptotic MCL-1 protein. *Bioorganic & Medicinal Chemistry Letters*, *24*(6), 1484–1488. doi:10.1016/j.bmcl.2014.02.010

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., … Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. doi:10.1002/jcc.20289

Pons, J.-L., & Labesse, G. (2009). @TOME-2: A new pipeline for comparative modeling of protein-ligand complexes. *Nucleic Acids Research*, *37*(suppl_2), W485–491. doi:10.1093/nar/gkp368

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327–341. doi:10.1016/0021-9991(77)90098-5

Schanda, P., & Brutscher, B. (2005). Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, *127*(22), 8014–8015. doi:10.1021/ja051306e

Shore, G. C., & Warr, M. R. (2008). Unique biology of Mcl-1: Therapeutic opportunities in cancer. *Current Molecular Medicine*, *8*(2), 138–147. doi:10.2174/156652408783769580

Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., & Vuister, G. W. (2016). CcpNmr AnalysisAssign: A flexible platform for integrated NMR analysis. *Journal of Biomolecular NMR*, *66*(2), 111–124. doi:10.1007/s10858-016-0060-y

Song, T., Wang, Z., Ji, F., Feng, Y., Fan, Y., Chai, G., … Zhang, Z. (2016). Deactivation of Mcl-1 by dual-function small-molecule inhibitors targeting the Bcl-2 homology 3 domain and facilitating Mcl-1 ubiquitination. *Angewandte Chemie International Edition*, *55*(46), 14250–14256. doi:10.1002/anie.201606543

Sopkova-de Oliveira Santos, J., Voisin-Chiret, A. S., Burzicki, G., Sebaoun, L., Sebban, M., Lohier, J.-F., … Rault, S. (2012). Structural characterizations of oligopyridyl foldamers, α-helix mimetics. *Journal of Chemical Information and Modeling*, *52*(2), 429–439. doi:10.1021/ci200424a

Stewart, M. L., Fire, E., Keating, A. E., & Walensky, L. D. (2010). The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nature Chemical Biology*, *6*(8), 595–601. doi:10.1038/nchembio.391

Tanaka, Y., Aikawa, K., Nishida, G., Homma, M., Sogabe, S., Igaki, S., … Ishikawa, T. (2013). Discovery of potent Mcl-1/Bcl-xL dual inhibitors by using a hybridization strategy based on structural analysis of target proteins. *Journal of Medicinal Chemistry*, *56*(23), 9635–9645. doi:10.1021/jm401170c

Tron, A. E., Belmonte, M. A., Adam, A., Aquila, B. M., Boise, L. H., Chiarparin, E., … Hird, A. W. (2018). Discovery of Mcl-1-specific inhibitor AZD5991 and preclinical activity in multiple myeloma and acute myeloid leukemia. *Nature Communications*, *9*(1), 5341. doi:10.1038/s41467-018-07551-w

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., … Mackerell, A. D. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, *31*, 671–690. doi:10.1002/jcc.21367

Wan, Y., Dai, N., Tang, Z., & Fang, H. (2018). Small-molecule Mcl-1 inhibitors: Emerging anti-tumor agents. *European Journal of Medicinal Chemistry*, *146*, 471–482. doi:10.1016/j.ejmech.2018.01.076

Wang, A., Song, T., Wang, Z., Liu, Y., Fan, Y., Zhang, Y., & Zhang, Z. (2016). Mechanism of Mcl-1 conformational regulation upon small molecule binding revealed by molecular dynamic simulation. *Chemical Biology & Drug Design*, *87*(4), 551–561. doi:10.1111/cbdd.12679

Waudby, C. A., Ramos, A., Cabrita, L. D., & Christodoulou, J. (2016). Two-dimensional NMR lineshape analysis. *Scientific Reports*, *6*, 24826.doi:10.1038/srep24826

Wiesner, S., & Sprangers, R. (2015). Methyl groups as NMR probes for biomolecular interactions. *Current Opinion in Structural Biology*, *35*, 60–67. doi:10.1016/j.sbi.2015.08.010

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *73*, 1–16. doi:10.1016/j.pnmrs.2013.02.001

Yu, W., He, X., Vanommeslaeghe, K., & MacKerell, A. D. (2012). Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry*, *33*(31), 2451–2468. doi:10.1002/jcc.23067

Zhang, Z., Song, T., Li, X., Wu, Z., Feng, Y., Xie, F., … Chen, H. (2013). Novel soluble myeloid cell leukemia sequence 1 (Mcl-1) inhibitor (E,E)-2-(benzylaminocarbonyl)-3-styrylacrylonitrile (4g) developed using a fragment-based approach. *European Journal of Medicinal Chemistry*, *59*, 141–149. doi:10.1016/j.ejmech.2012.10.050

SUPPORTING INFORMATION

Figure S1. Modification of dihedral angle parametrization between adjacent pyridine rings applied during molecular dynamics simulations. Red curve CGENFF parametrization, blue curve potential energy profile calculated by *ab initio* simulation (Gaussian: HF/3-21G) and green curve modified parametrization.

Table S1. The definition of pockets in different Mcl-1:Inhibitor complex models (Cut off=5Å)

Figure S2. The NMR ^{1}H-^{15}N HSQC spectra show chemical shift perturbations caused by the direct interaction of Pyridoclax with Mcl-1. An overlay of the control spectrum (blue) with the spectrum of Mcl-1 bound to Pyridoclax at ratio 1:1 (red) ([^{15}N Mcl-1] = 50 µM and [Pyridoclax] = 50 µM, 298 k, PBS -pH7, 5% DMSO-d_6, 600MHz)

Figure S3. Overlay of ^{1}H-^{13}C SOFAST HMQC of ^{15}N-^{13}C Mcl-1 recorded in absence (blue) and in presence of Pyridoclax at ratio 1:5 (red) ([^{15}N-^{13}C Mcl-1] = 50 µM and [Pyridoclax] = 250 µM, 298 k, PBS -pH7, 5% DMSO-d_6, 600MHz)

Table S2. Estimated distances from NMR data (NOE) of Pyridoclax in solution (DMSO d_6)

Table S3. (A) Twist angle values of the ten lowest energy conformations of Pyridoclax from molecular dynamics simulations with NOE constraints; (B) The observed twist angle values between the Pyridoclax pyridine rings in X-ray structure and NMR one compared to the theoretical prediction.

Figure S4. Mcl-1 backbone root-mean-square deviations of the studied Pyridoclax/Mcl-1 complexes with respect to initial structure along the whole trajectory; pose 1 pink, pose 2 cyan and pose 3 orange curve.

Figure S5. Mcl-1 backbone root-mean-square fluctuations averaged along the simulation trajectory. Black lines boxes in bottom of panel highlight the eight α-helical regions of the protein.

Figure S6. Clustering presentation over time.

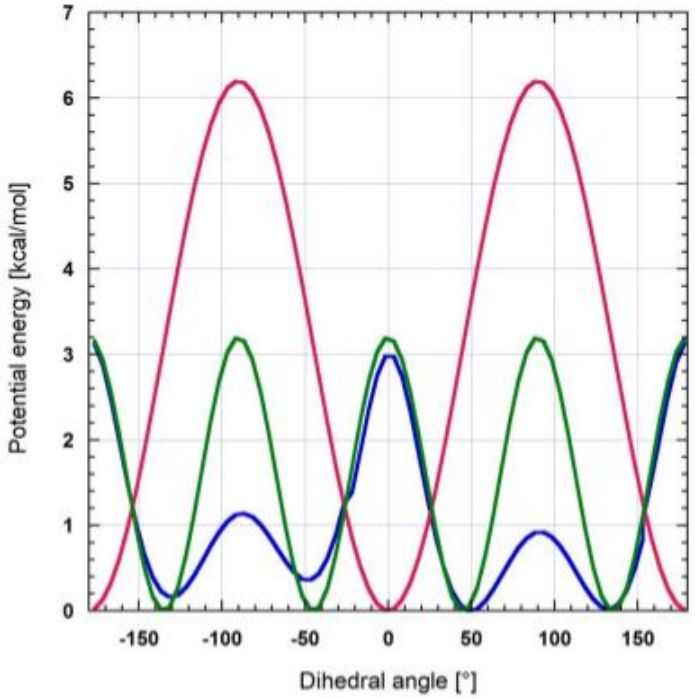


Figure S1. Modification of dihedral angle parametrization between adjacent pyridine rings applied during molecular dynamics simulations. Red curve CGENFF parametrization, blue curve potential energy profile calculated by ab initio simulation (Gaussian: HF/3-21G) and green curve modified parametrization.

Table S1. The definition of pockets in different Mcl-1:Inhibitor complex models (Cut off=5Å)

Code PDB	Mcl-1	Partner	Structure resolution	Key residues of partner (h1-h4)	Key residues of hydrophobic pockets of Mcl-1 protein (P1-P4)	Ref
3MK8	Human 158 residues (170-327)	hMcl-1- BH3 21 residues (208-228)	X-ray Diffraction (2,321 Å)	**h1** : L210 **h2** : L213 **h3** : V216 **h4** : V220	**P1** : V249, M231 **P2** : F270, L267, V253, F228 **P3** : A227, H224, T266 **P4** : V220, V265, V216	1
3PK1	Human 153 residues (174-326)	hBax-BH3 34 residues (48-81)	X-ray Diffraction (2,486 Å)	**h1** : L59 **h2** : L63 **h3** : I66 **h4** : L70	**P1** : K234, L235, M231, V249 **P2** : F270, L267, V53, **P3** : A227, F228, H224, T266 **P4** : V216, V220, V265	2
5C3F	Human 156 residues (172-327)	hBid-BH3 24 residues (79-102)	X-ray Diffraction (1,430 Å)	**h1** : I86 **h2** : L90 **h3** : V93 **h4** :NLE97	**P1** : L235, K234, V249 **P2** : M231, F270, F228, L267, V253 **P3** : H224, T266 **P4** : V220, V216	3
2KBW	Human 164 residues (163-326)	hBid-BH3 35 residues (72-106)	NMR solution	**h1** : I86 **h2** : L90 **h3** : V93 **h4** : M97	**P1** : L235, K234, V249, M231 **P2** : F270, L267, V253 **P3** : H224, A227, T266 **P4** : V220, V216, V265	4
2PQK	Human 158 residues (170-327)	Bim-BH3 29 residues (49-77)	X-ray Diffraction (2,000 Å)	**h1** : I58 **h2** : L62 **h3** : I65 **h4** : F69	**P1** : L235, K234 **P2** : M231, F270, L267, V253 **P3** : H224, A227, F228, T266 **P4** : V220, V216, V265	5

(1) Stewart, M. L.; Fire, E.; Keating, A. E.; Walensky, L. D. The MCL-1 BH3 Helix Is an Exclusive MCL-1 Inhibitor and Apoptosis Sensitizer. *Nat. Chem. Biol.* **2010**, *6* (8), 595–601.

(2) Czabotar, P. E.; Lee, E. F.; Thompson, G. V.; Wardak, A. Z.; Fairlie, W. D.; Colman, P. M. Mutation to Bax beyond the BH3 Domain Disrupts Interactions with Pro-Survival Proteins and Promotes Apoptosis. *J.Biol.Chem.* **2011**, *286*, 7123–7131.

(3) Miles, J. A.; Yeo, D. J.; Rowell, P.; Rodriguez-Marin, S.; Pask, C. M.; Warriner, S. L.; Edwards, T. A.; Wilson, A. J. Hydrocarbon Constrained Peptides - Understanding Preorganisation and Binding Affinity. *Chem Sci* **2016**, *7*, 3694–3702.

(4) Liu, Q.; Moldoveanu, T.; Sprules, T.; Matta-Camacho, E.; Mansur-Azzam, N.; Gehring, K. Apoptotic Regulation by MCL-1 through Heterodimerization. *J. Biol. Chem.* **2010**, *285* (25), 19615–19624.

(5) Fire, E.; Gulla, S. V.; Grant, R. A.; Keating, A. E. Mcl-1-Bim Complexes Accommodate Surprising Point Mutations via Minor Structural Changes. *Protein Sci* **2010**, *19*, 507–519.

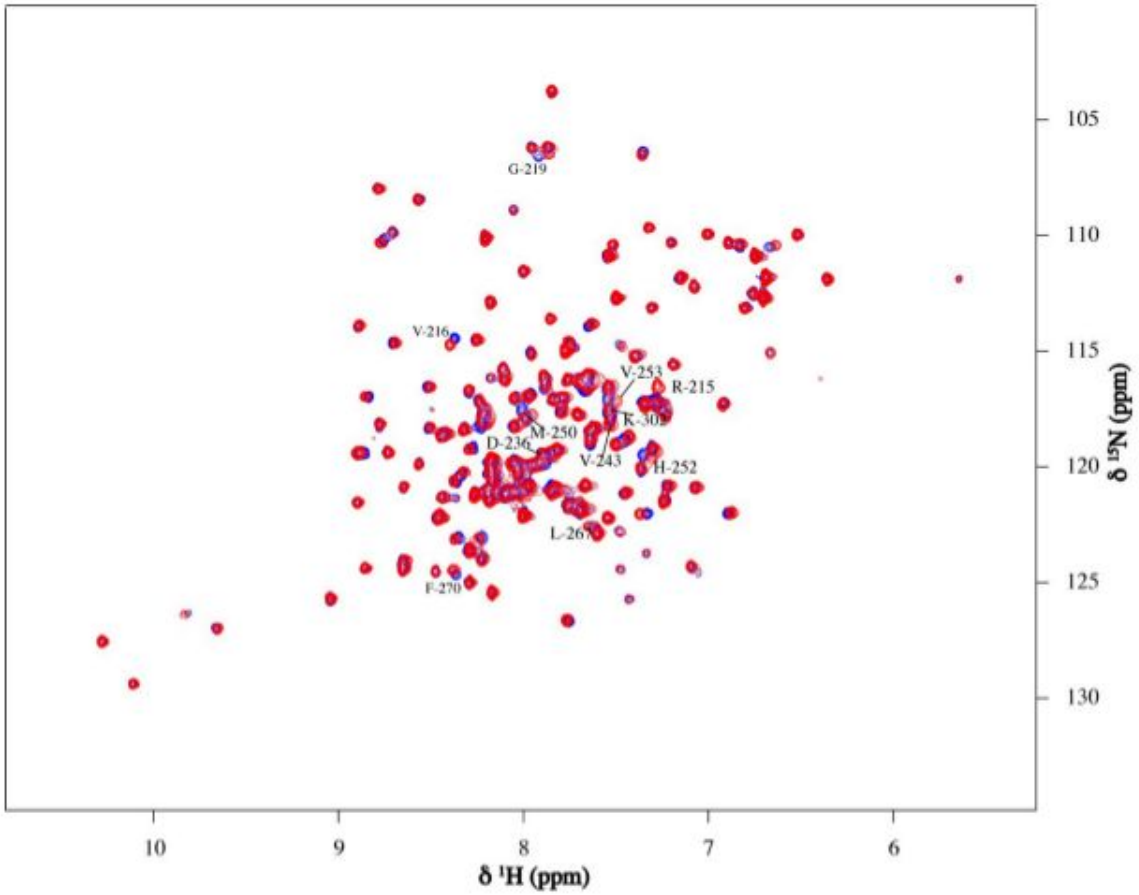


Figure S2. The NMR ^{1}H-^{15}N HSQC spectra show chemical shift perturbations caused by the direct interaction of Pyridoclax with Mcl-1. An overlay of the control spectrum (blue) with the spectrum of Mcl-1 bound to Pyridoclax at ratio 1:1 (red) ([^{15}N Mcl-1] = 50 µM and [Pyridoclax] = 50 µM, 298 k, PBS -pH7, 5% DMSO-d_6, 600MHz)

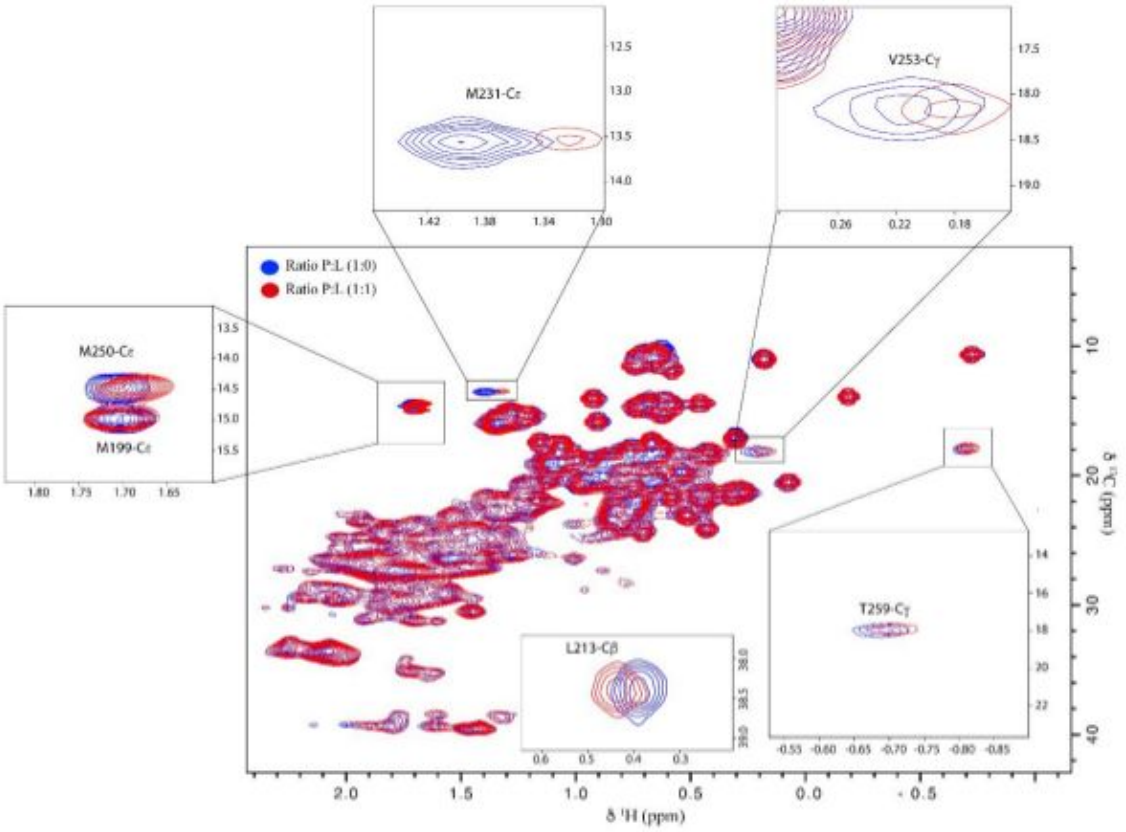


Figure S3. Overlay of ^{1}H-^{13}C SOFAST HMQC of ^{15}N-^{13}C Mcl-1 recorded in absence (blue) and in presence of Pyridoclax at ratio 1:5 (red) ([^{15}N-^{13}C Mcl-1] = 50 µM and [Pyridoclax] = 250 µM, 298 k, PBS -pH7, 5% DMSO-d_6, 600MHz)

Table S2. Estimated Distances from NMR data (NOE) of Pyridoclax in solution (DMSO d_6)

NOE	NOE estimated distances (Å)
D5/C6	2.3 ± 0.5
C3/C6	2.3 ± 0.5
C5/H10	2.7 ± 0.5
C3/H10	2.8 ± 0.6
H17/H10	2.4 ± 0.5
H11/H10	3.2 ± 0.6
A1/A2	2.5 ± 0.5
D1/D2	2.5 ± 0.5
B3/H11	2.5 ± 0.5
D5/D3	4.3 ± 0.8
A5/A3	4.0 ± 0.8
B5/A3	2.9 ± 0.6
B3/A3	2.9 ± 0.6
A5/B3	2.8 ± 0.6
A3/B5	3.5 ± 0.7

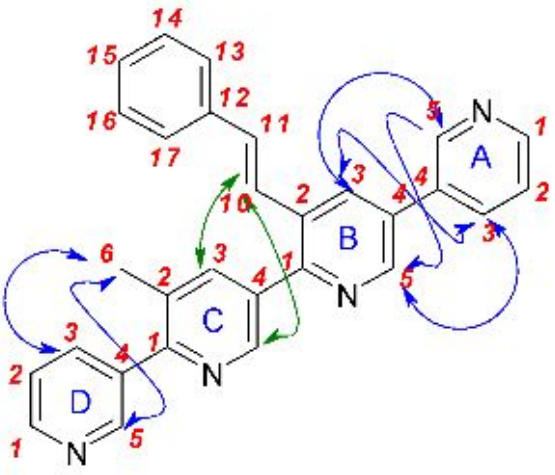


Table S3. (A) Twist angle values of the ten lowest energy conformations of Pyridoclax from molecular dynamics simulation with NOE constraints.

Energy [kcal/mol]	**Twist Angles**		
	∠ AB	∠ BC	∠ CD
128,35835	-71.47	-49.22	44.22
128,35835	-71.48	-49.21	44.18
128,35835	-71.46	-49.18	44.18
128,35835	-71.46	-49.23	44.18
128,35835	-71.46	-49.21	44.20
128,35835	-71.47	-49.22	44.21
128,35835	-71.47	-49.22	44.18
128,35835	-71.46	-49.23	44.18
128,35835	-71.47	-49.20	44.19
128,35835	-71.47	-49.22	44.18

(B) The observed twist angle values between the Pyridoclax pyridine rings in X-ray structure and NMR one compared to the theoretical prediction.

	X-ray conformer1	X-ray conformer2	NMR conformer	Theoretical[6,7] prediction
∠ Ring A-Ring B	-26.33°	+26.33°	-71.47°°±0.01°	±40-±60
∠ Ring B-Ring C	+39.31°	-39.31°	-49.21°±0.02°	±30-±60
∠ Ring C-Ring D	-49.27°	+49.27°	+44.19°±0.01°	±30-±60

(6) Sopkova-de Oliveira Santos, J.; Voisin-Chiret, A. S.; Burzicki, G.; Sebaoun, L.; Sebban, M.; Lohier, J.-F.; Legay, R.; Oulyadi, H.; Bureau, R.; Rault, S. Structural Characterizations of Oligopyridyl Foldamers, α-Helix Mimetics. *J. Chem. Inf. Model*. **2012**, 52 (2), 429–439.

(7) Perato, S.; Fogha. J.; Sebban, M. ; Voisin-Chiret, A. S.; Sopkova-de Oliveira Santos, J.; Oulyadi, H.; and Rault, S. Conformation Control of Abiotic α-Helical Foldamers. *J. Chem. Inf. Model.* **2013**, 53 (10), 2671–2680.

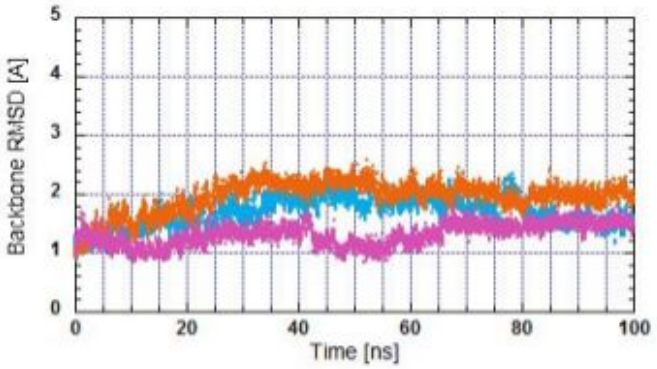


Figure S4. Mcl-1 backbone root-mean-square deviations of the studied Pyridoclax/Mcl-1 complexes with respect to initial structure along the whole trajectory; pose 1 pink, pose 2 cyan and pose 3 orange curve.

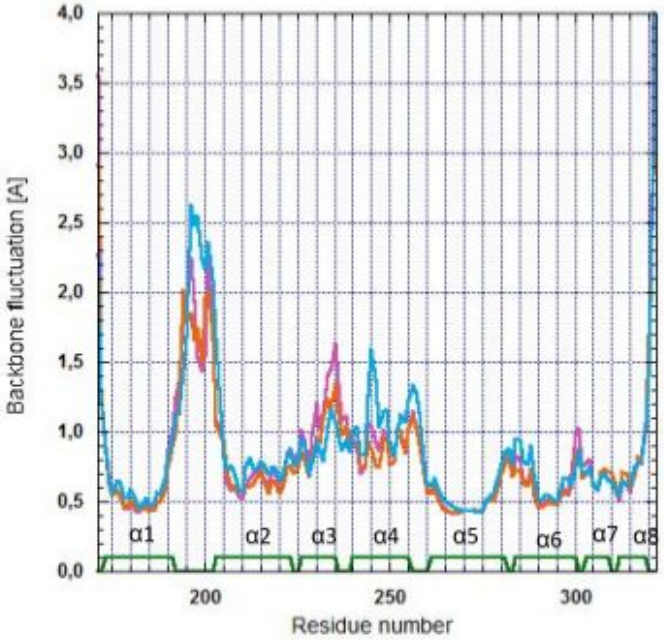
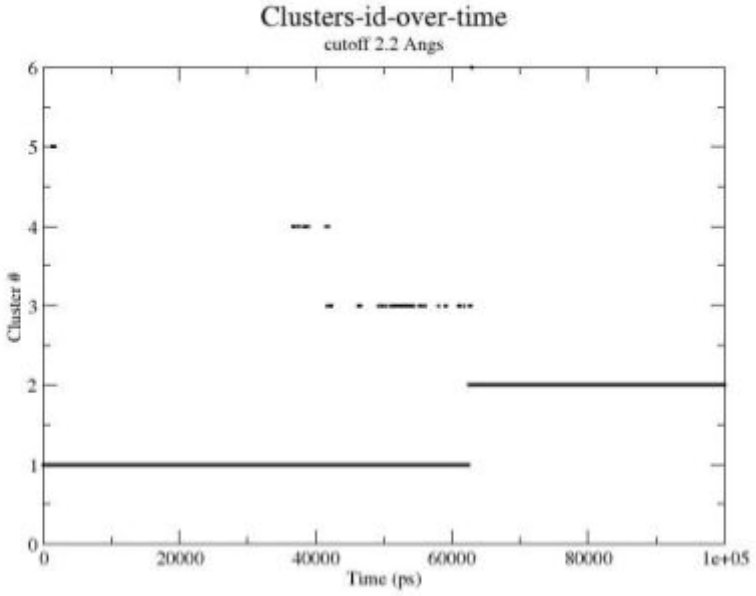


Figure S5. Mcl-1 backbone root-mean-square fluctuations averaged along the simulation trajectory. Black lines boxes in bottom of panel highlight the eight α-helical regions of the protein.



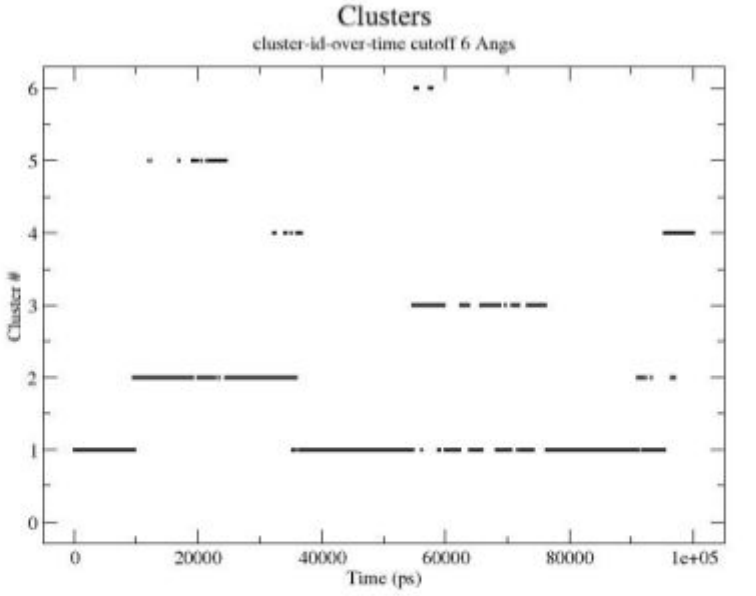


Figure S6. Clustering presentation over time.

3. Conclusion et perspectives

L'étude de l'association entre le Pyridoclax et Mcl-1 par spectroscopie RMN et simulations de dynamique moléculaire, a montré que le Pyridoclax se fixe sur le site orthostérique de Mcl-1. Le Pyridoclax établit des contacts hydrophobes avec les résidus de la poche P2 via son groupement styryl. Une interaction avec la Lys234 a également été relevée et pourrait expliquer la sélectivité du Pyridoclax à Mcl-1.

Des études complémentaires seront prochainement effectuées, notamment pour caractériser le processus d'interaction, et estimer des constantes d'affinité via des dynamiques moléculaires biaisées ("MD Binding approach" ou "Funnel Metadynamics").

4. Références

(1) Kerr, J. F.; Wyllie, A. H.; Currie, A. R. Apoptosis: A Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics. *Br. J. Cancer* **1972**, *26* (4), 239–257.

(2) Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell*. 2011, pp 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

(3) Gelinas, C. BH3-Only Proteins in Control: Specificity Regulates MCL-1 and BAK-Mediated Apoptosis. *Genes & Development*. 2005, pp 1263–1268. https://doi.org/10.1101/gad.1326205.

(4) Warr, M. R.; Shore, G. C. Unique Biology of Mcl-1: Therapeutic Opportunities in Cancer. *Curr. Mol. Med.* **2008**, *8* (2), 138–147.

(5) Petros, A. M.; Nettesheim, D. G.; Wang, Y.; Olejniczak, E. T.; Meadows, R. P.; Mack, J.; Swift, K.; Matayoshi, E. D.; Zhang, H.; Fesik, S. W.; Thompson, C. B. Rationale for Bcl-XL/Bad Peptide Complex Formation from Structure, Mutagenesis, and Biophysical Studies. *Protein Science*. 2000, pp 2528–2534. https://doi.org/10.1110/ps.9.12.2528.

(6) Delbridge, A. R. D.; Grabow, S.; Strasser, A.; Vaux, D. L. Thirty Years of BCL-2: Translating Cell Death Discoveries into Novel Cancer Therapies. *Nature Reviews Cancer*. 2016, pp 99–109. https://doi.org/10.1038/nrc.2015.17.

(7) Abid, M.; Sonawane, Y. A.; Contreras, J. I.; Rana, S.; Natarajan, A. Recent Advances in Cancer Drug Development: Targeting Induced Myeloid Cell Leukemia-1 (Mcl-1) Differentiation Protein. *Curr. Med. Chem.* **2017**, *24* (40), 4488–4514.

(8) Abulwerdi, F.; Liao, C.; Liu, M.; Azmi, A. S.; Aboukameel, A.; Mady, A. S. A.; Gulappa, T.; Cierpicki, T.; Owens, S.; Zhang, T.; Sun, D.; Stuckey, J. A.; Mohammad, R. M.; Nikolovska-Coleska, Z. A Novel Small-Molecule Inhibitor of Mcl-1 Blocks Pancreatic Cancer Growth In Vitro and In Vivo. *Molecular Cancer Therapeutics*. 2014, pp 565–575. https://doi.org/10.1158/1535-7163.mct-12-0767.

(9) Abulwerdi, F. A.; Liao, C.; Mady, A. S.; Gavin, J.; Shen, C.; Cierpicki, T.; Stuckey, J. A.; Showalter, H. D. H.; Nikolovska-Coleska, Z. 3-Substituted-N-(4-Hydroxynaphthalen-1-Yl)arylsulfonamides as a Novel Class of Selective Mcl-1 Inhibitors: Structure-Based Design, Synthesis, SAR, and Biological Evaluation. *J. Med. Chem.* **2014**, *57* (10), 4111–4133.

(10) Ashkenazi, A.; Fairbrother, W. J.; Leverson, J. D.; Souers, A. J. From Basic Apoptosis Discoveries to Advanced Selective BCL-2 Family Inhibitors. *Nat. Rev. Drug Discov.* **2017**, *16* (4), 273–284.

(11) Beekman, A. M.; Howell, L. A. Small-Molecule and Peptide Inhibitors of the Pro-Survival Protein Mcl-1. *ChemMedChem*. 2016, pp 802–813. https://doi.org/10.1002/cmdc.201500497.

(12) Belmar, J.; Fesik, S. W. Small Molecule Mcl-1 Inhibitors for the Treatment of Cancer. *Pharmacol. Ther.* **2015**, *145*, 76–84.

(13) Caenepeel, S.; Brown, S. P.; Belmontes, B.; Moody, G.; Keegan, K. S.; Chui, D.; Whittington, D. A.; Huang, X.; Poppe, L.; Cheng, A. C.; Cardozo, M.; Houze, J.; Li, Y.; Lucas, B.; Paras, N. A.; Wang, X.; Taygerly, J. P.; Vimolratana, M.; Zancanella, M.; Zhu, L.; Cajulis, E.; Osgood, T.; Sun, J.; Damon, L.; Egan, R. K.; Greninger, P.; McClanaghan, J. D.; Gong, J.; Moujalled, D.; Pomilio, G.; Beltran, P.; Benes, C. H.; Roberts, A. W.; Huang, D. C. S.; Wei, A.; Canon, J.; Coxon, A.; Hughes, P. E. AMG 176, a Selective MCL1 Inhibitor, Is Effective in Hematological Cancer Models Alone

and in Combination with Established Therapies. *Cancer Discovery*. 2018, p CD – 18. https://doi.org/10.1158/2159-8290.cd-18-0387.

(14) Kotschy, A.; Szlavik, Z.; Murray, J.; Davidson, J.; Maragno, A. L.; Le Toumelin-Braizat, G.; Chanrion, M.; Kelly, G. L.; Gong, J.-N.; Moujalled, D. M.; Bruno, A.; Csekei, M.; Paczal, A.; Szabo, Z. B.; Sipos, S.; Radics, G.; Proszenyak, A.; Balint, B.; Ondi, L.; Blasko, G.; Robertson, A.; Surgenor, A.; Dokurno, P.; Chen, I.; Matassova, N.; Smith, J.; Pedder, C.; Graham, C.; Studeny, A.; Lysiak-Auvity, G.; Girard, A.-M.; Gravé, F.; Segal, D.; Riffkin, C. D.; Pomilio, G.; Galbraith, L. C. A.; Aubrey, B. J.; Brennan, M. S.; Herold, M. J.; Chang, C.; Guasconi, G.; Cauquil, N.; Melchiore, F.; Guigal-Stephan, N.; Lockhart, B.; Colland, F.; Hickman, J. A.; Roberts, A. W.; Huang, D. C. S.; Wei, A. H.; Strasser, A.; Lessene, G.; Geneste, O. The MCL1 Inhibitor S63845 Is Tolerable and Effective in Diverse Cancer Models. *Nature* **2016**, *538* (7626), 477–482.

(15) Petros, A. M.; Swann, S. L.; Song, D.; Swinger, K.; Park, C.; Zhang, H.; Wendt, M. D.; Kunzer, A. R.; Souers, A. J.; Sun, C. Fragment-Based Discovery of Potent Inhibitors of the Anti-Apoptotic MCL-1 Protein. *Bioorg. Med. Chem. Lett.* **2014**, *24* (6), 1484–1488.

(16) Tron, A. E.; Belmonte, M. A.; Adam, A.; Aquila, B. M.; Boise, L. H.; Chiarparin, E.; Cidado, J.; Embrey, K. J.; Gangl, E.; Gibbons, F. D.; Gregory, G. P.; Hargreaves, D.; Adam Hendricks, J.; Johannes, J. W.; Johnstone, R. W.; Kazmirski, S. L.; Kettle, J. G.; Lamb, M. L.; Matulis, S. M.; Nooka, A. K.; Packer, M. J.; Peng, B.; Rawlins, P. B.; Robbins, D. W.; Schuller, A. G.; Su, N.; Yang, W.; Ye, Q.; Zheng, X.; Paul Secrist, J.; Clark, E. A.; Wilson, D. M.; Fawell, S. E.; Hird, A. W. Discovery of Mcl-1-Specific Inhibitor AZD5991 and Preclinical Activity in Multiple Myeloma and Acute Myeloid Leukemia. *Nature Communications*. 2018. https://doi.org/10.1038/s41467-018-07551-w.

(17) Wan, Y.; Dai, N.; Tang, Z.; Fang, H. Small-Molecule Mcl-1 Inhibitors: Emerging Anti-Tumor Agents. *European Journal of Medicinal Chemistry*. 2018, pp 471–482. https://doi.org/10.1016/j.ejmech.2018.01.076.

(18) Gloaguen, C.; Voisin-Chiret, A. S.; Sopkova-de Oliveira Santos, J.; Fogha, J.; Gautier, F.; De Giorgi, M.; Burzicki, G.; Perato, S.; Pétigny-Lechartier, C.; Simonin-Le Jeune, K.; Brotin, E.; Goux, D.; N'Diaye, M.; Lambert, B.; Louis, M.-H.; Ligat, L.; Lopez, F.; Juin, P.; Bureau, R.; Rault, S.; Poulain, L. First Evidence That Oligopyridines, α-Helix Foldamers, Inhibit Mcl-1 and Sensitize Ovarian Carcinoma Cells to Bcl-xL-Targeting Strategies. *J. Med. Chem.* **2015**, *58* (4), 1644–1668.

(19) Cala, O.; Guillière, F.; Krimm, I. NMR-Based Analysis of Protein–ligand Interactions. *Analytical and Bioanalytical Chemistry*. 2014, pp 943–956. https://doi.org/10.1007/s00216-013-6931-0.

(20) Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed Engl.* **2003**, *42* (8), 864–890.

(21) Orts, J.; Gossert, A. D. Structure Determination of Protein-Ligand Complexes by NMR in Solution. *Methods*. 2018, pp 3–25. https://doi.org/10.1016/j.ymeth.2018.01.019.

(22) Friberg, A.; Vigil, D.; Zhao, B.; Nathan Daniels, R.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of Potent Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods and Structure-Based Design. *Journal of Medicinal Chemistry*. 2013, pp 15–30. https://doi.org/10.1021/jm301448p.

(23) Song, T.; Wang, Z.; Ji, F.; Feng, Y.; Fan, Y.; Chai, G.; Li, X.; Li, Z.; Zhang, Z. Cover

Picture: Deactivation of Mcl-1 by Dual-Function Small-Molecule Inhibitors Targeting the Bcl-2 Homology 3 Domain and Facilitating Mcl-1 Ubiquitination (Angew. Chem. Int. Ed. 46/2016). *Angewandte Chemie International Edition*. 2016, pp 14179–14179. https://doi.org/10.1002/anie.201609560.

(24) Liu, G.; Poppe, L.; Aoki, K.; Yamane, H.; Lewis, J.; Szyperski, T. High-Quality NMR Structure of Human Anti-Apoptotic Protein Domain Mcl-1(171-327) for Cancer Drug Design. *PLoS One* **2014**, *9* (5), e96521.

(25) Grzesiek, S.; Bax, A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein. *Journal of Magnetic Resonance (1969)*. 1992, pp 432–440. https://doi.org/10.1016/0022-2364(92)90099-s.

(26) Grzesiek, S.; Bax, A. Amino Acid Type Determination in the Sequential Assignment Procedure of Uniformly 13C/15N-Enriched Proteins. *Journal of Biomolecular NMR*. 1993. https://doi.org/10.1007/bf00178261.

(27) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics*. 2015, pp 1325–1327. https://doi.org/10.1093/bioinformatics/btu830.

(28) Mori, S.; Abeygunawardana, C.; Johnson, M. O.; Vanzijl, P. C. M. Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation. *Journal of Magnetic Resonance, Series B*. 1995, pp 94–98. https://doi.org/10.1006/jmrb.1995.1109.

(29) Marion, D.; Wüthrich, K. Application of Phase Sensitive Two-Dimensional Correlated Spectroscopy (COSY) for Measurements of 1H-1H Spin-Spin Coupling Constants in Proteins. *Biochemical and Biophysical Research Communications*. 1983, pp 967–974. https://doi.org/10.1016/0006-291x(83)91093-8.

(30) Skinner, S. P.; Fogh, R. H.; Boucher, W.; Ragan, T. J.; Mureddu, L. G.; Vuister, G. W. CcpNmr AnalysisAssign: A Flexible Platform for Integrated NMR Analysis. *Journal of Biomolecular NMR*. 2016, pp 111–124. https://doi.org/10.1007/s10858-016-0060-y.

(31) Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J. Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*. 2016. https://doi.org/10.1038/srep24826.

(32) Williamson, M. P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2013, pp 1–16. https://doi.org/10.1016/j.pnmrs.2013.02.001.

(33) Schanda, P.; Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *Journal of the American Chemical Society*. 2005, pp 8014–8015. https://doi.org/10.1021/ja051306e.

(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry*. 2009, p NA – NA. https://doi.org/10.1002/jcc.21367.

(35) Pelz, N. F.; Bian, Z.; Zhao, B.; Shaw, S.; Tarr, J. C.; Belmar, J.; Gregg, C.; Camper, D. V.; Goodwin, C. M.; Arnold, A. L.; Sensintaffar, J. L.; Friberg, A.; Rossanese, O. W.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of 2-Indole-Acylsulfonamide Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods. *J. Med. Chem.* **2016**, *59* (5), 2054–2066.

(36) Bruncko, M.; Wang, L.; Sheppard, G. S.; Phillips, D. C.; Tahir, S. K.; Xue, J.; Erickson, S.; Fidanze, S.; Fry, E.; Hasvold, L.; Jenkins, G. J.; Jin, S.; Judge, R. A.; Kovar, P. J.;

Madar, D.; Nimmer, P.; Park, C.; Petros, A. M.; Rosenberg, S. H.; Smith, M. L.; Song, X.; Sun, C.; Tao, Z.-F.; Wang, X.; Xiao, Y.; Zhang, H.; Tse, C.; Leverson, J. D.; Elmore, S. W.; Souers, A. J. Structure-Guided Design of a Series of MCL-1 Inhibitors with High Affinity and Selectivity. *J. Med. Chem.* **2015**, *58* (5), 2180–2194.

(37) Jenson, J. M.; Ryan, J. A.; Grant, R. A.; Letai, A.; Keating, A. E. Epistatic Mutations in PUMA BH3 Drive an Alternate Binding Mode to Potently and Selectively Inhibit Anti-Apoptotic Bfl-1. *Elife* **2017**, *6*. https://doi.org/10.7554/eLife.25541.

(38) Pons, J.-L.; Labesse, G. @TOME-2: A New Pipeline for Comparative Modeling of Protein-Ligand Complexes. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W485–W491.

(39) Clifton, M. C.; Dranow, D. M.; Leed, A.; Fulroth, B.; Fairman, J. W.; Abendroth, J.; Atkins, K. A.; Wallace, E.; Fan, D.; Xu, G.; Ni, Z. J.; Daniels, D.; Van Drie, J.; Wei, G.; Burgin, A. B.; Golub, T. R.; Hubbard, B. K.; Serrano-Wu, M. H. A Maltose-Binding Protein Fusion Construct Yields a Robust Crystallography Platform for MCL1. *PLoS One* **2015**, *10* (4), e0125010.

(40) Burke, J. P.; Bian, Z.; Shaw, S.; Zhao, B.; Goodwin, C. M.; Belmar, J.; Browning, C. F.; Vigil, D.; Friberg, A.; Camper, D. V.; Rossanese, O. W.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of Tricyclic Indoles That Potently Inhibit Mcl-1 Using Fragment-Based Methods and Structure-Based Design. *Journal of Medicinal Chemistry*. 2015, pp 3794–3805. https://doi.org/10.1021/jm501984f.

(41) Tanaka, Y.; Aikawa, K.; Nishida, G.; Homma, M.; Sogabe, S.; Igaki, S.; Hayano, Y.; Sameshima, T.; Miyahisa, I.; Kawamoto, T.; Tawada, M.; Imai, Y.; Inazuka, M.; Cho, N.; Imaeda, Y.; Ishikawa, T. Discovery of Potent Mcl-1/Bcl-xL Dual Inhibitors by Using a Hybridization Strategy Based on Structural Analysis of Target Proteins. *J. Med. Chem.* **2013**, *56* (23), 9635–9645.

(42) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *Journal of Molecular Biology*. 1995, pp 43–53. https://doi.org/10.1016/s0022-2836(95)80037-9.

(43) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking 1 1Edited by F. E. Cohen. *Journal of Molecular Biology*. 1997, pp 727–748. https://doi.org/10.1006/jmbi.1996.0897.

(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(45) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *Journal of Chemical Theory and Computation*. 2012, pp 3257–3273. https://doi.org/10.1021/ct300400x.

(46) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods*. 2017, pp 71–73. https://doi.org/10.1038/nmeth.4067.

(47) Santos, J. S. O.; Voisin-Chiret, A. S.; Burzicki, G.; Sebaoun, L.; Sebban, M.; Lohier, J.-F.; Legay, R.; Oulyadi, H.; Ronan Bureau; Rault, S. Structural Characterizations of Oligopyridyl Foldamers, α-Helix Mimetics. *Journal of Chemical Information and Modeling*. 2012, pp 429–439. https://doi.org/10.1021/ci200424a.

(48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.

Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics*. 1983, pp 926–935. https://doi.org/10.1063/1.445869.

(49) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.

(50) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: AnN · log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics*. 1993, pp 10089–10092. https://doi.org/10.1063/1.464397.

(51) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *Journal of Computational Physics*. 1977, pp 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(52) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*. 2009, pp 1545–1614. https://doi.org/10.1002/jcc.21287.

(53) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*. 1999, pp 236–240. https://doi.org/3.0.co;2-m.">10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.0.co;2-m.

(54) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1-3), 43–56.

(55) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.

(56) Czabotar, P. E.; Colman, P. M. Crystal Structure of Mcl-1 in Complex with the BaxBH3 Domain, 2010. https://doi.org/10.2210/pdb3pk1/pdb.

(57) Bare, E.; Grant, R. A.; Keating, A. E. X-Ray Crystal Structure of Human Mcl-1 in Complex with Bim BH3, 2007. https://doi.org/10.2210/pdb2pqk/pdb.

(58) Liu, Q.; Moldoveanu, T.; Sprules, T.; Matta-Camacho, E.; Mansur-Azzam, N.; Gehring, K. Apoptotic Regulation by MCL-1 through Heterodimerization. *J. Biol. Chem.* **2010**, *285* (25), 19615–19624.

(59) Miles, J. A.; Yeo, D. J.; Rowell, P.; Rodriguez-Marin, S.; Pask, C. M.; Warriner, S. L.; Edwards, T. A.; Wilson, A. J. Crystal Structure of Mcl-1 Bound to BID-MM, 2016. https://doi.org/10.2210/pdb5c3f/pdb.

(60) Stewart, M. L.; Fire, E.; Keating, A. E.; Walensky, L. D. The MCL-1 BH3 Helix Is an Exclusive MCL-1 Inhibitor and Apoptosis Sensitizer. *Nat. Chem. Biol.* **2010**, *6* (8), 595–601.

(61) Cohen, N. A.; Stewart, M. L.; Gavathiotis, E.; Tepper, J. L.; Bruekner, S. R.; Koss, B.; Opferman, J. T.; Walensky, L. D. A Competitive Stapled Peptide Screen Identifies a Selective Small Molecule That Overcomes MCL-1-Dependent Leukemia Cell Survival. *Chem. Biol.* **2012**, *19* (9), 1175–1186.

(62) Modi, V.; Sankararamakrishnan, R. Binding Affinity of pro-Apoptotic BH3 Peptides for the Anti-Apoptotic Mcl-1 and A1 Proteins: Molecular Dynamics Simulations of Mcl-1

and A1 in Complex with Six Different BH3 Peptides. *J. Mol. Graph. Model.* **2017**, *73*, 115–128.

(63) Wang, A.; Song, T.; Wang, Z.; Liu, Y.; Fan, Y.; Zhang, Y.; Zhang, Z. Mechanism of Mcl-1 Conformational Regulation Upon Small Molecule Binding Revealed by Molecular Dynamic Simulation. *Chem. Biol. Drug Des.* **2016**, *87* (4), 551–561.

(64) Krzeminski, M.; Fuentes, G.; Boelens, R.; Bonvin, A. M. J. J. MINOES: A New Approach to Select a Representative Ensemble of Structures in NMR Studies of (partially) Unfolded States. Application to Delta25-PYP. *Proteins* **2009**, *74* (4), 895–904.

(65) Wiesner, S.; Sprangers, R. Methyl Groups as NMR Probes for Biomolecular Interactions. *Curr. Opin. Struct. Biol.* **2015**, *35*, 60–67.

(66) Masson, J.-F.; Battaglia, T. M.; Cramer, J.; Beaudoin, S.; Sierks, M.; Booksh, K. S. Reduction of Nonspecific Protein Binding on Surface Plasmon Resonance Biosensors. *Anal. Bioanal. Chem.* **2006**, *386* (7-8), 1951–1959.

(67) Brotin, E.; Meryet-Figuière, M.; Simonin, K.; Duval, R. E.; Villedieu, M.; Leroy-Dudal, J.; Saison-Behmoaras, E.; Gauduchon, P.; Denoyelle, C.; Poulain, L. Bcl-XL and MCL-1 Constitute Pertinent Targets in Ovarian Carcinoma and Their Concomitant Inhibition Is Sufficient to Induce Apoptosis. *Int. J. Cancer* **2010**, *126* (4), 885–895.

(68) Perato, S.; Fogha, J.; Sebban, M.; Voisin-Chiret, A. S.; Sopkova-de Oliveira Santos, J.; Oulyadi, H.; Rault, S. Conformation Control of Abiotic α-Helical Foldamers. *J. Chem. Inf. Model.* **2013**, *53* (10), 2671–2680.

CHAPITRE 3.

Espace conformationnel et inhibition allostérique de Mcl-1

1. Motivation

La caractérisation de la structure et de la dynamique des protéines est une étape primordiale pour comprendre leur fonction.

Pour notre cible d'étude, la protéine Mcl-1, plusieurs structures ont été résolues expérimentalement, notamment dans le cadre du développement d'outils thérapeutiques. Cependant, une caractérisation fine de sa dynamique n'est pas décrite dans la littérature.

Dans ce chapitre, trois objectifs ont été considérés :

- Apporter une meilleure compréhension des éléments structuraux contrôlant la dynamique de Mcl-1, à travers une analyse des ensembles structuraux.
- Affiner la description des états conformationnels que la protéine occupe, notamment sous sa forme apo.
- Détailler le mécanisme moléculaire permettant l'inhibition allostérique de Mcl-1.

2. Article 2

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

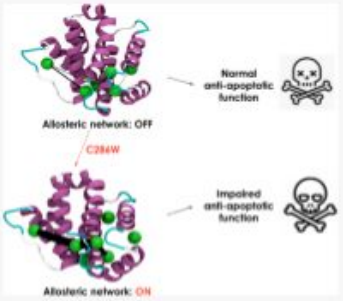
Insights into Mcl-1 Conformational States and Allosteric Inhibition Mechanism from Molecular Dynamics Simulations, Enhanced Sampling, and Pocket Crosstalk Analysis

Mohammed Benabderrahmane, Ronan Bureau, Anne Sophie Voisin-Chiret, and Jana Sopkova-de Oliveira Santos*

Cite This: J. Chem. Inf. Model. 2020, 60, 3172−3187 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this study, we explored the structural dynamics of Mcl-1, an anti-apoptotic protein. On the basis of structural ensembles, the essential dynamics was extracted and showed two major axes of variability: a breathing motion at the binding interface and a correlated motion through the internal loops. A free energy surface characterizing the breathing motion at the binding interface was generated and suggested an equilibrium between a closed conformation and a "ready to bind" conformation as the predominant states of Mcl-1 in solution. Moreover, the analysis of the dynamics along the internal loops revealed a hidden communication network of transient and cryptic pockets controlling the allosteric inhibition of Mcl-1. A detailed model joining the pocket crosstalk and salt bridge networks along the internal loops was proposed and allowed us to shed light on the key interactions governing Mcl-1's allosteric inhibition.



1. INTRODUCTION

Proteins are the workhorse of life at the molecular level, and their conformational dynamics is intimately related to their function.[1,2] Protein motions and molecular events such as binding and folding are widely modeled using the free energy landscape theory. The energy landscape was defined by Frauenfelder et al.[3] as the landscape describing the potential energy of the protein as a function of the conformational coordinates. It is a hypersurface in the high-dimensional space of the coordinates of all atoms in the protein. Taking myoglobin as an example, Frauenfelder et al.[3] elucidated a relation between the free energy landscape features and myoglobin function.

A set of computational techniques providing insights into protein conformational dynamics is now available. Molecular dynamics (MD) simulations and especially the enhanced sampling methods are suitable tools to determine the free energy landscape along a selected reaction coordinate that better describes the motion or event of interest. These methods are efficiently able to capture global conformational tendencies and the shifts in conformational populations underlying for instance allosteric modulation, protein−ligand binding, and promiscuity within protein−protein interaction (PPI) networks. For chemical hits, they can be used for optimization via estimation of binding rates.[4]

Bcl-2 family proteins are key and central switches for cell life and death. By direct PPIs between their members, they regulate the mitochondrial outer membrane permeabilization (MOMP) and hence contribute to the intrinsic apoptosis pathway. The Bcl-2 family consists of a number of evolutionarily conserved proteins that share conserved Bcl-2 homology (BH) domains, and its members are subdivided into three groups on the basis of their function: (1) multidomain anti-apoptotic proteins (Mcl-1, Bcl-2, Bcl-x_L, Bcl-W, and Bfl-1/A1), (2) pro-apoptotic multi-domain pore-forming proteins (BAX, BAK, and BOK) and (3) pro-apoptotic BH3-only proteins (BAD, BID, BIK, BIM, BMF, HRK, NOXA, PUMA, etc.). The occurrence of MOMP is the result of a subtle regulation network, and interaction models of these proteins have been proposed and discussed elsewhere.[5]

Here we focus our attention on a member of this family: Mcl-1 (Myeloid cell leukemia-1). Functionally, this member is unique among the anti-apoptotic proteins in that it is necessary for early embryonic development[6] and the survival of multiple cell lines, including lymphocytes,[7,8] hematopoietic stem cells,[9] neutrophils,[10,11] and neurons.[12] Furthermore, Mcl-1 is overexpressed in multiple types of human malignancies and is considered as a potential therapeutic target.[13]

From a structural point of view, Mcl-1 is different from the other pro-survival proteins in its larger size: 350 residues against 239 and 233 residues for Bcl-2 and Bcl-x_L, respectively. Also,

Received: March 30, 2020
Published: May 11, 2020

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

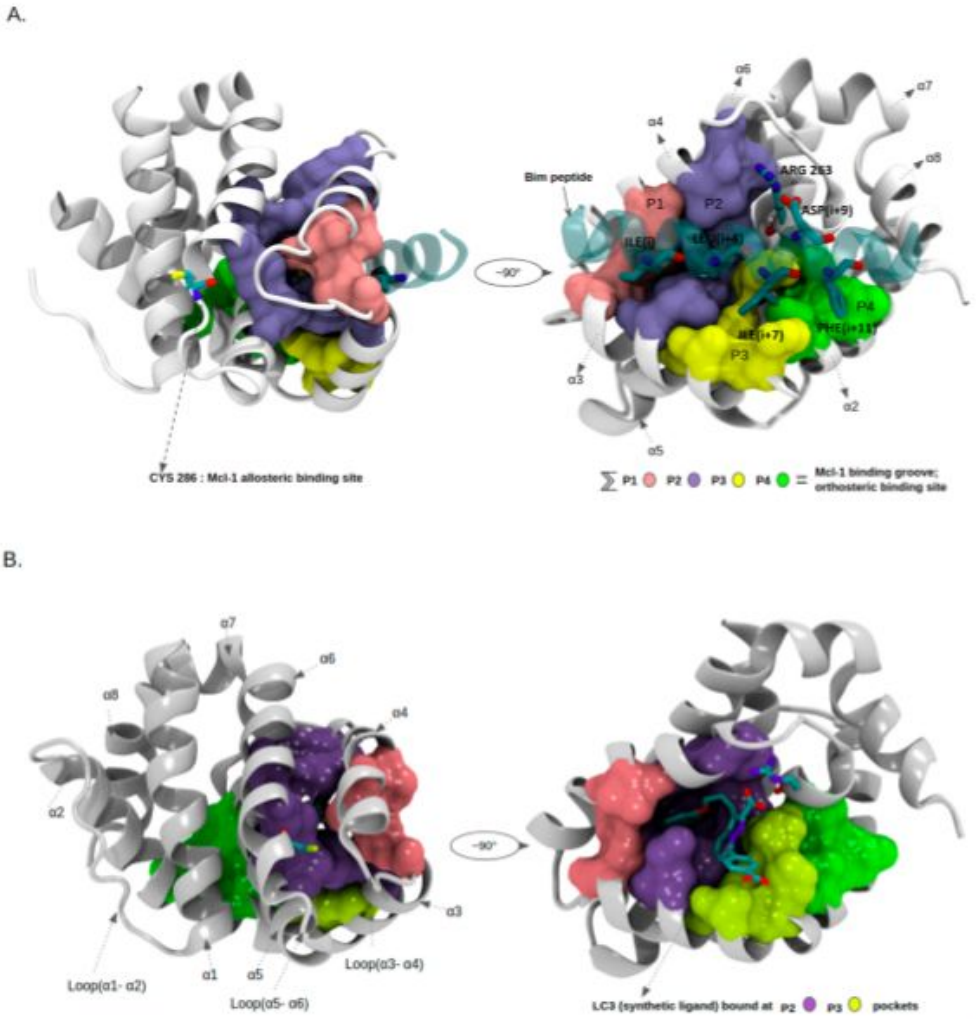


Figure 1. Key structural elements and binding sites of Mcl-1. (A) Mcl-1 binding groove hosting the BIM-BH3 peptide (PDB ID 2PQK).[19] Pockets forming the binding groove are represented as colored surfaces; the BIM-BH3 peptide is represented as a new cartoon and colored in transparent cyan. The Mcl-1 allosteric hotspot Cys286 is highlighted as well. (B) Mcl-1 binding with a synthetic ligand (LC3) (PDB ID 3WIX)[20] at the P2 and P3 pockets; the Mcl-1 internal loops are also shown.

Mcl-1 harbors a large N-terminal region enriched with proline-glutamic acid-serine-threonine (PEST) motifs related to its regulation and function,[14] although this N-terminal region (residues 1−170) is truncated in the structural models available for Mcl-1. To date, the structural data available for Mcl-1 concerns residues 170−320. This part of the protein is a well-structured region of eight α-helices with three highly conserved BH domains, BH1 through BH3, forming the canonical hydrophobic binding groove that sequesters its binding partners (Figure 1). For instance, in the case of peptide partners (Figure 1A), four conserved hydrophobic residues at positions i, $i+4$, $i+7$, and $i+11$ on the contact face of the α-helical peptide are projected into the hydrophobic cavities of the Mcl-1 binding groove. The receiving cavities of these four hydrophobic residues are usually named P1(i), P2(i+4), P3(i+7), and P4(i+11), respectively. The synthetic ligands bind in the same binding groove. Nevertheless, we should highlight the high plasticity of these pockets, especially the P2 and P3 pockets, which are the main pockets[15] targeted by the synthetic ligands (Figure 1B).

Despite the numerous structures of Mcl-1 that are publicly available, the structural dynamics and conformational ensembles of Mcl-1 in water and in apo form are less characterized, and only three human Mcl-1 (hMcl-1) apo structural models are available: one NMR model (PDB ID 2MHS)[16] and two X-ray models in which Mcl-1 was coexpressed with maltose-binding protein (PDB ID 4WMS)[17] or antibody fragments (PDB ID 6QB3)[18] to enhance the crystal growth.

Aside from the comparative study by Maity et al.[21] that demonstrated the existence of three different conformational states of Mcl-1 using ensemble structures, the conformational landscape of Mcl-1 is still less understood, and several questions are still open regarding (i) the structural elements of Mcl-1 leading the adaptation to different conformational states (apo, peptide-bound, and synthetic-ligand-bound) and (ii) the propensity and likelihood of the conformational states in

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

solution and the free energy landscape of the breathing motion at Mcl-1's binding interface.

Furthermore, Lee et al.[22] demonstrated the existence of an allosteric mode to disrupt the canonical BH3-binding activity of Mcl-1 by covalent decoration of the Cys286 residue with a small organic moiety. The authors showed as well that the C286W variant conferred this inhibitory effect and proposed an allosteric inhibition mode through rigidification of the Mcl-1 structure, leading to a shift of the Mcl-1 conformational population toward the unbound state. However, a detailed understanding of the structural mechanism behind this allosteric inhibition and the way by which the allosteric signal is transduced from the allosteric binding site toward the orthosteric binding groove of Mcl-1 is still to be uncovered.

Hence, our aim in this work is threefold:

1. to propose a better understanding of the key structural elements leading Mcl-1 to adapt to its different binding partners using ensemble structures and essential dynamics;
2. to quantify the breathing motion at the binding interface of Mcl-1 and provide a free energy surface that better describes the likelihood of the conformational states explored by Mcl-1 in solution;
3. to provide a detailed understanding of how Mcl-1 allosteric inhibition works by exposing the conformational population shift and highlighting the existence of an allosteric communication network through pocket cross-talk analysis.[23]

2. MATERIALS AND METHODS

2.1. PCA and Hierarchical Clustering of Mcl-1 X-ray Data. A starting set of chain A of 41 Mcl-1 crystallographic structures were gathered from the RCSB Protein Data Bank[24] (see Figure S1 for the full list of PDB IDs) and analyzed using the R Bio3D package.[25,26] The first set was refined from structures with missing residues and gave a reduced set of structures (Figure S2). Prior to the conformational ensemble analysis, the structures were superimposed iteratively until the most structurally invariant region or "core" was identified (refer to the Jupyter notebook Mcl-1_ens.ipynb at https://osf.io/4fqph/ for more information).[27]

Hierarchical clustering was performed on the pairwise Cα root-mean-square deviation (RMSD) matrix computed from the data set to get a first idea about the clusters formed by considering the distance between the coordinate sets. Principal component analysis (PCA) was performed afterward in order to gain insights into interconformer relationships. PCA is a dimensionality reduction algorithm that in the case of proteins captures the low-frequency modes (essential dynamics)[28] based on the diagonalization of the variance−covariance matrix built from the Cartesian coordinates of the Cα atoms. The principal components (PCs) of the diagonalized matrix represent a linear basis set on which the projection of the structural ensemble provides the major differences between the conformers. Cα residue contributions were computed using the PCA on the reduced data set because of the sensitivity of the results to missing residues. The full R script used for the analysis to produce the figures is available in the Jupyter notebook Mcl-1_ens.ipynb at https://osf.io/4fqph/.

2.2. Unbiased Molecular Dynamics Simulations. The BIM-bound Mcl-1 crystal structure (PDB ID 2PQK)[19] with the BIM-BH3 peptide removed was used as a starting structure for all of the MD simulations. Schrödinger's Protein Preparation Wizard (PrepWizard)[29] was used to prepare the structure, and the missing loop was reconstructed using Prime. All of the molecular dynamics simulations were performed using NAMD 2.12[30] with the CHARMM36m force field.[31] Two unbiased MD simulations, each for 100 ns, were produced: one with wild-type Mcl-1 (WT-MD) and the other with Cys286 mutated to Trp using the PyMOL mutagenesis tool[32] (C286W-MD). To simulate the aqueous solvent environment, each system was surrounded by a rectangular box of TIP3P explicit water model,[33] and 0.15 M NaCl was added to the system using the CHARMm-GUI solvator.[34] The chosen box size ensured that the simulated system was at a minimum distance of 10 Å from the edge. The van der Waals interactions were truncated using a force switching function between 10 and 12 Å, and the particle mesh Ewald method[35] was used to calculate the long-range electrostatic interactions. The SHAKE algorithm was applied to restrain all bonds involving hydrogen atoms.[36] The systems underwent energy minimization for 10 000 steps. Next, the minimized systems were heated to 303.15 K, and the dynamics was temperature-equilibrated during 50 ps via heating reassignment under *NVT* (canonical ensemble) conditions. Finally, the systems ran freely for 100 ns under *NPT* (isothermal−isobaric ensemble) conditions. Langevin dynamics[37] with a damping coefficient of 1 ps^{-1} was used to maintain the system temperature, and the Nosé−Hoover Langevin piston method was used to control the pressure at 1 atm. The Visual Molecular Dynamics package[38] was used to visually inspect and analyze the produced trajectories.

2.3. Metadynamics Simulation. The free energy landscape representing Mcl-1's breathing motion was computed using a metadynamics (MetaD) simulation[39,40] on the well-tempered ensemble (WT-MetaD).[41]

MetaD is an efficient and well-established method to enhance the conformational sampling for MD simulations and to estimate the free energy surface underlying protein functional dynamics along a selected collective variable (CV) that better captures the studied phenomenon. The Thr226(Cα)−Ser255(Cα) Euclidean distance and the Thr226(Cα)−Asp241(Cα)−Ser255(Cα) angle were found to capture the breathing motion at the Mcl-1 binding interface and were used as local CVs for the WT-MetaD simulation of 317 ns under the *NVT* ensemble starting from the 2PQK model with the peptide ligand removed. The same protocol as described above was used to prepare the starting model for the WT-MetaD simulation.

The WT-MetaD starting parameters were estimated from the first unbiased MD simulation (WT-MD). Gaussian widths of 0.02 nm and 0.6° were used for the distance and angle CVs respectively. During the simulation, Gaussians were deposited every 500 steps (PACE = 500), and a bias factor of 12 and a temperature of 300 K were used. The value of the Gaussian width for each collective variable is equivalent to one-fifth of the standard deviation of the same collective variable during the unbiased run. The WT-MetaD simulation was performed using NAMD 2.12 patched with the open-source, community-developed PLUMED library 2.3.2.[42−44] The PLUMED input is given in the file plumed_geom.dat at https://osf.io/4fqph/.

The Gaussian hills produced by the WT-MetaD simulation were analyzed using the Metadynminer R package.[45] The convergence assessment of the free energy surface and profiles was performed using a well-established protocol.[46] In a nutshell, one needs to check that (a) many recrossing events (see Figure S3) between the free energy minima are observed during the

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model 2020, 60, 3172−3187

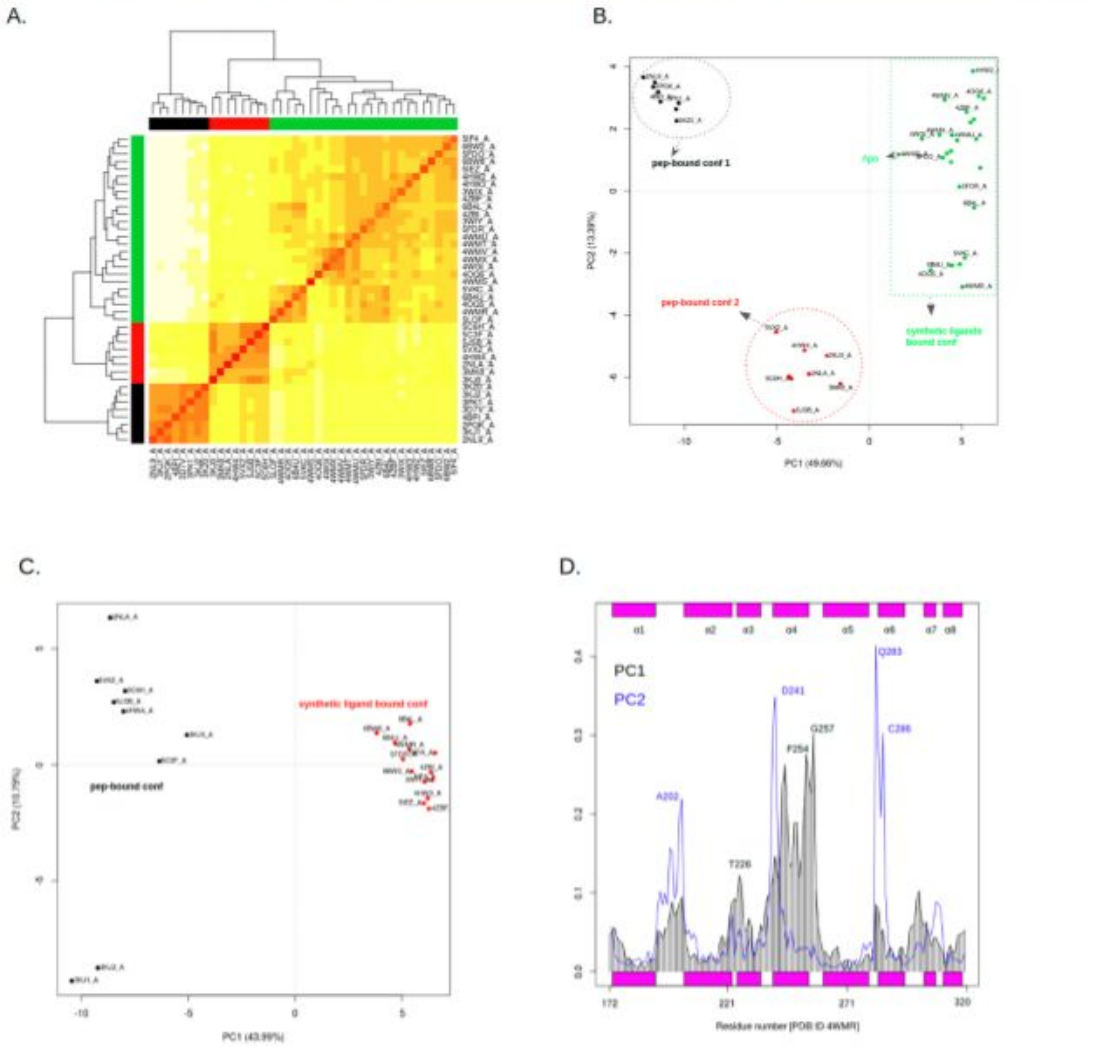


Figure 2. continued

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

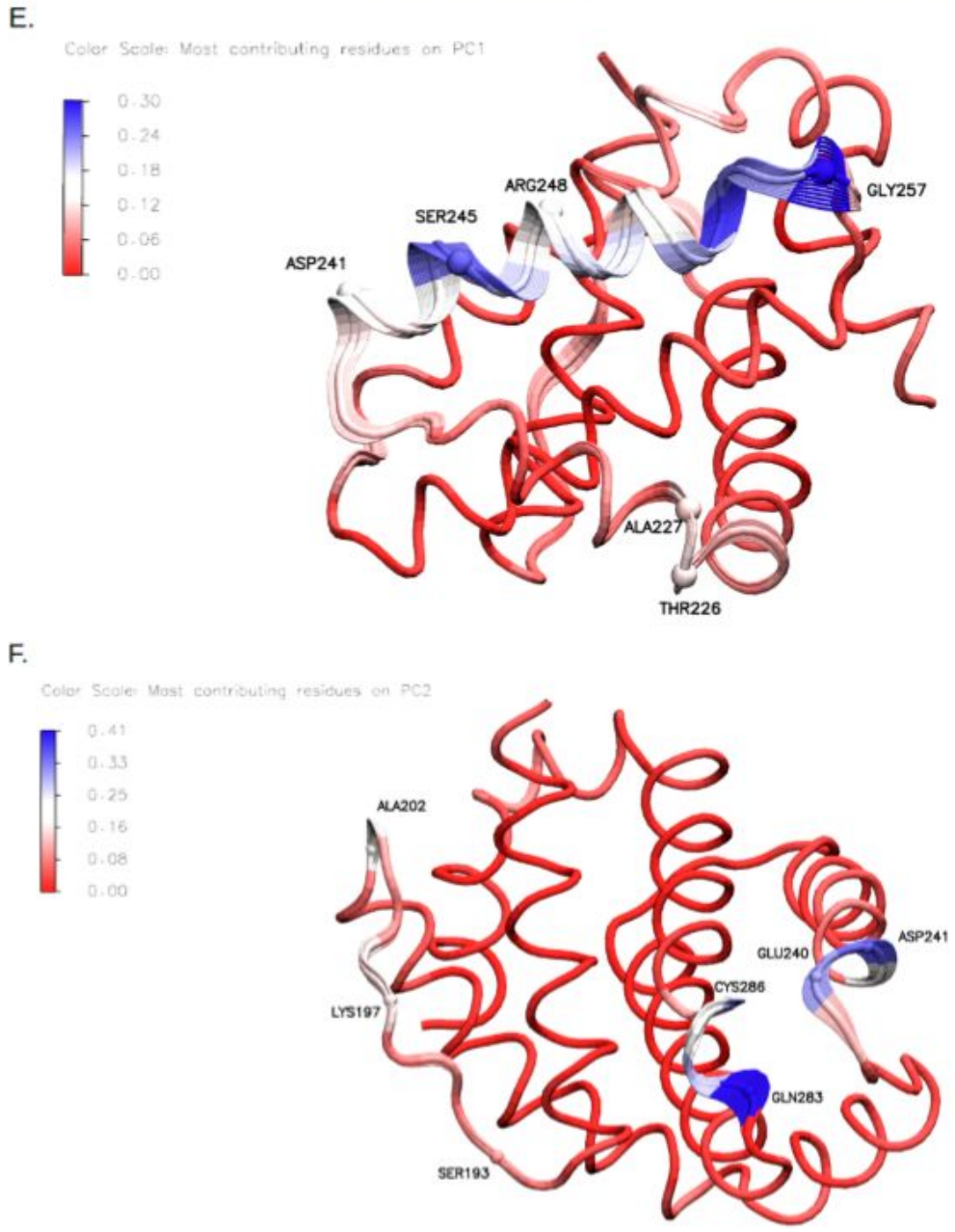


Figure 2. Mcl-1 conformational ensembles derived from X-ray structures (A) Heat map of the hierarchical clustering of the full data set. Three clusters are detected: apo and synthetic ligands (in green) and two peptide-bound conformation clusters (in red and black). (B, C) Principal component analysis of the full and reduced datasets. (D) Residue-wise contributions along PC1 (black shaded lines) and PC2 (blue). Mcl-1 helices are coded at the top and bottom in magenta. (E, F) Residue-wise contributions coded as B-factors on Mcl-1's structure. From the residue contributions along the PC1:PC2 space, two major axes of low-frequency dynamics (essential dynamics) are deduced: (E) the breathing motion at the binding interface, captured by PC1, and (F) the motion joining three remote regions of Mcl-1 that concerns the three loops $\alpha1-\alpha2$, $\alpha3-\alpha4$, and $\alpha5-\alpha6$, captured by PC2.

WT-MetaD simulation (i.e., the minima are visited several times) and (b) the free energy difference between minima reaches stability (see Figure S4). This convergence analysis was further supplemented by monitoring the evolution of the free energy profile along CV2 (angle) during the simulation time (Figure S5) and by fitting the WT-MetaD trajectory on the space of PCs derived from the X-ray ensemble (Figure S6).

The free energy barriers reported were computed using the nudged elastic band (NEB) method (Figures S7 and S8). The analysis workflow is reproduced in the R markdown notebook analyze_fes.Rmd at https://osf.io/4fqph/.

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

2.4. Pocket Crosstalk Analysis. In order to unveil the hidden communication networks for Mcl-1 within the MD simulations, we used the Pocketron algorithm proposed recently by La Salla et al.,[23] available through the Biki Life Sciences suite.[47] The Pocketron algorithm tracks the spatiotemporal evolution of all of the pockets detected during an MD simulation. By taking into account the atoms that are part of each pocket during two subsequent steps in an MD simulation, the algorithm derives an ensemble of splitting and merging events, corresponding to the exchange of atoms between adjacent pockets during the MD simulation. This exchange of atoms between pockets is termed "pocket crosstalk" and is used as a fingerprint of communication between two neighboring pockets. A network representation with the pockets as nodes and the communication frequencies between them as vertices provides a semiquantitative view of the amount of information and exchange between two distant sites within a protein. The two unbiased MD simulations (WT-MD and C286W-MD) were used as input to the Pocketron algorithm in order to investigate the existence of hidden communication networks. For each analysis, water and ions were removed to avoid artifacts, and only the protein was kept in the trajectory. The time persistency threshold of residues forming the pockets was set to 20% or 10% in order to account for the transient nature of the pockets within the framework of proteins involved in protein−protein interactions, which is the case of Mcl-1. To avoid very small pockets, a minimum volume of a detectable pocket was set equivalent to 4 water molecules.

3. RESULTS/DISCUSSION

3.1. Mcl-1 Conformational Space from Principal Component Analysis of the Ensemble X-ray Structures Dataset. The Mcl-1 conformational space derived from the X-ray structures dataset reveals, as expected, three main conformational ensembles from the hierarchical clustering (hclust) analysis (Figure 2A): apo and synthetic-ligand-bound conformations (colored in green) and two peptide-bound conformations (colored in red and black). In order to highlight the structural differences between these conformers from an ensemble point of view, we performed a PCA on the two data sets: the full data set comprising all 41 Mcl-1 X-ray structures (Figure 2B) and the reduced data set comprising only structures with no missing residues (Figure 2C).

The PCA on the full data set reproduced the clusters found through the hierarchical clustering analysis. The first two PCs accounted for 63% of the structural displacement encoded within the data set. PC1 captures the transition between the peptide-bound conformations and the apo and synthetic ligand conformations. PC2 encodes the structural variability within the clusters.

The basis set (PC1:PC2) formed by the PCA on the reduced data set accounts for 54% of the displacement within the conformations and reproduces the separation between the peptide-bound conformations and the synthetic-ligand-induced conformations on PC1 and an interconformer variance within the peptide-bound cluster on PC2. The latter result is important since we aim to generalize residue contributions from the reduced data set to the full data set.

Residue-wise contributions (Figure 2D−F) computed from the reduced set show two major axes of low-frequency dynamics (essential dynamics): PC1 captures the breathing motion (Figure 2E and PC1.mpg at https://osf.io/4fqph/) at Mcl-1's binding interface (helices α3 and α4) leading to the transition from a closed form (apo or synthetic-ligand-bound form) to the more open conformation (peptide-bound conformations). The major contributions on PC1 are generated by the α5−α4 loop (residue Gly257) separating helices α4 and α5 and propagated to the adjacent α4 helix (residue Ser245). On the other hand, PC2 captures the intracluster variations (variability within the peptide-bound conformations and synthetic-ligand−bound conformations). Residue contributions to PC2 (Figure 2D,F and PC2.mpg at https://osf.io/4fqph/) show the importance of three contributing sites: (i) the loop separating helices α5 and α6 (residue Gln283) and the first part of the α6 helix (residue Cys286); (ii) the α3−α4 loop with the main contribution from Glu240 and Asp241, and finally (iii) the α1−α2 loop (residue Lys197).

It is noteworthy that these three sites are functionally relevant for Mcl-1. In fact, the α1−α2 loop is important for Mcl-1 antisenescence activity in vitro and in vivo.[48] The Cys286 at the beginning of α6 was pointed out in the study of Lee et al.[22] as an allosteric modulation site for Mcl-1. The α3−α4 loop is closely linked to dynamical fluctuations of the canonical binding site of Mcl-1 responsible for the anti-apoptotic activity, and we consider it as a pivot region bridging the essential dynamics between PC1 and PC2 (the mostly shared residue contributions between PC1 and PC2 are linked to this loop). These results show another successful example of using a dimensionality reduction algorithm such as PCA to filter the low-frequency modes and to link these essential dynamics with the protein function.

To resume, we extracted the essential dynamics of Mcl-1 through an unsupervised learning approach (PCA) starting from an ensemble of Mcl-1 X-ray structures. Two main axes of dynamics stood out in our analysis: a breathing motion along the PC1 axis, leading the transition from the closed form (apo/synthetic-ligand-bound form) to the peptide-bound conformations. A second axis was highlighted as well and linked the internal loops of Mcl-1 along the second principal component axis. Our results are in good agreement with what Maity et al.[21] presented. However, the main difference lies in the fact that in their work, they focused mainly on the dynamics of residues forming the orthosteric binding site. Here we propose instead a more holistic view of the essential dynamics space by considering all of the Cα atoms, and we provide as well the key residues contributing to each essential dynamics axis.

The conformational landscape of Mcl-1 as presented, using the principal components as the space, is informative but lacks the detailed description that the free energy surface determination might provide regarding the likelihood of the conformations under different conditions (apo, synthetic-ligand-bound, or peptide-bound conformations). For that purpose, in the next section we provide a free energy surface describing the breathing motion at the Mcl-1 binding interface that was obtained using metadynamics simulations as an enhanced sampling technique and simple geometric collective variables that could recognize the three conformational states described above.

3.2. Free Energy Surface for Mcl-1's Breathing Motion. Because of the reduced number of Mcl-1 apo models in the literature, we wanted to propose a free energy landscape describing the likelihood of the conformational states explored by Mcl-1 in solution. To the best of our knowledge, the free energy surface characterizing the breathing motion at Mcl-1's binding interface has never been proposed. To do so, collective variables capturing the phenomenon are needed. For that

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

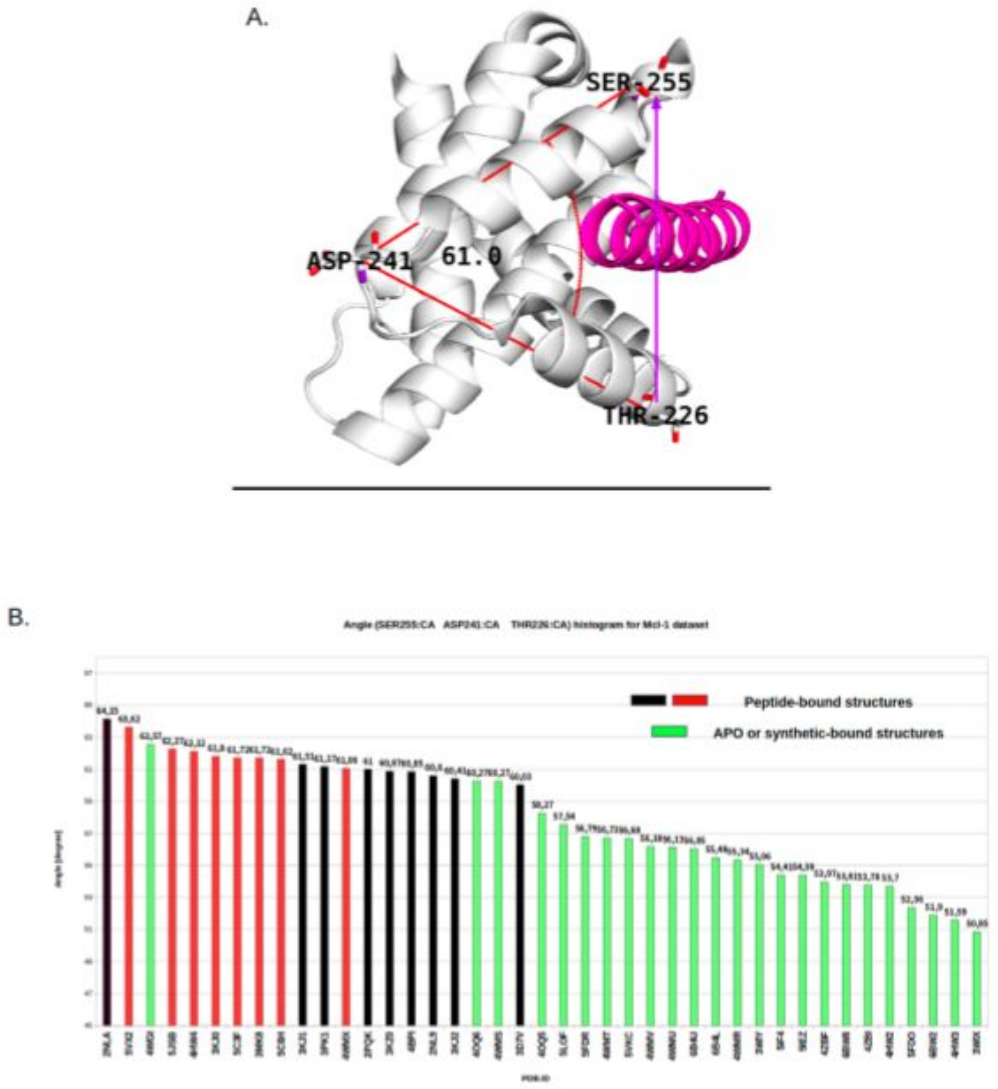


Figure 3. Collective variables used for the WT-MetaD simulation. (A) Ser255(Cα)−Asp241(Cα)−Thr226(Cα) angle and Ser255(Cα)−Thr226(Cα) distance CVs illustrated using the X-ray structure of the Mcl-1:BIM-BH3 peptide complex (PDB ID 2PQK).[19] (B) The distribution of the angle in the full data set shows the separative ability of the angle to distinguish between the conformational states of apo Mcl-1 (PDB ID 4WMS)[17]/synthetic-ligand-induced (in green) and peptide-binding-induced conformations (black and red). An exception (outlier) needs to be mentioned for the PDB ID 4WGI[49] structure, where the size of the synthetic ligand is large, which induces a shift toward the peptide-bound structures.

purpose, we analyzed the previously gathered Mcl-1 full dataset for the protein ensemble analysis and found that the Ser255-(Cα)−Asp241(Cα)−Thr226(Cα) angle and the linked distance Ser255(Cα)−Thr226(Cα) (Figure 3A) were able to distinguish between the peptide-binding-induced conformations and the apo/small-ligand-induced conformations (Figure 3B).

A well-tempered metadynamics simulation (317 ns) in the [angle, distance] collective variable space was used to construct the free energy surface. We report below the free energy surface and representative structures describing the breathing motion along the collective variables.

The analysis of the free energy surface (Figure 4A) and the profile as a function of the angle (Figure 4B) show the existence of two main free energy wells, A and B, of 18 and 12.8 kJ/mol located around angle values of 34° and 50°, respectively. Transition barriers between minima were estimated using the nudged elastic band method,[50] and the error was estimated as the standard deviation of the free energy difference between two minima (see the R markdown notebook analyze_fes.Rmd at https://osf.io/4fqph/ for more details).

The deepest free energy minimum (well A) corresponds to a closed conformation compared with the other free energy minimum (well B) and is not yet covered by the few experimentally available data concerning *h*Mcl-1 apo state. This free energy minimum is centered around 34° and spreads toward a value of 45°. The other free energy minimum (well B) corresponds to a "ready to bind" intermediate state with an angle varying from 47° to 56°. Interestingly, this free energy minimum B spans the domain covered by the only NMR apo ensemble

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

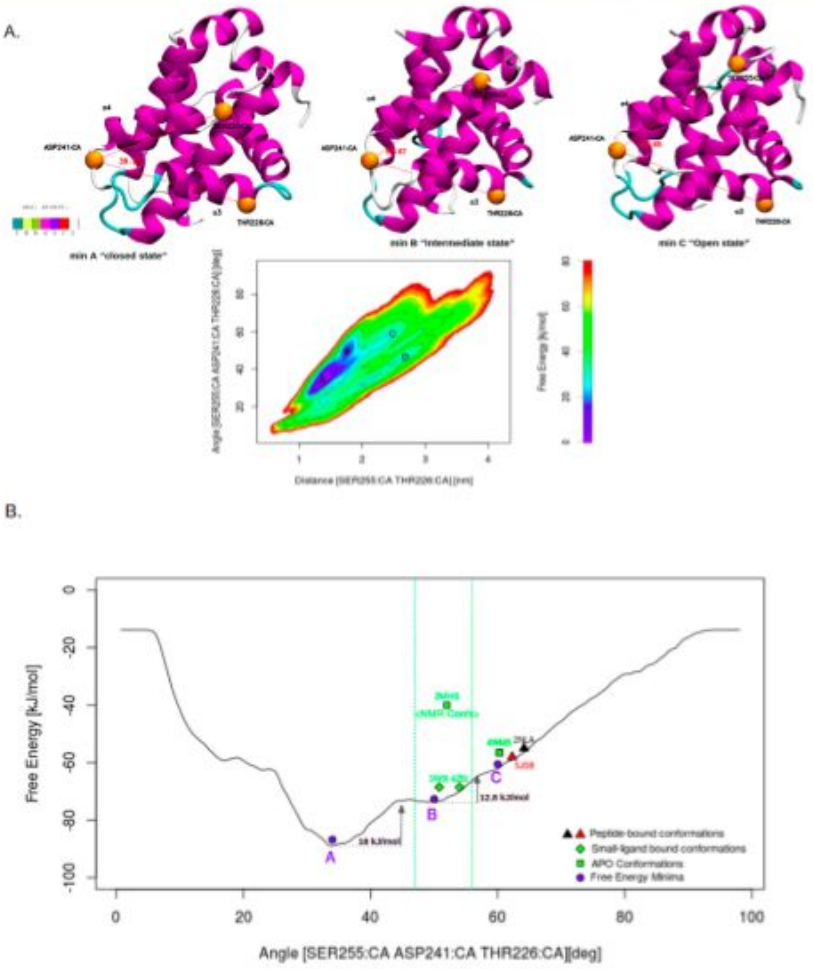


Figure 4. The free energy surface for Mcl-1's breathing motion suggests an equilibrium between a closed conformation and a "ready to bind" conformation as the predominant apo states of Mcl-1 in solution. (A) The free energy surface of Mcl-1's breathing motion and most populated free energy wells. Representative structures for the three main states (A, B, and C) are shown at the top; colors represent the secondary structures using the STRIDE color scheme. Cα atoms forming the Ser255−Asp241−Thr226 angle are represented as VDW spheres and colored in orange. (B) Free energy profile as a function of the Ser255−Asp241−Thr226 angle CV. Free energy wells are represented as blue circles, apo models as green squares, and peptide-bound structures as black and red triangles. The angle domain [min, max] = [47°, 56°] spanned by the Mcl-1 apo NMR models (PDB ID 2MHS) is represented by the green dashed lines.

currently available for *h*Mcl-1 (PDB ID 2MHS)[16] and covers as well the synthetic-ligand-bound conformations. This finding is a good parameter to assess the convergence and reliability of our metadynamics simulation since the free energy wells should correspond to energetically favored conformations.

Concerning the low depth (0.72 kJ/mol) of free energy "well" C, which corresponds to a more open conformation very similar to what the peptide-bound conformations adopt, we consider it as a transition state toward a deeper free energy well that might be induced and stabilized by binding with a peptide (deepening of a free energy well after binding to a peptide or a synthetic ligand).

Theoretically, the transition from a peptide-bound conformation toward an apo (closed or "ready to bind") conformation should be easily observed during an unbiased run starting from a conformation in which the peptide is removed. Moreover, the unbiased run should be able to visit the most favorable states shown above. To test this hypothesis, we computed the evolution of the angle during the unbiased run (WT-MD) starting from the 2PQK structure[19] with the BIM-BH3 peptide removed, and the results are shown in Figure 5.

As expected, during the unbiased run, the two most populated wells (A and B) were sampled with a shift toward the closed conformation at the end of the simulation. We should also pinpoint that the starting conformation (peptide-binding-induced conformation) (angle of ~62° around "well" C) was never reached again after the beginning of the simulation. This result demonstrates as well the reliability of the free energy profile proposed and its ability to capture the conformational transitions underpinning Mcl-1's breathing motion in the apo state. At first view, the free energy profile (Figure 4B) suggests a conformational selection for synthetic ligands (similar to minimum B). In fact, the conformational changes at the backbone level due to binding of synthetic ligands to Mcl-1 are very subtle (due mainly to the breathing motion) and do not allow the induced-fit model and a conformational selection

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

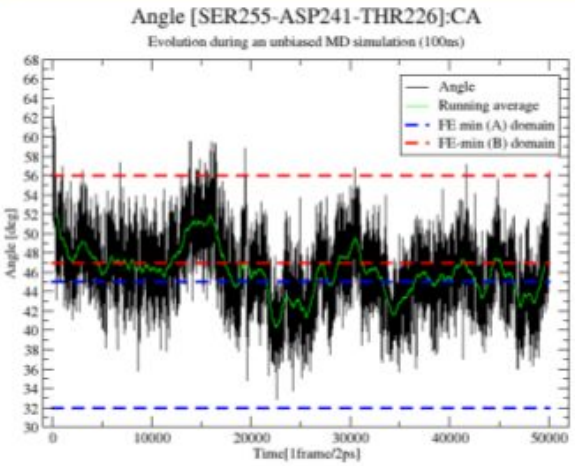


Figure 5. Evolution of the Ser255(Cα)−Asp241(Cα)−Thr226(Cα) angle during an unbiased run (WT-MD). The angle time series from the WT-MD simulation is shown in black, and the angle running average (smoothing) is represented in green. The domains covering free energy (FE) wells A and B are represented by the dashed blue and red lines, respectively.

model to be clearly distinguished. The separation is found at the side-chain level, where the pockets, especially P2 (see Figure 1), are induced by the presence of the synthetic ligands.

However, the separation becomes more clear for the peptide-bound conformations (similar to minimum C): since the peptide-bound conformations are induced by the binding with the peptide and not covered by the dynamics in Mcl-1's apo state (transitions between wells A and B), the free energy surface that we provide sustains a subtle induced-fit hypothesis. The free energy surface and free energy barriers computed are in overall good agreement with those obtained by Chu et al.[51] In their work, instead of focusing on Mcl-1's apo state, they characterized the combined folding−binding process of the PUMA peptide to Mcl-1 using the intermolecular contacts as collective variables. It is known that the BH3-only peptides, such as PUMA, are mainly unfolded under resting conditions and fold along the encounter pathway with Mcl-1. Interestingly, the free energy barrier of the transition from the unbound state to the bound state that they reported was estimated to be around 4 kJ/mol. The difference between the two free energy barriers (the transition from well B toward the transition state near "well" C, ~12.8 kJ/mol) is not surprising since the peptide is not taken into account in our simulation. However, it is well-known that the binding and encounter between a ligand and a receptor induces a flattening of the free energy surface and the formation of new free energy wells[52] due to stabilizing interactions during the binding process, which explains the difference observed.

Combined together, these results suggest that Mcl-1's breathing motion is a result of an equilibrium of transitions between a closed form (free energy well A) and a "ready to bind" state (free energy well B). The encounter with a peptide through non-native electrostatic interactions, as proposed by Chu et al.,[51] shifts the conformational states toward the bound state through an induced-fit mechanism.

After describing the breathing motion expressed along the first axis of essential dynamics and proposing a free energy surface, we wanted to focus on the dynamics along the second axis.

3.3. Mcl-1 Allosteric Inhibition: A Global Structural Viewpoint from MD Simulations and Allosteric Communication Networks. The PCA analysis showed that the second axis of essential dynamics expressed correlated motions along the internal loops of Mcl-1 and that Cys286 was detected as one of the principal contributors on the PC2 axis. Interestingly, this Cys286 was highlighted in the previously published work by Lee et al.[22] as an allosteric inhibition hotspot for Mcl-1. They proposed an allosteric mode to disrupt the canonical BH3-binding activity of Mcl-1 by covalent decoration of the Cys286 residue with a small organic moiety. The authors demonstrated as well that the C286W variant conferred this inhibitory effect and proposed an allosteric inhibition mode through rigid-

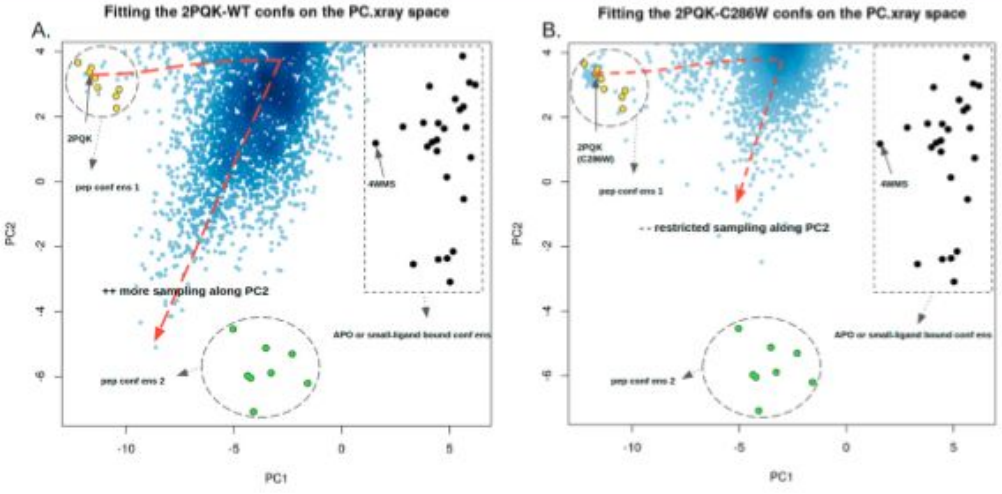


Figure 6. Conformational restriction along the PC2 essential dynamics axis controls the conformational ensemble shift underlying the allosteric effect of the Mcl-1 mutant isoform (C286W). Shown are the fits of (A) the WT-MD trajectory and (B) the MD-C286W trajectory (blue dots) on the PC1:PC2 space constructed using the full Mcl-1 X-ray structures dataset. The peptide-bound ensembles 1 and 2 are colored in yellow and green, respectively; the apo and synthetic-ligand-bound conformations are colored in black. The starting structure (PDB ID 2PQK) is indicated in both cases.

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

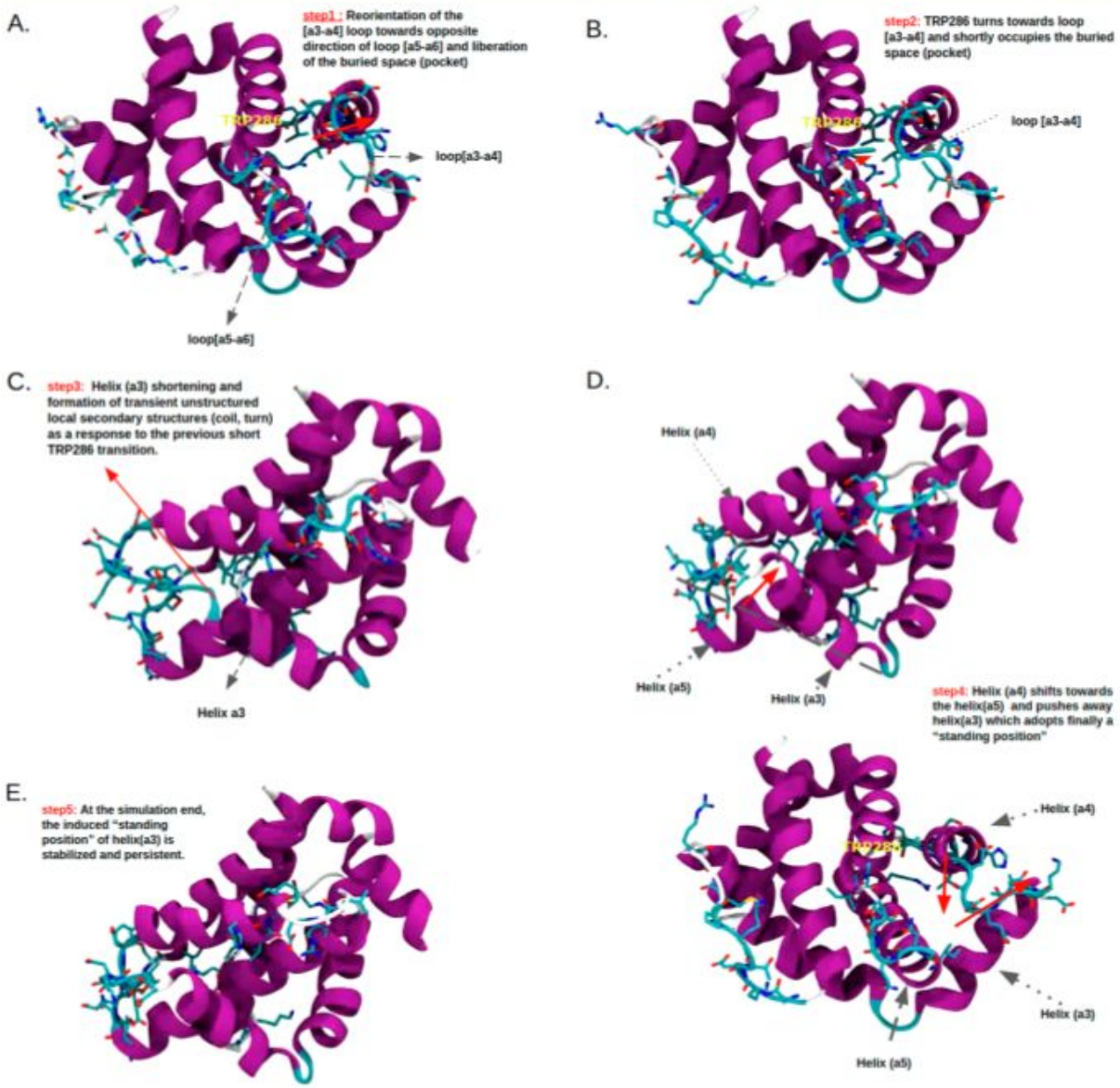


Figure 7. Major steps during the C286W-MD simulation leading to the conformational reorientation of helix $\alpha 3$: (A) $t = 16$ ns; (B) $t = 20$ ns; (C) $t = 21$ ns; (D) $t = 31$ ns; (E) $t = 100$ ns. The protein is colored using the STRIDE color keys for secondary structures. Helices are indicated with black dashed arrows. Events are marked with solid red arrows.

ification of the Mcl-1 structure and a shift toward the unbound state. However, a detailed understanding of the structural elements leading to this allosteric inhibition was still lacking.

It is now established that the dynamics of the conformational ensembles of proteins plays a central role in allostery.[53] From an ensemble point of view, two components are needed to understand the allosteric inhibition of a protein. In the first place, the shift in the conformational populations of the protein upon binding to the allosteric ligand or mutation needs to be assessed, and the structural elements driving this shift must be revealed. Second, the communication network conducting the signal from the allosteric site toward the orthosteric site needs to be provided as well.

In order to highlight the conformational dynamics of Mcl-1 upon allosteric modifications (the C286W mutation in this case), we used two unbiased 100 ns molecular dynamics simulations: the WT-MD simulation (the previously used one) and the C286W-MD simulation in which Cys286 was replaced by a Trp. After that, the trajectory frames from both simulations were projected onto the previously built PC1:PC2 conformational space computed from the full X-ray structures dataset. The results are presented in Figure 6.

The projection of the two trajectories (WT-MD and C286W-MD) onto the PC1:PC2 space shows a sampling and conformational dynamics restriction along the PC2 essential dynamics axis.

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

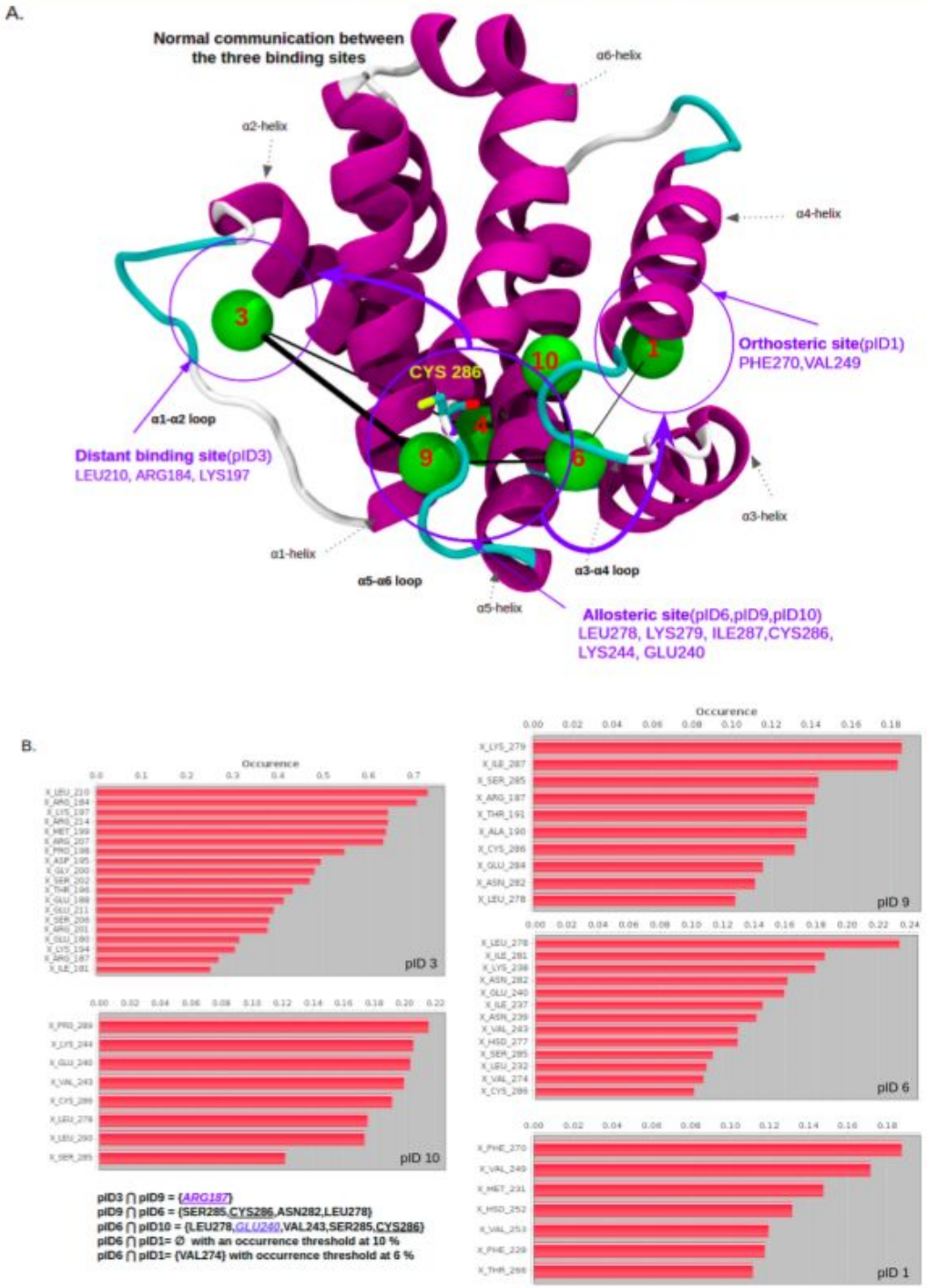


Figure 8. continued

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

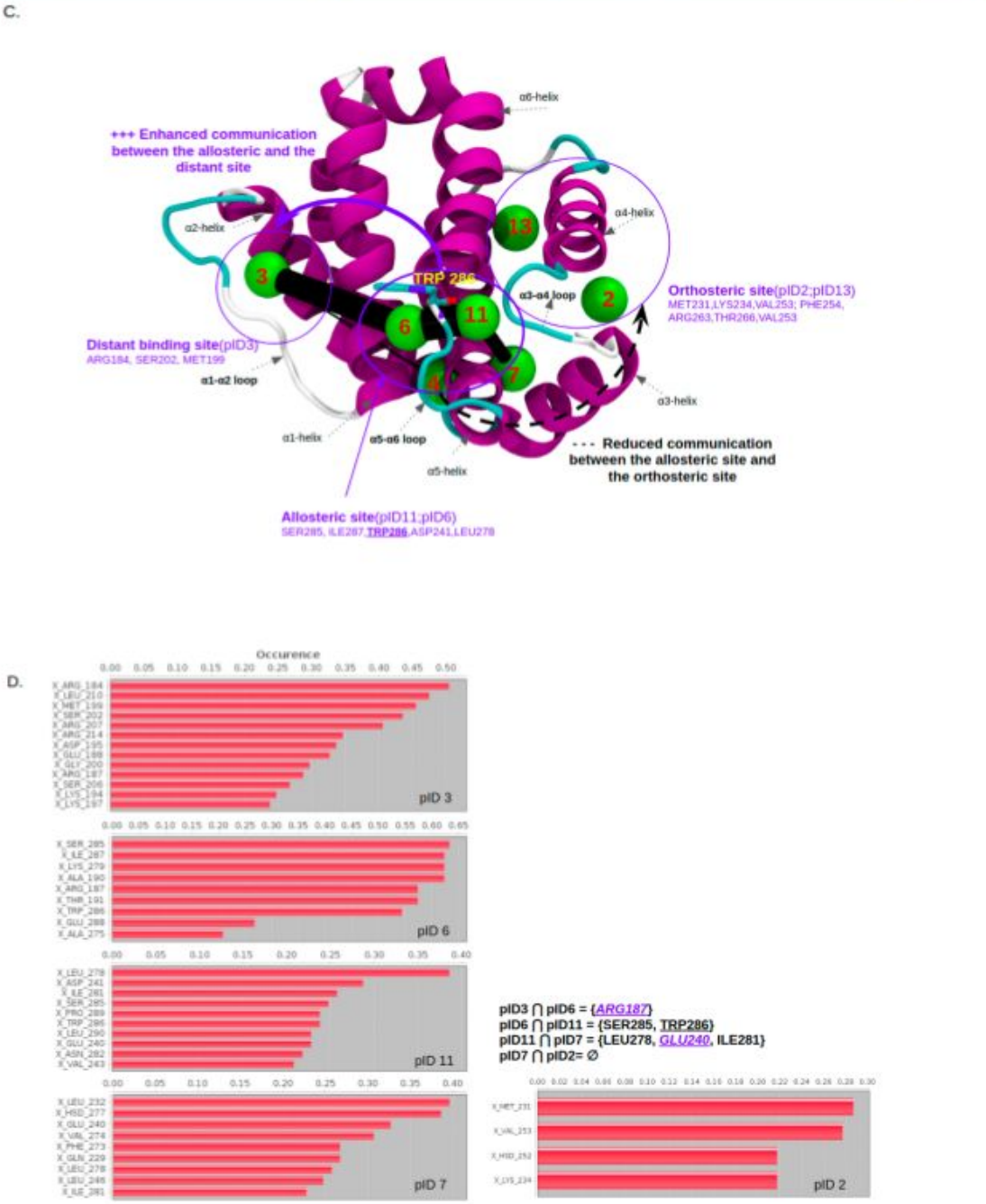


Figure 8. Mcl-1 hidden allosteric communication networks revealed through pocket crosstalk analysis. (A) Network for the WT-MD simulation. (B) Residue occurrence for the WT-MD simulation network. (C) Network for the C286W-MD simulation. (D) Residue occurrence for the C286W-MD network. The green vertices of the network represent the pockets, with pocket identifiers (pIDs) shown in red. The network edges are shown in black, and their thickness is proportional to the amount of communication between the corresponding pockets. Residue occurrences are computed for the pockets along the PC2 axis, and shared residues are highlighted.

By recalling the results presented in the first section (see Figure 2E,F), we have shown that the dynamics along the PC2 essential dynamics axis involves the three internal loops ($\alpha1-\alpha2$, $\alpha3-\alpha4$, and $\alpha5-\alpha6$). Under the light of this observation, we suggest that the dynamics of these three internal loops control the ensemble shift and hence lead the allosteric response resulting from the C286W mutation.

In order to gain more insights into the dynamics of the protein and especially the role played by the dynamics near the allosteric site in both WT-MD and C286W-MD, we analyzed and

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model 2020, 60, 3172−3187

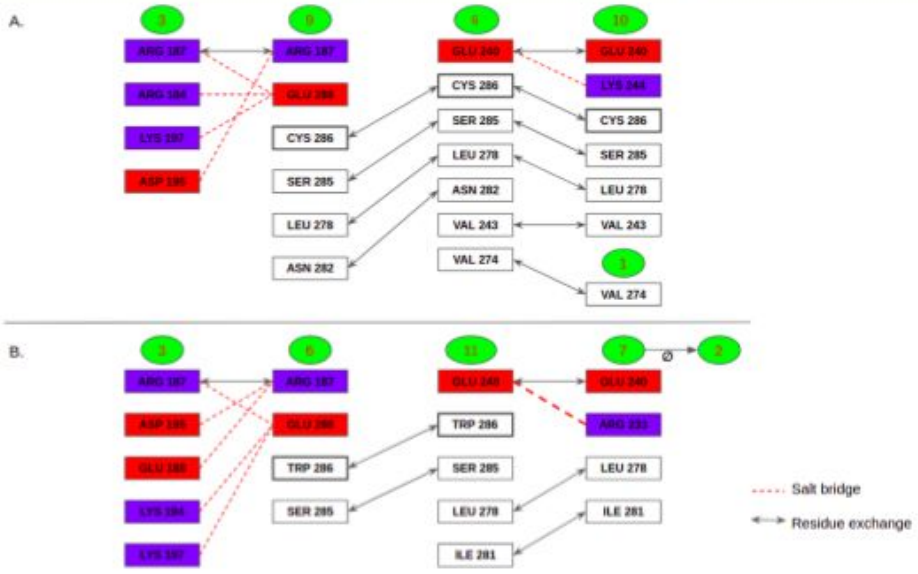


Figure 9. Schematic representation combining the pocket crosstalk and interpocket salt bridges for (A) the WT-MD simulation and (B) the C286W-MD simulation. Intrapocket salt bridges have been omitted in order to simplify the representation, and only interpocket salt bridges are shown. Pockets are represented as green spheres. Salt bridges are represented as dashed red lines. Residue exchange between pockets is shown by solid black arrows. Electronegative residues (Asp, Glu) are colored in red, and positively charged residues (Arg, Lys) are colored in purple.

compared the two trajectories (for the full comparison and a visual inspection between the WT-MD and C286W-MD simulations, see the video compare_MDs.mp4 at https://osf.io/4fqph/). Figure 7 presents and summarizes the major events observed in the C286W-MD simulation.

From our analysis, four major steps underlie the conformational response of the α3 helix to the mutation at the allosteric site (C286W). First, a shift of the α3−α4 internal loop toward the opposite direction of the α5−α6 loop liberates a buried surface (pocket) (Figure 7A). Then the Trp286 side chain turns toward the α3−α4 loop and briefly occupies the lately liberated space (Figure 7B). As a response to that, the C-terminal end of the α3 helix undergoes transient random-coil and turn transitions (Figure 7C). Finally, for a second time, the Trp286 side chain is buried briefly (around t = 31 ns), and that leads the α4 helix to shift toward the α5 helix, creating a hindrance (resulting from the buried state of the Trp286 side chain as well) that pushes away the α3−α4 loop and the α3 helix. This α3 helix ends by adopting a "standing-up" conformation almost perpendicular to the α4 helix (Figure 7D).

However, we should mention that a complete and permanent buried state of the Trp286 side chain was not observed and that the brief reorientations and burials (~1 ns each) were sufficient to induce the conformational change of the α3 helix and the α3−α4 loop. The finally adopted conformation is clearly not favorable for a BH3 peptide binding event since the binding groove hosting the binding of BH3-only peptides is deformed by the reorientation of the α3 helix.

In the WT-MD simulation, the Cys286 side chain adopted two main states: a solvent-exposed state and a buried state (similar to what the Trp286 briefly adopts at step 2). The structural changes induced by the Cys286 buried-side-chain state are not comparable to the ones undergone during the brief Trp286 side-chain-burial events, since no major changes were noticed at the BH3-binding interface (α3−α4 helices). Therefore, we deduce that a bulky hydrophobic side chain at this position is a requirement to induce the structural changes necessary for reorientation of the α3 helix.

After the existence of the conformational population shift is asserted and the structural differences observed during the two simulations are highlighted, another question remains, concerning the existence of any hidden allosteric communication networks controlling the signaling along the PC2 essential dynamics axis. To address this question, we used an innovative approach proposed recently by La Salla et al.[23] named Pocketron to unveil the existence of hidden allosteric communication networks.

Both of the unbiased MD simulations (WT-MD and C286W-MD) were used as input for the Pocketron algorithm. The networks, pocket occurrences, and shared residues between pockets are presented in Figure 8. The pocket analysis revealed in addition to the orthosteric and allosteric binding sites a distant binding site formed in the vicinity of the α1 and α2 helices and the α1−α2 loop. The pocket crosstalk fingerprint strategy allowed us to mine the communication between these different binding sites.

For the WT-MD simulation (Figure 8A), a communication network was detected and linked the three binding sites. Similarly, a network was found for the C286W-MD simulation (Figure 8C), but intriguingly, the communication was enhanced between the distant binding site (pID3) and the allosteric binding site (pID6 and pID11) and lost between the latter and the orthosteric site. In order to get more details about this observation, we analyzed the shared residues between the pockets for the WT-MD simulation (Figure 8B) and for the C286W-MD simulation (Figure 8D) and their respective occurrences. Two residues stood out from the occurrence analysis, and did so for the two simulations equally: Arg187 and Glu240. Arginine and glutamate residues are well-known for their ability to establish salt bridges (SBs).[54] Therefore, we complemented our analysis by fetching and comparing SB formations during the two simulations. In order to identify a salt

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

bridge, a restrictive distance cutoff of 4 Å between the N−O atoms was used, as defined by Barlow and Thornton.[55] The results were combined with the pocket crosstalk analysis presented above, and a schematic representation is provided in Figure 9.

In the WT-MD network, the distant binding site (pID3) communicated with the allosteric binding site (pID9) through an exchange of the Arg187 residue. Arg187 and the two other positively charged residues Arg184 and Lys197 formed SBs with Glu288, and a final SB linked Arg187 with Asp195. The communication within the allosteric site (pID9, pID6, and pID10) was represented by the exchange of four residues (Ser285, Cys286, Asn282, and Leu278) between pockets pID9 and pID6 and five residues (Val243, Ser285, Cys286, Leu278, and Glu240) between pockets pID6 and pID10. This exchange is due mainly to the transition of the Cys286 side chain from a solvent-exposed state at pID9 toward a buried state at pID6. Glu240 interacted with Lys244. The communication between the allosteric site and the orthosteric site was due only to the exchange of Val274 between pID6 and pID1.

On the other side, for the C286W-MD-derived network, five SBs were formed between pockets pID3 (distant site) and pID6 (allosteric site): Glu288 interacted by three SBs with Arg187, Lys197, and Lys194, and Arg187 formed two SBs with Asp195 and Glu188. This subtle difference in the number of SBs explains the enhanced communication between the distant site (pID3) and the allosteric site (pID6) in the C286W-MD network. The pocket crosstalk within the allosteric site for the C286-MD network was due to the brief burial states observed for the Trp286 side chain. These burial states of Trp286 induced a hindrance that caused the shift of the $\alpha3-\alpha4$ loop. In fact, the exchange of the Glu240 residue between pID11 and pID7 is directly linked to the presence of Trp286 and was not observed for the WT-MD network. Most importantly, and as a direct consequence, Glu240 established a unique salt bridge with Arg233, a residue in the $\alpha3$ helix. This salt bridge was persistent and is the key interaction that stabilizes the reorientation of the $\alpha3$ helix (see the video compare_sb.mp4 at https://osf.io/4fqph/ for a comparison between the SB networks).

By joining the pocket crosstalk due to residue exchange with salt bridge networks, we clearly unveil the key structural features underlying Mcl-1 allosteric inhibition due to the C286W mutation.

It is necessary to mention that the allosteric inhibition within the Bcl-2 family of proteins is not a unique feature of Mcl-1. In fact, BAX, a pro-apoptotic member of the family, is also allosterically sensitized in a similar binding site, as shown by the work of Pritz et al.[56] However, Bcl-x_L, an anti-apoptotic member was shown by the work of Lee et al.[22] to be insensitive to the C286W allosteric mutation. A structural mechanism clarifying this difference was not provided and is still to be explored.

We believe that a detailed understanding of the differences in the structural dynamics along the internal loops might be a key step toward the selective inhibition of these proteins and could constitute an alternative or a complementary way to the classical targeting through the BH3-binding groove.

Finally, through this study we have provided another example of the importance of the internal loops in the conformational dynamics and allosteric control of proteins.[57]

4. CONCLUSION

In this study, we explored the structural dynamics of Mcl-1, an anti-apoptotic protein, through MD simulations, enhanced sampling, and pocket crosstalk analysis. Three main aspects were covered.

First, the PCA analysis allowed us to highlight the existence of two main axes of essential dynamics for Mcl-1: a breathing motion at the binding interface and a correlated motion bridging its internal loops. The key residues contributing to each functional movement were identified. Then a free energy surface based on a metadynamics simulation was generated for the breathing motion. We showed that Mcl-1 adopts an equilibrium between a closed conformation and a "ready to bind" conformation in the apo state in solution (explicit solvent). However, it is necessary to mention that the active and functional conformations of Mcl-1 are located in a membrane environment and that a detailed description of the conformational states within this environment is not provided in the literature and is yet to be determined. A detailed understanding of the conformational states available to the protein within its functional environment is of great importance, especially in a structure-based drug discovery setting, and efforts toward this direction are to be considered.

Finally, driven by the essential dynamics along the internal loops of Mcl-1 revealed by PCA, the intrinsic details of the allosteric inhibition of this protein were highlighted: the shift of the conformational states was assessed, and a detailed model for the allosteric signal transmission between the distant, allosteric, and orthosteric sites was revealed. Key residues involved are Arg187 and Glu288 connecting the distant and allosteric sites through salt bridges. Trp286, which occupies the allosteric pockets, leads the structural change at the $\alpha3$ helix. A stable salt bridge linking Arg233 at the C-terminal end of the $\alpha3$ helix with Glu240 in the $\alpha3-\alpha4$ loop explains the overall rigidification of the Mcl-1 structure and the Mcl-1 orthosteric binding site.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00315.

Structural ensemble analysis and PCA; metadynamics simulation (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jana Sopkova-de Oliveira Santos − *Normandy Univ, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), 14000 Caen, France;* orcid.org/0000-0002-4829-8120; Email: jana.sopkova@unicaen.fr

Authors

Mohammed Benabderrahmane − *Normandy Univ, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), 14000 Caen, France*

Ronan Bureau − *Normandy Univ, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), 14000 Caen, France*

Anne Sophie Voisin-Chiret − *Normandy Univ, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), 14000 Caen, France;* orcid.org/0000-0001-5564-2244

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00315

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model 2020, 60, 3172−3187

Author Contributions

M.B. conceived and conducted the study and wrote the paper. R.B. and J.S.-d.O.S. conceived and supervised the study and provided input for the drafted manuscript. A.S.V.-C. supervised Mcl-1-related projects within the PPI research axis in the laboratory. All of the authors approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.
Additional files and datasets related to this work are publicly available at https://osf.io/4fqph/.

■ ACKNOWLEDGMENTS

This work was supported by a doctoral fellowship (MENRT) held by M.B. The present work was performed thanks to the generous computing resources provided by the CRIANN (Normandy, France). The authors gratefully acknowledge the European Community (European Regional Development Fund, ERDF) for the molecular modeling software. We are also grateful to Dr. Bogdan Marekha for early discussions of this work.

■ REFERENCES

(1) Saleh, T.; Kalodimos, C. G. Enzymes at Work Are Enzymes in Motion. *Science* **2017**, *355*, 247−248.
(2) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503−6515.
(3) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598−1603.
(4) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.
(5) Shamas-Din, A.; Kale, J.; Leber, B.; Andrews, D. W. Mechanisms of Action of Bcl-2 Family Proteins. *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a008714.
(6) Rinkenberger, J. L.; Horning, S.; Klocke, B.; Roth, K.; Korsmeyer, S. J. Mcl-1 Deficiency Results in Peri-Implantation Embryonic Lethality. *Genes Dev.* **2000**, *14*, 23−27.
(7) Opferman, J. T.; Letai, A.; Beard, C.; Sorcinelli, M. D.; Ong, C. C.; Korsmeyer, S. J. Development and Maintenance of B and T Lymphocytes Requires Antiapoptotic MCL-1. *Nature* **2003**, *426*, 671−676.
(8) Dzhagalov, I.; Dunkle, A.; He, Y.-W. The Anti-Apoptotic Bcl-2 Family Member Mcl-1 Promotes T Lymphocyte Survival at Multiple Stages. *J. Immunol.* **2008**, *181*, 521−528.
(9) Opferman, J. T.; Iwasaki, H.; Ong, C. C.; Suh, H.; Mizuno, S.-I.; Akashi, K.; Korsmeyer, S. J. Obligate Role of Anti-Apoptotic MCL-1 in the Survival of Hematopoietic Stem Cells. *Science* **2005**, *307*, 1101−1104.
(10) Dzhagalov, I.; St. John, A.; He, Y.-W. The Antiapoptotic Protein Mcl-1 Is Essential for the Survival of Neutrophils but Not Macrophages. *Blood* **2007**, *109*, 1620−1626.
(11) Steimer, D. A.; Boyd, K.; Takeuchi, O.; Fisher, J. K.; Zambetti, G. P.; Opferman, J. T. Selective Roles for Antiapoptotic MCL-1 during Granulocyte Development and Macrophage Effector Function. *Blood* **2009**, *113*, 2805−2815.
(12) Arbour, N.; Vanderluit, J. L.; Le Grand, J. N.; Jahani-Asl, A.; Ruzhynsky, V. A.; Cheung, E. C. C.; Kelly, M. A.; MacKenzie, A. E.; Park, D. S.; Opferman, J. T.; et al. Mcl-1 Is a Key Regulator of Apoptosis during CNS Development and after DNA Damage. *J. Neurosci.* **2008**, *28*, 6068−6078.
(13) Akgul, C. Mcl-1 Is a Potential Therapeutic Target in Multiple Types of Cancer. *Cell. Mol. Life Sci.* **2009**, *66*, 1326−1336.
(14) Thomas, L. W.; Lam, C.; Edwards, S. W. Mcl-1; the Molecular Regulation of Protein Function. *FEBS Lett.* **2010**, *584*, 2981−2989.
(15) Denis, C.; Sopková-de Oliveira Santos, J.; Bureau, R.; Voisin-Chiret, A. S. Hot-Spots of Mcl-1 Protein. *J. Med. Chem.* **2020**, *63*, 928−943.
(16) Liu, G.; Poppe, L.; Aoki, K.; Yamane, H.; Lewis, J.; Szyperski, T. High-Quality NMR Structure of Human Anti-Apoptotic Protein Domain Mcl-1(171−327) for Cancer Drug Design. *PLoS One* **2014**, *9*, No. e96521.
(17) Clifton, M. C.; Dranow, D. M.; Leed, A.; Fulroth, B.; Fairman, J. W.; Abendroth, J.; Atkins, K. A.; Wallace, E.; Fan, D.; Xu, G.; et al. A Maltose-Binding Protein Fusion Construct Yields a Robust Crystallography Platform for MCL1. *PLoS One* **2015**, *10*, No. e0125010.
(18) Luptak, J.; Bista, M.; Fisher, D.; Flavell, L.; Gao, N.; Wickson, K.; Kazmirski, S. L.; Howard, T.; Rawlins, P. B.; Hargreaves, D. Antibody Fragments Structurally Enable a Drug-Discovery Campaign on the Cancer Target Mcl-1. *Acta Crystallogr. D Struct Biol.* **2019**, *75*, 1003−1014.
(19) Fire, E.; Gullá, S. V.; Grant, R. A.; Keating, A. E. Mcl-1−Bim Complexes Accommodate Surprising Point Mutations via Minor Structural Changes. *Protein Sci.* **2010**, *19*, 507−519.
(20) Tanaka, Y.; Aikawa, K.; Nishida, G.; Homma, M.; Sogabe, S.; Igaki, S.; Hayano, Y.; Sameshima, T.; Miyahisa, I.; Kawamoto, T.; et al. Discovery of Potent Mcl-1/Bcl-xL Dual Inhibitors by Using a Hybridization Strategy Based on Structural Analysis of Target Proteins. *J. Med. Chem.* **2013**, *56*, 9635−9645.
(21) Maity, A.; Majumdar, S.; Ghosh Dastidar, S. Flexibility Enables to Discriminate between Ligands: Lessons from Structural Ensembles of Bcl-Xl and Mcl-1. *Comput. Biol. Chem.* **2018**, *77*, 17−27.
(22) Lee, S.; Wales, T. E.; Escudero, S.; Cohen, D. T.; Luccarelli, J.; Gallagher, C. G.; Cohen, N. A.; Huhn, A. J.; Bird, G. H.; Engen, J. R.; et al. Allosteric Inhibition of Antiapoptotic MCL-1. *Nat. Struct. Mol. Biol.* **2016**, *23*, 600−607.
(23) La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **2017**, *3*, 949−960.
(24) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 899−907.
(25) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695−2696.
(26) Yao, X.-Q.; Scarabelli, G.; Skjærven, L.; Grant, B. J. The Bio3D Package: New Interactive Tools for Structural Bioinformatics. *Biophys. J.* **2014**, *106*, 406a.
(27) Gerstein, M.; Altman, R. B. Average Core Structures and Variability Measures for Protein Families: Application to the Immunoglobulins. *J. Mol. Biol.* **1995**, *251*, 161−175.
(28) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193−226.
(29) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(30) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283−312.
(31) Huang, J.; MacKerell, A. D., Jr. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(32) DeLano, W. L. *PyMOL*, 2002.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Refined TIP3P Model for Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(34) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM,

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405−413.
(35) Darden, T.; York, D.; Pedersen, L. Partide Mesh Ewald: An *N*·log(*N*) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(36) Forester, T. R.; Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *J. Comput. Chem.* **1998**, *19*, 102−111.
(37) Coffey, W.; Kalmykov, Y. P. *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*; World Scientific, 2012.
(38) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(39) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(40) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(41) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603.
(42) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(43) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(44) PLUMED Consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670−673.
(45) Hošek, P.; Spiwok, V. Metadyn View: Fast Web-Based Viewer of Free Energy Surfaces Calculated by Metadynamics. *Comput. Phys. Commun.* **2016**, *198*, 222−229.
(46) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358−6363.
(47) Decherchi, S.; Bottegoni, G.; Spitaleri, A.; Rocchia, W.; Cavalli, A. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.* **2018**, *58*, 219−224.
(48) Demelash, A.; Pfannenstiel, L. W.; Tannenbaum, C. S.; Li, X.; Kalady, M. F.; DeVecchio, J.; Gastman, B. R. Structure-Function Analysis of the Mcl-1 Protein Identifies a Novel Senescence-Regulating Domain. *J. Biol. Chem.* **2015**, *290*, 21962−21975.
(49) Fang, C.; D'Souza, B.; Thompson, C. F.; Clifton, M. C.; Fairman, J. W.; Fulroth, B.; Leed, A.; McCarren, P.; Wang, L.; Wang, Y.; et al. Single Diastereomer of a Macrolactam Core Binds Specifically to Myeloid Cell Leukemia 1 (MCL1). *ACS Med. Chem. Lett.* **2014**, *5*, 1308−1312.
(50) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113*, 9978−9985.
(51) Chu, W.-T.; Clarke, J.; Shammas, S. L.; Wang, J. Role of Non-Native Electrostatic Interactions in the Coupled Folding and Binding of PUMA with Mcl-1. *PLoS Comput. Biol.* **2017**, *13*, No. e1005468.
(52) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, *8*, 2276.
(53) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508*, 331−339.
(54) Donald, J. E.; Kulp, D. W.; DeGrado, W. F. Salt Bridges: Geometrically Specific, Designable Interactions. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 898−915.
(55) Barlow, D. J.; Thornton, J. M. Ion-Pairs in Proteins. *J. Mol. Biol.* **1983**, *168*, 867−885.
(56) Pritz, J. R.; Wachter, F.; Lee, S.; Luccarelli, J.; Wales, T. E.; Cohen, D. T.; Coote, P.; Heffron, G. J.; Engen, J. R.; Massefski, W.; et al. Allosteric Sensitization of Proapoptotic BAX. *Nat. Chem. Biol.* **2017**, *13*, 961−967.
(57) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391−6423.

https://dx.doi.org/10.1021/acs.jcim.0c00315
J. Chem. Inf. Model. 2020, 60, 3172−3187

Supporting Information

Insights into Mcl-1 conformational states and allosteric inhibition mechanism from molecular dynamics simulations, enhanced sampling, and pocket crosstalk analysis

Mohammed Benabderrahmane[†], Ronan Bureau[†], Anne Sophie Voisin-Chiret[†], Jana Sopkova-de Oliveira Santos[†]

[†] *Normandy University, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), Caen, France*

* Corresponding authors

Email: jana.sopkova@unicaen.fr

I. Structural ensemble analysis and PCA

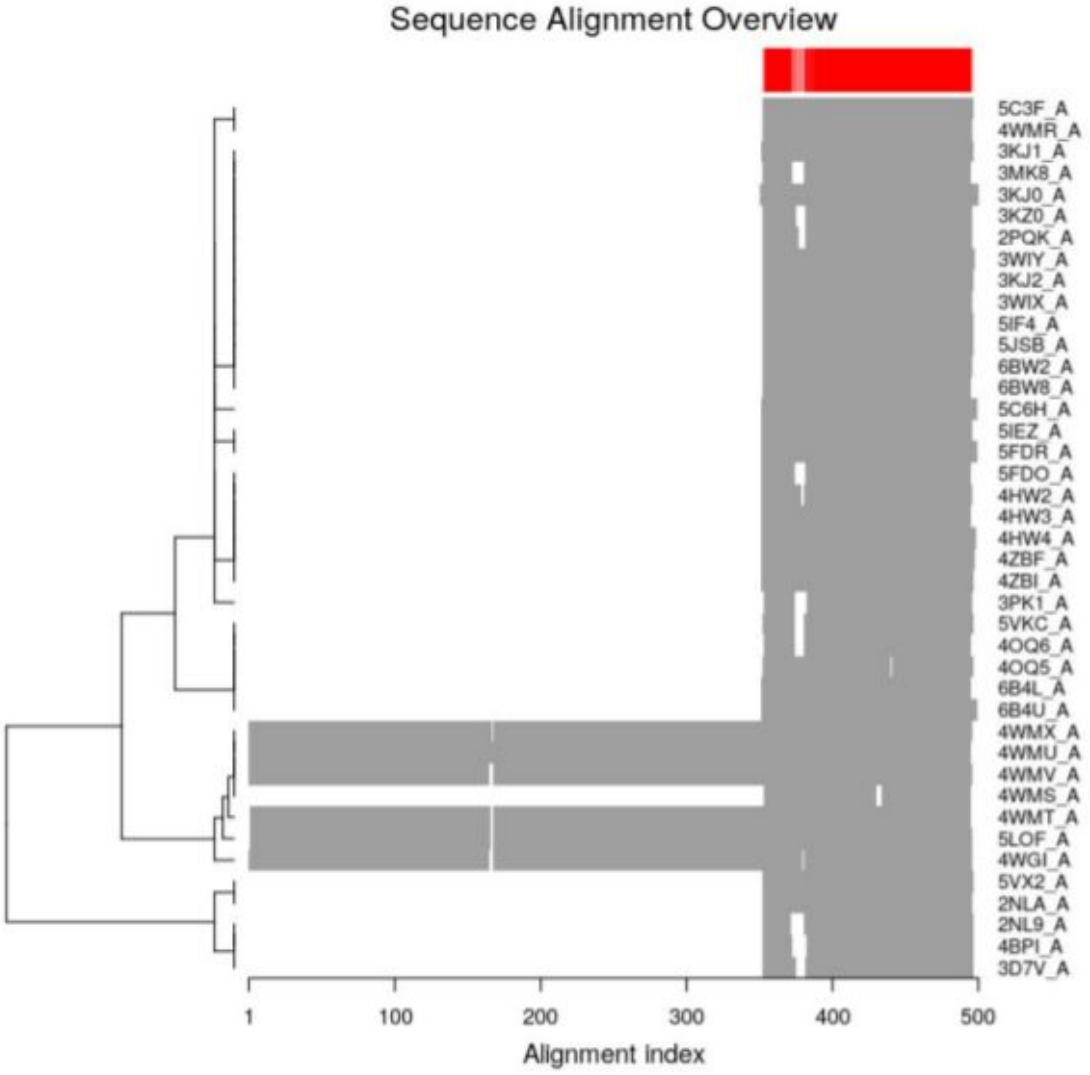


Figure S1. Mcl-1 full X-ray dataset (41 entry) and sequence alignment. Gap positions are shown in white, conserved regions in grey. Residue conservation, set-wise, is highlighted by the red ribbon on top. The entries 4WM[XUVT] are co-expressed with the maltose-binding protein (alignment index 1~350).

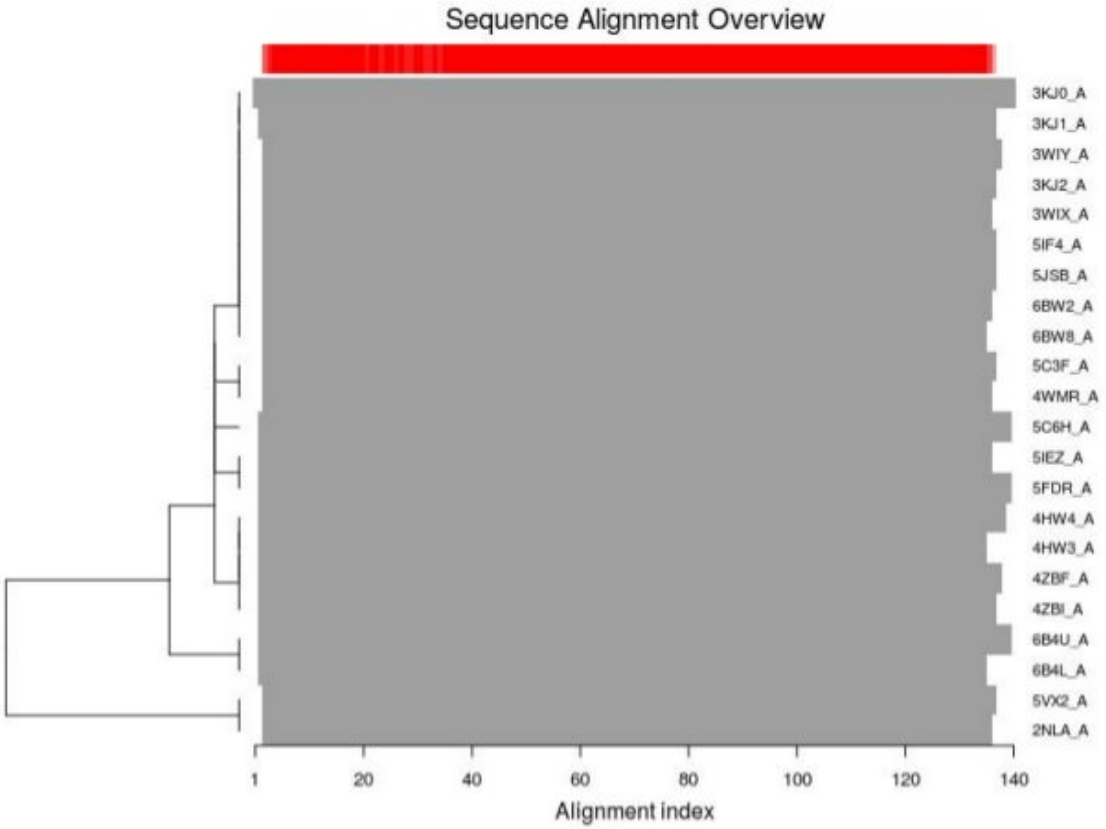


Figure S2. Mcl-1 reduced dataset (no missing residues) (22 entry) and sequence alignment.

* The full analysis procedure is included in the R Jupyter notebook (**Mcl-1_ens.ipynb**) See also (**Mcl-1_ens.pdf**) for a rendered version.

II. Metadynamics simulation

II.1 convergence assessment

In order to check the convergence, we :

1. show that many transitions are observed between different free energy wells (Figure S3)

2. compute the free-energy difference and error* between two minima (Figure S4)

3. compute the evolution of the Free Energy Profile along CV2 during the simulation (Figure S5)

4. Project the WT-MetaD Trajectory on the PCs.xray space (Figure S6).

*Error is estimated as the standard deviation of the free-energy distance between two minima.

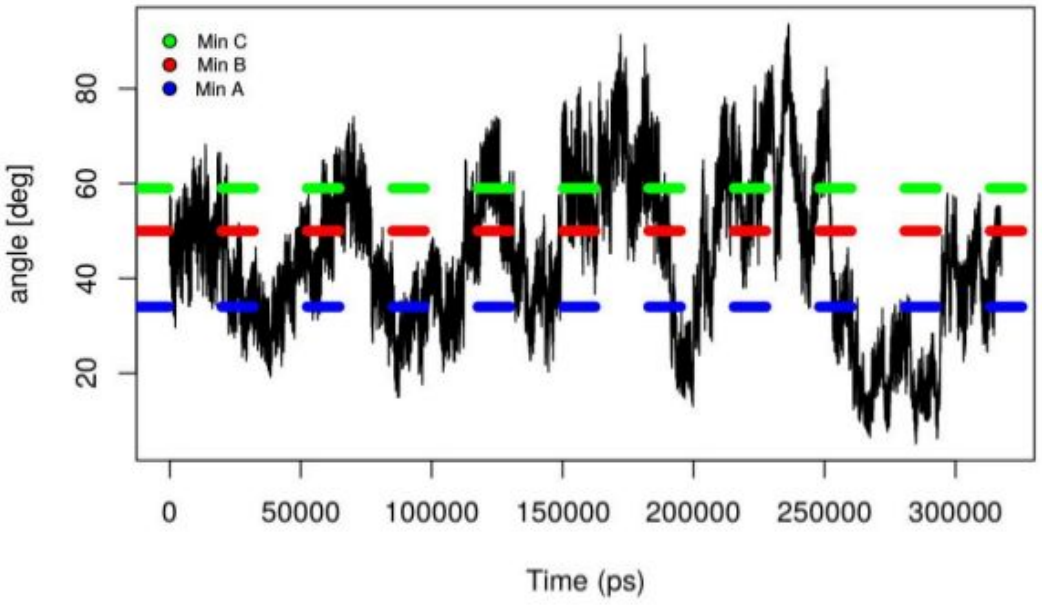


Figure S3. WT-MetaD convergence assessment by monitoring the evolution of the (angle)-CV [Thr226(Cα)-Asp241(Cα)-Ser255(Cα)] vs time vs minima.

The three most populated free-energy wells (A, B, and 'C') are sampled several times during the WT-MetaD simulation.

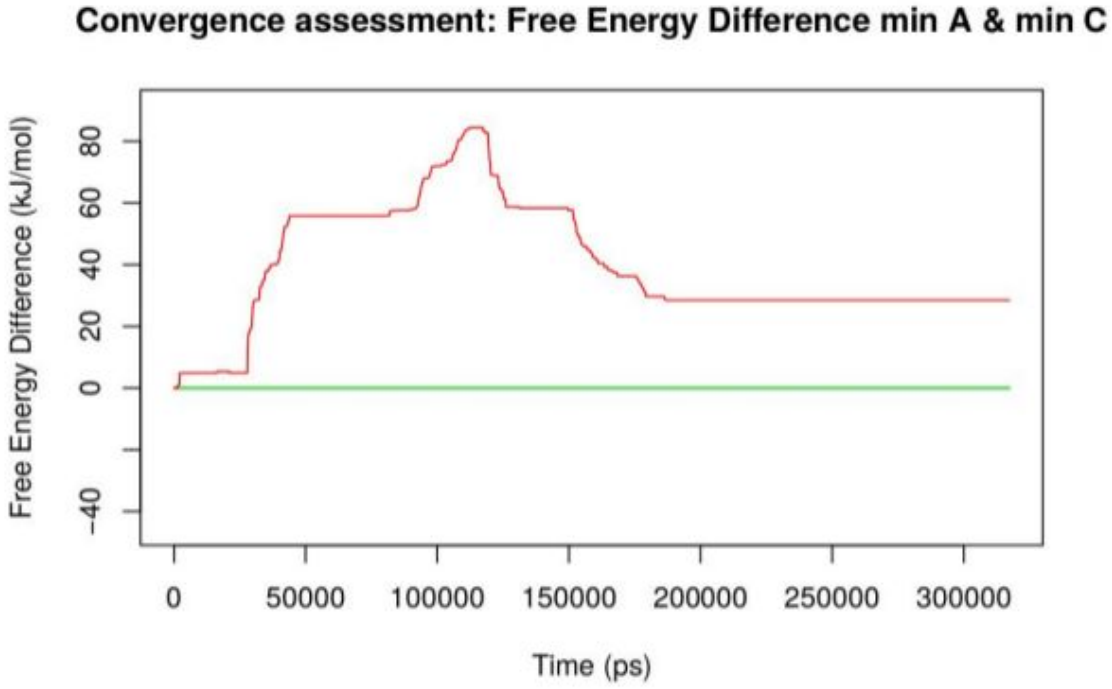


Figure S4. Free-energy difference between minimum A (green) and minimum C (in red).

From S4. we can see that the free-energy difference between min A and min C (and hence the free-energy surface) reaches a plateau of stability after ~ 180 ns of simulation time.

The error estimated as the standard deviation of this free-energy difference (180-317ns) equals 0.25 kJ/mol.

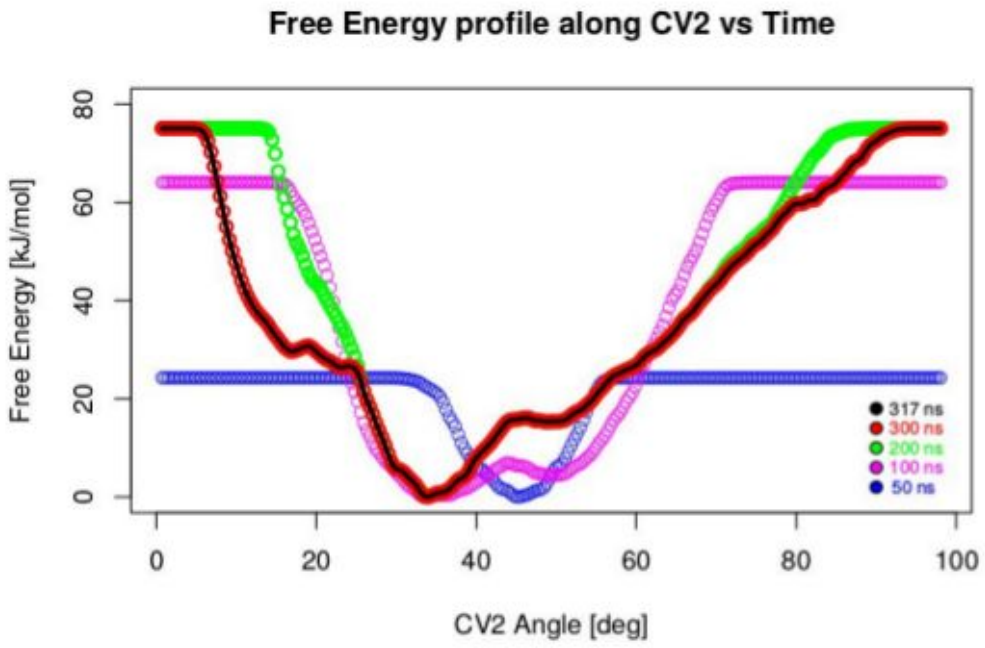


Figure S5. Evolution of the Free Energy Profile along CV2 (angle) vs Time.

In agreement with the data presented in Figure S4. the free energy profile around minima converges starting from 200 ns.

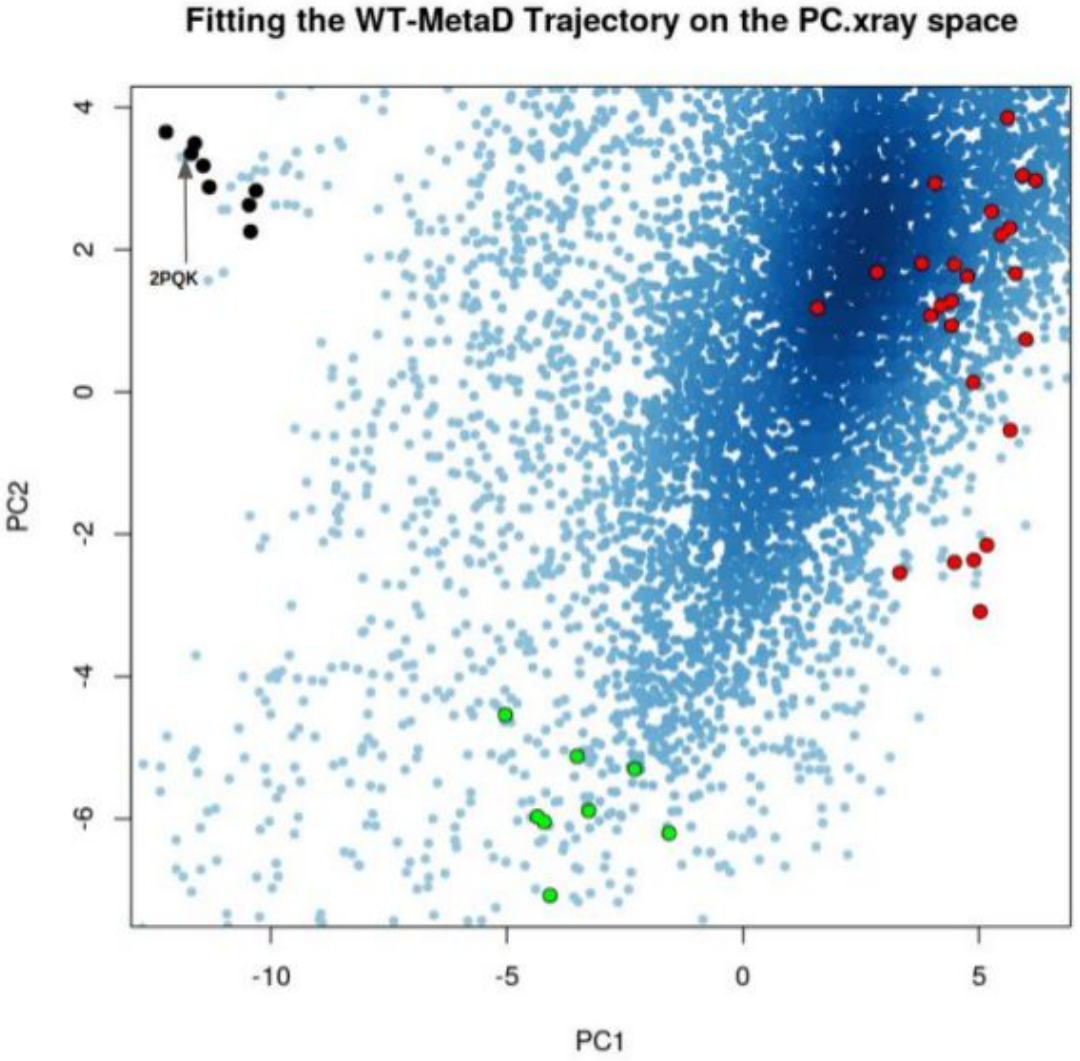


Figure S6. Projecting the Well-tempered Metadynamics trajectory on the PC.xray space. Black and green points represent peptide-bound structures. Red points represent (APO and synthetic-ligands bound structures) cluster. (complements Figure 6 in the main text)

The projection of the WT-MetaD on the PCs space derived from the Mcl-1 X-ray dataset shows the convergence and the sampling efficiency that we monitored previously.

II.2 Nudged Elastic Band (NEB) analysis: FE barrier and TS estimation

The transition barriers between free-energy wells, reported in the main text were estimated via the Nudged Elastic Band Analysis [1].

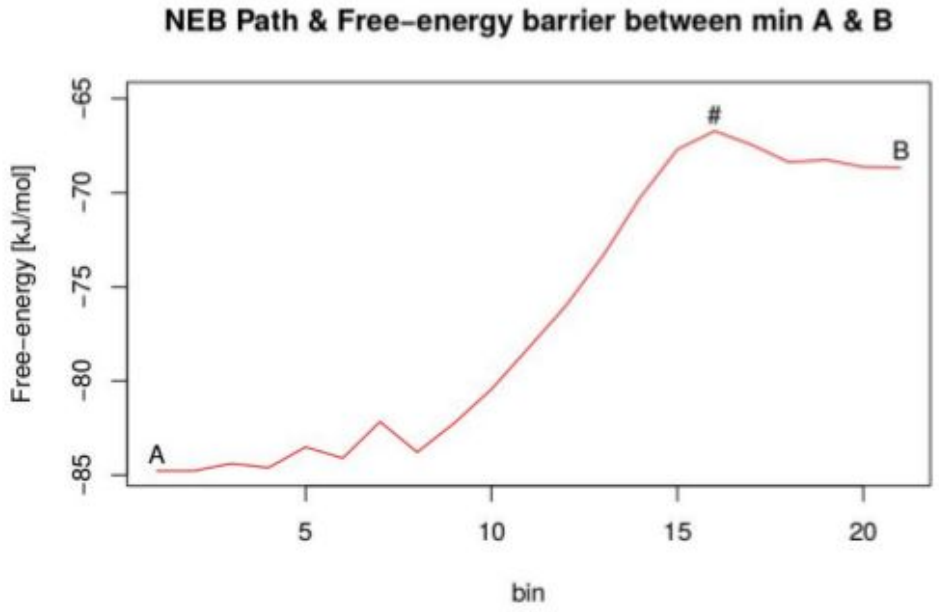


Figure S7. NEB analysis and transition state. A free-energy barrier of 18 kJ/mol (min A -> min B) was estimated.

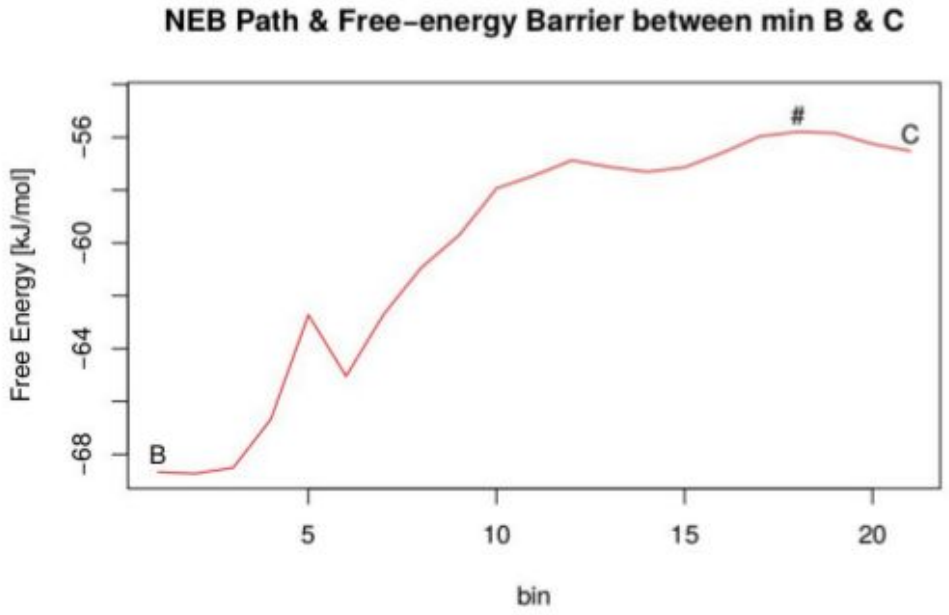


Figure S8. NEB analysis and transition state. A free-energy barrier of 12.8 kJ/mol (min B -> min C) was estimated.

** The full analysis procedure is included in the R markdown notebook (**analyze_fes.Rmd**)

See also (**fes_analyze.pdf**) for a rendered version.

*** All the supporting information, additional files and datasets needed are made publicly available at https://osf.io/4fqph/

References

(1) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *The Journal of Chemical Physics*. 2000, pp 9978–9985. https://doi.org/10.1063/1.1323224.

3. Conclusion et perspectives

Dans cette étude, la dynamique de Mcl-1 a été caractérisée en considérant trois volets principaux:

D'abord, l'analyse des ensembles structuraux a permis de dégager deux axes de dynamique essentielle pour Mcl-1. Un premier axe, concerne le mouvement de 'respiration' des hélices formant son site orthostérique. Un deuxième axe représente un mouvement corrélé de ses boucles internes. Les résidus clés contribuant aux mouvements essentiels ont été identifiés.

En deuxième lieu, à travers une caractérisation d'une surface d'énergie libre, nous montrons que Mcl-1 adopte un équilibre conformationnel entre une conformation fermée et une conformation intermédiaire ("ready to bind").

Finalement, le mécanisme d'inhibition allostérique a été révélé et un modèle de transmission du signal allostérique a été proposé.

Toutefois, il est à noter que la protéine Mcl-1 exerce ses fonctions au sein de la membrane externe de la mitochondrie et non dans le cytosol. De ce fait, sa structure et ses états conformationnels au sein d'un environnement membranaire reste une question à adresser.

4. Références

(1) Saleh, T.; Kalodimos, C. G. Enzymes at Work Are Enzymes in Motion. *Science* **2017**, *355*, 247–248.
(2) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503–6515.
(3) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598–1603.
(4) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061.
(5) Shamas-Din, A.; Kale, J.; Leber, B.; Andrews, D. W. Mechanisms of Action of Bcl-2 Family Proteins. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a008714.
(6) Rinkenberger, J. L.; Horning, S.; Klocke, B.; Roth, K.; Korsmeyer, S. J. Mcl-1 Deficiency Results in Peri-Implantation Embryonic Lethality. *Genes Dev.* **2000**, *14*, 23–27.
(7) Opferman, J. T.; Letai, A.; Beard, C.; Sorcinelli, M. D.; Ong, C. C.; Korsmeyer, S. J. Development and Maintenance of B and T Lymphocytes Requires Antiapoptotic MCL-1. *Nature* **2003**, *426*, 671–676.
(8) Dzhagalov, I.; Dunkle, A.; He, Y.-W. The Anti-Apoptotic Bcl-2 Family Member Mcl-1 Promotes T Lymphocyte Survival at Multiple Stages. *J. Immunol.* **2008**, *181*, 521–528.
(9) Opferman, J. T.; Iwasaki, H.; Ong, C. C.; Suh, H.; Mizuno, S.-I.; Akashi, K.; Korsmeyer, S. J. Obligate Role of Anti-Apoptotic MCL-1 in the Survival of Hematopoietic Stem Cells. *Science* **2005**, *307*, 1101–1104.
(10) Dzhagalov, I.; St. John, A.; He, Y.-W. The Antiapoptotic Protein Mcl-1 Is Essential for the Survival of Neutrophils but Not Macrophages. *Blood* **2007**, *109*, 1620–1626.
(11) Steimer, D. A.; Boyd, K.; Takeuchi, O.; Fisher, J. K.; Zambetti, G. P.; Opferman, J. T. Selective Roles for Antiapoptotic MCL-1 during Granulocyte Development and Macrophage Effector Function. *Blood* **2009**, *113*, 2805–2815.
(12) Arbour, N.; Vanderluit, J. L.; Le Grand, J. N.; Jahani-Asl, A.; Ruzhynsky, V. A.; Cheung, E. C. C.; Kelly, M. A.; MacKenzie, A. E.; Park, D. S.; Opferman, J. T.; et al. Mcl-1 Is a Key Regulator of Apoptosis during CNS Development and after DNA Damage. *J. Neurosci.* **2008**, *28*, 6068–6078.
(13) Akgul, C. Mcl-1 Is a Potential Therapeutic Target in Multiple Types of Cancer. *Cell. Mol. Life Sci.* **2009**, *66*, 1326–1336.
(14) Thomas, L. W.; Lam, C.; Edwards, S. W. Mcl-1; the Molecular Regulation of Protein Function. *FEBS Lett.* **2010**, *584*, 2981–2989.
(15) Denis, C.; Sopková-de Oliveira Santos, J.; Bureau, R.; Voisin-Chiret, A. S. Hot-Spots of Mcl-1 Protein. *J. Med. Chem.* **2020**, *63*, 928–943.
(16) Liu, G.; Poppe, L.; Aoki, K.; Yamane, H.; Lewis, J.; Szyperski, T. High-Quality NMR Structure of Human Anti-Apoptotic Protein Domain Mcl-1(171-327) for Cancer Drug Design. *PLoS One* **2014**, *9*, e96521.
(17) Clifton, M. C.; Dranow, D. M.; Leed, A.; Fulroth, B.; Fairman, J. W.; Abendroth, J.; Atkins, K. A.; Wallace, E.; Fan, D.; Xu, G.; et al. A Maltose-Binding Protein Fusion Construct Yields a Robust Crystallography Platform for MCL1. *PLoS One* **2015**, *10*, e0125010.
(18) Luptak, J.; Bista, M.; Fisher, D.; Flavell, L.; Gao, N.; Wickson, K.; Kazmirski, S. L.;

Howard, T.; Rawlins, P. B.; Hargreaves, D. Antibody Fragments Structurally Enable a Drug-Discovery Campaign on the Cancer Target Mcl-1. *Acta Crystallogr D Struct Biol* **2019**, *75*, 1003–1014.
(19) Fire, E.; Gullá, S. V.; Grant, R. A.; Keating, A. E. Mcl-1-Bim Complexes Accommodate Surprising Point Mutations via Minor Structural Changes. *Protein Sci.* **2010**, *19*, 507–519.
(20) Tanaka, Y.; Aikawa, K.; Nishida, G.; Homma, M.; Sogabe, S.; Igaki, S.; Hayano, Y.; Sameshima, T.; Miyahisa, I.; Kawamoto, T.; et al. Discovery of Potent Mcl-1/Bcl-xL Dual Inhibitors by Using a Hybridization Strategy Based on Structural Analysis of Target Proteins. *J. Med. Chem.* **2013**, *56*, 9635–9645.
(21) Maity, A.; Majumdar, S.; Ghosh Dastidar, S. Flexibility Enables to Discriminate between Ligands: Lessons from Structural Ensembles of Bcl-Xl and Mcl-1. *Comput. Biol. Chem.* **2018**, *77*, 17–27.
(22) Lee, S.; Wales, T. E.; Escudero, S.; Cohen, D. T.; Luccarelli, J.; Gallagher, C. G.; Cohen, N. A.; Huhn, A. J.; Bird, G. H.; Engen, J. R.; et al. Allosteric Inhibition of Antiapoptotic MCL-1. *Nat. Struct. Mol. Biol.* **2016**, *23*, 600–607.
(23) La Sala, G.; Decherchi, S.; De Vivo, M.; Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent Sci* **2017**, *3*, 949–960.
(24) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899–907.
(25) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695–2696.
(26) Yao, X.-Q.; Scarabelli, G.; Skjærven, L.; Grant, B. J. The Bio3D Package: New Interactive Tools for Structural Bioinformatics. *Biophys. J.* **2014**, *106*, 406a.
(27) Gerstein, M.; Altman, R. B. Average Core Structures and Variability Measures for Protein Families: Application to the Immunoglobulins. *J. Mol. Biol.* **1995**, *251*, 161–175.
(28) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226.
(29) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234.
(30) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
(31) Huang, J.; MacKerell, A. D., Jr. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.
(32) DeLano, W. L. PyMOL. 2002.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Refined TIP3P Model for Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
(34) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.

(35) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(36) Forester, T. R.; Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *J. Comput. Chem.* **1998**, *19*, 102–111.
(37) Coffey, W.; Kalmykov, Y. P. *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*; World Scientific, 2012.
(38) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, 27–28.
(39) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 1–22.
(40) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
(41) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(42) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
(43) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
(44) PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16*, 670–673.
(45) Hošek, P.; Spiwok, V. Metadyn View: Fast Web-Based Viewer of Free Energy Surfaces Calculated by Metadynamics. *Comput. Phys. Commun.* **2016**, *198*, 222–229.
(46) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.
(47) Decherchi, S.; Bottegoni, G.; Spitaleri, A.; Rocchia, W.; Cavalli, A. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.* **2018**, *58*, 219–224.
(48) Demelash, A.; Pfannenstiel, L. W.; Tannenbaum, C. S.; Li, X.; Kalady, M. F.; DeVecchio, J.; Gastman, B. R. Structure-Function Analysis of the Mcl-1 Protein Identifies a Novel Senescence-Regulating Domain. *J. Biol. Chem.* **2015**, *290*, 21962–21975.
(49) Fang, C.; D'Souza, B.; Thompson, C. F.; Clifton, M. C.; Fairman, J. W.; Fulroth, B.; Leed, A.; McCarren, P.; Wang, L.; Wang, Y.; et al. Single Diastereomer of a Macrolactam Core Binds Specifically to Myeloid Cell Leukemia 1 (MCL1). *ACS Med. Chem. Lett.* **2014**, *5*, 1308–1312.
(50) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
(51) Chu, W.-T.; Clarke, J.; Shammas, S. L.; Wang, J. Role of Non-Native Electrostatic Interactions in the Coupled Folding and Binding of PUMA with Mcl-1. *PLoS Comput. Biol.* **2017**, *13*, e1005468.
(52) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, *8*, 2276.

(53) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508*, 331–339.
(54) Donald, J. E.; Kulp, D. W.; DeGrado, W. F. Salt Bridges: Geometrically Specific, Designable Interactions. *Proteins* **2011**, *79*, 898–915.
(55) Barlow, D. J.; Thornton, J. M. Ion-Pairs in Proteins. *J. Mol. Biol.* **1983**, *168*, 867–885.
(56) Pritz, J. R.; Wachter, F.; Lee, S.; Luccarelli, J.; Wales, T. E.; Cohen, D. T.; Coote, P.; Heffron, G. J.; Engen, J. R.; Massefski, W.; et al. Allosteric Sensitization of Proapoptotic BAX. *Nat. Chem. Biol.* **2017**, *13*, 961–967.
(57) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391–6423.

CHAPITRE 4.

Détection de poches cryptiques par simulations de métadynamique sur l'espace de la dynamique essentielle: application sur Mcl-1

1. Motivation

Les protéines sont des entités dynamiques. Leur dynamique est souvent une signature de la fonction qu'elles exercent.
Les fonctions biochimiques qu'elles effectuent sont le fruit des interactions qu'elles établissent, soit en formant des complexes protéine-protéine ou protéine-ligand. Ces interactions requièrent des déformations au niveau de la surface protéique permettant la formation de poches.

Plusieurs types de poches sont répertoriés. Les poches cryptiques sont un type de poches majoritairement disponible à l'état lié de la protéine ("holo"). L'identification des poches cryptiques est une étape d'intérêt, notamment dans le cas des protéines difficiles à cibler par des approches classiques. Des approches expérimentales existent, mais elles sont souvent coûteuses et non applicables à très large échelle.

Dans ce contexte, les simulations de dynamique moléculaire (DM) représentent une approche de choix pour étudier et caractériser la dynamique de formation des poches cryptiques.
Néanmoins, l'application des DM dans la détection des poches cryptiques reste limitée pour deux raisons principales:

- L'ouverture des poches cryptiques nécessite souvent le passage de barrières énergétiques difficilement atteignables par DM non biaisée.
- L'application des DM biaisées nécessite un choix pertinent des variables collectives, capables de représenter efficacement les déterminants de l'ouverture des poches cryptiques.

Dans ce chapitre, la métadynamique sur l'espace de la dynamique essentielle, est combinée avec un algorithme de détection des poches (Pocketron), dans le but d'identifier et répertorier les poches cryptiques d'une protéine. Cette approche a été testée sur notre cible d'étude: la protéine Mcl-1.

2. Article 3

Cryptic pockets repository through pocket dynamics tracking and metadynamics on essential dynamics space: application to Mcl-1

Mohammed Benabderrahmane[†], Ronan Bureau[†] and Jana Sopkova-de Oliveira Santos[†,*]

[†] *Normandy Univ, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), 14000 Caen, FRANCE*

* Corresponding author
Email: jana.sopkova@unicaen.fr

In preparation

Abstract

Cryptic pockets (hidden protein pockets) detection is a hot topic in structure-based drug discovery, especially for drugging the still undruggable proteome. The experimental detection of cryptic pockets is still considered an expensive endeavor. Thus, computational methods, such as atomistic simulations are used instead. These simulation methods are able to provide a perspective on the protein dynamics that overpasses the static and average view provided by the experimental X-ray structures. Nonetheless, unbiased MD simulations fall short for the detection of transient and cryptic pockets requiring the crossing of high energy barriers. Enhanced sampling methods, such as Metadynamics, provide a solution to overcome the time-scale problem faced by unbiased molecular dynamics simulations. However, these methods are still limited by the availability of a collective variables space able to clearly capture the intricate parameters leading to cryptic pockets opening. Unfortunately, the design of such CVs requires *a priori* knowledge of the binding site, information that is by definition lacking for cryptic pockets. In this work, we evaluated the use of the Metadynamics biasing scheme on essential coordinates space as a general method for cryptic pocket detection. This approach was applied to an anti-apoptotic protein: Mcl-1 as a test model. The results obtained show the efficiency of this method to draw in an unsupervised way the full repository of Mcl-1's known and new cryptic pockets.

Keywords Mcl-1, Protein dynamics, Metadynamics on essential coordinates; cryptic pockets detection.

1. Introduction

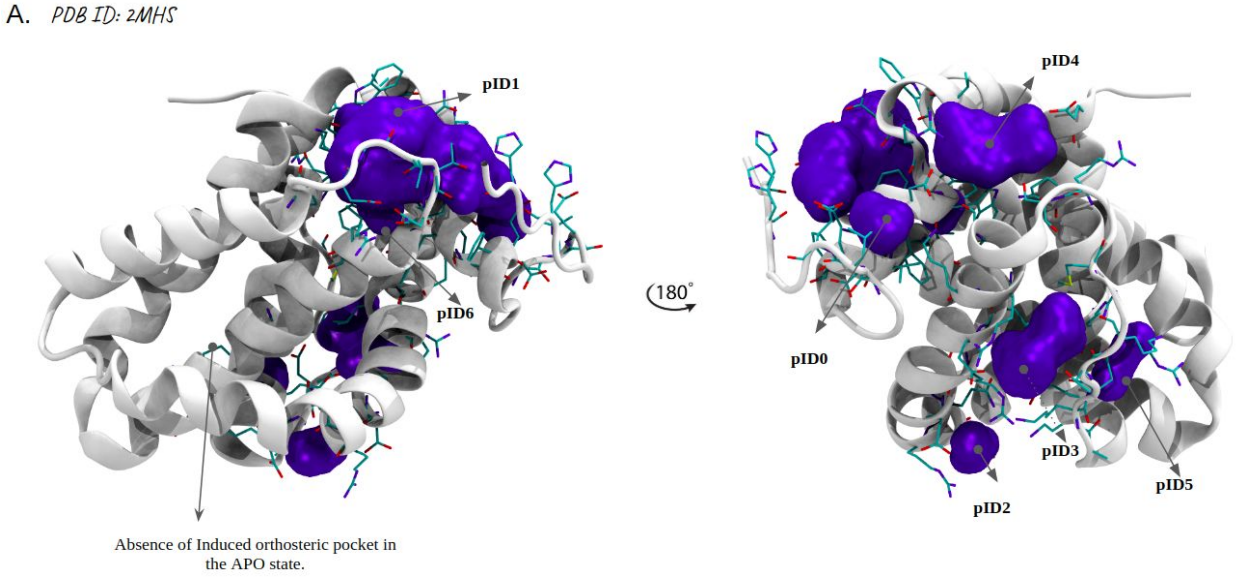
Proteins are biomolecular entities required for a large set of biological functions assured by living cells. The comprehension of their dynamics is necessary to unravel the full spectrum of the functions they assure.[1] Protein perform their function by interacting with each other via protein-protein interactions or by interacting with other molecules (small ligands, cofactors ...etc). This interaction requires a deformation at the protein's surface in order to favor the formation of the protein/ligand complex. These deformations at the surface are generally termed protein pockets. Many types of pockets were described in the literature.[2] Among them, cryptic pockets form a type of pockets available only at the holo state of the protein. The detection of this type of pockets is possible via site-directed tethering,[3] the use of antibodies[4], or large-scale fragment screening.[5] However, such experimental methods are still expensive, which limits their routinely use for an early drug discovery stage. In such settings, computer simulations of protein dynamics are able to provide key structural and conformational elements to guide the detection of the cryptic pockets.

The first requirement for protein dynamics simulations is the availability of a good starting experimental structure. Such data is usually acquired via X-ray and NMR experiments. The structures provided by X-ray experiments and hence the pockets computed from these structures are limited by the static and average nature of such data **(Figure 1)**.

The second challenge that faces the detection of cryptic pockets via unbiased molecular dynamics simulations is the time-scale problem and the conformational ensembles covered. In fact, even long time-scale simulations fail to detect transient and hidden pockets and the exploration of significant conformational changes is not reachable even by using modern computing architectures. Enhanced sampling methods such as Umbrella Sampling,[6] Metadynamics[7] were developed to deal with such reduced sampling problems. However, these methods require a collective variable (CV) space to explore. Unfortunately, for cryptic binding sites detection, the knowledge of the binding site is by definition lacking. Thus, the design of general CV spaces is of great interest, and efforts were made in this direction, by using for instance the hydrophobicity of the cryptic pockets as CV,[8] but still, these methods require an *a priori* knowledge of the binding site.

In this work, we evaluate the use of the Metadynamics scheme on essential coordinates space as a general approach to detect cryptic pockets. The essential coordinates space is computed via the essential dynamics[9,10] extracted from an unbiased MD simulation.

This approach was evaluated on the Bcl-2 protein family member: Mcl-1 **(Figure1)**. This protein is an anti-apoptotic member with an increasing therapeutic interest for anti-cancer drug development.[11] The results that we obtained show the efficiency of this strategy to draw the repository of the cryptic pockets at the protein surface and to enhance the exposure of the pockets.



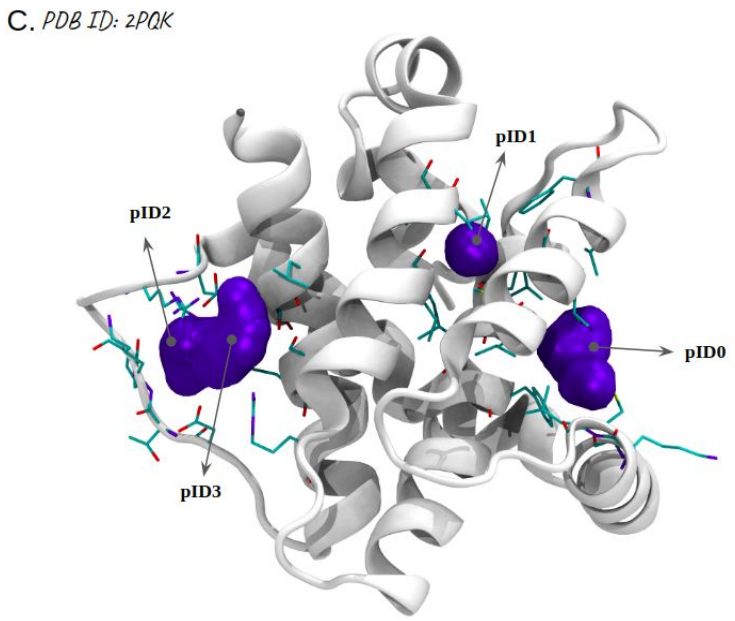


Figure 1. Cryptic pockets detection from experimental structures is limited by the average and static nature of the data. The pockets detected from three different conformational states of Mcl-1 are shown. **A.** APO state (from an NMR model, PDB ID 2MHS); **B.** Holo state synthetic ligand-bound PDB ID 3WIX; **C.** peptide-bound state (PDB ID 2PQK). The full list of residues forming the pockets and their respective average volumes are available in supplementary information Table (**Table T1**).

2. Material and methods

2.1 Unbiased molecular dynamics simulations

Two unbiased molecular dynamics simulations of 100 ns each were performed. The first, starting from the peptide-bound induced conformation (PDB ID: 2PQK)[12], where the Bim-BH3 peptide was removed (2PQK-MD simulation). The second, starting from the synthetic ligand-bound structure (PDB ID: 3WIX)[13], where the LC3 synthetic ligand was removed as well (3WIX-MD simulation). The Schrödinger's Protein Preparation Wizard (PrepWizard) was used to prepare the starting structures. All of the molecular dynamics simulations were performed using NAMD 2.12[14] with the Charmm36m forcefield.[15] To simulate the aqueous solvent environment, each system was surrounded by a rectangular box of TIP3P explicit water model,[16] and 0.15 M NaCl was added to the system using the CHARMm-GUI solvator.[17] The chosen box size ensured that the simulated system was at a minimum distance of 10 Å from the edge. The VDW interactions were truncated using a

force switching function between 10 and 12 Å, and the Particle Mesh Ewald (PME)[18] method was used to compute the long-range electrostatic interactions. The SHAKE[19] algorithm was applied to restrain all bonds involving hydrogen atoms. The systems underwent energy minimization for 10.000 steps. The systems were afterward heated up to 303.15 K, and the system was temperature-equilibrated during 50 ps via heating reassignment under NVT (canonical ensemble) conditions. Finally, the system ran freely for 100 ns under NPT (isothermal-isobaric ensemble) conditions. Langevin dynamics with a damping coefficient of 1 ps^{-1} was used to maintain the system temperature, and the Nosé-Hoover Langevin[20,21] piston method was used to control the pressure at 1 atm.

2.2 Metadynamics on essential coordinates space

A well-tempered Metadynamics[22] simulation was used to enhance the sampling of the configurational space. The Metadynamics scheme requires a collective variable (CV) space to explore. This step is critical since the observables are dependent on the choice of the CV space. The essential dynamics (soft modes)[23] captured via principal components analysis were used in this work. The 2PQK-MD simulation (100 ns) was used as a learning ensemble to extract the essential dynamics space (PC1 vector space is illustrated in **Figure 2**). This choice was made because the 2PQK-MD simulation explores the transition between the peptide-bound conformation (the starting one) and the APO state, as we demonstrated in our last work.[24] The Metadynamics starting parameters were estimated from the same unbiased MD simulation (2PQK-MD). A Gaussian width of 0.1 was used for both pc1.eig and pc2.eig (see ***plumed_bs12.dat*** file for more details). The Gaussian width value was equivalent to the standard deviation of the collective variables during the unbiased run (2PQK-MD). During the simulation, Gaussians were deposited every 500 steps (PACE=500), with an initial height of 0.4 and a bias factor of 12. A total of 750 ns was produced at 300 K under the NVT ensemble. The Metadynamics simulation was performed using NAMD 2.12 patched with PLUMED 2.3.2.[25] The Gaussian hills produced were analyzed using the Metadynminer R package.[26] The convergence of the simulation was assessed qualitatively by monitoring the free-energy surface and profile during the time (**Figure S3**).

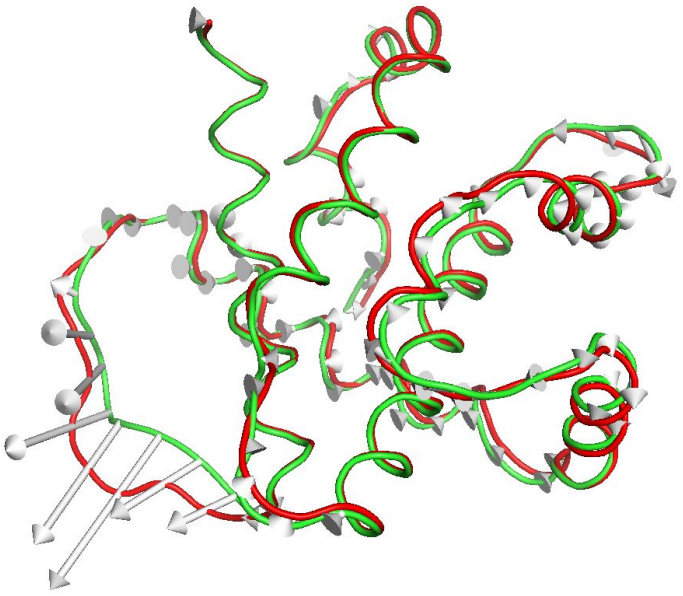


Figure 2. Porcupine plot illustrating the PC1 vector space. Two structures interpolated from the PC1 are shown (green and red). The arrows show the direction of the variation for each Cα-atom during the unbiased MD simulation (2PQK-MD).

2.3 Pocket's identification and tracking

Static pocket detection was performed using the NanoShaper algorithm.[27] NanoShaper's definition of a static pocket is based on the difference between the solvent excluded volumes, computed via the Connolly−Richards[28] definition by considering two probe-radii. The results reported in this paper were obtained by considering a 3 Å for the big probe radius and 1.4 Å for the small probe radius.

For pocket tracking during the MD simulations, we used the Pocketron algorithm.[29] Pocketron tracks the pockets during an MD run and provides atom (hence residues) exchange between the pockets. This information is crucial for the detection of transient and cryptic pockets. On top of providing observables such as pocket volumes, this algorithm is able also to derive a semi-quantitative view of the dynamics of the pockets detected by providing a network-based overview. For each simulation, water and ions were removed to avoid artifacts. A minimum volume of a detectable pocket was set to 4 water molecules to avoid the detection of very small pockets.

3. Results and discussion

3.1 Cryptic pocket networks are MD conformational-ensemble dependent and restricted by the sampling limitations

In order to overcome the limitation of pocket detection due to the average nature of the protein structures provided by X-ray experiments,[30,31] we performed two unbiased molecular dynamics simulations. Moreover, to explore the effect of the conformational ensembles sampled, two different starting structures were selected. The resulting trajectories were submitted to the pocketron algorithm to derive the set of pockets detected during the dynamics. The pockets and the networks representing the dynamics are presented in **Figure 3**.

For the 3WIX-MD network (**Figure 3.A**), seven pockets were found: three pockets at the orthosteric binding interface (pID7, pID1, pID11) and four pockets (pID13, pID2, pID9, pID3) at the distal binding site. The pockets (pID7, pID1, and pID11) have an average volume of 56, 87, and 20 $Å^3$ respectively. The pockets at the distal binding site were slightly more open, with an average volume varying from 156 $Å^3$ for pID9 to 30 $Å^3$ for the pocket pID2.

Concerning the 2PQK-MD network (**Figure 3.B**), six pockets were found: two pockets near the orthosteric binding site (pID1, pID4); three pockets at the allosteric binding site (pID10, pID9, pID6), and one pocket (pID3) at the distal binding site. The pocket volume time series fluctuated around an average of 177 $Å^3$ (with a maximum volume of 700 $Å^3$) for the distal binding site pocket, 20 $Å^3$ (with a maximum of 350 $Å^3$ for pID9) for the allosteric pockets, and 16 $Å^3$ (maximum of 200 $Å^3$) for the orthosteric pocket (see supporting information Table T2 for the full set of pockets/residues/volumes).

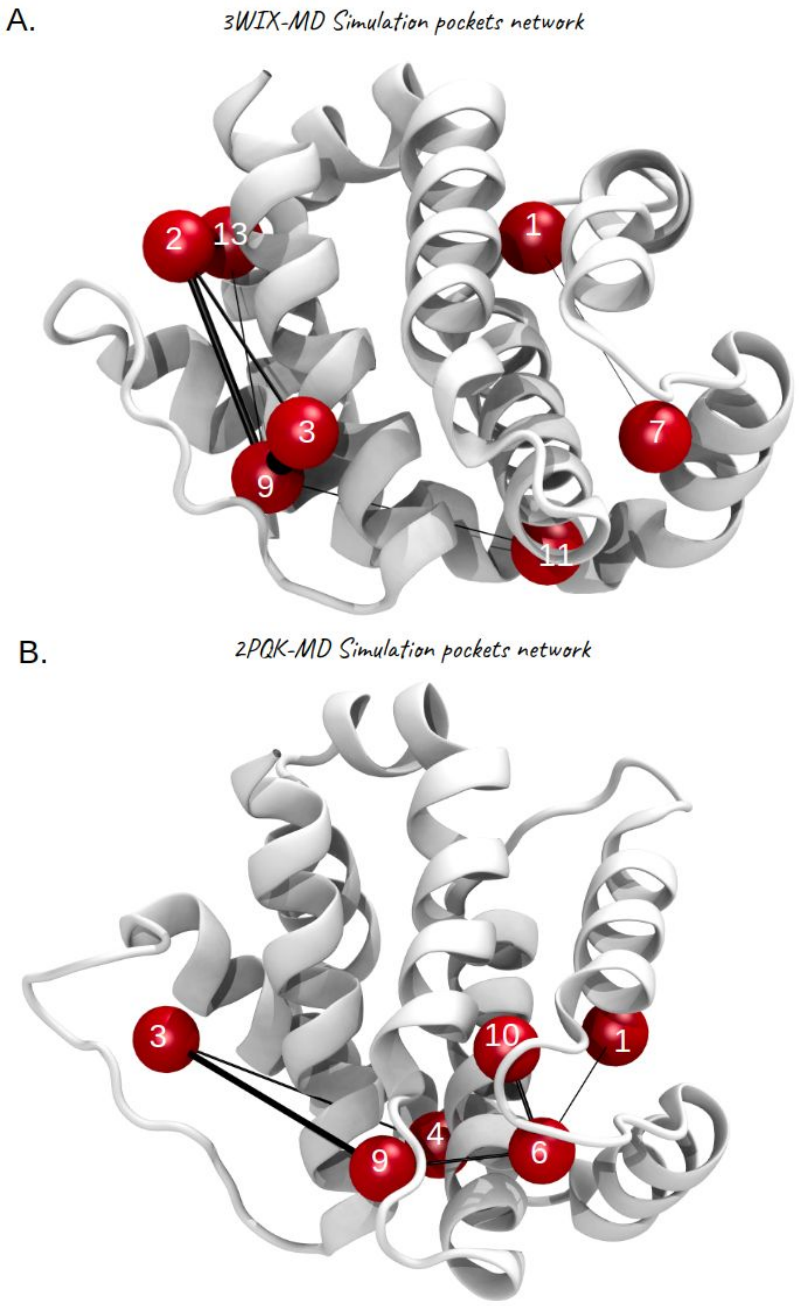


Figure 3. Pocket network for two independent MD simulations. A. MD-3WIX; B. 2PQK-MD. The cryptic pocket sets detected from the unbiased MD simulations are strongly related to the initial structure and the conformational ensemble available during the MD simulation.

The sets of pockets detected were different. Indeed, the cryptic pockets forming the allosteric binding site (near the Cys286) were not detected in the 3WIX-MD simulation. This is mainly due to the fact that during the 2PQK-MD simulation, the protein relaxes from the holo-like conformation (peptide-bound induced conformation) towards the APO state by

engaging the backbone atoms (required for the internal loops). However, for the 3WIX-MD simulation, the removal of the synthetic ligand engages a slight reorientation of the protein side chains at the orthosteric pocket. We should highlight as well that the allosteric cryptic pockets detected in the 2PQK-MD simulation were not persistent and majorly closed at the simulation end (average volume around 20 $Å^3$). This is not surprising since the cryptic pockets require a binding event with a ligand (or fragment) to be fully stabilized (cryptic pockets are described to be binding pockets with less hydrophobicity than normal pockets). These observations are directly linked to the fact that through the dynamics, the conformational population shifts toward the APO state (where the cryptic pockets are mainly supposed to be closed) but using different pathways (in the high dimensional free energy surface).

All these elements limit the use of conventional MD simulation in cryptic pockets discovery and more efficient approaches are sought after.

3.2 Solving the conformational ensemble and sampling limitation by introducing Metadynamics on essential coordinates space

We have shown in the last section that the set of transient and cryptic pockets detected were very dependent on the starting structure used and the conformational ensemble covered during the MD simulation. This drawback is a well-known problem of conventional MD simulations, where large conformational transitions requiring a significant entropic component are very costly to sample (the time-scale is hardly reachable even by modern architectures).

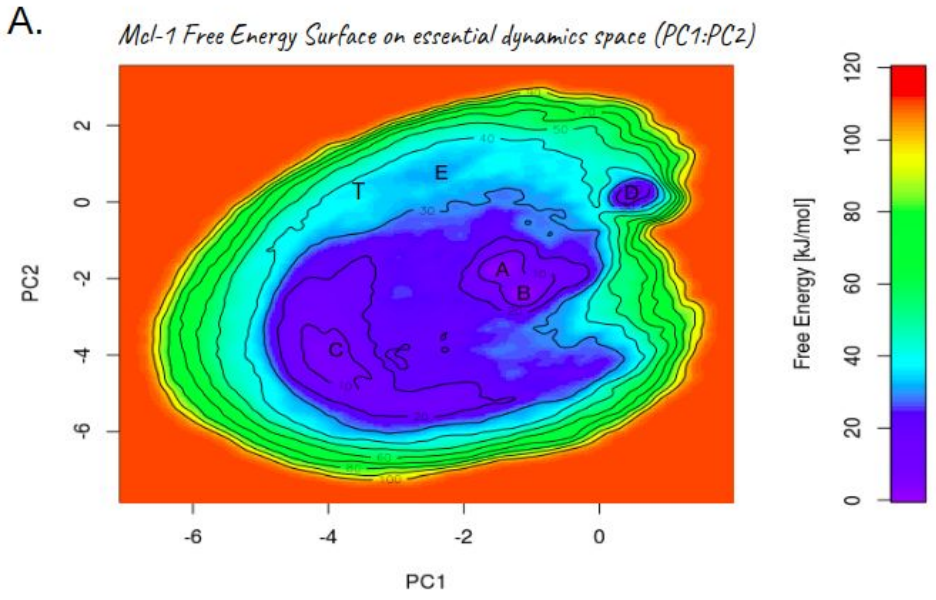
Enhanced sampling techniques, such as Metadynamics, have proven to be effective in sampling conformational states of proteins. The use of such methods requires the identification of a set of general collective variables able to capture the intricate parameters leading to the pocket opening. Efforts were made in this direction but they still assume an *a priori* knowledge of the binding site.[8] Unfortunately, such knowledge is rarely available for cryptic pockets.

To resume, two challenges face the identification of cryptic pockets. The first is a limitation related to the exploration of the phase space and the second concerns a general CV space able to focus the sampling on the protein binding sites. To solve the sampling limitation inherent to unbiased MD simulations, we enhanced the conformational sampling by adding a well-tempered Metadynamics bias. Moreover, to restrict the sampling to "functional" space, we considered the essential dynamics space (PC1: PC2) as collective variables. This space is physically consistent and does not depend on *a priori* knowledge of the system binding sites (CV space is learned in an unsupervised way from the unbiased MD).

We present hereafter the free energy surface obtained and the main conformational states sampled during the dynamics.

3.2.1 Free-energy surface

To expose the diversity of Mcl-1 structures explored during the metadynamics simulation, we reconstructed the free energy surface on the CV space (**Figure 4.A**) and extracted the structures corresponding to free energy wells. These structures are presented in **Figure 4.B**.



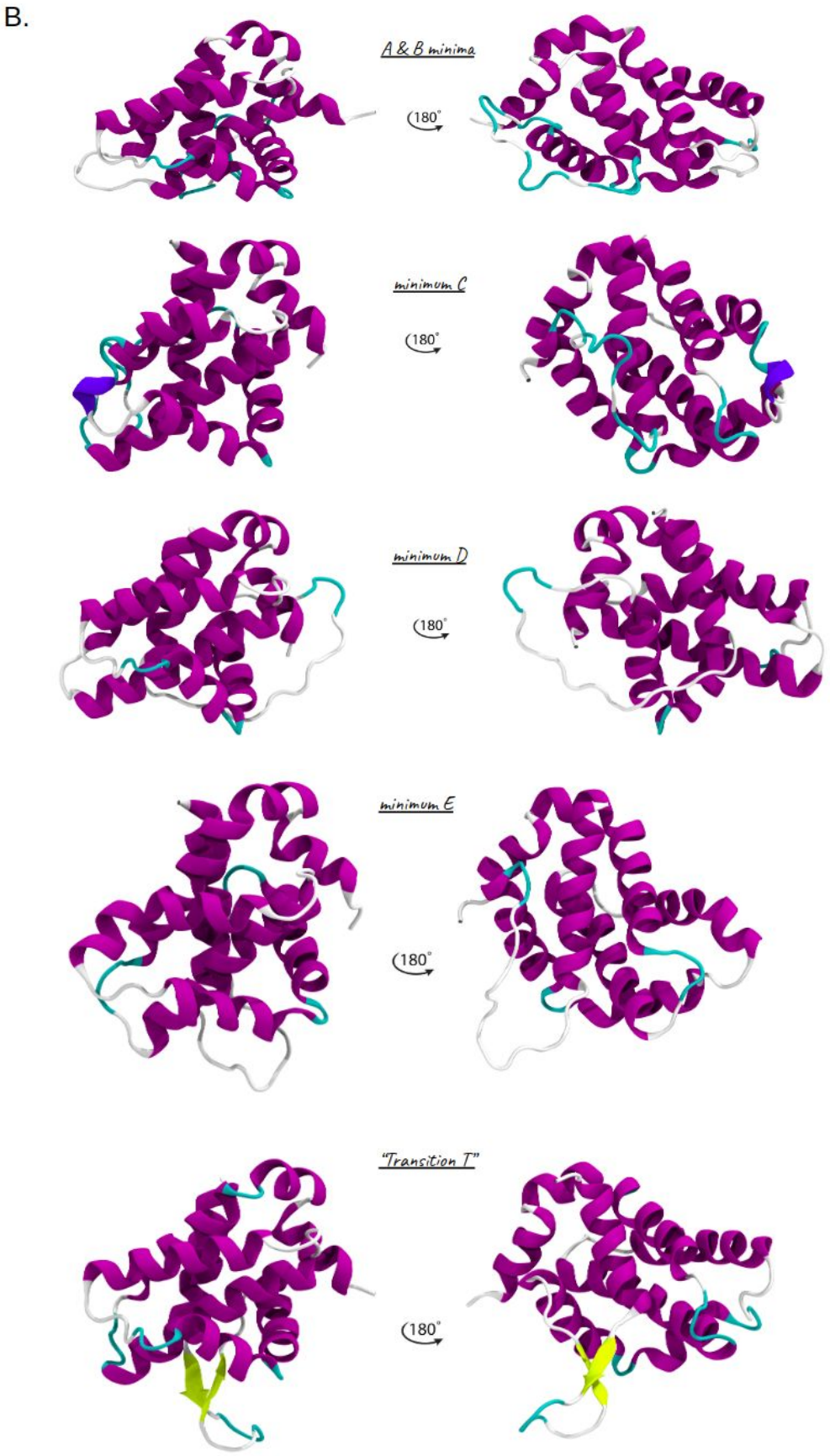

B.
A & B minima
180°
minimum C
180°
minimum D
180°
minimum E
180°
"Transition T"
180°

Figure 4. A. Mcl-1 free energy surface **B.** main representative structures from different minima. Secondary structures are represented using the STRIDE color scheme.

The free energy landscape reconstructed in **Figure 4.A** shows a wide conformational basin whose dynamics are mainly governed by the swing of the α1−α2 loop and local secondary structure variations at the orthosteric binding groove.

The minimum A and B represent the main populated wells (94 % of the conformational population in the APO state) with low free energy barriers (~ 8 kJ/mol) separating them. The free energy barriers separating the minima were computed using the nudged elastic band method (see SI **Figure S4**) and are estimated as follows:

- The transition minimum A → minimum C is equivalent to 23 kJ/mol
- The transition minimum A → minimum D is equivalent to 38 kJ/mol
- The transition minimum A → minimum E is equivalent to 33 kJ/mol

From a structural viewpoint **(Figure 4.B)**, the minima A and B represent the "equilibrium state" of Mcl-1 in its APO state, considering the position of the α1−α2 loop and the closed conformation of the orthosteric binding groove (α3 and α4 helices), in good agreement with the data we have recently reported.[24]

The transition from minimum A to minimum C requires the shift of the α1−α2 loop toward the α5−α6 loop and the formation of a 3_{10} helix (colored in blue) at the N-terminal region of the α3 helix. However, for minimum D, the α1−α2 loop shifts in the opposite direction (towards the helix α8) (transition from minimum C to minimum D). For the minimum E, the α1−α2 loop shifts toward the C-terminal part of helix α1.

We report as well, the formation of transient antiparallel beta-sheets in the α1-α2 loop (at residues Thr191...Gly203) (transition T in the FES). Even though such a local folding of the α1−α2 loop was never reported, this might be required for the functional recognition of this binding site. Such a transition state might as well be stabilized by interacting with another putative protein or by the phosphorylation state of Mcl-1's PEST region), especially knowing that this site is important for Mcl-1 anti-senescence activity and its degradation.

The dynamics of the α1-α2 loop are highly influenced by the conformational transitions at the C-terminal end of the α1 helix (residues Arg187 Glu188 Gln189 Ala190 Thr191) and the N-terminal end of α2 helix (Gly203 Ala204 Thr205 Ser206 Arg207 Lys208). These

residues undergo helix to coil transitions (see SI **Figure S5**) leading to the shortening and extension of the loop.

In fact, even small changes in the local conformational states of Mcl-1 is known to be important for its function. A good example of that is the conformational switch at residues (Q221 R222 N223), proposed by the Zhang group,[32] that controls the ubiquitination state of Mcl-1: a key process engaged in its degradation and apoptosis.

In parallel, we recently reported the Arg187 and Glu188 (part of the C-terminal end of the α1 helix) residues as an important component of Mcl-1's allosteric network.

Combining these observations, we hypothesize that the C-terminal end of the α1 helix and the N-terminal end of the α2 helix might play the role of a functional switch as well for Mcl-1, contributing by shifting its conformational populations.

To resume, the free-energy surface and structures provided shows that Mcl-1 breathing motion is well conserved in large time-scales dynamics and that the protein conserves globally its folding pattern. This FES complements the data that we have recently proposed and highlights additionally, the major role of the α1−α2 loop in the conformational dynamics of Mcl-1.

After exposing the FES and the conformations explored during the simulation, we applied the pocket detection procedure on the Metadynamics trajectory to see the gain provided by such an enhanced sampling strategy for cryptic pocket detection.

3.2.2 Full cryptic pockets network derived from Metadynamics simulation

The trajectory obtained from the well-tempered metadynamics simulation was subject to the pocketron algorithm. The set of pockets discovered and their network is represented in **Figure 5**. The network derived shows the existence of three clusters of pockets, forming 13 pockets in total. These pockets correspond to the three binding sites of Mcl-1: a distal binding site near the α1−α2 loop, a central cluster representing the allosteric binding site, and finally a set of cryptic pockets distributed all along the α3−α4 loop and the orthosteric site. (refer to SI **Table T3** for the full list of residues forming the pockets and their volumes).

Figure 5. Cryptic and transient pockets detected from the Metadynamics simulation trajectory. The nodes represent pockets; edges represent the communication between pockets and their thickness is related to the amount of exchange between neighboring pockets. Three node clusters are detected, corresponding to Mcl-1 three binding sites.

The Pockets pID[17, 3, 4 and 5] represent cryptic pockets forming the distal binding site; pockets pID[6, 7, 1] represent cryptic pockets forming the allosteric binding site; pockets pID[9, 2, 12, 8, 26, 11] form the set of cryptic pockets related to Mcl-1 orthosteric binding groove. The allosteric cryptic pockets, especially pID7, form a hub for this network (the number of connections is relatively higher than the other nodes).

On top of accounting for the already known pockets of Mcl-1 (such as the pID9 which represents the traditional orthosteric pocket of Mcl-1), other cryptic pockets such as **(pID2 and pID1)** are newly found and to the best of our knowledge, were never described in the literature. The pID2 pocket is formed by residues (Ser293 Lys244 Asp296 Ser247 Val243 Pro289 Glu292) and sits between the α4 and α5 helices. We report as well the pID1 pocket formed by the residues (Lys276 **Gln221** Ala272 Phe273 Tyr185 Ser269 Gln189 Asp218 Glu225 Gly217 Phe228): this pocket connects the functional switch [Gln221, Arg222, Asn223] through the Gln221 to the rest of Mcl-1 binding sites.

We further analyzed the volumetric data (**Figure 6**) concerning the pockets exposure. Interestingly, pockets exposure was highly improved compared to unbiased MD simulations. If we consider Mcl-1's allosteric pocket (pID7), an average volume of 67 $Å^3$ was achieved, with a maximum opening volume at 920 $Å^3$; an open state (volume > 500 $Å^3$) was available more than six times during the Metadynamics simulation.

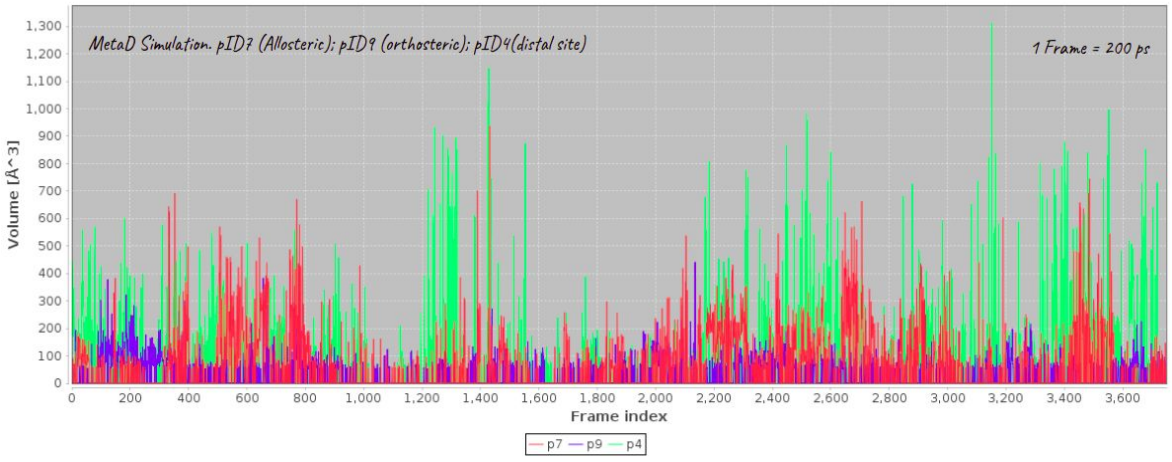


Figure 6. Evolution of the pocket volumes for three cryptic pockets during the Metadynamics simulation.

This data shows the ability of this approach to effectively sample the open conformations of Mcl-1's cryptic pockets and to derive the full cryptic pockets repository.

However, this comes with a computational cost. In fact, 750 ns were necessary to achieve a qualitative convergence of the free-energy surface. This is an expected problem when using general CV spaces such as the principal coordinates. One of the remedies to such situations is the combination of Metadynamics with a Parallel-Tempering[33] scheme to help accelerate the convergence of the free-energy surface. We are currently exploring such schemes in our group. Indeed, an approach using the Hamiltonian Replica exchange (HREX) was recently used to effectively sample the open cryptic pockets (the "SWISH" approach)[34] for a set of relevant proteins. Finally, we believe that a combination of a general CV space with a mixed solvent approach might as well stabilize the cryptic pockets and reduce the amount of sampling required to unravel them.

4. Conclusion

Proteins are dynamical entities essential for a large set of vital cellular functions. Their interaction with the environment (other proteins, small ligands ...etc) requires deformations at the surface leading to the formation of binding pockets. Many types of "pockets" are available ranging from large binding grooves formed by a combination of small pockets to cryptic or "hidden" surface pockets. The experimental detection of cryptic pockets is still a costly endeavor. Atomistic simulations provide an accessible tool to detect cryptic pockets but they suffer from the time-scale problem. In this work, we evaluated the use of Metadynamics simulations based on essential dynamics space as an enhanced sampling strategy to detect cryptic pockets. Metadynamics provides an efficient sampling necessary to overcome high energy barriers, while the essential dynamics CV space allows the system to focus on "functional" space. This approach does not rely on *a priori* knowledge of the binding sites.

Finally, this approach allowed us to fully characterize the cryptic pockets repository for Mcl-1.

5. Acknowledgments

This work was supported by a doctoral fellowship (MENRT) held by M.B. The present work was performed thanks to the generous computing resources provided by the CRIANN (Normandy, France). The authors gratefully acknowledge the European Community (European Regional Development Fund, ERDF) for the molecular modeling software. We are also grateful to Dr. Sergio Decherchi for the insightful discussions about Pocketron.

6. Note

All the supporting information will be made available at https://osf.io/fxmj9/

Supporting Information

Cryptic pockets repository through pocket dynamics tracking and metadynamics on essential dynamics space: application to Mcl-1

Mohammed Benabderrahmane[†], Ronan Bureau[†], Jana Sopkova-de Oliveira Santos *[†]

[†] *Normandy University, UNICAEN, Centre d'Etude et Recherche sur le Médicament de Normandie (CERMN), Caen, France*

I. Supporting Figures

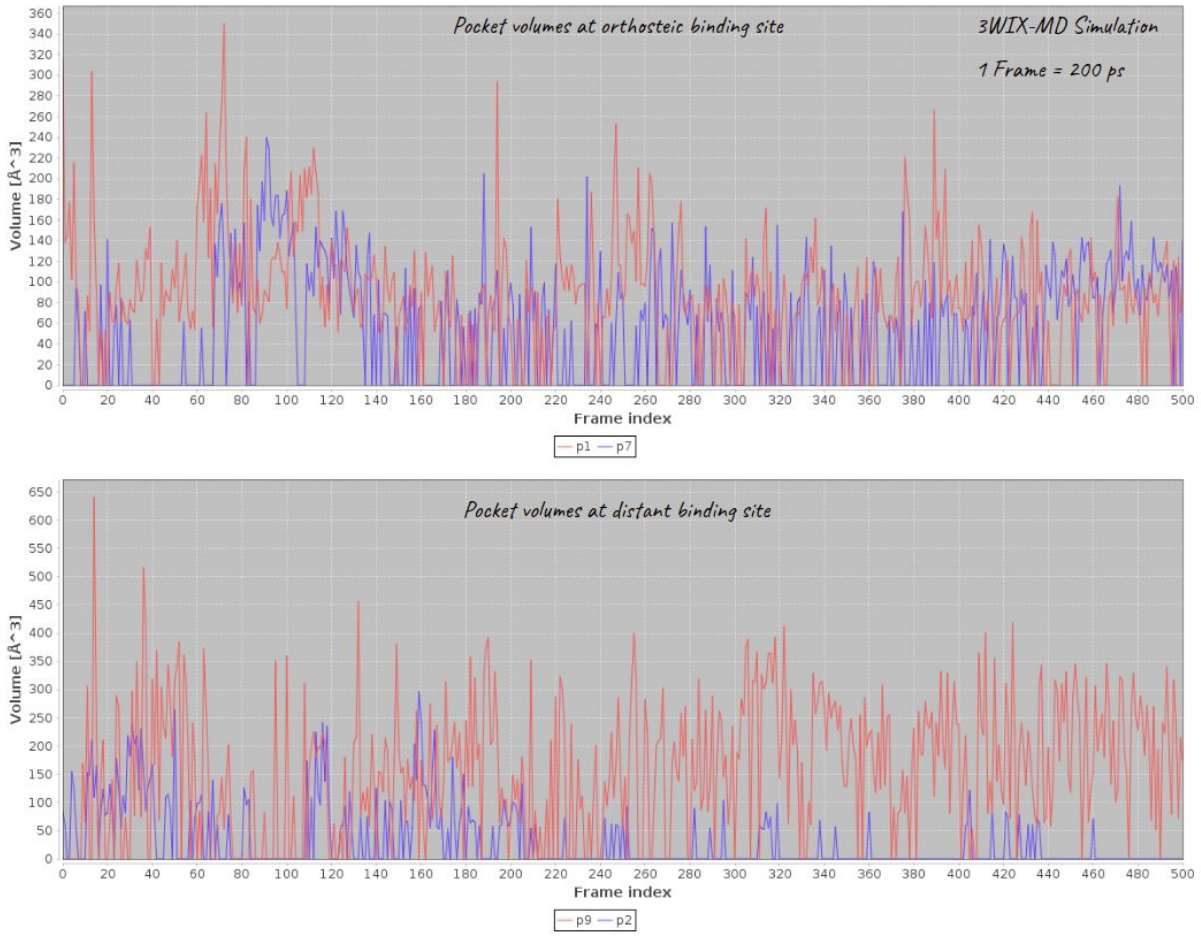
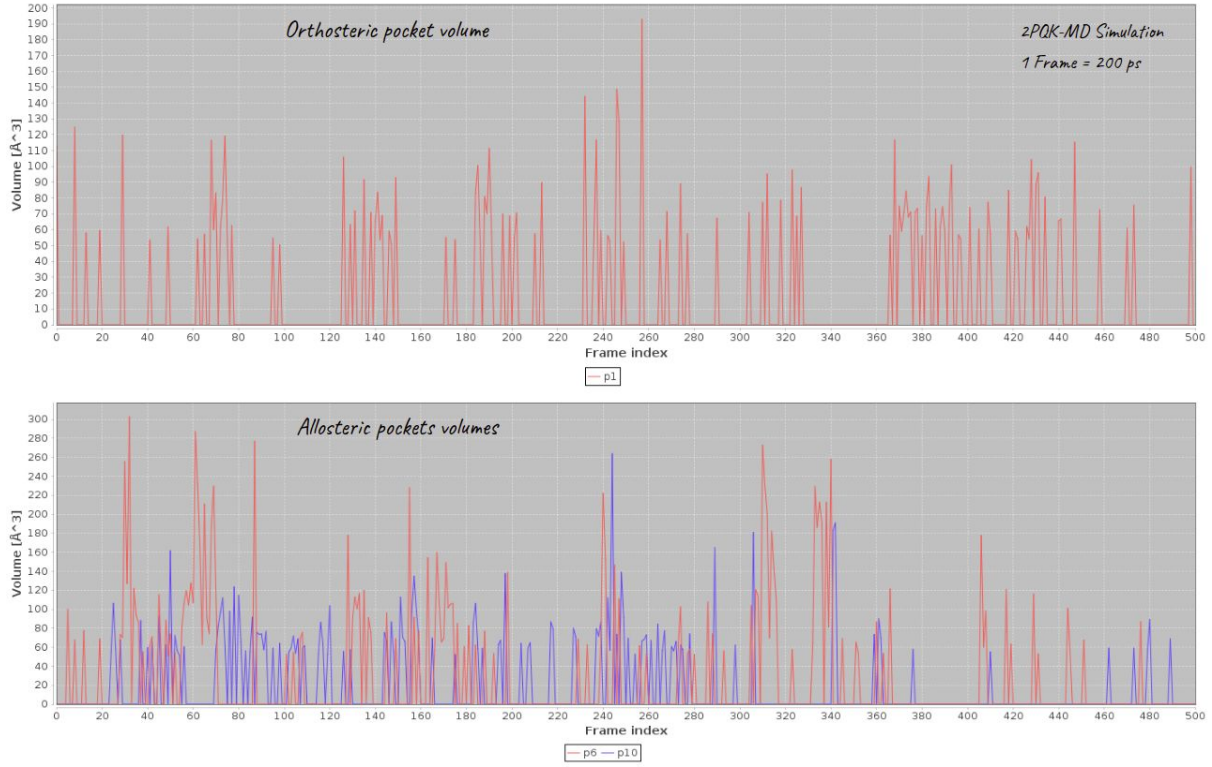


Figure S1. Pocket volumes time series for the 3WIX-MD simulation.



F

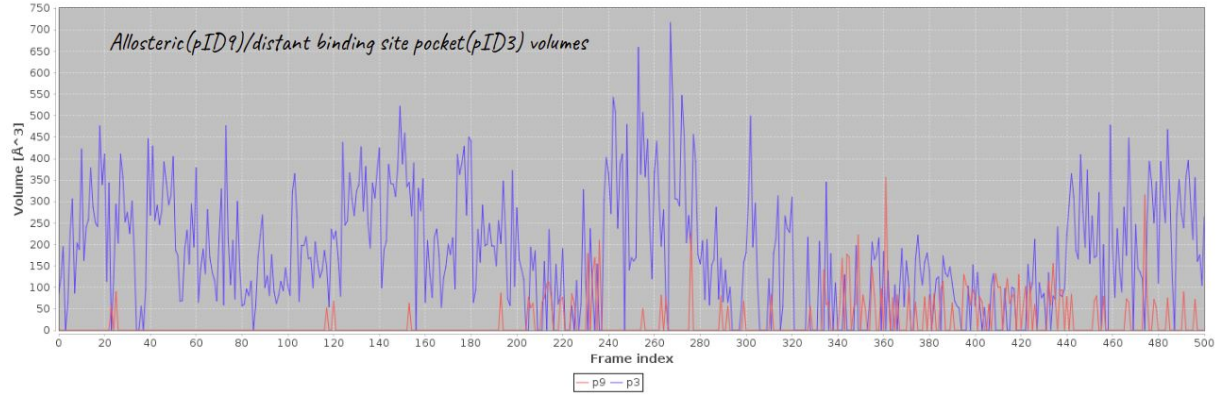


Figure S2. Pocket volumes time series for the 2PQK-MD simulation.

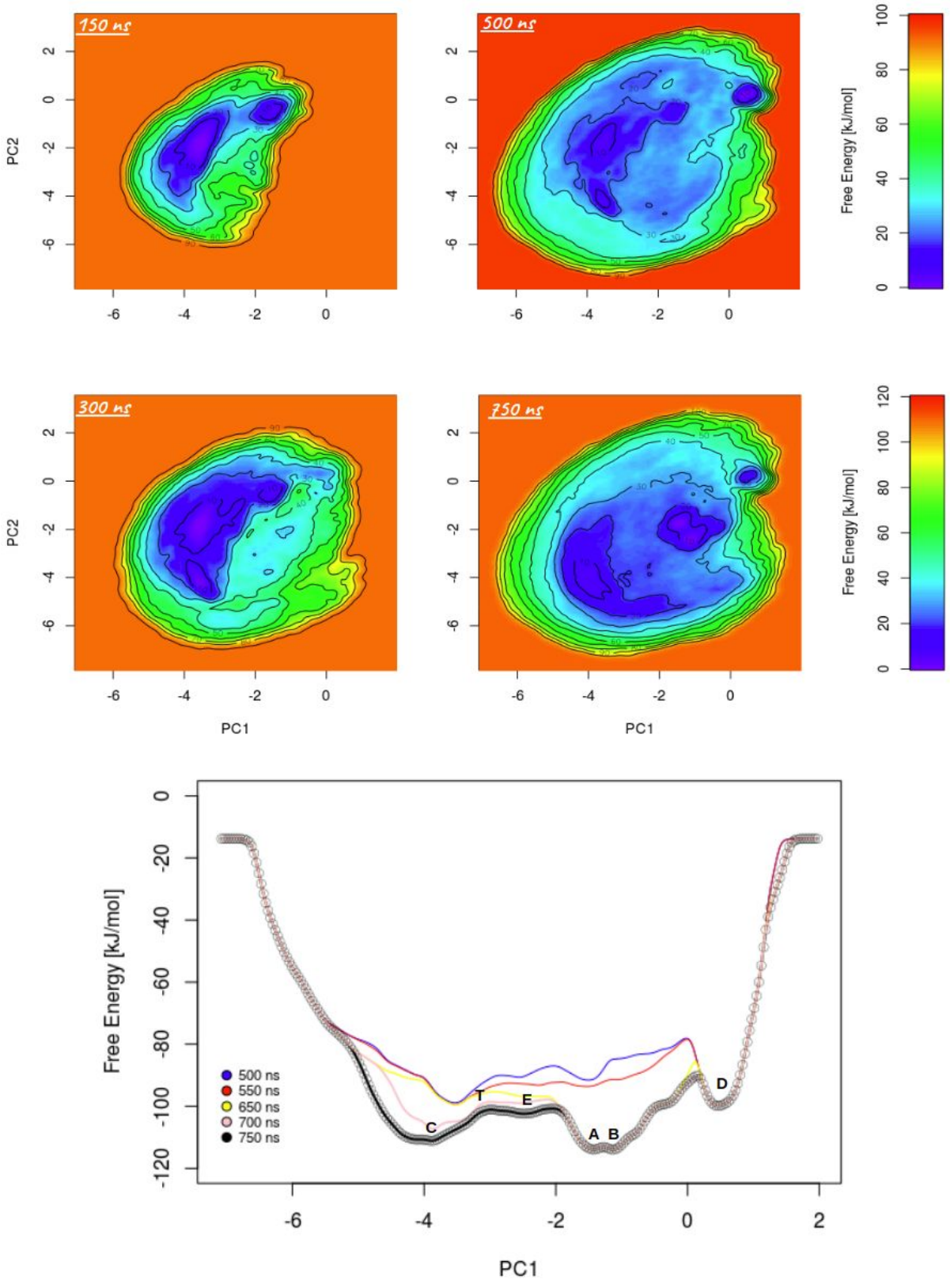


Figure S3. Qualitative assessment of the FES convergence. **A.** FES evolution during time. **B.** Time evolution of the free energy profile along the first collective variable (PC1).

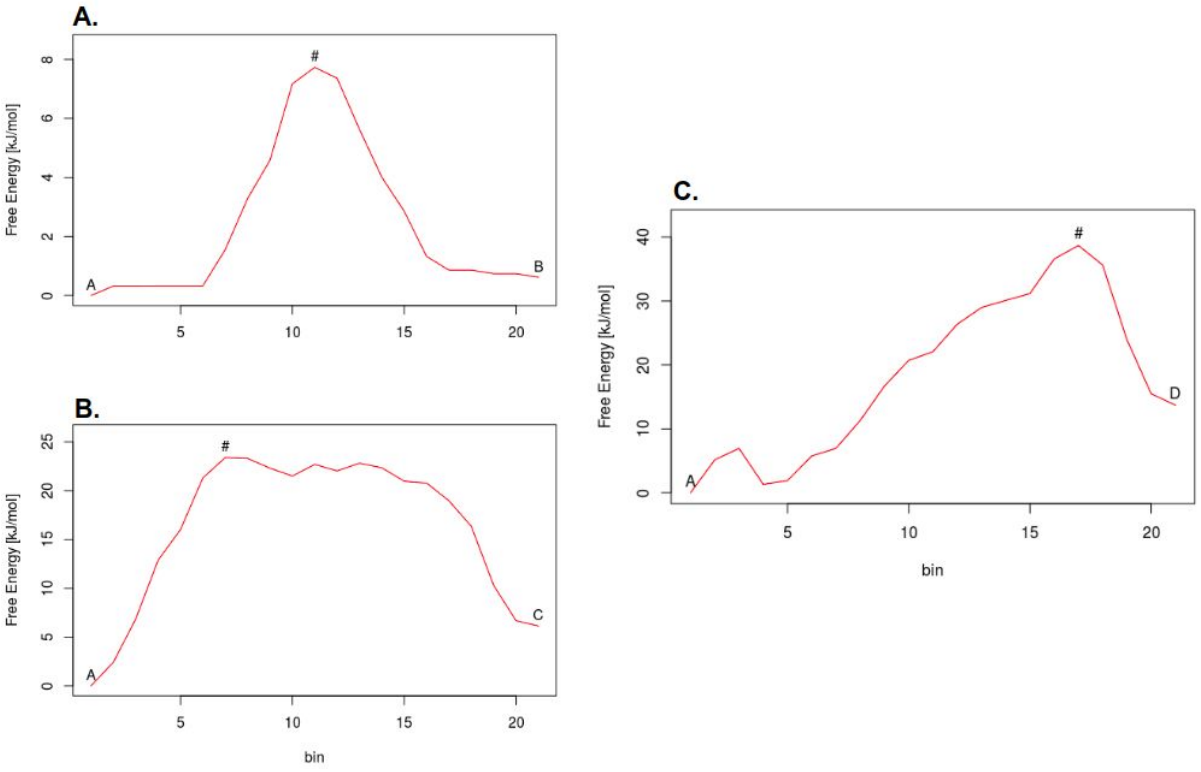
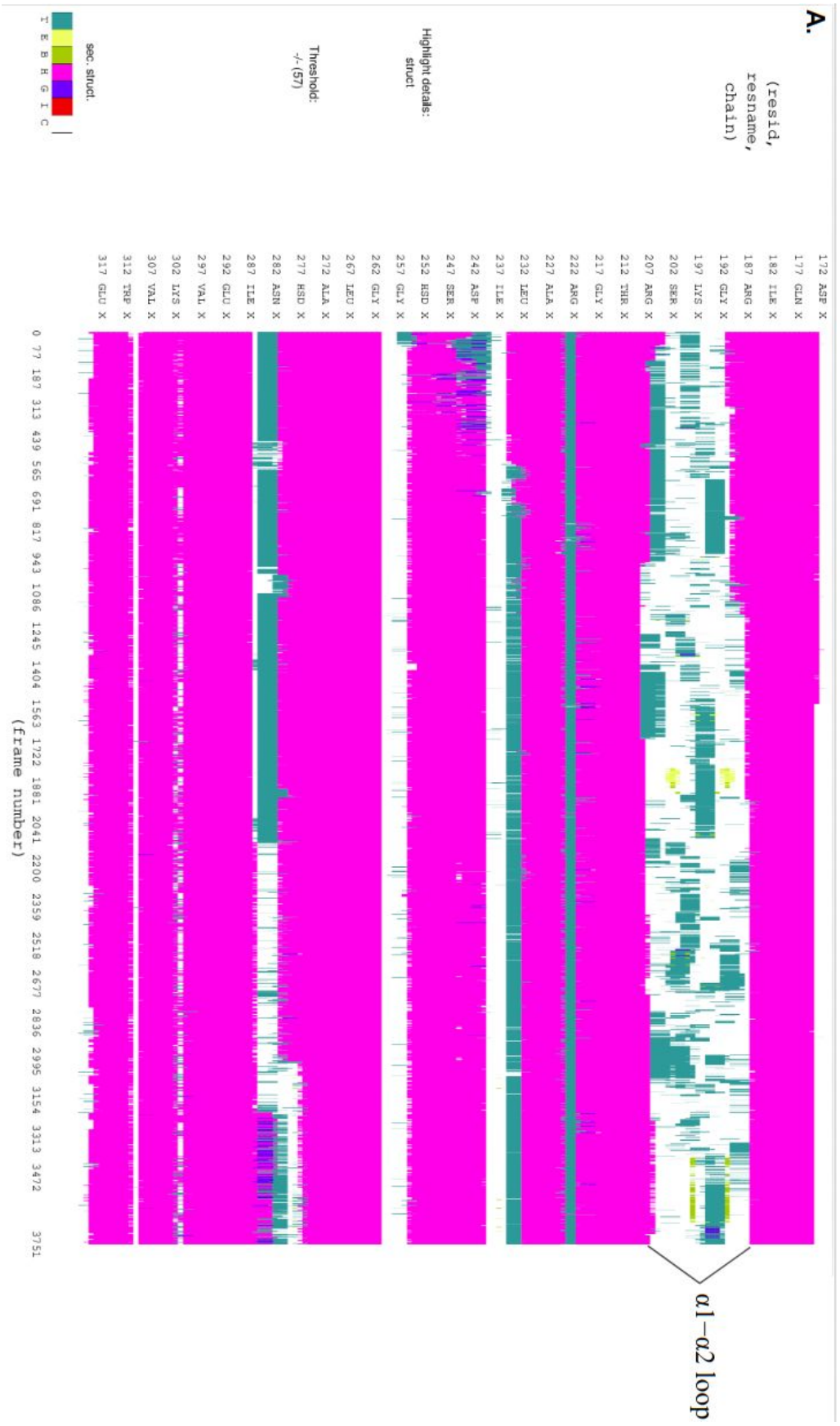


Figure S4. Free energy barriers computed via the Nudged Elastic Band (NEB) Analysis. **A.** Transition A -> B : ~ 8 kJ/mol ; **B.** Transition A -> C : 23 kJ/mol ; **C.** Transition A -> D : 33 kJ/mol.


A.
(resid, resname, chain)
Highlight details: struct
Threshold: -/- (57)
sec. struct.
T E B H G I C
α1–α2 loop
(frame number)

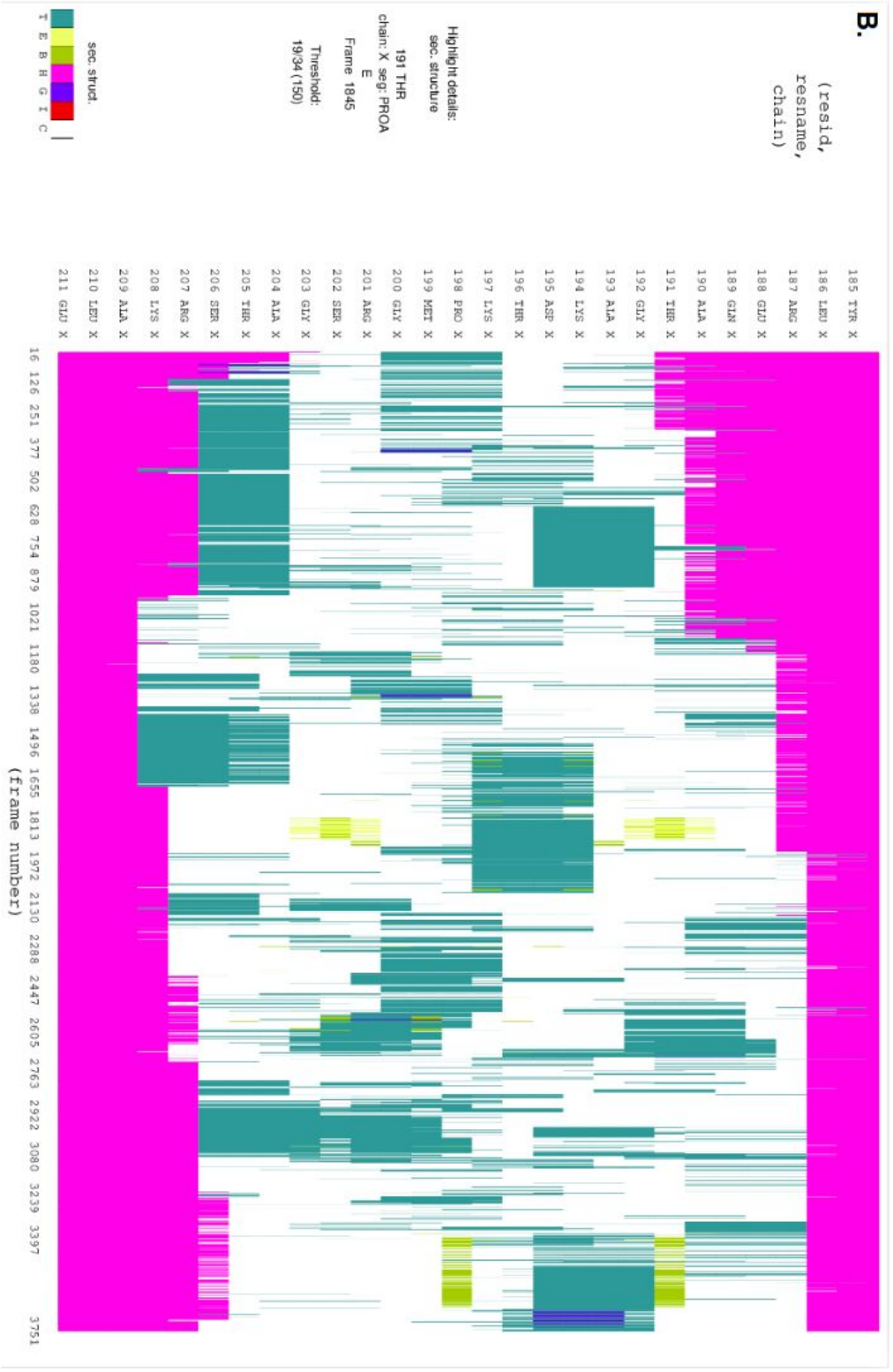


Figure S5. The dynamics of the α1-α2 loop are highly influenced by the conformational transitions at the C-terminal end of the α1 helix (residues Arg187 Glu188 Gln189 Ala190 Thr191) and the N-terminal end of α2 helix (Gly203 Ala204 Thr205 Ser206 Arg207 Lys208). Secondary structures (computed using STRIDE) evolution during the Metadynamics simulation. **A.** For all the protein; **B.** For the α1−α2 loop.

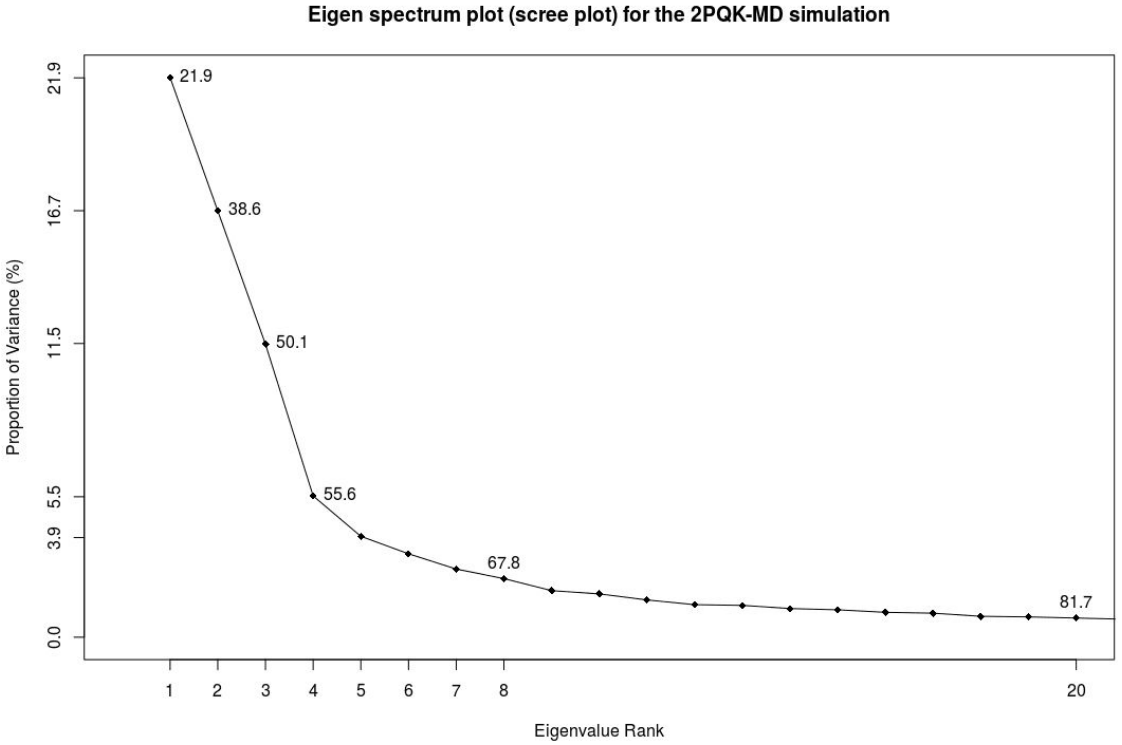


Figure S6. Eigen spectrum plot (scree plot) for the 2PQK-MD simulation. The first and the second eigenvector account for ~ 39 % of the variability (accumulated variance) in the trajectory.

The Eigen spectrum plot shows that the two eigenvectors (extracted from the 2PQK-MD simulation) used as collective variables space account for ~ 39 % of the accumulated variance in the trajectory. Using the "Elbow" criterion, it would be more convenient to choose a large number of eigenvectors (up to 4 here) but this will require a much more sampling effort to converge the free-energy surface. In the classical form of the well-tempered Metadynamics, a rule of thumb is to choose two separate collective variables. However, other flavors of the Metadynamics (ex. Bias-exchange Metadynamics) allow the use of more than two collective variables, but still, the interpretation of the free-energy surface resulting from such schemes is far from being intuitive.

Two natural ways to enhance the amount of variance explained by the eigenvectors is to compute the eigenspace from a "merged" trajectory starting from different conformations or to use a much longer simulation. Other dimensionality reduction methods that do not assume

the linearity of the protein dynamics (unlike the PCA), such as ISOMAPs (see for example the work of Spiwok et. *al*) might be used as well.

A.

pID	Volume(Å^3)	Residues
0	61	Lys208 Asp313 Val316 Glu317 Asp323 Leu324 Glu325
1	454	Phe254 Asp256 Gly257 Val258 Thr259 Asn260 Trp261 Arg 263 Ile264 Lys302 Trp305 Gln309 Gly314 Glu317 Phe318 His329 His331
2	54	Asp218 Gln221 Arg222 Glu225 Lys276
3	169	Arg184 Glu188 Lys194 Asp195 Thr196 Lys 197 Met199 Arg207 Leu210 Glu211 Arg214
4	173	Glu173 Leu174 Gln177 Arg201 Ser202 Thr205 Leu306 Asp307 Arg310 Gly311 Trp312 Asp313
5	44	Arg184 Arg187 Glu188 Ala193 Lys194 Asp195 Lys197
6	26	Trp261 Ile264 Leu306 Gly311 Trp312 Phe315

B.

pID	Volume(Å^3)	Residues
0	100	Ile183 Arg184 Arg187 Asp195 Lys197 Glu288
1	319	Ala227 Phe228 Met231 Leu235 Leu246 Val249 Met250 Val253 Phe254 Arg263 Thr266 Leu267 Phe270 Gly271 Val274 Leu290 Ile294
2	81	Arg176 Met199 Gly200 Arg201 Ser202 Gly203 Ser206
3	38	Lys208 Glu211 Thr212 Arg215 Val316 His320
4	50	Lys194 Asp195 Thr207 Leu210 Glu211 Arg214

C.

pID	Volume(Å^3)	Residues
0	106	Met231 Lys234 Leu235 Leu246 Val249 Val253 Thr266 Leu267 Phe270
1	32	Met250 Phe254 Ile264 Leu267 Ile268 Ile294 Val297 Leu298
2	79	Arg184 Thr196 Lys 197 Pro198 Ser202 Arg207 Leu210
3	75	Leu179 Glu180 Ser183 Arg184 Arg187 Asp195 Arg201

Tables T1. Pocket residues and volumes from experimental structures computed via Nanoshaper. **A.** PDB ID 2MHS; **B.** PDB ID 3WIX; **C.** PDB ID 2PQK

A.

pID	Avg.Vol (Å^3)	Residues
9	156	Glu188 Leu210 Arg207 Lys194 Asp195 Arg184 Arg187 Ala193
1	87	Phe254 Arg 263 Lys302 Leu267 Thr259 Ile264 Asp256 Val258
7	56	Ile237 His277 Leu232 Ile281
3	53	Arg187 Arg184 Glu288 Ser183 Asp195 Glu188 Ala193 Lys197
13	35	Trp312 Gln177 Asp313 Arg201 Leu174 Gly311
2	30	Arg201 Gln177 Gln180 Gly200 Arg176
11	20	**Gln221** Lys276 **Arg222** Phe273 Glu225 Asp218

B.

pID	Avg.Vol ($Å^3$)	Residues
3	177	Leu210 **Arg184** Lys 197 Arg214 Met199 Arg207 Pro198
4	78	**Gln221** Gln189 Ala272 Ser269 Phe273 Tyr185 Lys276 Asp218 Gly217
6	27	Leu278 Ile281 Lys238 Glu240 **Cys286**
9	19	Lys279 Ile287 Ser285 **Arg187** Thr191 **Cys286**
10	18	Pro289 Lys244 Glu240 Val243 **Cys286** Leu278 Leu290 Ser285
1	16	**Phe270** Val249 Met231 Hsd252 Val253 Phe228 Thr266

Table T2. Pocket residues and average volumes for the unbiased MD simulations: **A.** 3WIX-MD; **B.** 2PQK-MD.

A. orthosteric site

pID	Avg.Vol ($Å^3$)	Main residues
9	31	**Phe254 Arg263** Val253 Asp256 Thr266
2	14	Ser293 Lys244 Asp296 Ser247 Val243 Pro289 Glu292
8	30	Asn282 Ser285 Leu278 Lys238 Ile281 Ile237
26	15	Hsd277 Leu232 Ile237 Val274 Asp236 Leu278 Lys279 Leu235
11	16	Leu232 Hsd277 Ile237
1	53	Lys276 **Gln221** Ala272 Phe273 Tyr185 Ser269 Gln189 Asp218 Glu225 Gly217 Phe228

B. Allosteric site

pID	Avg.Vol ($Å^3$)	Main residues
6	44	**Glu188** Gln189 Arg214 Tyr185
7	67	Ile287 **Arg187** Lys279 Ser285 Ala275 Leu186 **Glu188** Glu284 **Cys286** Gln283 Leu278
12	30	Pro289 Val243 Leu278 **Cys286** Leu290 Glu240 Ile237 Ser285

C. Distant binding site

17	68	Asp313 Trp312 Gln177 Ala209 Ile181 Ser206 Val316 Lys208 Thr205 Arg207 Glu180
3	40	Glu180 Gln177 Ile181 Ser206 Thr205 Arg184 Ala204
4	77	Arg184 Leu210 Arg207 Glu180 Ser206 Arg201 Ser202 Thr205 Ala204 Gly203
5	19	Thr212 Arg215 Lys208 Glu211 Val316 Hsd320 Val321 Phe319

Table T3. Pocket residues and average volumes for the Metadynamics simulation: **A.** Orthosteric pockets; **B.** Allosteric pockets; **C.** Distant binding site pockets.

3. Conclusion et perspectives

Dans ce chapitre, la dynamique globale de Mcl-1 a été caractérisée en utilisant une simulation de métadynamique sur l'espace de la dynamique essentielle. La surface d'énergie libre proposée permet de mettre en lumière les différents états conformationnels accessibles à la protéine.

Les données générées au cours de la métadynamique, ont été utilisées dans le but d'évaluer une approche de détection des poches cryptiques.
Le réseau de communication entre les poches des différents sites fonctionnels de Mcl-1 a été révélé.

Cette approche pourrait être utilisée comme une étape préalable pour générer des ensembles conformationnels larges (surtout pour des protéines où les données structurales ne sont pas ou peu disponibles). Ceci, pourrait enrichir l'ensemble des conformations de départ utilisées dans des expériences de docking ensembliste (screening virtuel).

Cette approche sera évaluée davantage sur d'autres cibles thérapeutiques, étudiées au laboratoire.

4. Références

(1) Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J. The Role of Protein Dynamics in the Evolution of New Enzyme Function. *Nat. Chem. Biol.* **2016**, *12* (11), 944–950.

(2) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. Protein Binding Pocket Dynamics. *Acc. Chem. Res.* **2016**, *49* (5), 809–815.

(3) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. Site-Directed Ligand Discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (17), 9367–9372.

(4) Lawson, A. D. G. Antibody-Enabled Small-Molecule Drug Discovery. *Nat. Rev. Drug Discov.* **2012**, *11* (7), 519–525.

(5) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of Secondary Binding Sites in Proteins Using Fragment Screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (52), 15910–15915.

(6) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.

(7) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71* (12), 126601.

(8) Cuchillo, R.; Pinto-Gil, K.; Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *J. Chem. Theory Comput.* **2015**, *11* (3), 1292–1307.

(9) Daidone, I.; Amadei, A. Essential Dynamics: Foundation and Applications. *WIREs Comput Mol Sci* **2012**, *2* (5), 762–770.

(10) David, C. C.; Jacobs, D. J. Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins. *Methods Mol. Biol.* **2014**, *1084*, 193–226.

(11) Kelly, G. L.; Strasser, A. Toward Targeting Antiapoptotic MCL-1 for Cancer Therapy. *Annu. Rev. Cancer Biol.* **2020**, *4* (1), 299–313.

(12) Fire, E.; Gullá, S. V.; Grant, R. A.; Keating, A. E. Mcl-1-Bim Complexes Accommodate Surprising Point Mutations via Minor Structural Changes. *Protein Sci.* **2010**, *19* (3), 507–519.

(13) Tanaka, Y.; Aikawa, K.; Nishida, G.; Homma, M.; Sogabe, S.; Igaki, S.; Hayano, Y.; Sameshima, T.; Miyahisa, I.; Kawamoto, T.; Tawada, M.; Imai, Y.; Inazuka, M.; Cho, N.; Imaeda, Y.; Ishikawa, T. Discovery of Potent Mcl-1/Bcl-xL Dual Inhibitors by Using a Hybridization Strategy Based on Structural Analysis of Target Proteins. *J. Med. Chem.* **2013**, *56* (23), 9635–9645.

(14) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.

(15) Huang, J.; MacKerell, A. D., Jr. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.

(16) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960.

(17) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., 3rd;

MacKerell, A. D., Jr; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405–413.

(18) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(19) Forester, T. R.; Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *J. Comput. Chem.* **1998**, *19* (1), 102–111.

(20) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613–4621.

(21) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177–4189.

(22) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.

(23) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins* **1993**, *17* (4), 412–425.

(24) Benabderrahmane, M.; Bureau, R.; Voisin-Chiret, A. S.; Sopková-de Oliveira Santos, J. Insights into Mcl-1 Conformational States and Allosteric Inhibition Mechanism from Molecular Dynamics Simulations, Enhanced Sampling, and Pocket Crosstalk Analysis. *J. Chem. Inf. Model.* **2020**. https://doi.org/10.1021/acs.jcim.0c00315.

(25) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

(26) Trapl, D.; Spiwok, V. Analysis of the Results of Metadynamics Simulations by Metadynminer and metadynminer3d. *arXiv [q-bio.BM]*, 2020.

(27) Decherchi, S.; Spitaleri, A.; Stone, J.; Rocchia, W. NanoShaper-VMD Interface: Computing and Visualizing Surfaces, Pockets and Channels in Molecular Systems. *Bioinformatics* **2019**, *35* (7), 1241–1243.

(28) Richards, F. M. Areas, Volumes, Packing and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.

(29) Decherchi, S.; Bottegoni, G.; Spitaleri, A.; Rocchia, W.; Cavalli, A. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.* **2018**, *58* (2), 219–224.

(30) Cooper, D. R.; Porebski, P. J.; Chruszcz, M.; Minor, W. X-Ray Crystallography: Assessment and Validation of Protein-Small Molecule Complexes for Drug Discovery. *Expert Opin. Drug Discov.* **2011**, *6* (8), 771–782.

(31) Kuzmanic, A.; Pannu, N. S.; Zagrovic, B. X-Ray Refinement Significantly Underestimates the Level of Microscopic Heterogeneity in Biomolecular Crystals. *Nat. Commun.* **2014**, *5*, 3220.

(32) Song, T.; Wang, Z.; Ji, F.; Feng, Y.; Fan, Y.; Chai, G.; Li, X.; Li, Z.; Zhang, Z. Deactivation of Mcl-1 by Dual-Function Small-Molecule Inhibitors Targeting the Bcl-2 Homology 3 Domain and Facilitating Mcl-1 Ubiquitination. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (46), 14250–14256.

(33) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141–151.

(34) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138* (43), 14257–14263.

CONCLUSION GÉNÉRALE

Le cancer est un enjeu de santé publique. La carcinogenèse est intrinsèquement liée au dérèglement de l'apoptose: un processus finement contrôlé par les protéines de la famille Bcl-2.

La protéine Mcl-1, un membre anti-apoptotique de la famille Bcl-2, joue un rôle essentiel dans ce processus. En effet, elle est souvent surexprimée dans plusieurs types de cancers hématologiques et solides. Elle contribue également à la résistance des cellules cancéreuses aux chimiothérapies existantes. Elle est de ce fait considérée comme une cible thérapeutique d'intérêt en cancérologie.

Les protéines sont des entités dynamiques. Elles interagissent entre elles pour former un réseau d'interactions permettant l'expression d'une fonction biochimique. De plus, ces biomolécules possèdent une dynamique intrinsèque, souvent dictée par leur structure.

Dans le cadre de ce travail, la structure et la dynamique de Mcl-1 ont été décrites. Dans un premier temps, l'interaction de Mcl-1 avec le Pyridoclax (un inhibiteur BH3-mimétique) est caractérisée par des approches expérimentales (RMN) et théoriques (docking et simulations de dynamique moléculaire).

Le deuxième volet de ce travail a été consacré à la caractérisation de l'espace conformationnel de Mcl-1, et notamment à l'étude de son mode d'inhibition allostérique. Des simulations de dynamique moléculaire biaisées (métadynamique) ont été utilisées pour générer des surfaces d'énergie libre permettant une meilleure compréhension des états conformationnels de la protéine.
De plus, une approche d'analyse des réseaux de communication entre poches, a été utilisée pour proposer le premier modèle d'inhibition allostérique de Mcl-1. Cela permettra *in fine* de proposer des approches alternatives pour cibler sélectivement Mcl-1.

La détection des poches cryptiques d'une protéine est souvent une étape nécessaire pour cibler des protéines dites ‘’undruggables”.
Le troisième volet de ce travail, concerne l'évaluation d'une approche de détection de poches cryptiques à la surface d'une protéine. Les simulations de métadynamique sur l'espace de la dynamique essentielle, ont été utilisées et ont montré leur pertinence dans ce contexte. Cette approche sera davantage testée sur d'autres cibles pour évaluer de façon plus large son domaine d'applicabilité.

ANNEXES

Notebooks relatifs au chapitre 3.

Mcl-1_ens

April 26, 2020

###

Mcl-1 structural ensemble analysis notebook

This notebook regroups the R script used to analyze the Mcl-1 structural ensemble data and the Molecular Dynamics Simulations conducted for the paper #####

Myeloid cell leukemia-1 (Mcl-1) conformational dynamics: insights into Mcl-1 conformational states and allosteric inhibition mechanism from molecular dynamics simulations, enhanced sampling and pocket crosstalk analysis.

M.Benabderrahmane et al.

[19]:
```r
# Install the bio3d package and load
#install.packages("bio3d")
library(bio3d)

# Point to the dir where the pdbs are stored ...
#A folder 'pdbs' including all the pdbs is provided ...
setwd(".")
# MCL-1 structures PDB:IDs
ids <- c("5JSB","5C6H","2PQK","4HW4","3WIX","5C3F",
         "3PK1","4HW2","4WMR","4WMS","5FDO","5IEZ",
         "5FDR","4OQ6","4OQ5","2NL9","3MK8","3KJ0","4WMT",
         "3KJ1","3KJ2","3WIY","5IF4","4ZBI","4HW3","4WMU",
         "2NLA","5VX2","3D7V","3KZ0",
         "4BPI","5LOF","4WGI","4WMV",
         "4ZBF","5VKC","6B4U","6B4L","6BW8","6BW2","4WMX")

ids_A <- c("5JSB_A","5C6H_A","2PQK_A","4HW4_A","3WIX_A","5C3F_A",
           "3PK1_A","4HW2_A","4WMR_A","4WMS_A","5FDO_A","5IEZ_A",
           "5FDR_A","4OQ6_A","4OQ5_A","2NL9_A","3MK8_A","3KJ0_A","4WMT_A",
           "3KJ1_A","3KJ2_A","3WIY_A","5IF4_A","4ZBI_A","4HW3_A","4WMV_A",
           "2NLA_A","5VX2_A","3D7V_A","3KZ0_A",
           "4BPI_A","4WMX_A","5LOF_A","4WGI_A",
           "4ZBF_A","5VKC_A","6B4U_A","6B4L_A","6BW8_A","6BW2_A",
           "4WMU_A")

# annotate
```

```
anno <- pdb.annotate(ids)

# Load data
files <- list.files(path=".", pattern=".pdb")

# Split and keep Chain=="A" (using the ids)
files_splitted = pdbsplit(files, ids_A, path ="sp") # Ignore the unmatched␣
 ↪structures ...
```

```
  |======================================================================| 100%
Warning message in pdbsplit(files, ids_A, path = "sp"):
"unmatched pdb files: 2KBW, 2MHS, 4BPJ, 5FC4, 5MES, 5MEV, 5UUM, 5W89, 5W8F, pc1,
pc2, protein_C286W, protein-WT, step3_pdb_no_water_Meta"
```

[20]:
```
# Read and align
pdbs.all <- pdbaln(files_splitted, fit=TRUE)
pdbs = read.fasta("aln.fa")
sel=store.atom()
pdbs_aa = read.all(pdbs, rm.wat=TRUE, rm.ligand=TRUE, sel=sel)
#pdbs_aa$all
atm <- colnames(pdbs_aa$all)
l=unlist(strsplit(basename(pdbs_aa$id), split=".pdb"))
```

```
Reading PDB files:
sp/2NL9_A.pdb
sp/2NLA_A.pdb
sp/2PQK_A.pdb
sp/3D7V_A.pdb
sp/3KJ0_A.pdb
sp/3KJ1_A.pdb
sp/3KJ2_A.pdb
sp/3KZ0_A.pdb
sp/3MK8_A.pdb
sp/3PK1_A.pdb
sp/3WIX_A.pdb
sp/3WIY_A.pdb
sp/4BPI_A.pdb
sp/4HW2_A.pdb
sp/4HW3_A.pdb
sp/4HW4_A.pdb
sp/4OQ5_A.pdb
sp/4OQ6_A.pdb
sp/4WGI_A.pdb
sp/4WMR_A.pdb
sp/4WMS_A.pdb
sp/4WMT_A.pdb
sp/4WMU_A.pdb
```

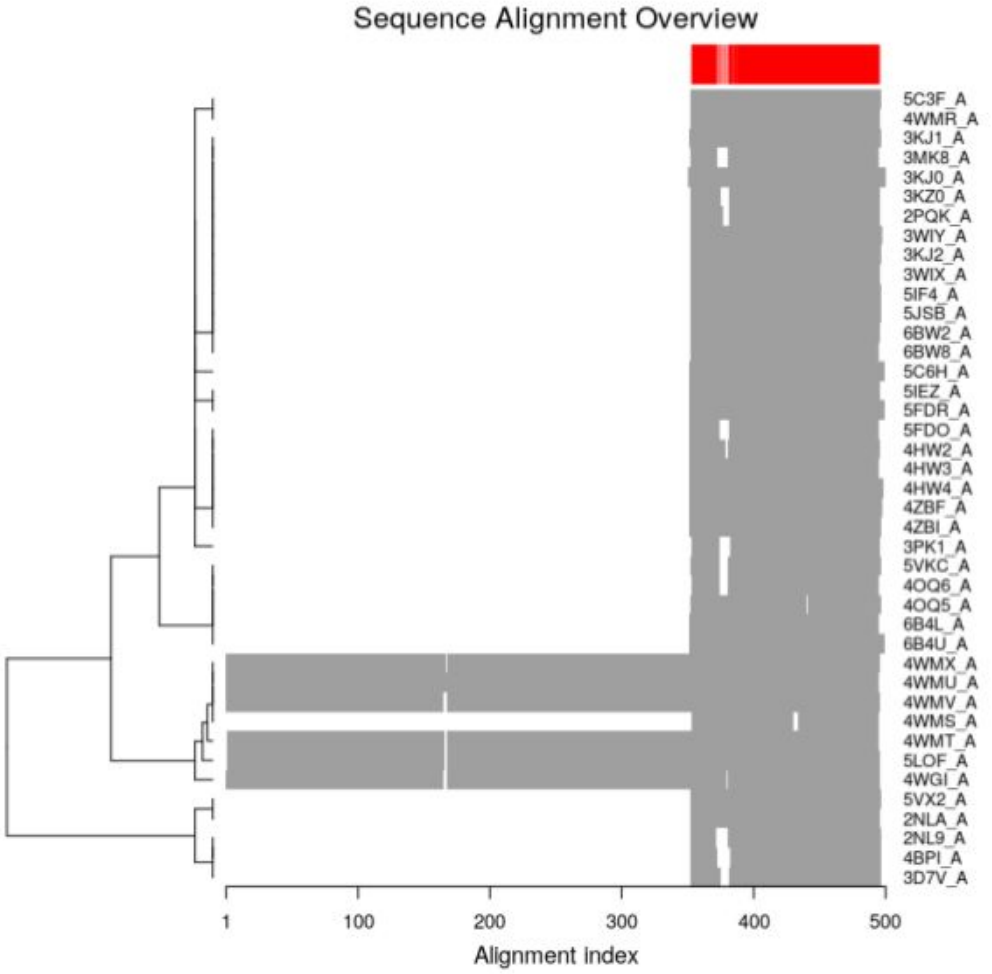
```
sp/4WMV_A.pdb
sp/4WMX_A.pdb
sp/4ZBF_A.pdb
sp/4ZBI_A.pdb
sp/5C3F_A.pdb
sp/5C6H_A.pdb
sp/5FDO_A.pdb
sp/5FDR_A.pdb
sp/5IEZ_A.pdb
sp/5IF4_A.pdb
sp/5JSB_A.pdb
sp/5LOF_A.pdb
sp/5VKC_A.pdb
sp/5VX2_A.pdb
sp/6B4L_A.pdb
sp/6B4U_A.pdb
sp/6BW2_A.pdb
sp/6BW8_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
..   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
.   PDB has ALT records, taking A only, rm.alt=TRUE
…   PDB has ALT records, taking A only, rm.alt=TRUE
…

Extracting sequences

pdb/seq: 1   name: sp/2NL9_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 2   name: sp/2NLA_A.pdb
pdb/seq: 3   name: sp/2PQK_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 4   name: sp/3D7V_A.pdb
pdb/seq: 5   name: sp/3KJ0_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6   name: sp/3KJ1_A.pdb
```

```
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 7   name: sp/3KJ2_A.pdb
pdb/seq: 8   name: sp/3KZ0_A.pdb
pdb/seq: 9   name: sp/3MK8_A.pdb
pdb/seq: 10   name: sp/3PK1_A.pdb
pdb/seq: 11   name: sp/3WIX_A.pdb
pdb/seq: 12   name: sp/3WIY_A.pdb
pdb/seq: 13   name: sp/4BPI_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 14   name: sp/4HW2_A.pdb
pdb/seq: 15   name: sp/4HW3_A.pdb
pdb/seq: 16   name: sp/4HW4_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 17   name: sp/4OQ5_A.pdb
pdb/seq: 18   name: sp/4OQ6_A.pdb
pdb/seq: 19   name: sp/4WGI_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 20   name: sp/4WMR_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 21   name: sp/4WMS_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 22   name: sp/4WMT_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 23   name: sp/4WMU_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 24   name: sp/4WMV_A.pdb
pdb/seq: 25   name: sp/4WMX_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 26   name: sp/4ZBF_A.pdb
pdb/seq: 27   name: sp/4ZBI_A.pdb
pdb/seq: 28   name: sp/5C3F_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 29   name: sp/5C6H_A.pdb
pdb/seq: 30   name: sp/5FDO_A.pdb
pdb/seq: 31   name: sp/5FDR_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 32   name: sp/5IEZ_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 33   name: sp/5IF4_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 34   name: sp/5JSB_A.pdb
pdb/seq: 35   name: sp/5LOF_A.pdb
pdb/seq: 36   name: sp/5VKC_A.pdb
pdb/seq: 37   name: sp/5VX2_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 38   name: sp/6B4L_A.pdb
pdb/seq: 39   name: sp/6B4U_A.pdb
pdb/seq: 40   name: sp/6BW2_A.pdb
```

```
pdb/seq: 41   name: sp/6BW8_A.pdb
```

Plotting Sequence Alignment for the PDB structures(full_dataset)

[21]:
```
plot.new()
plot(pdbs_aa,cex.lab = 0.75, labels=1)
```



[22]:
```
# pass to final variable pdbs_f (here we are using the full dataset)
pdbs_f = pdbs_aa
# a list of PDB codes of our final selection
ids <- unlist(strsplit(basename(pdbs_f$id), split=".pdb"))

# SSE annotations based on 3WIY_A
pdbs_f$sse = pdbs2sse(pdbs_f, pdb=TRUE, ind=33, rm.gaps=FALSE, resno = FALSE)
```

```
anno <- pdb.annotate(ids)
```

```
Re-reading PDB (3WIX_A) to extract SSE
Warning message in dssp.pdb(pdb.ref, …):
"Non-protein residues detected in input PDB: LC3, HOH"Warning message in
pdb.annotate(ids):
"ids should be standard 4 character PDB-IDs: trying first 4 characters…"
```

Analyzing the ensemble

1. Hierarchical RMSD clustering

compute core and fit

[23]:
```
# Compute the core
core <- core.find(pdbs_f)

#Core indices and residues ...
core.inds <- print(core, vol=0.5)
```

```
core size 132 of 133  vol = 23.46
core size 131 of 133  vol = 21.485
core size 130 of 133  vol = 19.649
core size 129 of 133  vol = 18.363
core size 128 of 133  vol = 17.504
core size 127 of 133  vol = 16.667
core size 126 of 133  vol = 15.744
core size 125 of 133  vol = 14.765
core size 124 of 133  vol = 13.99
core size 123 of 133  vol = 13.199
core size 122 of 133  vol = 12.483
core size 121 of 133  vol = 11.706
core size 120 of 133  vol = 10.968
core size 119 of 133  vol = 10.174
core size 118 of 133  vol = 9.548
core size 117 of 133  vol = 9.015
core size 116 of 133  vol = 8.475
core size 115 of 133  vol = 8.166
core size 114 of 133  vol = 7.843
core size 113 of 133  vol = 7.434
core size 112 of 133  vol = 7.113
core size 111 of 133  vol = 6.81
core size 110 of 133  vol = 6.467
core size 109 of 133  vol = 6.111
core size 108 of 133  vol = 5.816
core size 107 of 133  vol = 5.552
```

```
core size 106 of 133  vol = 5.186
core size 105 of 133  vol = 4.94
core size 104 of 133  vol = 4.678
core size 103 of 133  vol = 4.495
core size 102 of 133  vol = 4.317
core size 101 of 133  vol = 4.092
core size 100 of 133  vol = 3.902
core size 99 of 133  vol = 3.73
core size 98 of 133  vol = 3.549
core size 97 of 133  vol = 3.378
core size 96 of 133  vol = 3.236
core size 95 of 133  vol = 3.107
core size 94 of 133  vol = 2.987
core size 93 of 133  vol = 2.86
core size 92 of 133  vol = 2.741
core size 91 of 133  vol = 2.63
core size 90 of 133  vol = 2.53
core size 89 of 133  vol = 2.426
core size 88 of 133  vol = 2.32
core size 87 of 133  vol = 2.213
core size 86 of 133  vol = 2.133
core size 85 of 133  vol = 2.062
core size 84 of 133  vol = 1.99
core size 83 of 133  vol = 1.917
core size 82 of 133  vol = 1.844
core size 81 of 133  vol = 1.778
core size 80 of 133  vol = 1.712
core size 79 of 133  vol = 1.647
core size 78 of 133  vol = 1.591
core size 77 of 133  vol = 1.536
core size 76 of 133  vol = 1.481
core size 75 of 133  vol = 1.429
core size 74 of 133  vol = 1.384
core size 73 of 133  vol = 1.334
core size 72 of 133  vol = 1.285
core size 71 of 133  vol = 1.236
core size 70 of 133  vol = 1.184
core size 69 of 133  vol = 1.139
core size 68 of 133  vol = 1.094
core size 67 of 133  vol = 1.054
core size 66 of 133  vol = 1.012
core size 65 of 133  vol = 0.975
core size 64 of 133  vol = 0.936
core size 63 of 133  vol = 0.899
core size 62 of 133  vol = 0.864
core size 61 of 133  vol = 0.833
core size 60 of 133  vol = 0.8
core size 59 of 133  vol = 0.764
```

```
 core size 58 of 133  vol = 0.724
 core size 57 of 133  vol = 0.684
 core size 56 of 133  vol = 0.654
 core size 55 of 133  vol = 0.622
 core size 54 of 133  vol = 0.593
 core size 53 of 133  vol = 0.563
 core size 52 of 133  vol = 0.535
 core size 51 of 133  vol = 0.509
 core size 50 of 133  vol = 0.486
 FINISHED: Min vol ( 0.5 ) reached
# 51 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1   175 189     15
2   208 212      5
3   261 263      3
4   265 278     14
5   289 296      8
6   311 316      6
```

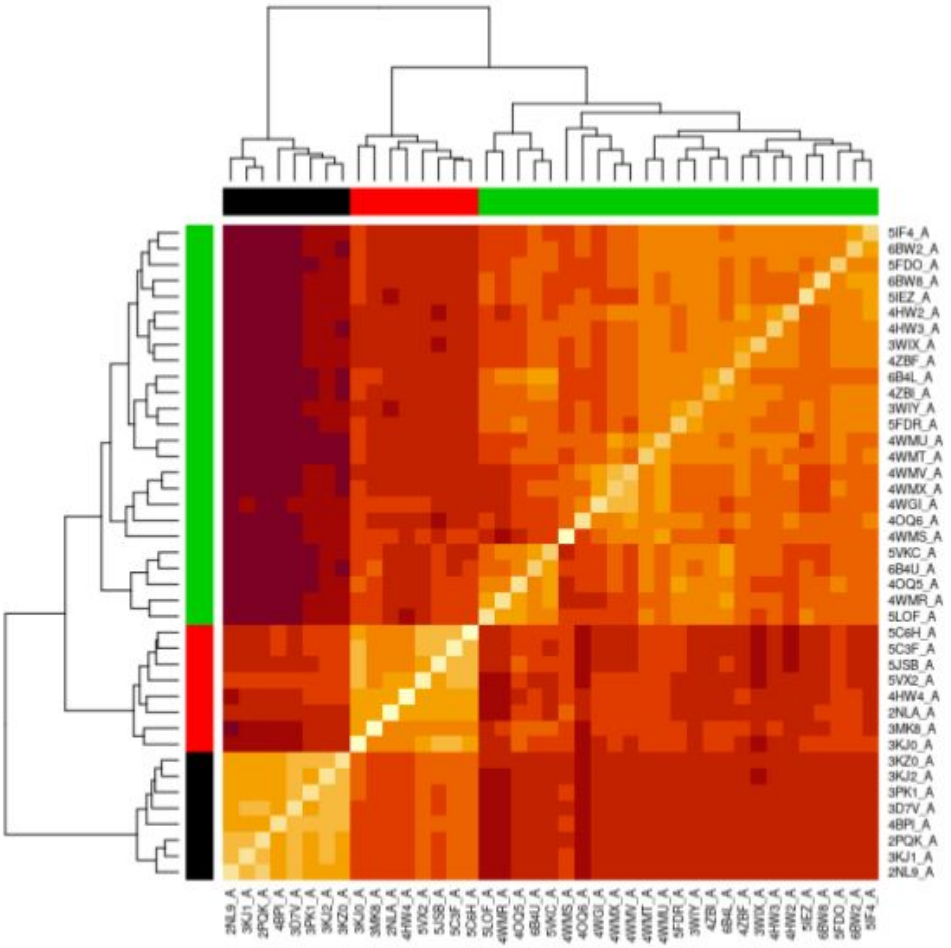
[24]:
```
## superpose all structures to core and save to fitted directory
pdbs_f$xyz = pdbfit(pdbs_f, core$c0.5A.xyz ,outpath = "fitted_core_final")
```

```
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
  PDB has ALT records, taking A only, rm.alt=TRUE
```

[25]:
```
# RMSD after superimpose on core ...
rd <- rmsd(pdbs_f$xyz)
```

```
Warning message in rmsd(pdbs_f$xyz):
"No indices provided, using the 133 non NA positions
"
```

[26]:
```
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, h=3)
heatmap(rd, labRow = c(ids[1:41]), labCol =c(ids[1:41]),
        ColSideColors=as.character(grps.rd),
        RowSideColors = as.character(grps.rd),
        cexRow = 0.85 ,cexCol = 0.85)
```

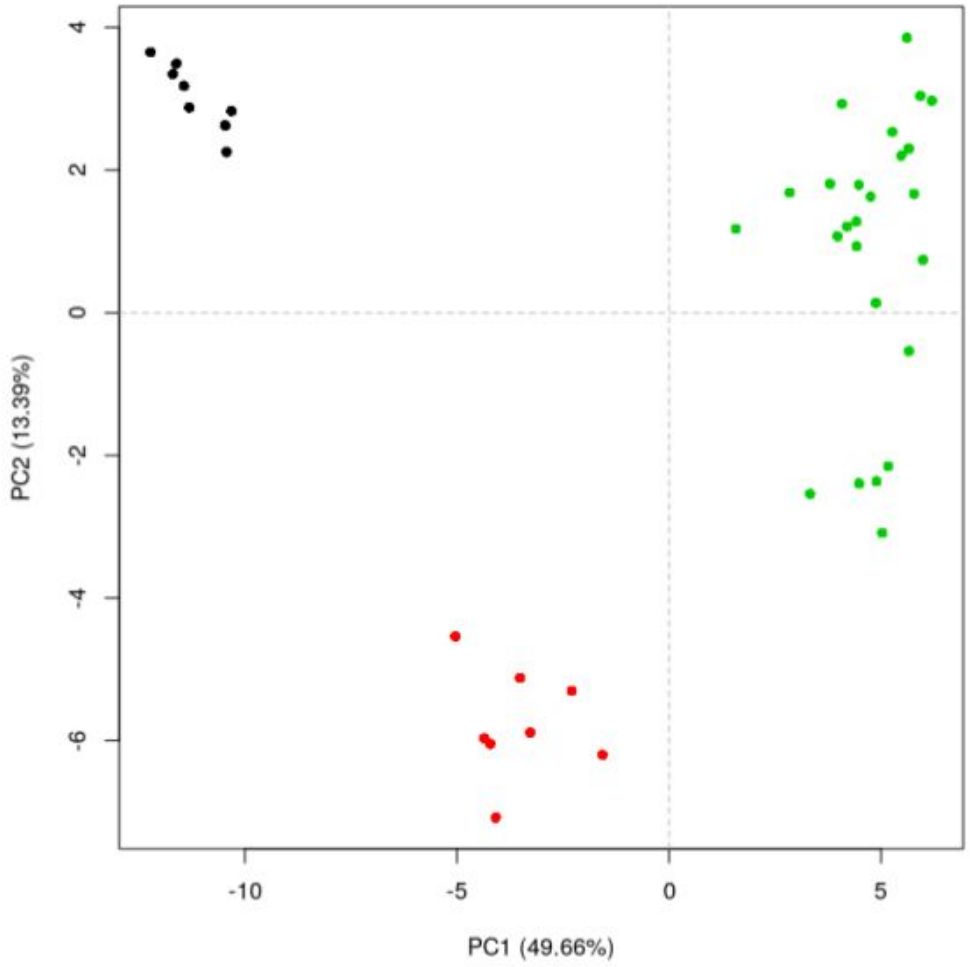


2. PCA analysis on ensemble data set

2.1 Full dataset

[27]:
```r
# identify gaps, and perform PCA
gaps.pos <- gap.inspect(pdbs_f$xyz)
gaps.res <- gap.inspect(pdbs_f$ali)
pc.xray <- pca.xyz(pdbs_f$xyz[,gaps.pos$f.inds])
pc.xray <- pca.xyz(pdbs_f$xyz,rm.gaps = TRUE)

print(pc.xray)
# plot PCA
plot.pca(pc.xray,col=grps.rd,pc.axes = 1:2)

# Use identify function in RStudio only ! to indentify the structures ...
#identify(pc.xray$z[,1:2], labels=basename.pdb(pdbs_f$id),offset = 0.5)
```

```
NOTE: Removing 389 gap positions with missing coordinate data
      retaining 133 non-gap positions for analysis.

Call:
  pca.xyz(xyz = pdbs_f$xyz, rm.gaps = TRUE)

Class:
  pca

Number of eigenvalues:
  399

       Eigenvalue Variance Cumulative
  PC 1     42.425   49.662     49.662
  PC 2     11.439   13.390     63.052
  PC 3      8.554   10.013     73.065
  PC 4      4.149    4.857     77.922
  PC 5      2.838    3.323     81.244
  PC 6      2.126    2.488     83.733

  (Obtained from 41 conformers with 399 xyz input values).

+ attr: L, U, z, au, sdev, mean, call
```



2. PCA analysis on ensemble dataset

2.2 Reduced dataset

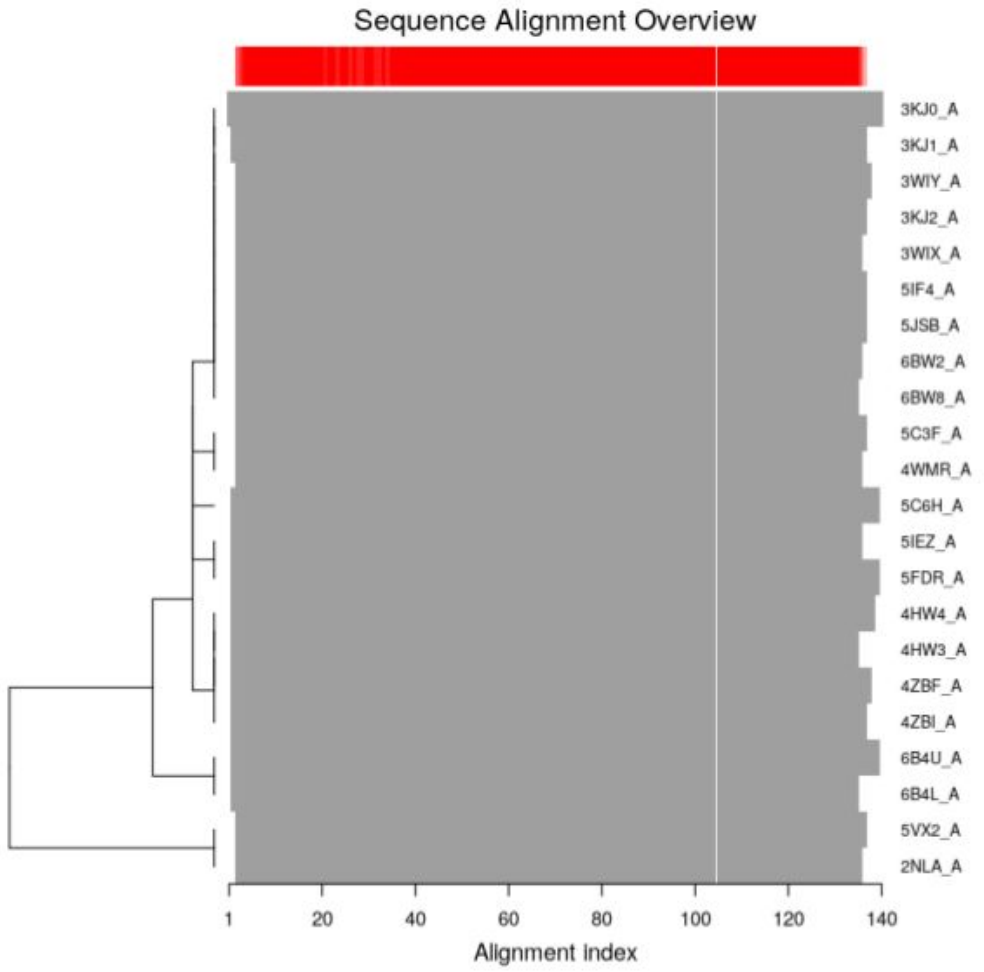
[28]:
```
# Remove structures with missing residues (reduced dataset)
conn <- inspect.connectivity(pdbs_aa, cut = 4.05)
pdbs_aa_trimmed <- trim(pdbs_aa, row.inds = which(conn))
# Which structures are removed ?
entries_with_missing_residues = as.list(which(!conn))

# Pass to the final variable pdbs_f (here we use the reduced dataset)
pdbs_f=pdbs_aa_trimmed
```

```
# a list of PDB codes of our final selection
ids <- unlist(strsplit(basename(pdbs_f$id), split=".pdb"))
plot.new()
plot(pdbs_f, labels=ids)

# SSE annotations based on 4WMR_A sequence
pdbs_f$sse = pdbs2sse(pdbs_f, pdb=T, ind=5, rm.gaps=T, resno=TRUE)
#pdbs_f$sse

# Compute the core
core <- core.find(pdbs_f)
core.inds <- print(core, vol=0.5)

## superimpose all structures to core and save to fitted directory
pdbs_f$xyz = pdbfit(pdbs_f, core$c0.5A.xyz ,outpath =␣
 ↪"fitted_core_final_reducedSet")
```

```
Re-reading PDB (4WMR_A) to extract SSE

   PDB has ALT records, taking A only, rm.alt=TRUE

Warning message in dssp.pdb(pdb.ref, …):
"Non-protein residues detected in input PDB: ZN, 865, POP, HOH"

 core size 148 of 149  vol = 40.375
 core size 147 of 149  vol = 34.365
 core size 146 of 149  vol = 31.179
 core size 145 of 149  vol = 28.709
 core size 144 of 149  vol = 26.496
 core size 143 of 149  vol = 24.79
 core size 142 of 149  vol = 23.205
 core size 141 of 149  vol = 21.868
 core size 140 of 149  vol = 20.717
 core size 139 of 149  vol = 19.769
 core size 138 of 149  vol = 18.665
 core size 137 of 149  vol = 17.667
 core size 136 of 149  vol = 16.799
 core size 135 of 149  vol = 15.893
 core size 134 of 149  vol = 15.116
 core size 133 of 149  vol = 14.377
 core size 132 of 149  vol = 13.555
 core size 131 of 149  vol = 12.851
 core size 130 of 149  vol = 12.167
 core size 129 of 149  vol = 11.571
 core size 128 of 149  vol = 10.967
 core size 127 of 149  vol = 10.409
 core size 126 of 149  vol = 9.904
 core size 125 of 149  vol = 9.385
```

```
core size 124 of 149  vol = 8.904
core size 123 of 149  vol = 8.4
core size 122 of 149  vol = 8.062
core size 121 of 149  vol = 7.722
core size 120 of 149  vol = 7.414
core size 119 of 149  vol = 7.116
core size 118 of 149  vol = 6.801
core size 117 of 149  vol = 6.523
core size 116 of 149  vol = 6.272
core size 115 of 149  vol = 6.061
core size 114 of 149  vol = 5.805
core size 113 of 149  vol = 5.588
core size 112 of 149  vol = 5.359
core size 111 of 149  vol = 5.135
core size 110 of 149  vol = 4.938
core size 109 of 149  vol = 4.734
core size 108 of 149  vol = 4.479
core size 107 of 149  vol = 4.286
core size 106 of 149  vol = 4.103
core size 105 of 149  vol = 3.907
core size 104 of 149  vol = 3.735
core size 103 of 149  vol = 3.575
core size 102 of 149  vol = 3.407
core size 101 of 149  vol = 3.251
core size 100 of 149  vol = 3.106
core size 99 of 149  vol = 2.96
core size 98 of 149  vol = 2.834
core size 97 of 149  vol = 2.704
core size 96 of 149  vol = 2.601
core size 95 of 149  vol = 2.493
core size 94 of 149  vol = 2.401
core size 93 of 149  vol = 2.307
core size 92 of 149  vol = 2.207
core size 91 of 149  vol = 2.11
core size 90 of 149  vol = 2.029
core size 89 of 149  vol = 1.961
core size 88 of 149  vol = 1.884
core size 87 of 149  vol = 1.809
core size 86 of 149  vol = 1.762
core size 85 of 149  vol = 1.683
core size 84 of 149  vol = 1.612
core size 83 of 149  vol = 1.546
core size 82 of 149  vol = 1.487
core size 81 of 149  vol = 1.437
core size 80 of 149  vol = 1.383
core size 79 of 149  vol = 1.328
core size 78 of 149  vol = 1.271
core size 77 of 149  vol = 1.222
```

```
 core size 76 of 149  vol = 1.181
 core size 75 of 149  vol = 1.137
 core size 74 of 149  vol = 1.093
 core size 73 of 149  vol = 1.052
 core size 72 of 149  vol = 1.018
 core size 71 of 149  vol = 0.98
 core size 70 of 149  vol = 0.939
 core size 69 of 149  vol = 0.902
 core size 68 of 149  vol = 0.868
 core size 67 of 149  vol = 0.835
 core size 66 of 149  vol = 0.8
 core size 65 of 149  vol = 0.761
 core size 64 of 149  vol = 0.726
 core size 63 of 149  vol = 0.694
 core size 62 of 149  vol = 0.67
 core size 61 of 149  vol = 0.645
 core size 60 of 149  vol = 0.622
 core size 59 of 149  vol = 0.598
 core size 58 of 149  vol = 0.576
 core size 57 of 149  vol = 0.55
 core size 56 of 149  vol = 0.524
 core size 55 of 149  vol = 0.5
 FINISHED: Min vol ( 0.5 ) reached
# 56 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1   178 191     14
2   208 213      6
3   217 217      1
4   260 279     20
5   287 287      1
6   289 296      8
7   311 316      6
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
   PDB has ALT records, taking A only, rm.alt=TRUE
```



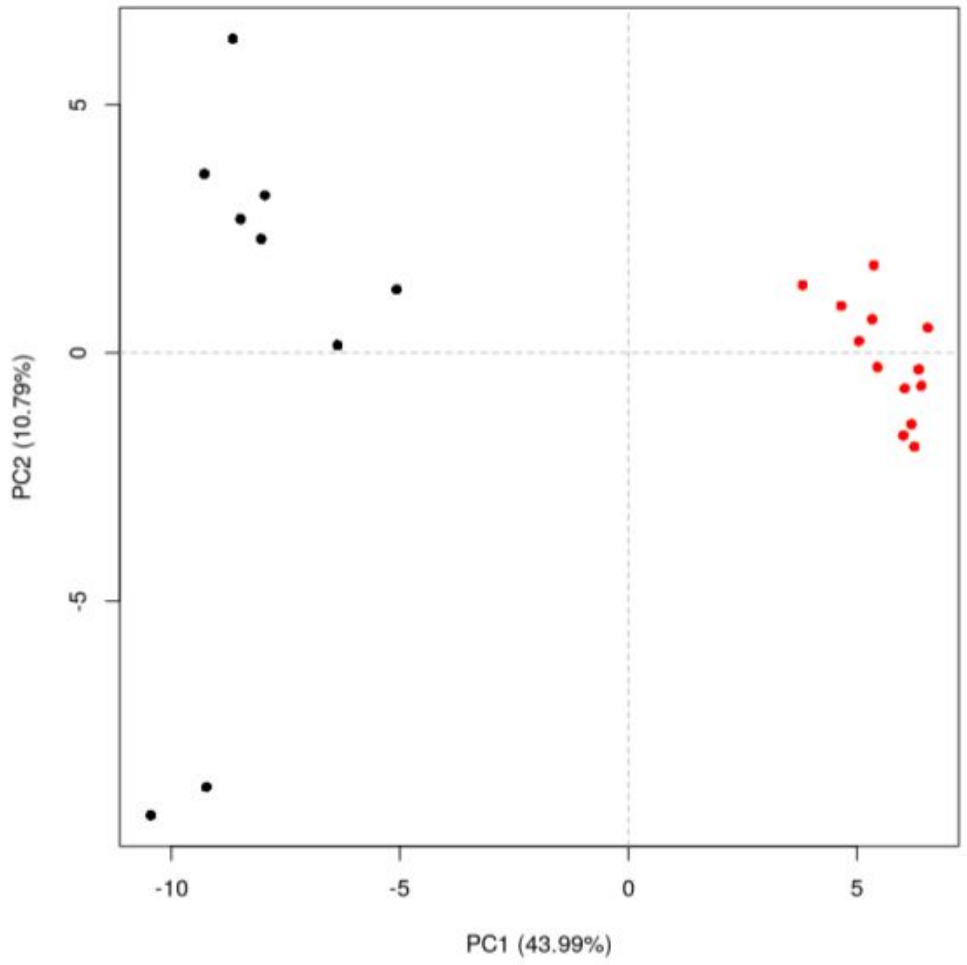
[29]:
```
# RMSD after superimpose on core ...
rd2 <- rmsd(pdbs_f$xyz)
hc.rd2 <- hclust(dist(rd2))
grps.rd2 <- cutree(hc.rd2, k = 2)
```

```
Warning message in rmsd(pdbs_f$xyz):
"No indices provided, using the 149 non NA positions
"
```

[30]:
```
#2. PCA analysis on ensemble data set ##########
# Perform PCA
gaps.pos <- gap.inspect(pdbs_f$xyz)
gaps.res <- gap.inspect(pdbs_f$ali)
```

```
pc.xray <- pca.xyz(pdbs_f$xyz[,gaps.pos$f.inds])
print(pc.xray)

# plot PCA
plot.pca(pc.xray, col=grps.rd2, pc.axes = 1:2)
#identify(pc.xray$z[,1:2], labels=basename.pdb(pdbs_f$id),offset = 0.3, cex=0.
 ↪7,pos = TRUE)
```

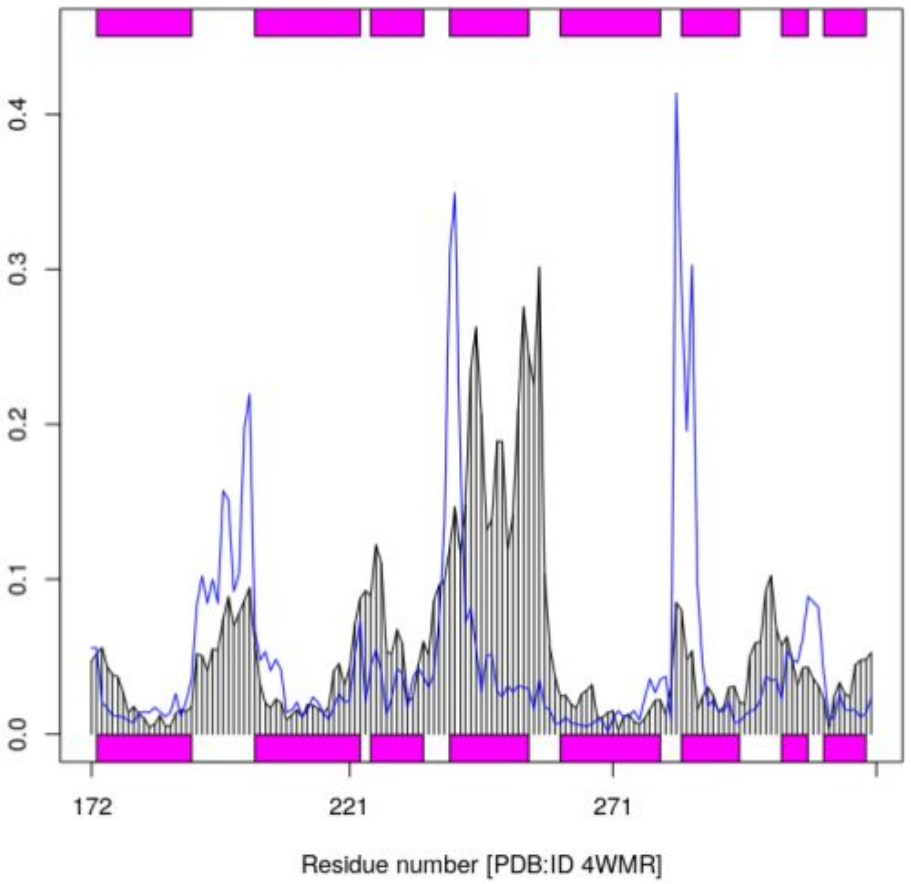
```
Call:
  pca.xyz(xyz = pdbs_f$xyz[, gaps.pos$f.inds])

Class:
  pca

Number of eigenvalues:
  447

       Eigenvalue Variance Cumulative
  PC 1     49.754   43.986     43.986
  PC 2     12.209   10.793     54.780
  PC 3     10.460    9.248     64.027
  PC 4      7.730    6.834     70.861
  PC 5      5.822    5.147     76.009
  PC 6      5.288    4.675     80.684

   (Obtained from 22 conformers with 447 xyz input values).

+ attr: L, U, z, au, sdev, mean, call
```



2.3 Residue contributions from reduced dataset

[31]:
```
# Residue contributions

#Extracting SSE from pdbs$sse attribute

plotb3(pc.xray$au[,1],sse=pdbs_f$sse, resno=pdbs_f$resno[1, gaps.res$f.inds],
       ylab="", sse.type='classic',
       col="black", xlab="Residue number [PDB:ID 4WMR]",
       ylim=c(0,0.45),helix.col = "magenta", sse.border = "black")
lines(pc.xray$au[,1], col="black")
lines(pc.xray$au[,2], col="blue")
```

```
# To Extract per residue contributions : here we View contrib on PC1/PC2
res_contrib_PC1=cbind(pdbs_f$resno[1, gaps.res$f.inds],pc.xray$au[,1])
#print(res_contrib_PC1)
#View(res_contrib_PC1)

#res_contrib_PC2=cbind(pdbs_f$resno[1, gaps.res$f.inds],pc.xray$au[,2])
#View(res_contrib_PC2)

#PCs trajectory: PC1:pc1.pdb ; PC2:pc2.pdb
#Only C-alpha atoms.
#Can be viewed using VMD tube representation and B-factors coloring scheme ...
mktrj.pca(pc.xray, pc=1, b=pc.xray$au[,1], file="pc1.pdb")
mktrj.pca(pc.xray, pc=2, b=pc.xray$au[,2], file="pc2.pdb")
```



Molecular dynamics simulations of 2PQK (WT) & C286W

1. WT-MD

[33]:
```
pdb3 = read.pdb("protein-WT.pdb")

# Mcl-1-WT trajectory (100ns)2PQK
trj3 <- read.dcd("2PQK_FULL_TRAJ_fit_backbone_stride10.dcd")

#1). trajectory fit
ca.inds <- atom.select(pdb3, "protein", elety="CA")
trj_for_pca <- fit.xyz(pdb3$xyz, trj3 ,
                       fixed.inds=ca.inds$xyz,
                       mobile.inds=ca.inds$xyz)

protpdb3 <- trim.pdb(pdb3, ca.inds)
prottrj3 <- trim(trj3, col.inds=ca.inds$xyz)
```

```
NATOM = 2432
NFRAME= 5004
ISTART= 0
last  = 5004
nstep = 5004
nfile = 5004
NSAVE = 1
NDEGF = 0
version 24
 |======================================================================| 100%
```

1) Project 2PQK-WT Trajectory on X-ray.PC space

[34]:
```
#Use the full dataset for the PCA ...
pdbs_f = pdbs_aa
ids <- unlist(strsplit(basename(pdbs_f$id), split=".pdb"))
# SSE annotations based on 3WIY_A !
pdbs_f$sse = pdbs2sse(pdbs_f, pdb=TRUE,
                      ind=33, rm.gaps=FALSE,
                      resno = FALSE)
anno <- pdb.annotate(ids)

core <- core.find(pdbs_f)
core.inds <- print(core, vol=0.5)

pdbs_f$xyz = pdbfit(pdbs_f, core$c0.5A.xyz)
# identify gaps, and perform PCA
gaps.pos <- gap.inspect(pdbs_f$xyz)
gaps.res <- gap.inspect(pdbs_f$ali)
pc.xray <- pca.xyz(pdbs_f$xyz[,gaps.pos$f.inds])
```

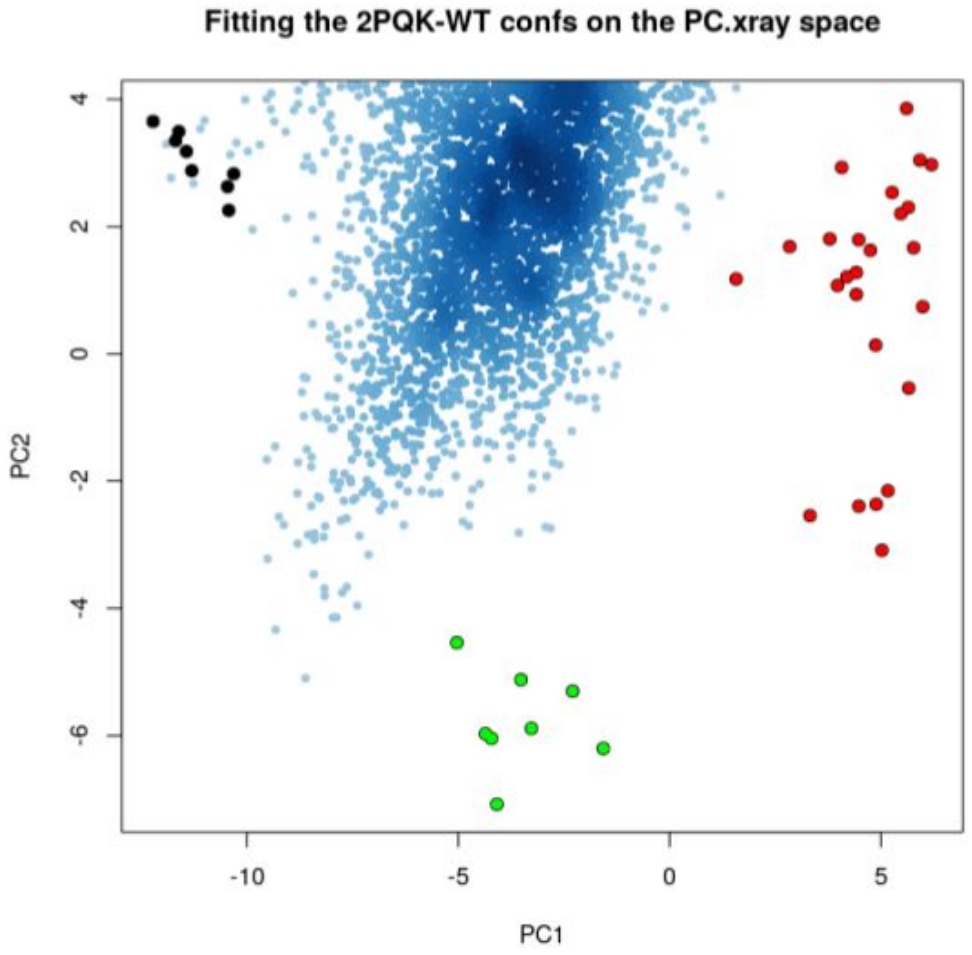
```
pc.xray <- pca.xyz(pdbs_f$xyz,rm.gaps = TRUE)

# Plot to check if same PC1:PC2
#plot.pca(pc.xray,col=grps.rd,pc.axes = 1:2)
###### OK, same PC1:PC2 space

inds3 <- pdb2aln.ind(pdbs_f, pdb3, gaps.res$f.inds)
inds3.core <- pdb2aln.ind(pdbs_f, pdb3, core$c0.5A.atom)

# inds$a is for the pdbs object
# inds$b is for the MD structure
inds3$a
inds3$b

# Here I'm fitting the trajectory
trj3.fit <- fit.xyz(pdbs_f$xyz[31,], trj3,
                    core$c0.5A.xyz,
                    inds3.core$b$xyz)
proj3 <- project.pca(trj3.fit[,inds3$b$xyz], pc.xray)
cols3 <- densCols(proj3[,1:2])

plot(proj3[,1:2], col=cols3,
     pch=20,main="Fitting the 2PQK-WT confs on the PC.xray space",
     ylab="PC2", xlab="PC1",
     xlim=range(pc.xray$z[,1]),
     ylim=range(pc.xray$z[,2]))


#show the xray structures
points(pc.xray$z[,1:2], col=1, pch=1, cex=1.1)
# Then color them depending on hclus grouping ...
cols <- c("black", "green", "red")[grps.rd]
points(pc.xray$z[,1:2], col=cols, pch=16)

# identify only on RStudio ...
#identify(pc.xray$z[,1:2], labels=basename.pdb(pdbs_f$id))
```

```
Re-reading PDB (3WIX_A) to extract SSE
Warning message in dssp.pdb(pdb.ref, …):
"Non-protein residues detected in input PDB: LC3, HOH"Warning message in
pdb.annotate(ids):
"ids should be standard 4 character PDB-IDs: trying first 4 characters…"

 core size 132 of 133  vol = 23.46
 core size 131 of 133  vol = 21.485
 core size 130 of 133  vol = 19.649
 core size 129 of 133  vol = 18.363
 core size 128 of 133  vol = 17.504
```

```
core size 127 of 133  vol = 16.667
core size 126 of 133  vol = 15.744
core size 125 of 133  vol = 14.765
core size 124 of 133  vol = 13.99
core size 123 of 133  vol = 13.199
core size 122 of 133  vol = 12.483
core size 121 of 133  vol = 11.706
core size 120 of 133  vol = 10.968
core size 119 of 133  vol = 10.174
core size 118 of 133  vol = 9.548
core size 117 of 133  vol = 9.015
core size 116 of 133  vol = 8.475
core size 115 of 133  vol = 8.166
core size 114 of 133  vol = 7.843
core size 113 of 133  vol = 7.434
core size 112 of 133  vol = 7.113
core size 111 of 133  vol = 6.81
core size 110 of 133  vol = 6.467
core size 109 of 133  vol = 6.111
core size 108 of 133  vol = 5.816
core size 107 of 133  vol = 5.552
core size 106 of 133  vol = 5.186
core size 105 of 133  vol = 4.94
core size 104 of 133  vol = 4.678
core size 103 of 133  vol = 4.495
core size 102 of 133  vol = 4.317
core size 101 of 133  vol = 4.092
core size 100 of 133  vol = 3.902
core size 99 of 133  vol = 3.73
core size 98 of 133  vol = 3.549
core size 97 of 133  vol = 3.378
core size 96 of 133  vol = 3.236
core size 95 of 133  vol = 3.107
core size 94 of 133  vol = 2.987
core size 93 of 133  vol = 2.86
core size 92 of 133  vol = 2.741
core size 91 of 133  vol = 2.63
core size 90 of 133  vol = 2.53
core size 89 of 133  vol = 2.426
core size 88 of 133  vol = 2.32
core size 87 of 133  vol = 2.213
core size 86 of 133  vol = 2.133
core size 85 of 133  vol = 2.062
core size 84 of 133  vol = 1.99
core size 83 of 133  vol = 1.917
core size 82 of 133  vol = 1.844
core size 81 of 133  vol = 1.778
core size 80 of 133  vol = 1.712
```

```
 core size 79 of 133  vol = 1.647
 core size 78 of 133  vol = 1.591
 core size 77 of 133  vol = 1.536
 core size 76 of 133  vol = 1.481
 core size 75 of 133  vol = 1.429
 core size 74 of 133  vol = 1.384
 core size 73 of 133  vol = 1.334
 core size 72 of 133  vol = 1.285
 core size 71 of 133  vol = 1.236
 core size 70 of 133  vol = 1.184
 core size 69 of 133  vol = 1.139
 core size 68 of 133  vol = 1.094
 core size 67 of 133  vol = 1.054
 core size 66 of 133  vol = 1.012
 core size 65 of 133  vol = 0.975
 core size 64 of 133  vol = 0.936
 core size 63 of 133  vol = 0.899
 core size 62 of 133  vol = 0.864
 core size 61 of 133  vol = 0.833
 core size 60 of 133  vol = 0.8
 core size 59 of 133  vol = 0.764
 core size 58 of 133  vol = 0.724
 core size 57 of 133  vol = 0.684
 core size 56 of 133  vol = 0.654
 core size 55 of 133  vol = 0.622
 core size 54 of 133  vol = 0.593
 core size 53 of 133  vol = 0.563
 core size 52 of 133  vol = 0.535
 core size 51 of 133  vol = 0.509
 core size 50 of 133  vol = 0.486
 FINISHED: Min vol ( 0.5 ) reached
# 51 positions (cumulative volume <= 0.5 Angstrom^3)
  start end length
1   175 189     15
2   208 212      5
3   261 263      3
4   265 278     14
5   289 296      8
6   311 316      6
NOTE: Removing 389 gap positions with missing coordinate data
      retaining 133 non-gap positions for analysis.

 Call:  NULL

   Atom Indices#: 133  ($atom)
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz
```

```
 Call:  NULL

   Atom Indices#: 133  ($atom)
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz
```



2) Project 2PQK-C286W Trajectory on X-ray.PC space

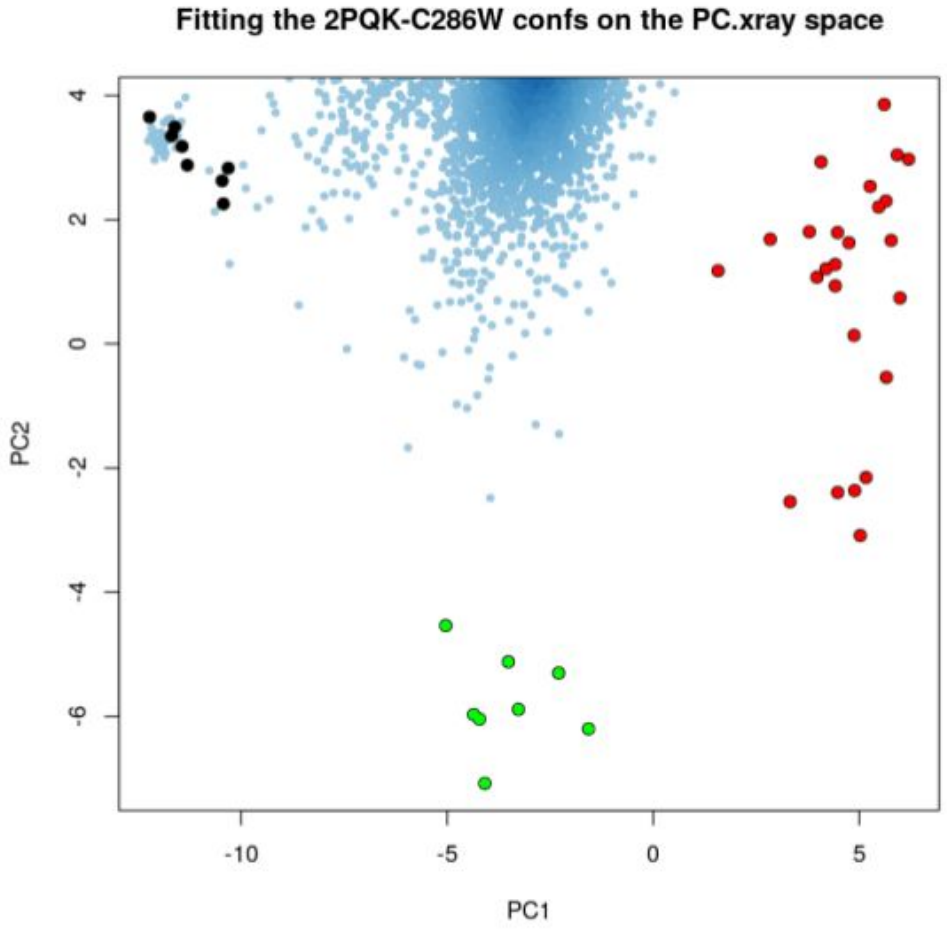
[35]:
```
#Change the path if needed ...
pdb3 = read.pdb("protein_C286W.pdb")

#C286W-MD traj ... (100ns) starting from 2PQK-C286W
#All the frames are used ...

### ------------------------------------------------------------ ###
# ... NB. => The sampling restriction along PC2 is not the result of #
# the number of frames used but a real conformational shift ... !    #
### ------------------------------------------------------------ ###

trj3 <- read.dcd("FULL_2PQK_C286W_eq_md_fit_backbone.dcd")

#1). Fit trajectory
ca.inds <- atom.select(pdb3, "protein", elety="CA")

trj_for_pca <- fit.xyz(pdb3$xyz, trj3 ,
                       fixed.inds=ca.inds$xyz,
                       mobile.inds=ca.inds$xyz)

protpdb3 <- trim.pdb(pdb3, ca.inds)
prottrj3 <- trim(trj3, col.inds=ca.inds$xyz)
```

```
NATOM = 2445
NFRAME= 10036
ISTART= 0
last  = 10036
nstep = 10036
nfile = 10036
NSAVE = 1
NDEGF = 0
version 24
 |======================================================================| 100%
```

[36]:
```
inds3 <- pdb2aln.ind(pdbs_f, pdb3, gaps.res$f.inds)
inds3.core <- pdb2aln.ind(pdbs_f, pdb3, core$c0.5A.atom)

# inds$a is for the pdbs object
# inds$b is for the MD structure
inds3$a
inds3$b

# Here I'm fitting the trajectory on Mcl-1 core.
trj3.fit <- fit.xyz(pdbs_f$xyz[31,], trj3,
                    core$c0.5A.xyz,
                    inds3.core$b$xyz)
```

```
proj3 <- project.pca(trj3.fit[,inds3$b$xyz], pc.xray)
cols3 <- densCols(proj3[,1:2])

plot(proj3[,1:2], col=cols3, pch=20,
     main="Fitting the 2PQK-C286W confs on the PC.xray space",
     ylab="PC2", xlab="PC1",
     xlim=range(pc.xray$z[,1]), ylim=range(pc.xray$z[,2]))


#show the xray structures
points(pc.xray$z[,1:2], col=1, pch=1, cex=1.1)
# Then color them depending on hclus grouping ...
cols <- c("black", "green", "red")[grps.rd]
points(pc.xray$z[,1:2], col=cols, pch=16)
#identify(pc.xray$z[,1:2], labels=basename.pdb(pdbs_f$id))
```

```
 Call:  NULL

   Atom Indices#: 133  ($atom)
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz


 Call:  NULL

   Atom Indices#: 133  ($atom)
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz
```



3) Project Well-tempered Metadynamics Trajectory on X-ray.PC space

[37]:
```
#Change the path if needed ...
pdb3 = read.pdb("step3_pdb_no_water_Meta.pdb")

#EQ+MetaD Trajectory 317 ns (stride10)
trj3 <- read.dcd("FULL_eq_and_Meta_prot_only_stride10.dcd")

# grep CA and project
ca.inds <- atom.select(pdb3, "protein", elety="CA")

trj_for_pca <- fit.xyz(pdb3$xyz, trj3 ,
                       fixed.inds=ca.inds$xyz,
```

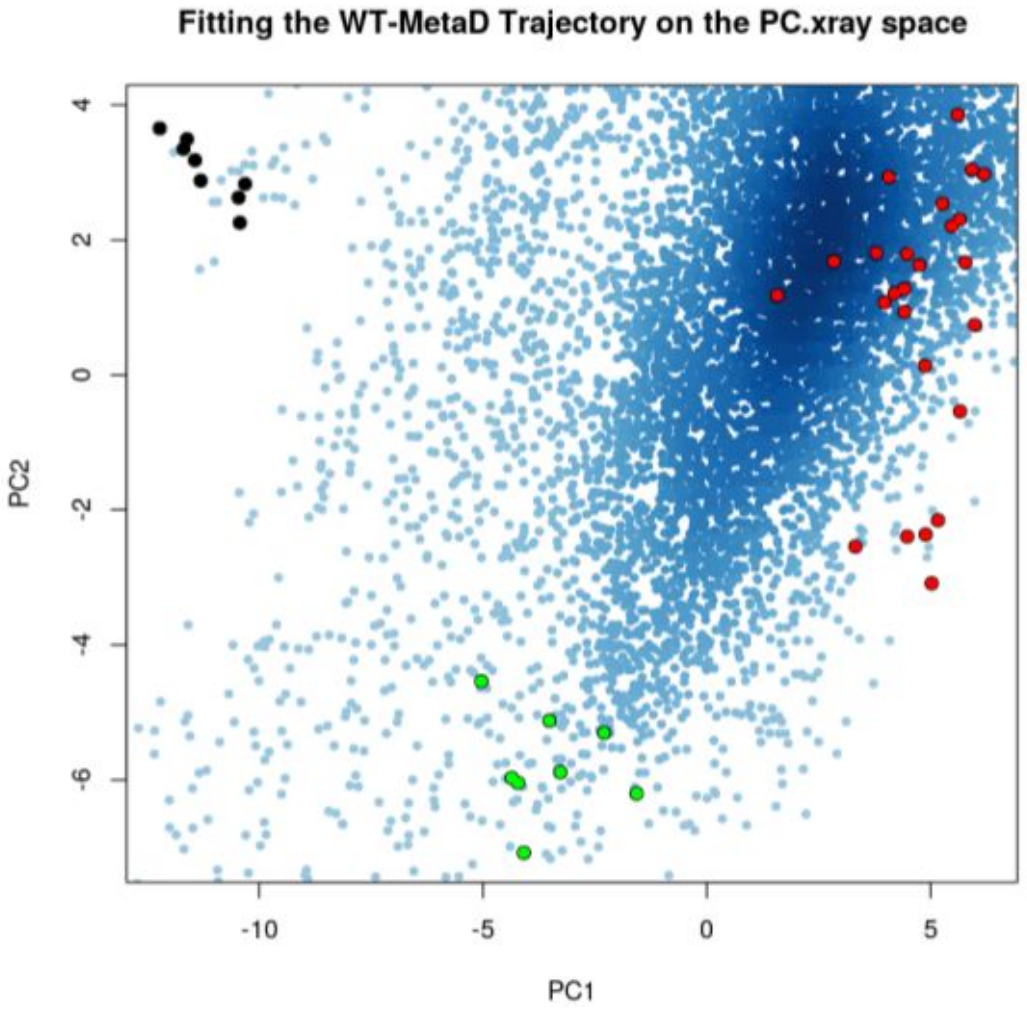
```
                        mobile.inds=ca.inds$xyz)

protpdb3 <- trim.pdb(pdb3, ca.inds)
prottrj3 <- trim(trj3, col.inds=ca.inds$xyz)
```

```
NATOM = 2432
NFRAME= 15925
ISTART= 0
last  = 15925
nstep = 15925
nfile = 15925
NSAVE = 1
NDEGF = 0
version 24
 |=================================================================| 100%
```

[39]:
```
inds3 <- pdb2aln.ind(pdbs_f, pdb3, gaps.res$f.inds)
inds3.core <- pdb2aln.ind(pdbs_f, pdb3, core$c0.5A.atom)

# inds$a is for the pdbs object
# inds$b is for the MD structure
inds3$a
inds3$b

# Here I'm fitting the trajectory on Mcl-1 core.
trj3.fit <- fit.xyz(pdbs_f$xyz[31,], trj3,
                    core$c0.5A.xyz,
                    inds3.core$b$xyz)

proj3 <- project.pca(trj3.fit[,inds3$b$xyz], pc.xray)
cols3 <- densCols(proj3[,1:2])

plot(proj3[,1:2], col=cols3, pch=20,
     main="Fitting the WT-MetaD Trajectory on the PC.xray space",
     ylab="PC2", xlab="PC1",
     xlim=range(pc.xray$z[,1]), ylim=range(pc.xray$z[,2]))


#show the xray structures
points(pc.xray$z[,1:2], col=1, pch=1, cex=1.1)
# Then color them depending on hclus grouping ...
cols <- c("black", "green", "red")[grps.rd]
points(pc.xray$z[,1:2], col=cols, pch=16)
#identify(pc.xray$z[,1:2], labels=basename.pdb(pdbs_f$id))
```

```
Call:  NULL

  Atom Indices#: 133  ($atom)
```

```
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz

 Call:  NULL

   Atom Indices#: 133  ($atom)
   XYZ  Indices#: 399  ($xyz)

+ attr: atom, xyz
```



Analyze_FES R markdown notebook

M.Benabderrahmane et al.

This R Markdown notebook was used to analyze the HILLS file produced by the well-tempered metadynamics simulation reported in the paper.

1. Install the metadynminer R package and load library

```
# Install required packages & load data #
#install.packages("metadynminer")
library(metadynminer)
#sessionInfo()
```

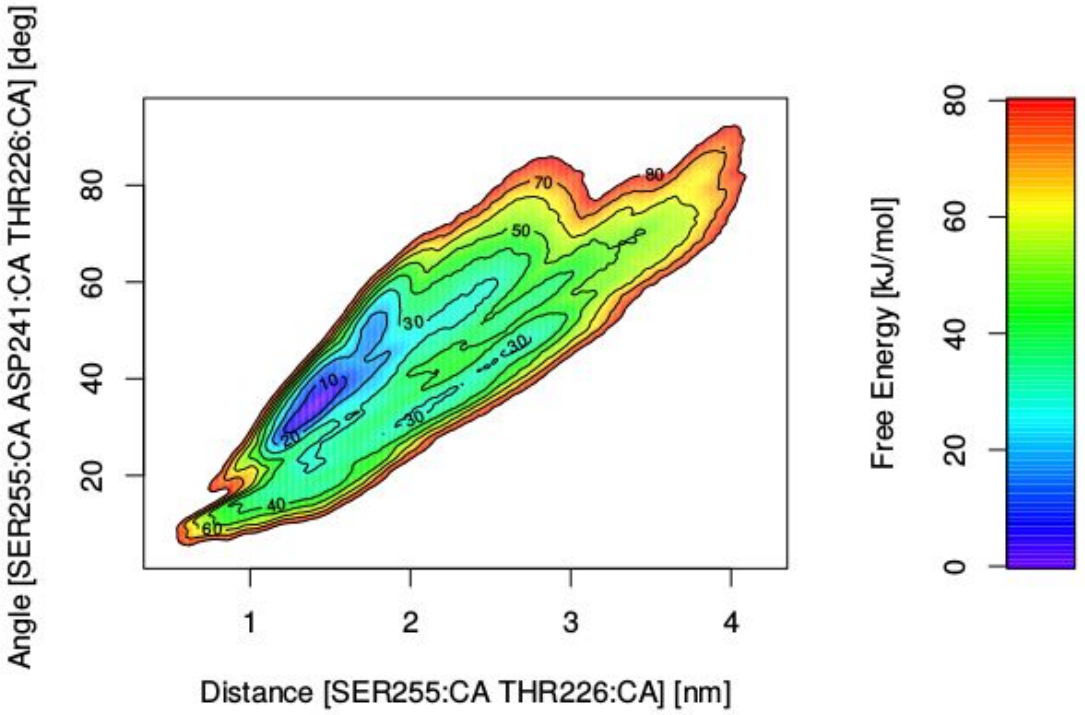
2. Read Hills file

```
# Read HILLS file
hillsf<-read.hills("HILLS")
```

```
## 2D HILLS file read
```

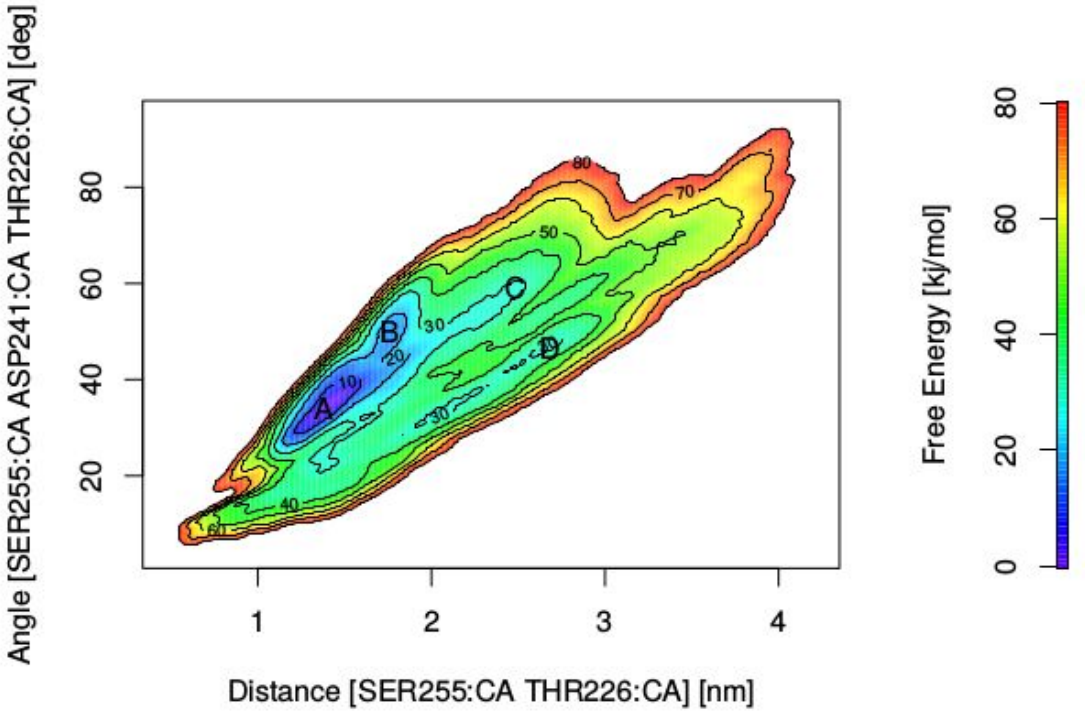
3. Compute the Free Energy Surface

```
# Compute FES
tfes<-fes2(hillsf)
tfes<-tfes-min(tfes)
plot(tfes, colscale = TRUE, plottype = "both",
     colscalelab ="Free Energy [kJ/mol]",
     xlab="Distance [SER255:CA THR226:CA] [nm]",
     ylab="Angle [SER255:CA ASP241:CA THR226:CA] [deg] ", zlim = c(0,80))
```



4. Compute minima & plot FES

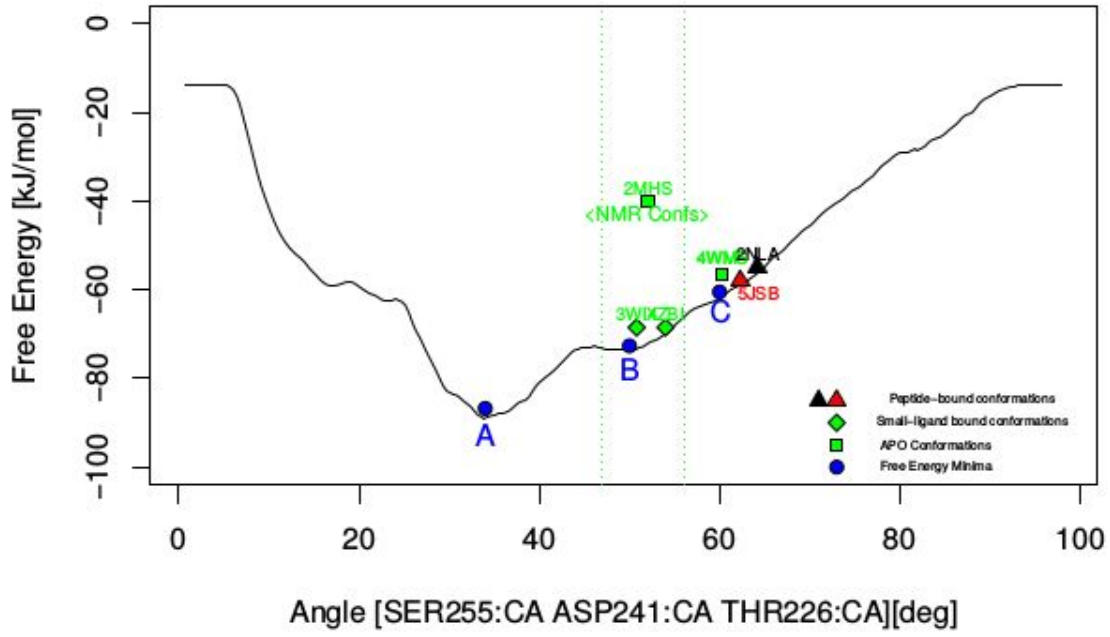
```
#compute minima, nbin=2 to focus on min of interest
minima<-fesminima(tfes, nbins = 2)
plot(minima, colscale = TRUE,
     colscalelab = "Free Energy [kj/mol]",
     xlab="Distance [SER255:CA THR226:CA] [nm]",
     ylab="Angle [SER255:CA ASP241:CA THR226:CA] [deg]",
     zlim = c(0,80), textcol = "black", lwd = 0.4)
```



5. Free Energy Profile along CV2(angle)

```
tfes2<-fes2d21d(hillsf, remdim=1)
plot(tfes2,
     xlab ="Angle [SER255:CA ASP241:CA THR226:CA][deg]",
     ylab ="Free Energy [kJ/mol]",
     col = "black",lty = "1")

#Minima as points on the profile
points(34,-86.77, type ="p", pch=21,bg="blue")
text(33,-92.77,"A",0.03,col="blue")
points(50,-72.66, type ="p", pch=21,bg="blue")
text(49,-78,"B",0.03,col="blue")
points(60,-60.52, type ="p", pch=21,bg="blue")
text(59,-65,"C",0.03,col="blue")

#Add the NMR (2MHS-APO) borders (min and max for the angle) on the profile
abline(v=47,lty=3,col="green")#Min
abline(v=56,lty=3,col="green")#Max
points(52,-40, type="p", pch=22, bg="green")#2MHS
```

```
text(52,-37,"2MHS", cex=0.57,col="green")
text(52,-43,"<NMR Confs>", cex=0.68,col="green")

#Add The X-ray APO structure (4WMS)
points(60.27,-56.52, type="p", pch=22, bg="green")
text(60.27,-53.0,"4WMS", cex=0.55, col="green")
points(60.27,-56.52, type="p", pch=22, bg="green")
text(60.27,-53.0,"4WMS", cex=0.55, col="green")

#Add synthetic-ligand structures
points(50.8, -68.52, type ="p", pch=23, bg="green") #3WIX
text(50.8,-65.52,"3WIX", cex=0.55, col="green")
points(54, -68.52, type ="p", pch=23, bg="green") #4ZBI
text(54,-65.52,"4ZBI", cex=0.55, col="green")

#Add The X-ray peptide-bound induced conformations on the profile
points(64.15, -55, type ="p", pch=24, bg="black")
text(64.15,-52,"2NLA", cex=0.55, col="black")
points(62.27, -58, type ="p", pch=24, bg="red")
text(64.27,-61,"5JSB", cex=0.55, col="red")

# Legend
points(71, -85, type ="p", pch=24, bg="black") # Pep
points(73, -85, type ="p", pch=24, bg="red") # Pep
points(73, -90, type ="p", pch=23, bg="green") # Synth
points(73, -95, type ="p", pch=22, bg="green") # APO
points(73, -100, type ="p", pch=21, bg="blue") # APO

text(88,-85,"Peptide-bound conformations", cex=0.4, col="black")
text(88,-90,"Small-ligand bound conformations", cex=0.4, col="black")
text(84,-95,"APO Conformations", cex=0.4, col="black")
text(84,-100,"Free Energy Minima", cex=0.4, col="black")
```



6. Convergence assessment

In order to check the convergence we :

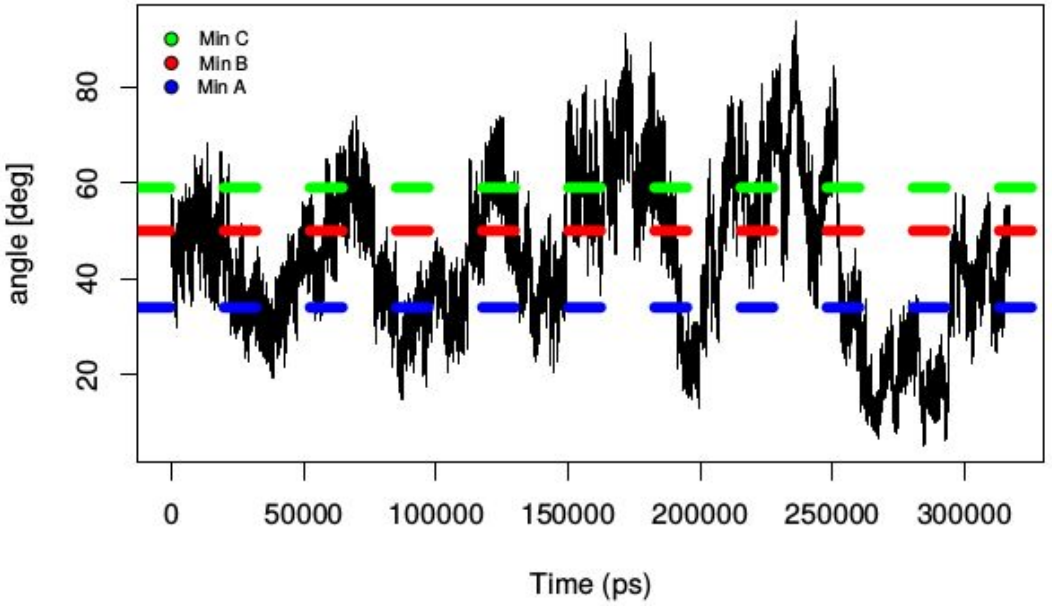
1. check the transitions observed between minima
2. compute the free-energy difference and error between two minima
3. compute the evolution of the Free Energy Profile along CV2 during the simulation

NB. Error is estimated as the standard deviation of the free-energy distance between two minima ...

```
#Compute FE-Profile
tfes<-fes2(hillsf)
minima<-fesminima(tfes, nbins = 2)
prof<-feprof(minima)
```

6.1 Check inter-minima transition events

```
plot(hillsf$time, hillsf$cv2,
     type ="l",
     xlab="Time (ps)",
     ylab="angle [deg]")
abline(h=34,lty=2,col="blue",lwd=6)#Min A
abline(h=50,lty=2,col="red",lwd=6)#Min B
abline(h=59,lty=2,col="green",lwd=6)#Min C
points(3, 80, type ="p", pch=21, bg="blue") # Min A
points(3, 85, type ="p", pch=21, bg="red") # Min B
points(3, 90, type ="p", pch=21, bg="green") # Min C
text(20000,80,"Min A ", cex=0.7, col="black")
text(20000,85,"Min B", cex=0.7, col="black")
text(20000,90,"Min C", cex=0.7, col="black")
```

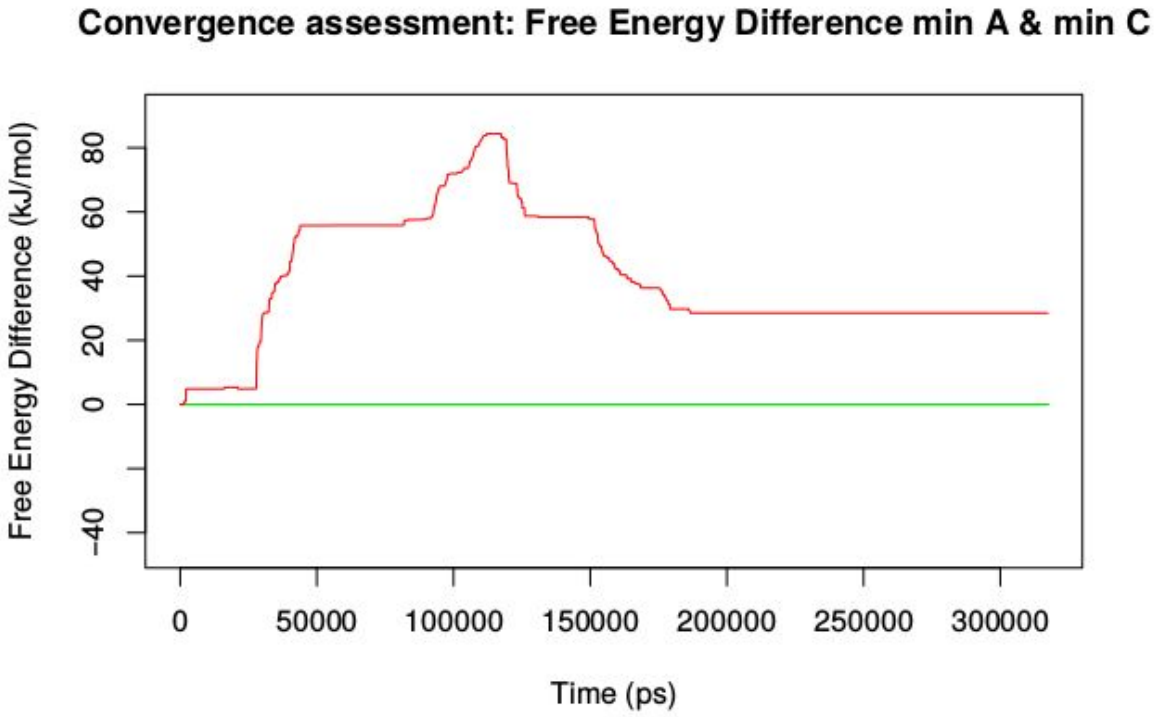


=> Many transitions(at least two) are observed between the three most populated minima A, B and C.

6.2 compute the free-energy difference & error

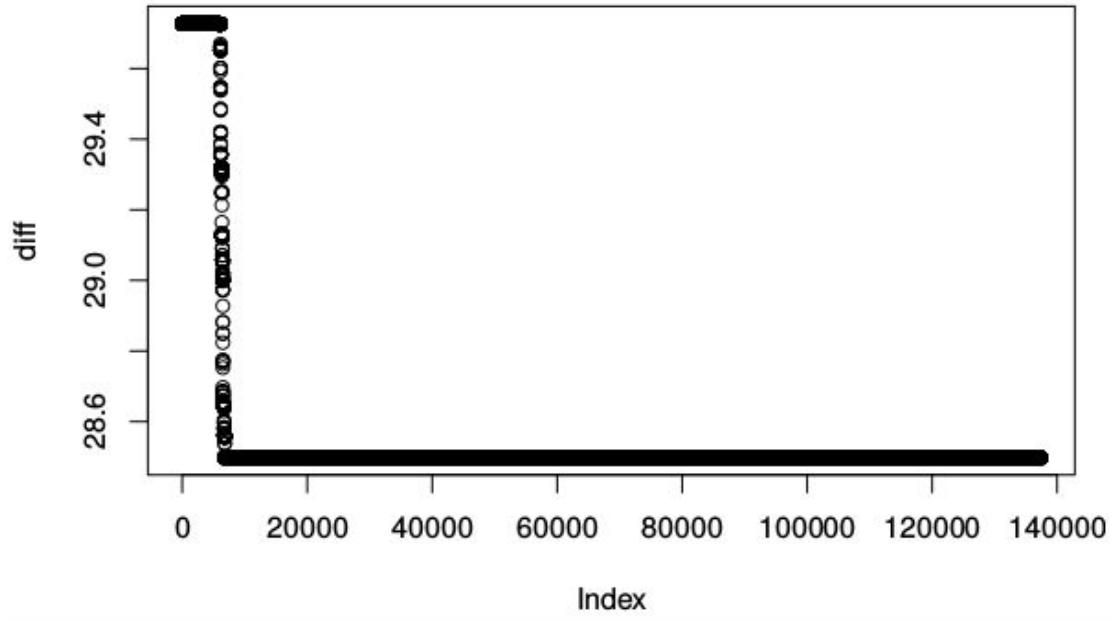
FE-profile of two minima (A & C)

```
# Plot the FE-profile of two minima (A&C)
plot(prof, which=c(1,3), xlab="Time (ps)",
     main = "Convergence assessment: Free Energy Difference min A & min C",
     xlim = c(0,317340)
)
```



FE-difference between two free-energy wells (A & C)

```
#Free-energy difference between min A & min C from 180 - 317ns
diff=prof$mms[180000:317340,4]-prof$mms[180000:317340,2]
#Plot the FE-distance minA,minC
plot(diff)
```



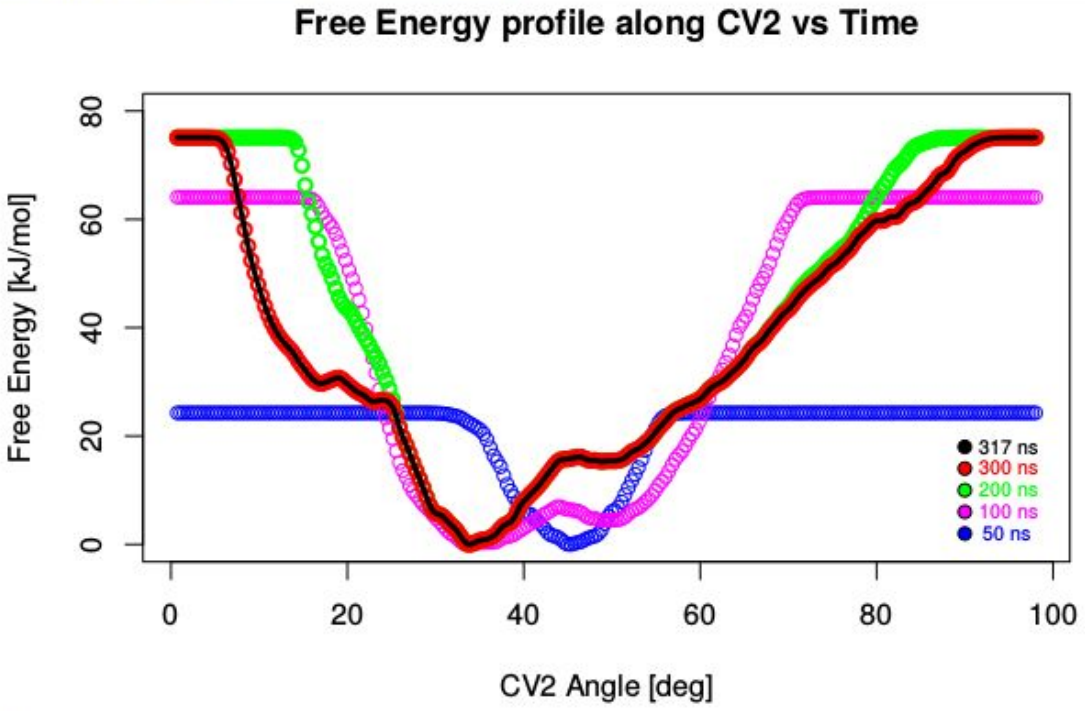

```
#Error as std. of FE-difference between min A & min C
sd(diff) #kJ/mol
```

```
## [1] 0.2562608
```

6.3 compute the evolution of the Free Energy Profile along CV2 during the simulation

```
#Interval convergence along CV2 (at 50, 100, 200, 300,317ns)
tfes2_50<-fes2d21d(hillsf, remdim=1,imax=5000)
tfes2_100<-fes2d21d(hillsf, remdim=1,imax=100000)
tfes2_200<-fes2d21d(hillsf, remdim=1,imax=200000)
tfes2_300<-fes2d21d(hillsf, remdim=1,imax=300000)
tfes2_317<-fes2d21d(hillsf, remdim=1,imax=317000)
plot.new()
plot(tfes2-min(tfes2),
     ylim=c(0,80), col="black",lwd=0.8,type="o",
     main="Free Energy profile along CV2 vs Time",
     xlab="CV2 Angle [deg]",
     ylab="Free Energy [kJ/mol]")
points(tfes2_50-min(tfes2_50), col="blue",lwd=0.4,type="o")
points(tfes2_100-min(tfes2_100), col="magenta",lwd=0.8,type="o")
points(tfes2_200-min(tfes2_200), col="green",lwd=2,type="o")
points(tfes2_300-min(tfes2_300), col="red",lwd=2,type="o")
lines(tfes2_317-min(tfes2_317), col="black",lwd=3)

#Labels
points(90, 2, type ="p", pch=21, bg="blue") # 50 ns
points(90, 6, type ="p", pch=21, bg="magenta") # 100 ns
points(90, 10, type ="p", pch=21, bg="green") # 200 ns
points(90, 14, type ="p", pch=21, bg="red") # 300 ns
points(90, 18, type ="p", pch=21, bg="black") # 317 ns
text(95,2,"50 ns ", cex=0.7, col="blue")
text(95,6,"100 ns", cex=0.7, col="magenta")
text(95,10,"200 ns", cex=0.7, col="green")
```

```
text(95,14,"300 ns", cex=0.7, col="red")
text(95,18,"317 ns", cex=0.7, col="black")
```



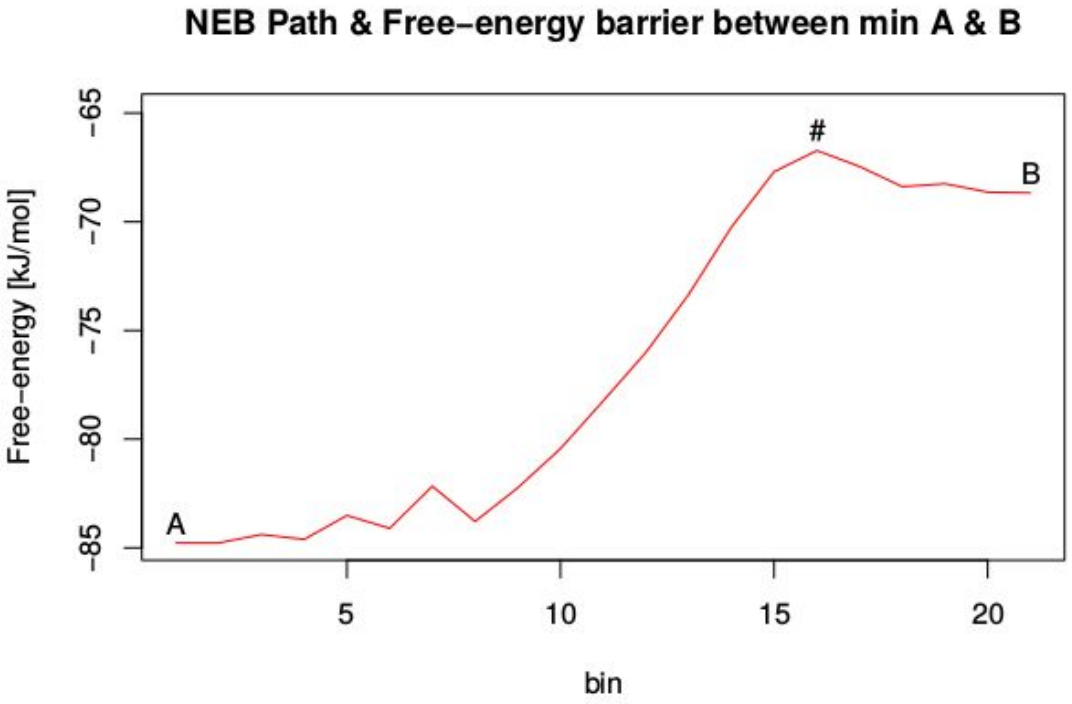

```
#
```

7. Transition barriers estimation via Nudged elastic band analysis

```
#NEB : min A & min B
nebAB_20 <- neb(minima, min1="A", min2="B", nbins = 20)
summary(nebAB_20)
```

```
## path between minima:
##   letter CV1bin CV2bin      CV1      CV2 free_energy
## 1      A     67     88 1.375968 33.92249   -84.77522
##   letter CV1bin CV2bin      CV1      CV2 free_energy
## 2      B     91    130 1.753737 49.94862   -68.66839
## with kinetics

##   direction    deltag halflife units
## 1        -> 18.050325 154.0485    ps
## 2        <-  1.943498 241.6927    fs
```

```
plot(nebAB_20, main="NEB Path & Free-energy barrier between min A & B",
     ylab="Free-energy [kJ/mol]")
```

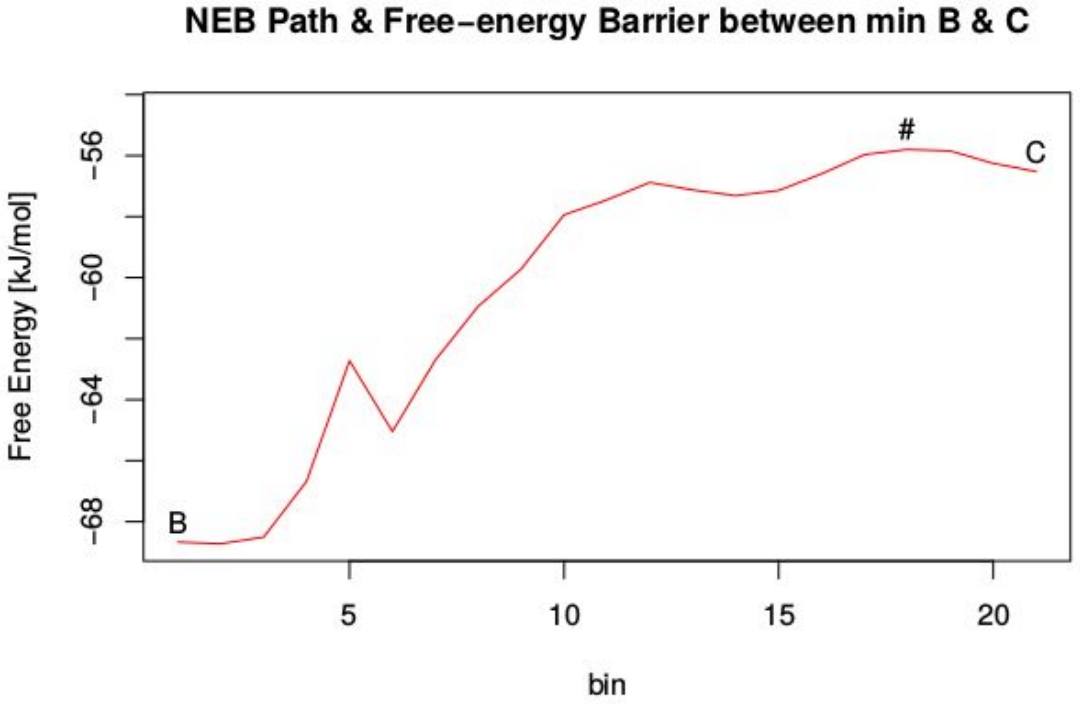



```
#NEB : min B & min C
nebBC_20 <- neb(minima, min1="B", min2="C", nbins = 20)
summary(nebBC_20)
```

```
## path between minima:
##   letter CV1bin CV2bin      CV1      CV2 free_energy
## 2      B     91    130 1.753737 49.94862   -68.66839
##   letter CV1bin CV2bin      CV1     CV2 free_energy
## 3      C    137    154 2.477793 59.1064   -56.52094
## with kinetics
```

```
##   direction     deltag  halflife units
## 1        -> 12.8714976  19.31794    ps
## 2        <-  0.7240476 148.23220    fs
```

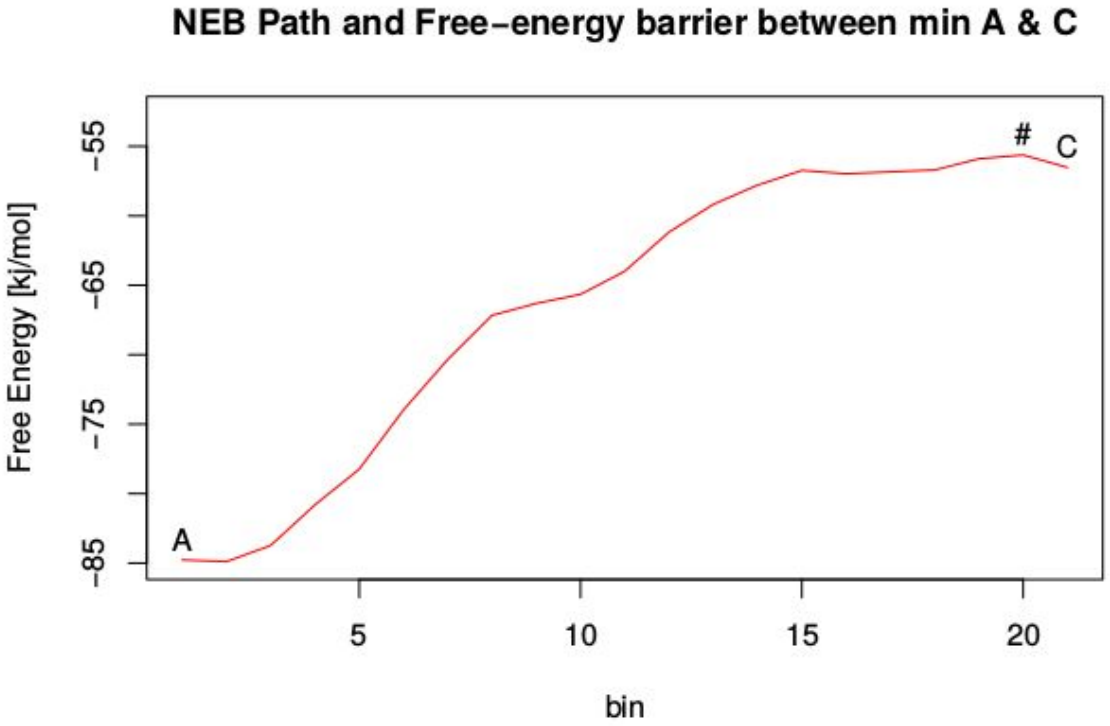
```
plot(nebBC_20,ylab = "Free Energy [kJ/mol]",
     main = "NEB Path & Free-energy Barrier between min B & C")
```




```
#NEB : min A & min C
nebAC_20 <- neb(minima, min1="A", min2="C", nbins = 20)
summary(nebAC_20)
```

```
## path between minima:
##   letter CV1bin CV2bin      CV1      CV2 free_energy
## 1      A     67     88 1.375968 33.92249   -84.77522
##   letter CV1bin CV2bin      CV1     CV2 free_energy
## 3      C    137    154 2.477793 59.1064   -56.52094
## with kinetics

##   direction     deltag  halflife units
## 1        -> 29.1485304  13.18225    ns
## 2        <-  0.8942535 158.70020    fs
```

```
plot(nebAC_20,ylab = "Free Energy [kj/mol]",
     main = "NEB Path and Free-energy barrier between min A & C")
```



Session informations

```
sessionInfo()
```

```
## R version 3.4.4 (2018-03-15)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.3 LTS
##
## Matrix products: default
## BLAS: /usr/lib/libblas/libblas.so.3.6.0
## LAPACK: /usr/lib/lapack/liblapack.so.3.6.0
##
## locale:
##  [1] LC_CTYPE=fr_FR.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=fr_FR.UTF-8        LC_COLLATE=fr_FR.UTF-8
##  [5] LC_MONETARY=fr_FR.UTF-8    LC_MESSAGES=fr_FR.UTF-8
##  [7] LC_PAPER=fr_FR.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] metadynminer_0.1.3
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.4.4  magrittr_1.5    tools_3.4.4     htmltools_0.3.6
##  [5] yaml_2.2.0      Rcpp_1.0.0      stringi_1.3.1   rmarkdown_1.11
##  [9] knitr_1.21      stringr_1.4.0   xfun_0.4        digest_0.6.18
```

Étude de la structure et la dynamique de Mcl-1 : application en cancérologie

Ce travail ayant pour objet l'étude de la structure et la dynamique de Mcl-1, une protéine anti-apoptotique d'intérêt en cancérologie, est scindé en trois parties.
Une première étude concerne une caractérisation du mode d'interaction du Pyridoclax (un inhibiteur BH3-mimétique) avec Mcl-1 par des approches expérimentales (RMN) et théoriques (simulations de dynamique moléculaire). Une deuxième partie est consacrée à l'étude et à la caractérisation des espaces conformationnels de Mcl-1 et son mode d'inhibition allostérique. Le troisième volet de ce travail, traite d'une approche d'analyse, basée sur des simulations de métadynamique avec comme application la détection du répertoire des poches cryptiques de Mcl-1.

Study of the structure and dynamics of Mcl-1: application in oncology

This work, which aims to study the structure and dynamics of Mcl-1, an anti-apoptotic protein of interest in cancer, was carried out in three parts.
A first study focused on a characterization of the interaction mode of Pyridoclax (a BH3-mimetic) with Mcl-1 by experimental (NMR) and theoretical approaches (molecular dynamics simulations). A second part is devoted to the study and characterization of the conformational space of Mcl-1 and its mode of allosteric inhibition. In the last part of this work, Metadynamics simulations on essential dynamics space as a general approach for Mcl-1's cryptic pockets detection were evaluated.

Mots clés. Cancer / Protéines anti-apoptotiques / Protéine Mcl-1 /
Dynamique moléculaire / Métadynamique / Allostérie /