



THESIS / THÈSE

DOCTOR OF SCIENCES

Biophysical and enzymological exploration of Mycobacterium tuberculosis phosphoserine phosphatase SerB2 self-assembly and regulation by L-serine

Pierson, Elise

Award date:
2022

Awarding institution:
University of Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



FACULTÉ
DES SCIENCES

**Biophysical and enzymological exploration
of *Mycobacterium tuberculosis*
phosphoserine phosphatase SerB2
self-assembly and regulation by L-serine**

A thesis submitted by **Elise Pierson**
in fulfillment of the requirements
for the degree of Doctor in Sciences (Ph.D.)

Thesis committee:

Prof. Marianne Fillet (Université de Liège)

Prof. René Wintjens (Université Libre de Bruxelles)

Dr. Catherine Michaux (Université de Namur, president of the jury)

Prof. Xavier De Bolle (Université de Namur)

Prof. Johan Wouters (Université de Namur, thesis advisor)

October 2022

Université de Namur
FACULTÉ DES SCIENCES
Rue de Bruxelles 61, 5000 Namur, Belgique

**Biophysical and enzymological exploration of *Mycobacterium tuberculosis*
phosphoserine phosphatase SerB2 self-assembly and regulation by L-serine**

by **Elise Pierson**

Abstract

Homo-oligomerization, or protein self-assembly, is a phenomenon governing a wide variety of cellular functions. The homo-oligomeric state of an enzyme can, for instance, influence its function and/or activity, and consequently all the biochemical mechanisms arising therefrom. An emerging therapeutic strategy therefore consists in regulating the activity of an enzyme by interfering with its quaternary structure(s) using small molecules. The implementation of such an approach requires the thorough characterization of homo-oligomerization in the targeted system.

In the context of research against tuberculosis and the ever-growing threat to public health caused by the emergence of multidrug-resistant strains, this thesis focuses on the phosphoserine phosphatase SerB2 of *M. tuberculosis* (*MtSerB2*). Not only essential to the survival of the pathogen through its metabolic function in L-serine biosynthesis, the enzyme is also an interesting therapeutic target in light of its suspected role in host invasion.

This work consists of the exploration of *MtSerB2* homo-oligomerization in the presence and absence of L-serine. The nature, structure, and activity of the oligomeric species formed by the enzyme and its homologs (natural or engineered) are studied by electrophoresis, size exclusion chromatography, light scattering in solution, X-ray crystallography, and enzyme kinetics. Combined, the results suggest that *MtSerB2* is a morpheein. Multiple alternative quaternary assemblies of the enzyme are evidenced, including an active domain-swapped dimer, an inactive tetramer, and a low-activity trimer formed in the presence of L-serine. Their structural characteristics and formation mechanisms indicate that the interconversion between these species can take place through a conformationally flexible monomeric state. In addition to establishing the basis for the rational design of selective allosteric inhibitors, this equilibrium may provide answers to the moonlighting properties of *MtSerB2*.

Ph.D. thesis in Sciences
Namur, October 3, 2022

Université de Namur
FACULTÉ DES SCIENCES
Rue de Bruxelles 61, 5000 Namur, Belgique

**Exploration biophysique et enzymologique de l'auto-assemblage et de la
régulation par la L-sérine de la phosphosérine phosphatase SerB2 de
*Mycobacterium tuberculosis***

par Elise Pierson

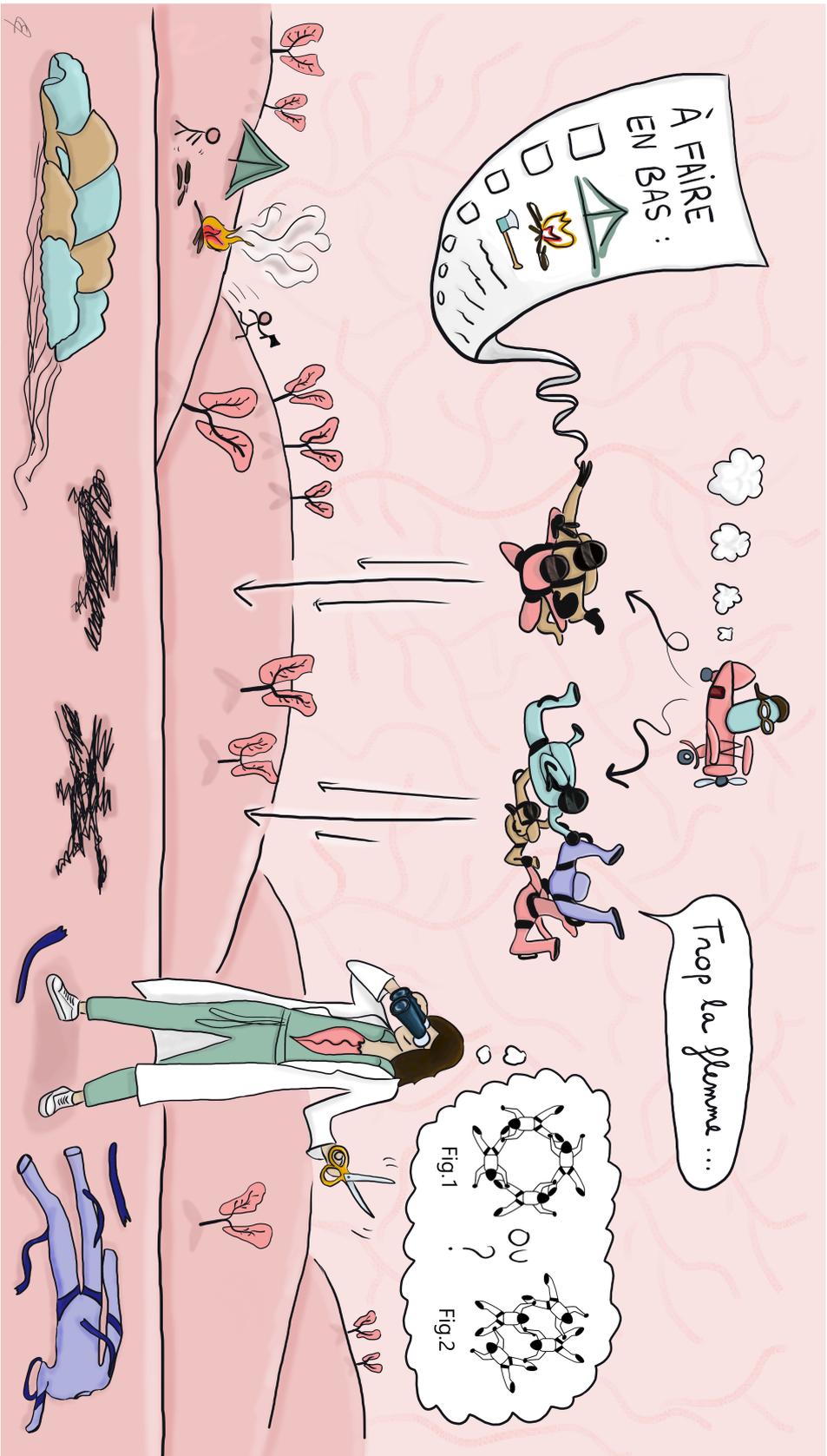
Résumé

L'homo-oligomérisation, ou auto-assemblage des protéines, est un phénomène régissant une grande variété de fonctions cellulaires. L'état homo-oligomérique d'une enzyme peut, entre autre, influencer sa fonction et/ou son activité, et par conséquent tous les mécanismes biochimiques en découlant. Une stratégie thérapeutique émergente consiste donc à réguler l'activité d'une enzyme en interférant avec sa ou ses structure(s) quaternaire(s) à l'aide de petites molécules. L'implémentation d'une telle approche nécessite la caractérisation détaillée de l'homo-oligomérisation du système ciblé.

Inscrite dans le contexte de la recherche contre la tuberculose, dont les souches multirésistantes constituent une menace croissante pour la santé publique, cette thèse a pour sujet la phosphosérine phosphatase SerB2 de *M. tuberculosis* (*MtSerB2*). Essentielle à la survie du pathogène de par sa fonction métabolique dans la biosynthèse de la L-sérine, cette enzyme est également une cible thérapeutique intéressante au regard de son rôle soupçonné dans l'invasion de l'hôte.

Ce travail consiste en l'exploration de l'homo-oligomérisation de *MtSerB2* en présence et absence de L-sérine. La nature, structure et activité des espèces oligomériques que l'enzyme et ses homologues (naturels ou conçus *in vitro*) forment sont étudiées par électrophorèse, chromatographie d'exclusion de taille, diffusion de la lumière en solution, cristallographie aux rayons-X et cinétique enzymatique. Combinés, les résultats obtenus suggèrent que *MtSerB2* est une morphéine. Plusieurs assemblages quaternaires alternatifs de l'enzyme sont mis en évidence, dont un dimère à domaines échangés actif, un tétramère inactif et un trimère peu actif formé en présence de L-sérine. Leurs caractéristiques structurales et mécanismes de formation indiquent que l'interconversion entre ces espèces peut avoir lieu au travers d'un état monomérique conformationnellement flexible. Outre le fait d'établir les bases pour la conception rationnelle d'inhibiteurs allostériques sélectifs, cet équilibre pourrait apporter des réponses quant aux propriétés multifonctionnelles ("*moonlighting*") de *MtSerB2*.

Dissertation doctorale en Sciences
Namur, 3 octobre 2022



Des parachutistes fainéants à la rescousse de vos poumons

par **Elise Pierson**

dans le cadre du concours " Ma thèse en 180 secondes "

Namur, 25 mars 2022

Résumé vulgarisé

C'est très contagieux et ça vous fait tousser. Et non, je ne vous parle pas ENCORE du coronavirus! Ce soir, la star c'est la tuberculose. " *Tibie* " pour les intimes. Et je vous assure que c'est une bactérie encore très à la mode. Elle dort dans les poumons d'un tiers de la population mondiale, et résiste de plus en plus aux antibiotiques actuels. Dans les labos, on cherche donc de nouveaux moyens de la mettre hors d'état de nuire!

Parce que *Tibie* a un but: habiter dans les cellules de vos poumons pour toujours. Pour y parvenir, elle y largue une escouade de minuscules parachutistes répondant au doux nom de SerB2. Les membres de cette escouade sont des enzymes, le genre de grosses molécules qui adorent bricoler dans vos cellules. Ces enzymes parachutistes, la première partie de ma thèse, c'était de les espionner.

Ici l'agent spécial Elise, j'ai un visuel! Les parachutistes sautent le plus souvent en tandem. Ils sont très efficaces pour préparer le terrain: ils débroussaillent, ils installent les tentes, ils font du feu... attendez, je vois aussi des parachutistes acrobates qui sautent par... quatre? Que font-ils? Euh, là, absolument rien. Ils sont en train de se faire griller des marshmallow... Hein?! Mais pourquoi une telle différence de motivation?!

Et bien, c'est lié à la figure effectuée pendant le saut. Chez les enzymes, l'architecture est très importante. La façon dont leurs atomes sont déployés dans l'espace détermine comment elles fonctionnent et si elles sont actives ou non. La flemme de l'équipe de 4 est donc expliquée par la forme de son acrobatie. Et c'est très intéressant, car *Tibie* aurait bien du mal à établir son campement dans vos poumons si on forçait toute l'escouade à sauter comme eux.

Mais pour y parvenir, il faut d'abord déterminer par quelle main, quelle jambe, quel petit doigt les parachutistes s'accrochent entre eux. Ca, c'était la deuxième partie de ma thèse. Pour observer, j'ai d'abord essayé de faire une photo en zoomant très fort. Pour faire une photo détaillée d'une enzyme au labo, il faut attendre qu'elle se fige en formant des cristaux. Qu'elle prenne la pose, quoi! Mais bon... est-ce que mes parachutistes ont bien voulu poser? Eh bien non! Donc pour contourner ça, j'ai utilisé la diffusion de la lumière pour étudier la forme et la taille de mes enzymes et fait de la mutagenèse dirigée pour identifier les contacts moléculaires importants à leur assemblage... Hum, bon, si on reprend, en fait c'est comme si je m'étais concentrée sur la silhouette des parachutistes en analysant l'ombre qu'ils projetaient au sol et puis méthodiquement, j'ai saboté une par une les poignées de leurs combinaisons pour voir quelles étaient celles qui leur permettait le maintien de leur formation de la flemme!

Au final, ces subterfuges m'ont permis de mieux comprendre la dynamique de l'escouade SerB2 et comment elle pourrait être utilisée dans le développement de nouveaux antituberculeux. Ma conclusion? Je sais pas si vous voyez où je veux en venir mais... c'est que la flemme, parfois, ça a du bon! Non?

<https://youtu.be/ot72Q2oij10>

“Art is solving problems that cannot be formulated before they have been solved. The shaping of the question is part of the answer.”

*“Problems worthy
of attack
prove their worth
by hitting back.”*

- Piet Hein

Remerciements

Il y a quatre ans, à quelques jours près, je m'installais à l'un des grands bureaux du laboratoire de Chimie Biologique Structurale pour me lancer dans cette aventure si particulière qu'est la thèse de doctorat. Loin d'imaginer les péripéties diverses et variées à venir - dont une pandémie mondiale tout de même (!) - je n'en étais pas moins convaincue que le chemin serait ponctué de rencontres enrichissantes!

Aujourd'hui, c'est toujours assise à l'un de ces grands bureaux (mais à un autre cette fois, j'ai translaté à un moment donné) que j'écris ces quelques lignes. Mon intuition n'a pas menti, et je voudrais remercier ici toutes les personnes qui m'ont accompagnée le long du chemin, et qui, de près ou de loin, ont contribué à l'élaboration de ce travail.

*Tout d'abord, je tiens à exprimer toute ma gratitude à mon promoteur de thèse, le professeur **Johan Wouters**, qui m'a ouvert les portes de son laboratoire et par la même occasion celles du monde fascinant de la chimie biologique structurale, et ce dès mon mémoire. Merci pour cette belle opportunité dans la recherche (pluridisciplinaire!), vos conseils, votre disponibilité et votre soutien tout au long de ces quatre années.*

*Je remercie également les professeurs **René Wintjens** et **Xavier De Bolle**, qui m'ont épaulée au fil du projet, **Marianne Fillet**, qui m'a chaleureusement accueillie quelques semaines au LAM à l'ULiège pour générer des résultats déterminants en SEC-MALS, et **Catherine Michaux**, qui m'a enseigné de précieuses connaissances tant en biochimie qu'en caractérisation des macromolécules durant mon cursus à Namur. Cher jury, merci à vous d'avoir accepté d'évaluer mon travail, pour l'intérêt porté à ce dernier et pour toutes les discussions constructives et bienveillantes que j'ai pu avoir avec chacun d'entre vous.*

*Un merci tout particulier au professeur **Jeffrey Keillor**. Bien que vous n'ayez pas été directement impliqué dans ce travail, mon stage dans votre laboratoire à Ottawa a grandement participé à mon envie de me lancer dans un doctorat!*

*Cette thèse n'aurait pas vu le jour sans le soutien financier du **F.R.S.-FNRS**, que je remercie pour sa confiance, ni la précieuse aide scientifique et logistique apportée par de nombreuses personnes auxquelles je suis très reconnaissante. Un grand merci aux docteurs **Javier Pérez** et **Aurélien Thureau** du synchrotron SOLEIL pour leur accueil sur la ligne SWING et le temps précieux qu'ils nous ont accordé pour mener à bien les expériences de BioSAXS. Pour une première, c'était réussi! Merci aussi aux "local contacts" les docteurs **Bill Sheppard**, **Pierre Montaville**, **Serena Sirigu**, et **Martin Savko** des lignes PX-1 et PX-2 pour leur support lors de ces nombreux runs synchrotron, de jour comme de nuit - même si la majorité de mes cristaux n'a pas diffracté! Heureusement que j'ai pu faire de la cinétique à la place... merci à la **cellule didactique** et à **Sarah-Davina Pierard** pour le prêt du spectro. Quelques*

expériences sont aussi restées à un stade très très préliminaire... mais merci à **Marc Dieu** pour les quelques essais en MS et au **département de pharmacie** pour l'accès au Zetasizer. Je voudrais ensuite remercier les **chercheuses du LAM**, **Alice** qui n'a pas hésité à laisser tomber sa rédaction quelques heures pour m'apprendre à dompter la bioinert, et **Marie, Cindy, et Aurore** pour le chouette accueil. Depuis mon mémoire, j'ai également toujours été bien accueillie par toute **l'équipe de l'URBM, PIs et chercheurs confondus**, pour les nombreuses manip de biologie moléculaire - et des "beer hour" bien cool certains vendredis soir. **Françoise, Mathieu, Aurélie**, je ne sais pas ce qu'on ferait sans vous! Une dédicace toute particulière à mes mentors en mutagenèse dirigée: **Pierre**, courageux pionnier dans cette expédition, amateur de templates VIDES mais aussi de soirées raclettes, **Adélie**, ayant réussi avec brio la mission de m'expliquer les PCR check lors de son mémoire, **Angy**, qui n'a jamais baissé les bras devant toutes ces boîtes de Petri désertes, mon "colléééé" **Elie**, toujours là pour me filer un protocole magique, une oreille attentive dans laquelle me plaindre, faire du problem solving ou papoter entre deux paillasses, et **Jérôme**, une référence aux conseils avisés et à l'humour douteux. **Caro, Francesco, Pauline, Aurore, Alex, Anthony, Gwen, Aish, Emeline, Elodie, Aravind**, je ne vous oublie pas, vous avez toujours fait de chacun de mes passages en bio un chouette moment! Merci! Pour le support informatique, toute ma reconnaissance à **Laurent et Fred**, qui, je suis sûre sont bien tristes d'avoir dû déménager dans le nouveau bâtiment... (hein?) et à la logistique, **Annick et Evelyne**. Merci de m'avoir gentiment laissé vous interrompre de nombreuses fois avec mes requêtes (et pour les sandwichs!)

Mes **chères et chers collègues du CBS**, actuels et alumni, je vous vois. Je commence à bien vous connaître, je sais que vous cherchez frénétiquement vos noms dans tout ce petit monde. C'est votre tour! Vous avez été là tout au long de cette espèce de montagne russe qu'est le doctorat. On aura partagé les montées, les descentes et les loopings de l'enfer (au sens propre comme au figuré!)... sans oublier d'innombrables mets salés et sucrés, nos goûts musicaux perso sur le baffle du labo, des nuits blanches devant l'écran du synchrotron, beaucoup de fou rires, des discussions sérieuses, un peu moins sérieuses, et j'en passe! Alors, pour cette ambiance au top, je remercie du fond du coeur **Mégane**, ma comparse de bureau désert avec qui les longues journées de rédaction deviennent (presque) un plaisir, **Manon**, excellente compétitrice à Just Dance et barmaid/cristallographe de talent, **Kali**, incroyable party organizer et amie fidèle au poste depuis nos premiers pas en chimie, **Laurie**, qui a toujours su sentir quand tout le monde avait besoin d'aller chez Glace Mania, mon ancienne encadrante **Marie**, qui m'a appris les bases dans la joie et la bonne humeur, **Tanguy et Gilles**, la relève, qui m'ont bien fait tourner en bourrique, piètres amis mais bons joueurs de badminton (et de "vin"-et-un). Merci aussi à **Eddy**, avec qui on a vraiment beaucoup ri que ce soit au bureau ou à la foire, à **Jeanne**, partenaire de pause puzzle et serial-colleuse de post-it, à **Nikolay** et à **Laurence** pour leur aide mais aussi pour les anecdotes (parfois du temps où le benzène se pipetait encore à la bouche) ponctuant le temps de midi, à **Andrea** pour ses recettes italiennes au tableau (avec jingle, s'il vous plaît), à **Martin** et son grand sourire, à **Seb, Thomas et Quentin** pour les balises au début du chemin, et à **Adeline, Adrien et Charlotte**, qui m'ont fait passer du côté "encadrante" de la force et pour leurs pierres à l'édifice!

Je n'ai pas tout à fait terminé avec le CBS... Fin 2020, un moussaillon arrivé sur sa petite barque a jeté son dévolu sur le projet SerB2, et je lui exprime ici ma plus profonde gratitude pour cet abordage plus que réussi. **Florian "Chef" De Pol**, mon co-navigateur de l'Olig-en-Mer, cette thèse ne ressemblerait pas à ce qu'elle est aujourd'hui sans ton aide inestimable tout au long de 2021 jusqu'à ce jour. Pour ta curiosité, nos deux épopées jusqu'au synchrotron, les failles spatio-temporelles, les tardives, les centaines de cafés et de goûters, tes relectures patientes, tes critiques constructives, les motd(b) - je n'ai toujours pas digéré ma défaite, les découvertes musicales, et bien d'autres choses (la liste est interminable, comme tu dis si bien dans ton mémoire)... merci mille fois!! Je suis fier de te voir reprendre ta petite barque pour aborder le navire suivant, repéré à 18 721 kilomètres d'ici. Le Roi des Pirates arrive, quelle chance ils ont!!

Je tiens aussi à remercier quelques personnes qui ont bien veillé sur moi, mon moral, mon estomac (et mon foie?) en dehors de la thèse: l'équipe de choc de MT180, **Aline, Jacques, Jia-Wei, Camille, Clara, Nico et Nico, Louise, Joëlle, et Florian**, pour cette parenthèse théâtrale géniale ; mon "deuxième" labo, le CMI, **Tarek, Val, Hongyan, Minghui, Mymy, Pengcheng, Iva, Thomas, Andrei, Chao, Li, Yingying, Maxence** (merci pour le nouveau vocabulaire mon reuf), **Louis, Eloïse, Maxime, Eva, Isa** et le professeur **Bao-Lian Su**, pour l'ambiance dont eux seuls ont le secret ; mes anciens cokotteurs, **Fanny, Sorenza et Thomas**, pour leurs encouragements et tous les bons souvenirs de notre cagibi présidentiel ; les potos de chimie que je n'ai pas encore cité ci-dessus, **Octave, Lio, Margaux et Stef** ; mes "chasseurs de papillons" préférés, **Mathilde et Thibaut**, pour nos aventures dans la nature (vive le gourmet trip!) et tous ces délicieux soupers ; les Gembloutois, **Pauline, Martin, Max, Auré, Danaé, Justine, Marceau, Mich et Emilie** ; les vieux de la vieille, **Alexis, Camille et Diego** ; my dear **Sydney**, who made sure I had a great time in Victoria this summer ; sans oublier **Lucas, Job, Damien et Nacho** qui sont très doués pour mettre de bonne humeur lorsqu'on les croise entre deux portes!

Enfin, et du fond du coeur, merci à **mes parents, mon frère, ma soeur** et toute **ma famille** d'avoir été à mes côtés tout au long du chemin. Vous m'avez toujours soutenue et encouragée, dans ce projet comme dans les autres, dans les hauts mais aussi dans les bas. Je suis tellement chanceuse de vous avoir. Et pour terminer, quelqu'un mérite ici une mention spéciale... **Marvin**, merci infiniment pour ta patience quotidienne, tes encouragements, ta personnalité solaire et positive que j'admire tellement, mais aussi ces voyages inoubliables et tous les beaux moments. J'ai pu grandir ces quatre dernières années, et tu n'y es pas pour rien.

A vous tous, et à ceux dont le nom ne serait malencontreusement pas imprimé noir sur blanc ici, encore merci!

Contents

List of abbreviations	xxi
List of amino acids	xxii
I Introduction	1
1 <i>MtSerB2</i>, an interesting oligomeric target for the development of new anti-tuberculosis agents	3
1.1 Oligomerization, the process of protein self-assembly	3
1.1.1 Why build homo-oligomeric proteins?	4
The morpheein model of allostery	6
1.1.2 What do homo-oligomeric proteins look like?	8
1.1.3 How does homo-oligomerization arise?	11
Oligomerization by domain-swapping	14
1.1.4 Why is it interesting to study the oligomerization of a protein? .	16
1.2 Tuberculosis	17
1.2.1 A treatable, but ever-present, enduring enemy	17
1.2.2 Targeting the amino acid metabolism of <i>M. tuberculosis</i>	20
Focus on the phosphorylated L-serine pathway	21
1.3 The phosphoserine phosphatase <i>MtSerB2</i>	25
1.3.1 A suspected secreted virulence factor	25
... and therefore a potential moonlighting protein	27
1.3.2 A structurally unusual haloacid dehalogenase phosphatase . . .	28
The PSP catalytic domain	28
The ACT regulatory domains	32

1.3.3	The regulation of <i>MtSerB2</i> activity: state of the art	37
II	Objectives and strategy	51
2	Studying <i>MtSerB2</i> self-assembly and regulation: why and how?	53
III	Results and discussions	57
3	Getting to know <i>MtSerB2</i> orthologs	59
3.1	Homologs selection	59
3.2	Sequence conservation analysis	60
3.2.1	Global alignment	60
3.2.2	Alignment of PSP domains	63
3.2.3	Alignment of (putative) ACT domains	65
3.2.4	Conclusion of the sequence conservation analysis	67
3.3	Biophysical and biochemical features of the orthologs	68
3.3.1	A word about enzyme production and analysis	68
3.3.2	PSP from <i>H. sapiens</i> , the human PSP (<i>HsPSP</i>)	70
	Investigation of <i>HsPSP</i> structure in solution	71
3.3.3	SerB from <i>M. avium</i> (<i>MaSerB</i>)	75
	Investigation of <i>MaSerB</i> structure in solution	76
3.3.4	SerB2 from <i>M. marinum</i> (<i>MmSerB2</i>)	78
	Enzymological characterization of <i>MmSerB2</i>	78
	Investigation of <i>MmSerB2</i> structure in solution	80
3.3.5	SerB from <i>B. melitensis</i> (<i>BmSerB</i>)	82
	Determination of the structure of <i>BmSerB</i> (on-hold PDB entry 7QPL)	82
	Investigation of <i>BmSerB</i> structure in solution	87
	Enzymological characterization of <i>BmSerB</i> and behavior in the presence of L-serine	90
3.4	Summary of the findings	93
4	Characterization of <i>MtSerB2</i> oligomeric states in the absence of L-serine	99
4.1	Observation and identification of the oligomeric states	99
4.1.1	<i>MtSerB2</i> elutes and migrates as two oligomeric species	99

4.1.2	<i>MtSerB2</i> can form a dimer and a tetramer	101
	Relative estimation of M_W by calibration of the SEC column . . .	101
	Absolute measure of M_W by SEC coupled to Multi-Angle Light Scattering (SEC-MALS)	103
4.1.3	<i>MtSerB2</i> dimer and tetramer populations interconvert through a slow equilibrium independent of the total protein concentration	106
	The dimer and the tetramer are stable species that can be sepa- rated and individually concentrated	106
	The dimer and tetramer populations rearrange over time	108
	The interconversion between dimer and tetramer is not affected by sample concentration	110
	Phosphate and pH do not seem to influence the dimer/tetramer equilibrium	112
4.2	Structural characterization of the oligomeric species	113
4.2.1	<i>MtSerB2</i> dimer and tetramer are difficult to crystallize	113
4.2.2	Investigation of the major species: <i>MtSerB2</i> dimer	114
	<i>MtSerB2</i> depends on its ACT1 domain to oligomerize	114
	<i>MtSerB2</i> forms a dimer similar to that of <i>MaSerB</i>	117
4.2.3	Investigation of the minor species: <i>MtSerB2</i> tetramer	118
	Gln92 is involved in <i>MtSerB2</i> propensity to form a tetramer . . .	118
	The N-term part of the second α -helix of ACT2 domain is a hotspot for tetramer formation	120
	<i>MtSerB2</i> tetramer is a rather globular, distinct oligomeric species	124
	<i>MtSerB2</i> tetramer could be an assembly of monomers with global D2 symmetry	126
4.3	Enzymological characterization of the oligomeric species	134
4.3.1	<i>MtSerB2</i> dimer is active and exhibits catalytic properties similar to those of <i>MmSerB2</i> and <i>MaSerB</i>	134
4.3.2	<i>MtSerB2</i> tetramer is inactive	137
4.3.3	<i>MtSerB2</i> in the monomeric form could be almost inactive	139
4.4	Summary of the findings	140
5	Characterization of <i>MtSerB2</i> oligomeric states in the presence of L-serine	145
5.1	Observation and identification of the oligomeric states	145
5.1.1	<i>MtSerB2</i> forms a new species in the presence of L-Ser	145

5.1.2	<i>MtSerB2</i> L-Ser induced form is a trimer	147
5.1.3	L-Ser-induced trimer formation is specific to <i>MtSerB2</i>	150
5.1.4	<i>MtSerB2</i> trimer is formed from the dimer	152
5.2	Structural characterization of <i>MtSerB2</i> L-Ser induced trimer	153
5.2.1	<i>MtSerB2</i> trimer has proven difficult to crystallize	153
5.2.2	<i>MtSerB2</i> trimer is an extended assembly	154
5.2.3	The trimeric interface of <i>MtSerB2</i> is probably formed by the PSP and ACT2 domains	157
5.3	Investigation of <i>MtSerB2</i> trimer formation mechanism	161
5.3.1	<i>MtSerB2</i> dimer possesses a second L-Ser binding pocket	161
5.3.2	Glu33, Arg103 and Thr136 are implicated in <i>MtSerB2</i> trimer for- mation	162
5.4	Enzymological characterization of the oligomeric species in the pres- ence of L-Ser	166
5.4.1	L-Ser acts as a mixed-type predominantly competitive inhibitor of <i>MtSerB2</i> dimer	166
5.4.2	L-Ser acts differently on <i>MtSerB2</i> and its dual-ACT domain or- thologs	168
5.5	Summary of the findings	173
IV	General discussion, conclusions and outlooks	177
6	<i>MtSerB2</i>, a potential morpheein	179
6.1	General discussion	179
6.2	Conclusions	191
6.3	Outlooks	195
V	Appendices	201
A	Brief overview of the techniques	203
B	Experimental procedures	219
B.1	Protein production	219
B.1.1	Plasmids	219
	Wild-type enzymes	219

	Mutant enzymes	219
B.1.2	Protein overexpression	221
B.1.3	Protein purification	222
B.2	Observation and identification of oligomeric states	223
B.2.1	Native polyacrylamide gel electrophoresis (Native-PAGE)	223
B.2.2	Mass photometry (MP)	223
B.2.3	Size exclusion chromatography (SEC)	223
	Semi-preparative SEC	223
	Preparative SEC	224
	Analytical SEC	224
B.2.4	SEC coupled to multi-angle light scattering (SEC-MALS)	225
B.3	Structural characterization	226
B.3.1	<i>In silico</i> modeling of oligomeric species	226
	Homology modeling of <i>MtSerB2</i> and <i>MmSerB2</i> dimers	226
	<i>In silico</i> modeling of <i>MtSerB2</i> theoretical closed monomer using Molecular Dynamics (MD)	226
	<i>In silico</i> modeling of <i>MtSerB2</i> tetramer and trimer using sym- metrical protein-protein docking	227
B.3.2	<i>In silico</i> study of L-Ser binding sites	227
	Ligand binding site prediction	227
	Induced-fit docking of L-Ser in <i>MtSerB2</i> dimer model	228
B.3.3	SEC coupled to small-angle X-ray scattering (SEC-SAXS) and SAXS-based model evaluation/modelisation	228
	SEC-SAXS data collection	228
	SAXS data treatment	229
	Evaluation of the models generated by symmetric protein-protein docking and comparison of crystallographic structures with solution structures	229
	Refining of models and crystallographic structures based on SAXS data	230
	Rigid body modeling	230
B.3.4	Crystallization assays	230
	Screening in 96-well plates	230
	<i>BmSerB</i> crystallization	231
B.3.5	X-ray data collection and analysis	231

B.4	Enzyme activity	231
B.4.1	Steady-state kinetics with and without L-Ser	231
	Malachite green-based phosphatase assay	231
	Evaluation of the kinetic parameters	232
C	Superimposition of <i>MaSerB</i> crystallographic structures	237
D	Biochemical characterization of phosphoserine phosphatase SerB2 from <i>Mycobacterium marinum</i>	239
E	Data collection and refinement statistics for <i>BmSerB</i> structure	241
F	Mass photometry measurement on unseparated <i>MtSerB2</i> sample	243
G	Crystals obtained from <i>MtSerB2</i> crystallization assays	245
H	Calculation of the assumed amount of active subunit in a solution of <i>MtSerB2</i> dimer and tetramer	259

List of abbreviations

<i>Bm</i>	<i>Brucella melitensis</i>
CFZ	Clofazimine
COVID-19	Corona Virus Disease appeared in 2019
Cryo-EM	Cryo-Electron Microscopy
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
FPLC	Fast Protein Liquid Chromatography
HAD	Halo-Acid Dehalogenase
HPLC	High Performance Liquid Chromatography
<i>Hs</i>	<i>Homo sapiens</i>
IC ₅₀	Half maximal Inhibitory Concentration
IL-8	Interleukin 8
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl- β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
<i>Ma</i>	<i>Mycobacterium avium</i>
MALS	Multi-Angle Light Scattering
MAPK p38	Mitogen Activated Protein Kinase p38
MDR	Multi Drug Resistant
<i>Mm</i>	<i>Mycobacterium marinum</i>
MP	Mass Photometry
<i>Mt</i>	<i>Mycobacterium tuberculosis</i>
NAD ⁺	Nicotinamide Adenine Dinucleotide (oxidized)
NF κ B	Nuclear Factor-Kappa B
PBGS	Porphobilinogen Synthase
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PGDH	Phosphoglycerate Dehydrogenase
PS	Phosphoserine
PSAT	Phosphoserine Aminotransferase
PSP	Phosphoserine phosphatase
PAGE	Poly-Acrylamide Gel Electrophoresis
RR	Rifampicin Resistant

SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SAXS	Small-Angle X-ray Scattering
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
TCEP	Tris(2-carboxyethyl)phosphine
TB	Tuberculosis
THP-1	Tamm-Horsfall Protein 1
UV	Ultra-Violet
WHO	World Health Organization
XDR	Extensively Drug Resistant
XRD	X-ray diffraction

List of amino acids

Name	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Part I

Introduction

Chapter 1

MtSerB2, an interesting oligomeric target for the development of new antituberculosis agents

1.1 Oligomerization, the process of protein self-assembly

Proteins are naturally occurring long amino acid polypeptides that perform a wide range of functions within the organisms of all kingdoms of life, from catalysis of metabolic reactions to providing cell structure, through small molecule transport. They are characterized by four structure levels:

- **The primary structure** or the amino acid sequence
- **The secondary structure** or the local folding of the amino acid chain on itself, stabilized by hydrogen bonds between atoms in the polypeptide backbone
- **The tertiary structure** or the spatial arrangement of the secondary structures into a stable three-dimensional shape, stabilized by hydrogen bonds, van der Waals forces, ionic bonds and lipophilic contacts

- **The quaternary structure** or the association of several folded protein chains

This thesis focuses on the last level of structure. Although the latter can describe interactions between different protein chains, giving rise to *hetero-oligomers*, we will focus specifically on *homo-oligomerization*, i.e. the self-association of proteins.

1.1.1 Why build homo-oligomeric proteins?

It is nowadays recognized, thanks to structural determination techniques, that a vast majority of existing proteins form homo-oligomeric complexes [1]. This finding is not due to chance: as well reviewed in references [2] and [3] the formation of homomers is advantageous for the cell at various levels.

To protect from denaturation and to save space - By definition, oligomerization creates *larger* proteins. These large complexes gain stability due to an extensive network of weak interactions at the interface separating the subunits. The resulting favorable enthalpic gain makes the oligomeric architecture less prone to denaturation than the subunits alone, in which the entropic cost of solvation and conformational restriction is sometimes not sufficiently compensated. Homomeric proteins also gain stability by burying hydrophobic clusters exposed to solvent at oligomeric interfaces, and possess a smaller surface area, which reduces the amount of solvent needed to hydrate them. Moreover, homo-oligomerization provides a genetically compact way to produce large macromolecules and a way to reduce translation errors by discarding faulty subunits.

To perform a diversity of physiological functions - Morphology plays a role in function, as the cell sometimes needs to create objects of a given size and shape to perform defined roles. This is well illustrated by the structure of the DNA clamp of the bacterial DNA polymerase III, which, thanks to its ring shape composed of two subunits of 3 domains each, can perfectly surround a DNA strand. The oligomeric state can also influence the enzymatic activity, when for example the active site is formed at the interface of two subunits or homomerization is required for the catalytic residues within the active site to adopt the proper 3D conformation. Moreover, some enzymes exhibit distinct functions (e.g. *transferase* vs. *hydrolase* [4]) in different oligomeric states.

To fine tune these physiological functions - Homomerization also provides a structural basis for allostery, a phenomenon referring to regulation at a distance from the active site. Historically, two models have been proposed for the allosteric regulation in multimeric enzymes: the sequential *KNF* model of Koshland, Nemethy, and Filmer [5], and the concerted *MWC* model of Monod, Wyman, and Changeux [6] (Figure 1.1). These models explain the regulation seen in homomeric enzymes by an equilibrium between two major conformational states, a low-affinity (inactive) form and a high-affinity (active) form, interchanged upon ligand binding through a concerted or sequential conformational change at the individual subunit level. These models, and their differences, are well discussed in the literature and will not be developed further in this thesis. However, an important postulate of these models is that the multiplicity of the quaternary structure does not change during the allosteric transition. This is not the case with the *morpheein model of allostery*, proposed by Jaffe in 2005 [7], which implies a dynamic behavior of the protein not only at the tertiary structure level but also at the quaternary level.

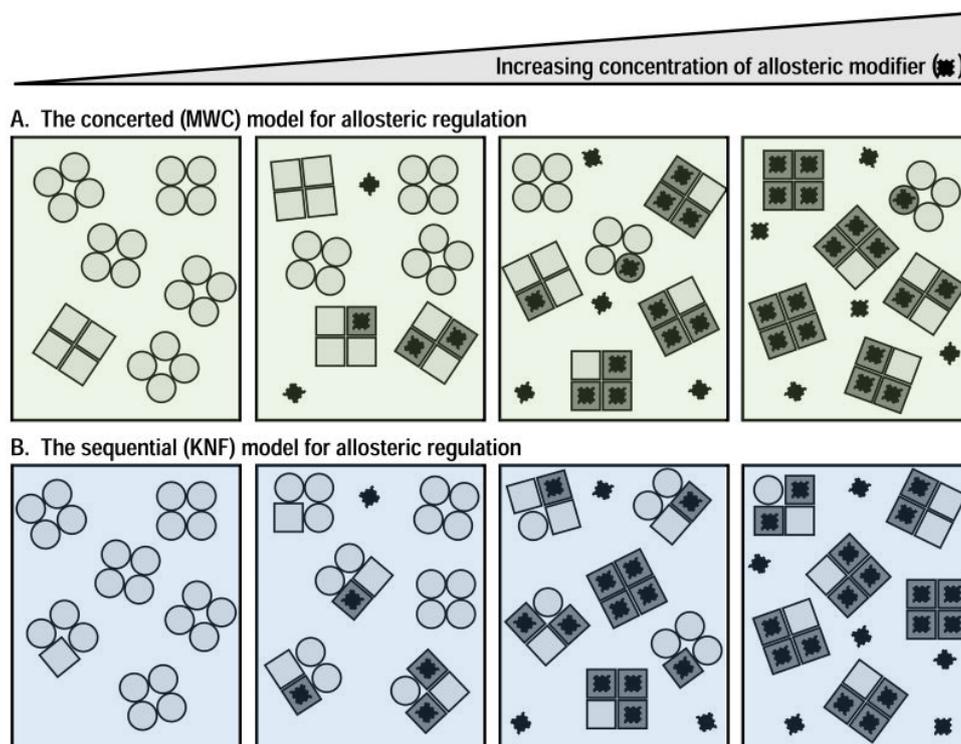


FIGURE 1.1: The MWC and KNF models of allostery. Round subunits are inactive and square subunits are active. The jagged black square represent an allosteric activator and subunits on which it is bound are shaded.

The figure is adapted from reference [8].

The morphein model of allostery

Although the historical models of allostery have been shown to explain the behavior of numerous enzymes, for other allosteric proteins, the structural basis for allosteric regulation lies in the interconversion between quaternary structures of different multiplicity exhibiting different functionality [7]. This is the case of the enzyme phosphobilinogen synthase (PBGS, EC 4.2.24), the prototype morphein.

PBGS was first described as an octamer and thought to dissociate into its component tetramers, dimers and monomers. A number of studies then showed that the enzyme could also be hexameric, which contradicted the classical view in which proteins exhibit a unique stable quaternary structure dictated by an unalterable tertiary structure. The subsequent detailed investigation of Jaffe and coworkers led to the finding that an equilibrium existed between the octameric and hexameric forms of PBGS from many species [9].

In the case of human PBGS (*HsPBGS*), Jaffe's team observed that the enzyme existed as an equilibrium of high activity octamers and low activity hexamers and that the interconversion between these two higher oligomer assemblies required their dissociation into a conformationally flexible dimer (Figure 1.2A and B). As evidenced by the crystallographic structures of the hexamer and the octamer, the dimer could adopt two distinct conformations at the level of the N-terminal arm: a "detached" dimer leading to the formation of the octamer, and a "hugging" dimer leading to the formation of the hexamer (Figure 1.2B and C). It was estimated that total dimer represented about 0.5% mole fraction and that the position of the equilibrium was dependent on pH, substrate concentration and total protein concentration. Small molecules have also been demonstrated to shift the equilibrium by stabilizing one or the other form, thus modulating the activity of the enzyme in such a way that the equilibrium can serve as a basis for allosteric regulation [7–9].

With the example of PBGS in mind, the morphein model of allosteric regulation can be defined as a "concerted dissociative model wherein the allosteric protein exists in a dynamic and reversible equilibrium of alternate oligomers, the interconversion of which requires a conformational change of composite subunits in the dissociated state prior to reassembly to an alternate oligomer" [9] (Figure 1.3). In that case, the allosteric protein is called a **morphein** and the oligomers are called morphein forms. The dissociated state exists in different conformations, each of which dictates the

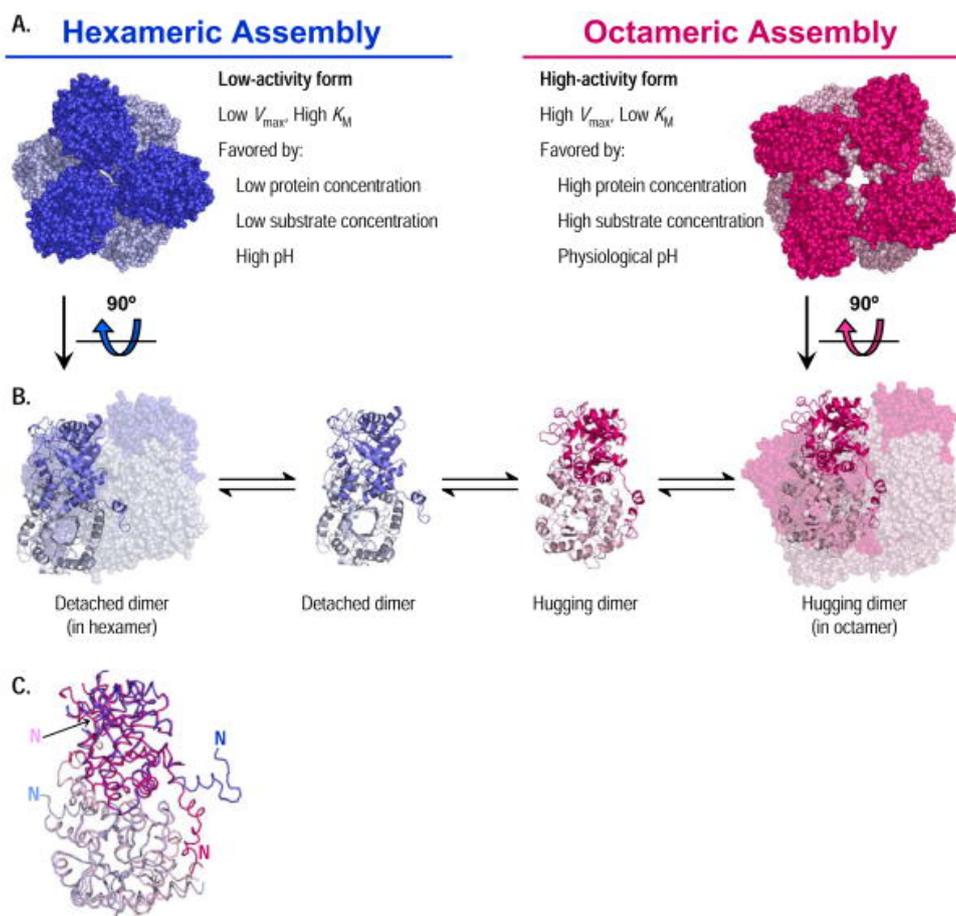


FIGURE 1.2: Human porphobilinogen synthase, the prototype morphein. (A) The low-activity hexamer and the high-activity octamer. (B) The morphein equilibrium model of *HsPBGs* where the hexamer and the octamer are in equilibrium with two structurally distinct dimer conformations. (C) Superimposition of the two dimer conformations, where a shift in the conformation of the N-terminal arm, guiding higher order oligomerization, can be seen. The figure is from reference [8].

assembly to a particular morpheein form, structurally and functionally distinct from the others. The interconversion between morpheein forms is fully reversible, in the sense that it does not require core structure refolding, and the equilibrium can be shifted depending on the buffer conditions or the presence of a ligand. The protein's function is then altered towards that of the morpheein form that is stabilized by the conditions/ligand, which provides a mechanism for allosteric regulation.

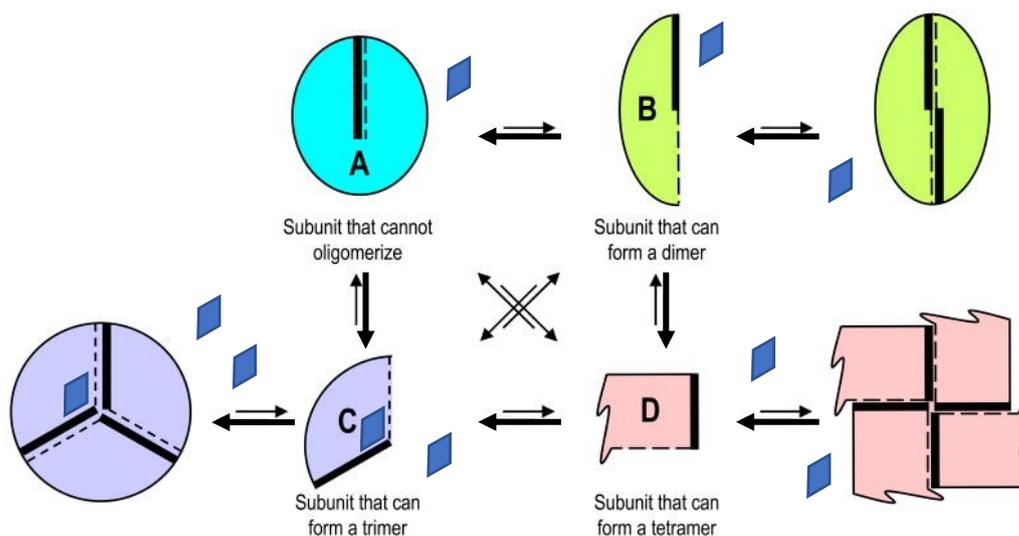


FIGURE 1.3: A schematic representation of an equilibrium between morpheein forms. The morpheein can form a dimer, a trimer and a tetramer from distinct conformations of a monomeric dissociated state. The “rule of engagement” for multimeric assembly is the association of a thick line with a dashed line. The blue trapezoid depicts an allosteric regulator molecule that has the appropriate geometry to only bind the pro-trimer subunit and the trimer, hence shifting the equilibrium in their favor, allosterically stabilizing the trimer and promoting its function. The figure is adapted from reference [8].

1.1.2 What do homo-oligomeric proteins look like?

One of the most widely used experimental techniques to directly access the atomic detail of proteins is X-ray crystallography. However, the latter used alone does not allow an unambiguous study of homo-oligomers because of the difficulty to discriminate crystalline contacts from biological contacts [10]. The combination of crystallographic data with the results of experimental techniques enabling the more detailed characterization in solution of oligomeric proteins of known structure has allowed over the

years to recognize geometric and physico-chemical characteristics of homo-oligomers and their interfaces.

Various interfaces with shared general characteristics - Apart from a few exceptions, biological interfaces tend to show larger contact areas than crystal-packing contacts [11, 12] but they are also characterized by features such as planarity, electrostatic and geometrical shape-complementarity and circularity [10]. Their amino acid composition, characterized by more frequent contacts between amino acids with nonpolar side chains [12, 13], differentiates them from crystalline contacts and solvent-exposed surfaces but varies according to the nature of the homo-oligomeric interaction. In the case of non-obligate, transient contacts, the composition of the interface is closer to that of the rest of the surface, including more polar and charged residues. On the other hand, contacts in obligate homo-oligomers, more intertwined and of larger surface area, exhibit a larger fraction of hydrophobic and aromatic residues giving them an intermediate composition between those of the surface and the protein core [1, 10, 13]. Although presenting recurrent general characteristics, homo-oligomeric interfaces are very diverse and their identification remains a challenge today.

The prevalence of symmetrical homodimers and homotetramers - As reviewed by Goodsell and Olson, oligomeric, especially homo-oligomeric, proteins are prevalent in protein cells [3]. Among them, as illustrated by a study of the natural occurrence of oligomeric proteins in *E. coli*, are found mostly homodimers, then homotetramers. Higher order homo-oligomers are less frequent, hexamers being slightly more observed, and oligomers with an odd number of subunits are relatively scarce. Nevertheless, the virtual totality of these homo-oligomers present a fundamental characteristic: their subunits are symmetrically arranged in space. The most observed symmetry groups are *cyclic groups* (C_n), containing a single axis of n -fold rotational symmetry and allowing for the formation of a ring of n symmetrically arranged subunits, and *dihedral groups* (D_n), containing an axis of n -fold rotational symmetry and n perpendicular axes of two-fold symmetry (Figure 1.4). Although rarer, some proteins playing specialized roles in storage and transport show *cubic symmetries* including tetrahedral, octahedral and icosahedral arrangements.

Goodsell and Olson also listed the various advantages of symmetry: symmetrical oligomers present a better stability of association and allow the construction of assemblies of finite stoichiometry, hence avoiding aggregation and the development

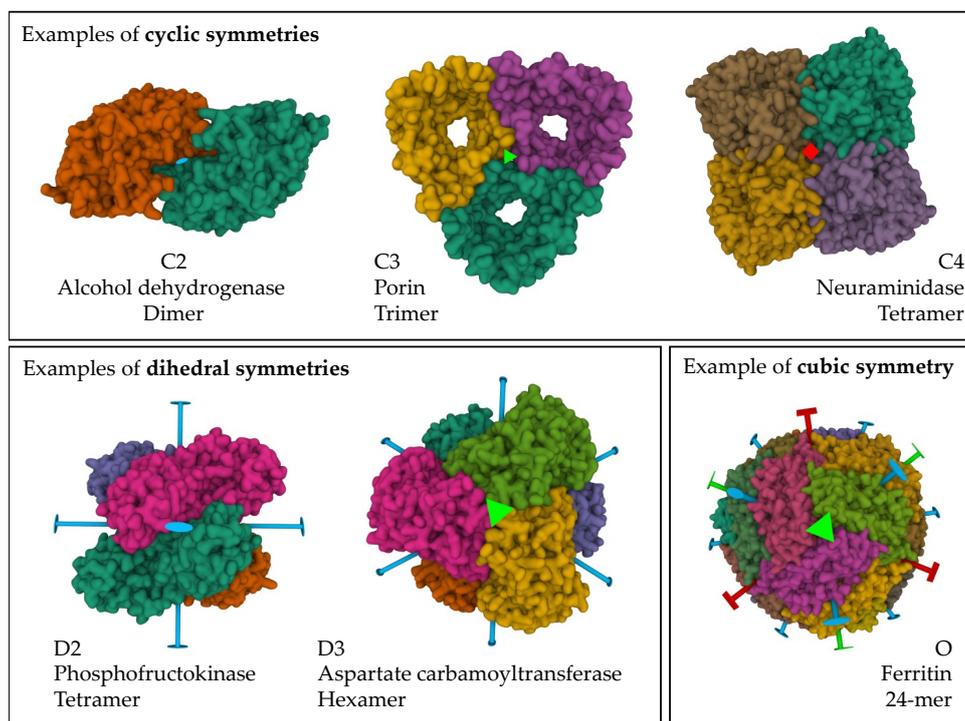


FIGURE 1.4: Example of crystallographic point group symmetries. Two-fold axes are depicted in blue, 3-fold in green and 4-fold in red. O = octahedral. The PDB accession codes for structures are the following from left to right and top to bottom: 1E3L, 1PRN, 1BJI, 4I36, 2RGW, 1GWG. The figure was adapted from reference [3].

of disease conditions, like Alzheimer's and Parkinson's diseases where open-ended protein polymers are formed from soluble monomers and lead to the precipitation of insoluble amyloid fibrils damaging nerves and brain tissues [14]. Folding of symmetrical assemblies would also be more efficient than for asymmetrical structures. Moreover, their particular molecular geometry, especially that observed in dihedral symmetry, enables the allosteric transmission of messages from one subunit to the rest of the protein [3].

Asymmetric by function - Even if symmetry is the rule in protein association, asymmetry is nevertheless observed at several levels (Figure 1.5). Homo-oligomers are said to be *locally asymmetric* when individual residues show different conformation when comparing different subunits. This local asymmetry can also extend to the conformation of a subunit, in proteins that present reciprocating mechanisms for example [3]. This would be the case of a homodimer with a subunit in conformation

A, optimized for catalysis, and a subunit in conformation B, ready to receive the substrate. The binding event would induce a conformational change from B to A, which would be reflected in the neighboring subunit, shifting from A to B conformation. On the other hand, homo-oligomers exhibiting *global asymmetry*, in which subunits are not related by any symmetry operation, are scarce [15]. Even though, such assemblies can be biologically relevant, as shown by Swapna *et al.* [16]. The authors studied 11 globally asymmetric homodimers, demonstrating that nature could use asymmetry to perform specialized functions such as 2:1 binding between homodimers and ligands or binding to DNA response elements.

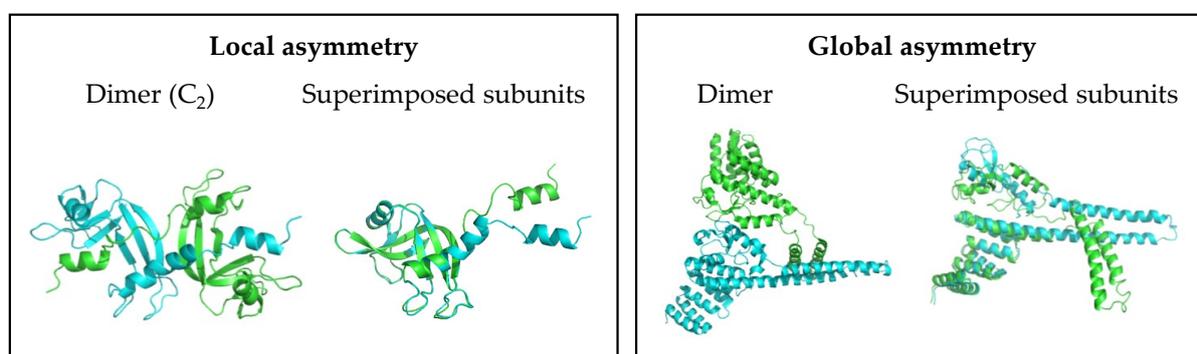


FIGURE 1.5: Difference between local and global asymmetry. The locally asymmetric protein is the bovine pancreatic ribonuclease N-term-swapped dimer (PDB: 1A2W) and the globally asymmetric protein is the murine CHIP-U-box E3 ubiquitin ligase (PDB: 2C2L). The figure is adapted from reference [15].

1.1.3 How does homo-oligomerization arise?

During evolution, proteins that were formerly monomeric have acquired the ability to oligomerize via several mechanisms. The diversity of these mechanisms is such that even homologous proteins, belonging to the same family, may have employed different oligomerization routes and may vary in their oligomeric state. The enzymes studied in this thesis will illustrate this point, but it is also well exemplified, among others, by the glycosyltransferase (GT) family. Depending on the organism they come from, enzymes from the GT family appear as monomers, dimers, or tetramers, and different homomeric binding modes engaging opposite sites on the molecule, for example, have been observed between very similar orthologs [17]. Hereafter are summarized the main evolutionary mechanisms of homo-oligomerization, further explained and exemplified in references [13] and [1].

Substitutions - Amino acid substitution in residues of a protein surface can create a new interface for oligomerization (Figure 1.6). Those mutations, usually involving the replacement of solvent exposed residues by more hydrophobic and larger protruding residues, can lead to the creation or extension of a hydrophobic patch that may favor the formation of an oligomer. Since polar interactions also contribute to the stabilization of oligomeric interfaces [18], mutations may also involve hydrophilic and charged residues creating electrostatic complementarity. Substitutions can also play a role in creating shape complementarity, not only directly at the interface but also at a distance from them. So-called "allosteric mutations" can induce conformational changes elsewhere in the protein leading to the appearance of an oligomeric interface [19].

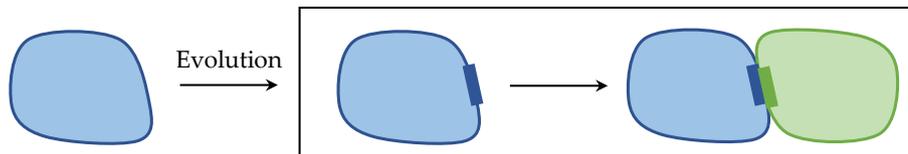


FIGURE 1.6: Dimerization of a monomer through the creation of an oligomeric interface by amino acid substitution.

Insertions and deletions - Amino acid insertions and deletion in the primary sequence of a protein can constitute "enabling regions" for the formation of a new oligomeric interface (Figure 1.7). Indeed, these insertions/deletions lengthen/shorten loops, alpha helices, or beta strands, thus modifying the topology of a surface and allowing interactions with another subunit that would have been impossible in the absence or presence of the inserted/deleted residues [20].

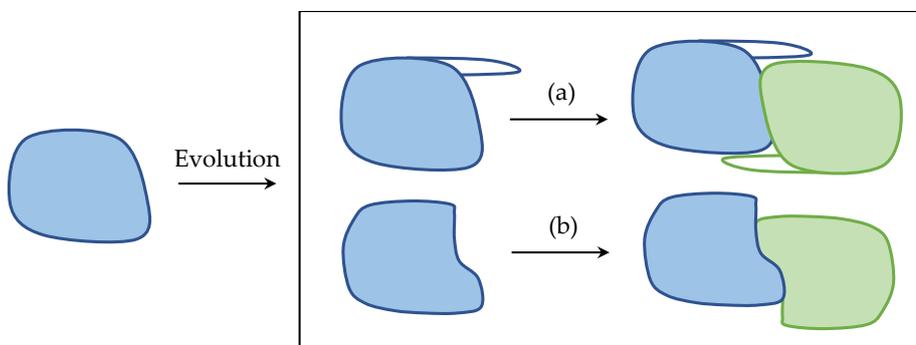


FIGURE 1.7: Dimerization of a monomer through amino acid (a) insertion or (b) deletion.

Structural motifs - Some monomeric proteins have acquired during evolution, via gene duplication events for example, particular structural motifs conferring them the ability to oligomerize (Figure 1.8). There are several well characterized motifs used by proteins to oligomerize, the most common of which being the alpha-helical coiled coil, where two to five α -helices are wrapped around each other into a left-handed helix to form a supercoil. The amino acid composition of the helices composing the coiled-coil is determinant of the oligomerization state, essentially at hydrophobic positions *a* and *d* of the seven-residue repeat $(abcdefg)_n$ [21]. Among other motifs mediating oligomerization, we can also mention the BTP/POZ domains, dimerizing in an intertwined fashion; the foldon domains, leading to the formation of stable trimers; and the ACT domains, which will be described in detail later in this thesis [22].

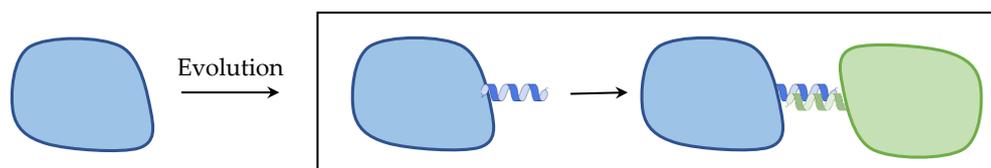


FIGURE 1.8: Dimerization of a monomer through the acquisition of a oligomerization motif.

Domain-swapping - As observed in some proteins that coexist in both monomeric and oligomeric states, monomers can oligomerize by exchanging identical secondary structural elements or domains between them (Figure 1.9). In the neighboring subunit, the exchanged part adopt a position identical to that which it occupied in the monomeric form. This results in the formation of an intertwined higher order homo-oligomer. Domain-swapping is a key concept in this thesis and we describe it in more detail in the following section.

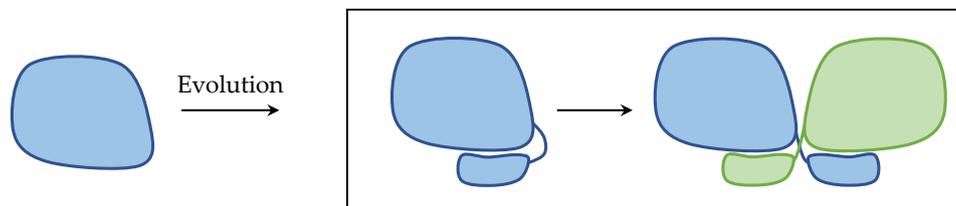


FIGURE 1.9: Dimerization of a monomer through the acquisition of new domain and domain-swapping.

Oligomerization by domain-swapping

The mechanism of oligomerization by domain-swapping was introduced in 1994 by Bennett, Choe and Eisenberg with the determination of the structures of monomeric and dimeric diphtheria toxin (DT) [23]. The authors noticed that DT dimerized through a substantial conformational rearrangement where the entire receptor binding "R" domain of a DT monomer had to be separated from the rest of the monomer by breaking noncovalent interactions, to undergo a 180° rotation and a displacement of at least 65 Å to finally be relocated in the same position in a second DT molecule. This resulted in a DT dimer in which the R domain from each DT molecule was exchanged for the R domain from the other (Figure 1.10A).

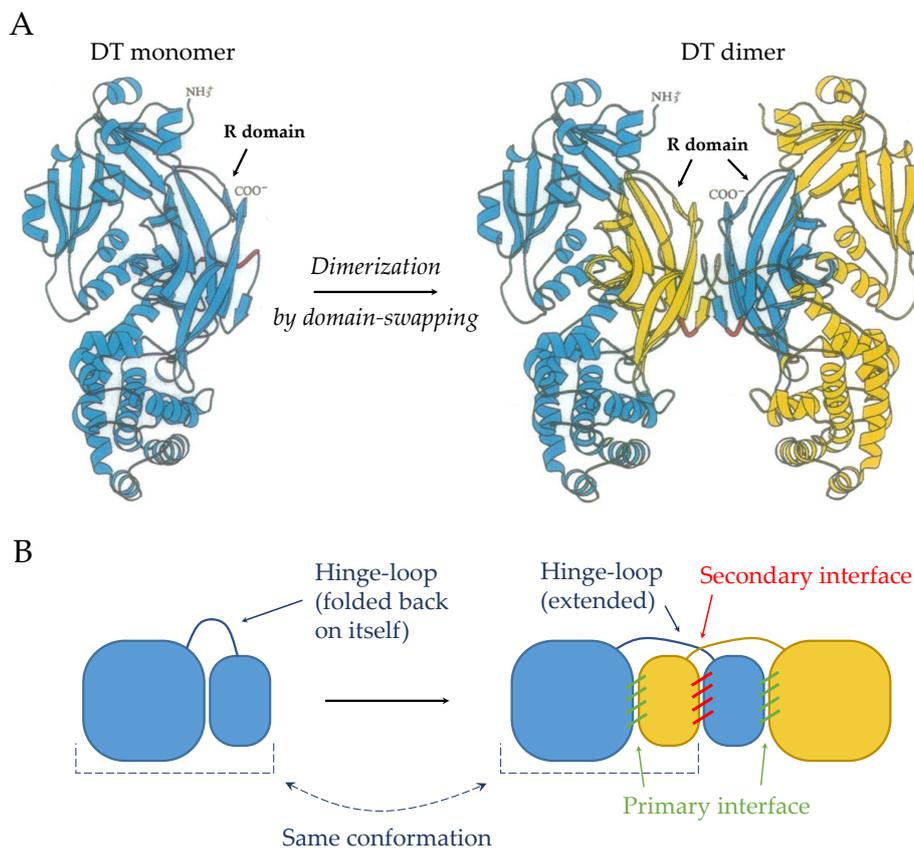


FIGURE 1.10: Domain-swapping in diphtheria toxin (DT). (A) Crystallographic structures of monomeric and dimeric DT. The figure is adapted from reference [23]. (B) Schematic representations of monomeric and domain-swapped dimeric DT to illustrate the determinant features of domain-swapping.

As illustrated with the schematic representations of DT structures in Figure 1.10B, the conformation of each subunit in the domain-swapped oligomer is identical to that of the monomer, except for one part of the molecule: the *hinge-loop*. The hinge-loop is the linker region that connects the exchanged domain to the rest of the molecule and the only part that undergoes a conformational change upon swapping: it is usually a loop or a turn, folded back on itself in the monomer and adopting an extended conformation in the oligomer.

The hinge-loop is the major determinant of domain-swapping as its structural and dynamic characteristics govern the equilibrium between the monomer, which has to open, and the domain-swapped oligomer. Its flexibility, either controlled by its length or amino acid composition, plays a significant role in domain-swapping propensity. As shown in various studies, insertions/deletions and substitutions of defined residues, especially prolines, can affect the oligomeric equilibrium [24, 25].

Another determinant of the oligomeric equilibrium is the formation of new interfaces in the domain-swapped form which are not present in the monomer, called *secondary interfaces* (as opposed to *primary interfaces*, the interfaces for which inter-protein interaction are identical to those present in the monomer). Studies have shown via site-directed mutagenesis that alteration of these secondary interfaces could change the propensity of proteins to domain-swap [26–28].

In 2012, about 60 domain-swapped¹ protein structures had been characterized [25]. Barnase (Figure 1.11A) belongs to the latter, and the resolution of its domain-swapped trimeric structure proved that domain-swapping was not only a dimerization mechanism but could also lead to the formation of higher-order cyclic oligomers [29] (Figure 1.11B).

It was also shown that open-ended oligomers (Figure 1.11C) could be the consequence of oligomerization by domain-swapping, leading to the formation of disease-causing fibrillar aggregates. However, besides this unfortunate feature, domain-swapping was also shown to be involved in the modulation of protein function

¹The term "domain-swapped" refers to proteins for which the structures of both the monomer and the oligomer are available. If only a domain-swapped structure is known, then the protein is a "candidate" for domain-swapping or "quasi" domain-swapped if one of the structures is from a homologous protein.

[25, 30]. In the α -isopropylmalate synthase of *M. tuberculosis*, for example, domain-swapping enables the feedback allosteric regulation by leucine, as an allosteric binding site is created at the primary interface, between the exchanged regulatory domain and the rest of the other subunit [31].

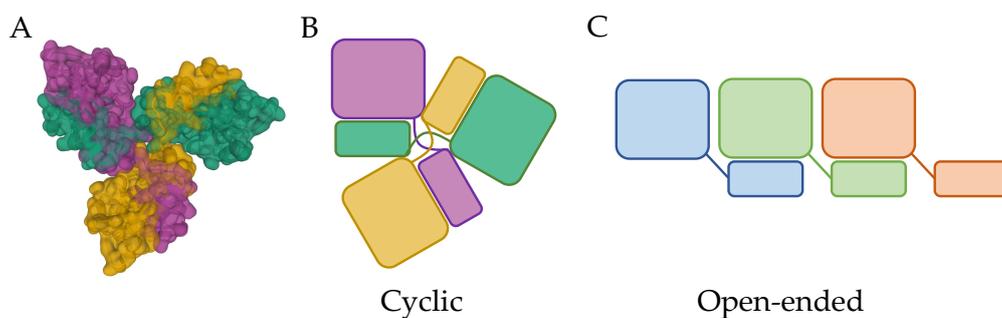


FIGURE 1.11: Domain-swapping in higher-order oligomers. (A) Bar-nase domain-swapped trimer (PDB:1YVS). (B) Schematic cyclic domain-swapped trimer. (C) Schematic open-ended domain-swapped oligomer.

1.1.4 Why is it interesting to study the oligomerization of a protein?

As developed above, homo-oligomers afford several important structural and functional benefits for the cell, whether eukaryotic or prokaryotic. In some cases, homo-oligomerization is also associated with aggregation and amyloid fibril formation in human neurodegenerative diseases. These reasons led to the emergence of the concept of homomeric disruption as a new therapeutic strategy [32]. In this context, the extensive characterization of protein quaternary structures and their equilibrium dynamics is providing valuable support to drug design.

Targeting the self-association of proteins offers several advantages compared to the usual strategy of inhibition through active site occupation, as summarized in reference [2]. For example, this approach allows to **extend the druggability** to proteins that lack an active site, such as transcription factors, or to **avoid compensatory mechanisms** usually encountered with active site inhibitors (increased substrate synthesis, target overexpression, mutations in the catalytic site,...). Disrupting homomeric interfaces can also lead to **protein degradation** through the promotion of protein misfolding or exposure of residues that may be ubiquitinated and recognized by the proteasome, and may present the advantage of **substoichiometric inhibition**, where only a single molecule is required to disrupt a multi-subunit complex.

In this thesis, we are mainly interested in the **increased selectivity** conferred by homo-oligomers as therapeutic targets. In contrast to active sites, oligomeric interfaces within proteins of the same family tend to show little conservation, and as mentioned earlier, even very similar enzymes may exhibit different oligomeric states. This property can therefore be used to precisely target a given enzyme among its close homologs and thus avoid deleterious off-target effects due to multi-enzyme inhibition.

The homo-oligomeric disruption approach is particularly interesting for the development of antibiotics. Classically, such treatments are designed to target enzymes specific to prokaryotic cells so as not to cause harm to the human host through the inhibition of a shared enzyme [33, 34]. Exploitation of weakly conserved oligomeric interfaces or species-specific oligomerization states of a given enzyme thus expands the pool of potential targets for the treatment of infectious diseases. To date, as outlined in reference [2], the disruption of the self-assembly of proteins from various pathogens such as *Staphylococcus aureus*, *Leishmania infantum*, *Vibrio cholerae*, and *Trypanosoma cruzi* has been the subject of several studies. In this thesis, we study a target from *Mycobacterium tuberculosis*, the causative agent of tuberculosis.

1.2 Tuberculosis

1.2.1 A treatable, but ever-present, enduring enemy

Tuberculosis (TB), the leading cause of death worldwide from a single infectious agent after COVID-19, is a contagious disease caused by *Mycobacterium tuberculosis* (Figure 1.12). The pathogen primarily affects the lungs, where it is phagocytosed by macrophages after entering the host by the airborne route. As estimated by the World Health Organization (WHO) and recently confirmed in a study of Cohen and coworkers [35] a quarter of the world's population is asymptotically infected with *M. tuberculosis*. This indicates a large reservoir of individuals at risk of developing the disease, since it is estimated that 5-10% of latent infections evolve into active TB at some point, sometimes years after the primary infection [36].

TB is an ancient scourge. The first historical record of *M. tuberculosis* is thought to date back three million years with the infection of early hominids in East Africa by a progenitor of the bacterium. Skeletal lesions characteristic of the disease have been

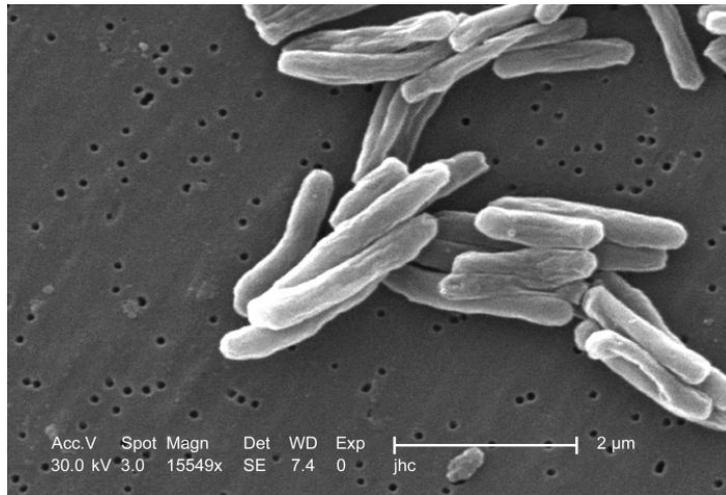


FIGURE 1.12: Scanning electron micrograph of *Mycobacterium tuberculosis* under x15 549 magnification [37].

found in Egyptian mummies dating from 2400 BC and writings testifying to the incidence of the disease, described under many names, punctuate the course of history from Ancient Greece ("*phthisis*") to the 19th century ("*White plague*", "*Consumption*", "*Captain of all these men of Death*") through the Middle Ages ("*King's evil*") [38].

It was only in 1882 that the pathogen was identified and isolated by Robert Koch, during a time where TB killed one out of every seven people living in the United States and Europe [39, 40]. Nowadays, fortunately, the prevalence of the disease is drastically lower, thanks to the continuous battle against TB since the first milestone set by Koch. Still, TB remains a major public health problem, with an estimated ten million new cases and 1.5 million deaths in 2020. With a combined strategy based on improving treatment, diagnosis and prevention, the World Health Organization (WHO) aims to end the global TB epidemic by 2035 by reducing the incidence rate by 90% and mortality by 95% [41].

The TB issue is even more topical in light of the recent health crisis related to the COVID-19 pandemic, which has reversed years of progress in the fight against the disease and put global TB targets mostly off-track. Sanitary restrictions all over the world reduced the access to diagnosis and treatment, resulting in the first year-over-year increase in TB deaths (by 5.6%) since 2005 and the total number of deaths returning to 2017 levels. TB incidence could also increase globally in the next years [41].

While restoring access to and provision of essential TB services is the priority to counteract the impact of COVID-19, increased research and innovation is part of the strategy to tackle antimicrobial drug resistance, one of the major burdens of TB care and public health threat.

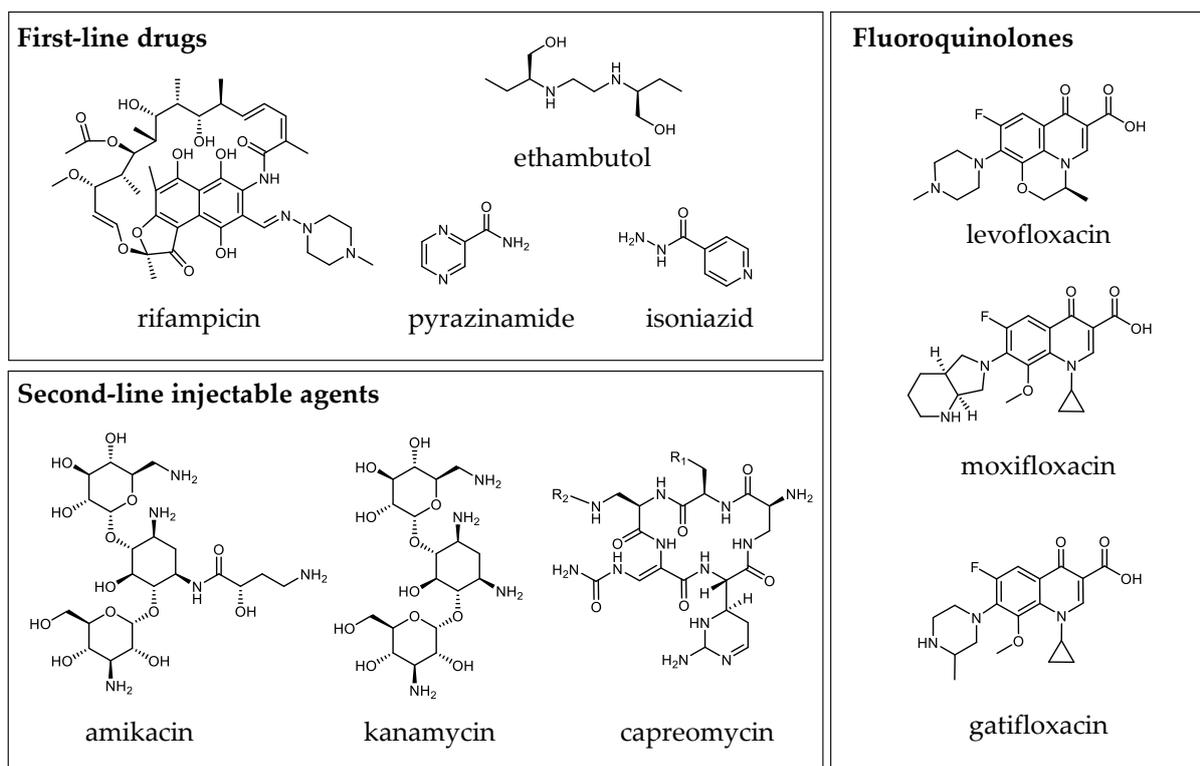


FIGURE 1.13: Two dimensional structures of anti-TB drugs.

In the case of drug-susceptible pulmonary tuberculosis, the patient is treated with a six-month regimen of four first-line antibiotics. The initial phase of treatment lasts two months and consists of daily isoniazid, rifampicin, pyrazinamide and ethambutol (Figure 1.13). If there is no resistance at this stage, treatment with isoniazid and rifampicin is continued for a further four months. However, some strains are partly resistant to this treatment (*multi-drug resistant or rifampicin resistant TB, MDR/RR-TB*) and greatly complicate the healing process. Eliminating RR-TB and MDR-TB strains requires longer, more expensive therapies with more side effects. In such cases, patients are treated for at least 9 months (up to 20 months) with second-line compounds, including fluoroquinolones and the injectable drugs amikacin, capreomycin and kanamycin. The length, complexity and toxicity of the treatment makes compliance difficult and results in much lower success rates compared to drug-susceptible

TB (50-75% compared to 85%). In addition, other strains, known as *extremely drug-resistant strains (XDR-TB)*, are MDR-TB strains that also resist to any fluoroquinolone and at least one of the three second-line injectable drugs, hence leaving few treatment options [41, 42].

In 2020, nearly 133,000 MDR/RR-TB cases and 26,000 pre-XDR-TB or XDR-TB cases were detected out of the 2.1 million people diagnosed with pulmonary TB and tested for resistance. These figures demonstrate the continuous need for progress in the search for new options in the TB treatment pipeline [41].

1.2.2 Targeting the amino acid metabolism of *M. tuberculosis*

One strategy implemented for the development of new anti-tuberculosis compounds is rooted in a target-based approach, where the focus of the drug-design effort is on a particular biomolecule involved in the disease. The sequencing of *M. tuberculosis* genome by Cole and coworkers in 1998 has enabled the identification of genes involved in the growth, survival and virulence of the pathogen [43–47].

Many of these genes encode for enzymes of the amino acid metabolism, as *M. tuberculosis* like many pathogens depend on these building blocks for protein synthesis and as a carbon source. Amino acids are the core of the host-pathogen interaction, acting as the mediators of metabolic cross-talk shaping the outcome of infection. Pathogens often rely on their host for the supply of these nutrients but the host can defend itself by a starvation process, making energy resources unavailable in the intracellular niche [48, 49] This is for example the case of the competition for histidine in *M. tuberculosis* where the host up-regulates its histidine catabolizing enzymes to starve the bacillus of intracellular free histidine. *M. tuberculosis* then highly depends on its *de novo* histidine biosynthesis to evade the host immune response and continue to proliferate [50]. Catabolic pathways can also be exploited to this end, as exemplified by the dehydrogenation of glutamate to 2-oxoglutarate in *M. tuberculosis*. The reaction, catalyzed by the glutamate dehydrogenase (GDH), produces buffering ammonium that is thought to offer protection against macrophage acidification [51]. These examples among others illustrate how enzymes involved in amino acid metabolic pathways can play an essential role in the survival of *M. tuberculosis* in macrophages and therefore serve as potential targets for anti-tubercular drug development. Reference [52] reviews the key enzymes involved in 17 amino acid pathways required for the growth

and pathogenesis of *M. tuberculosis* and describes the related inhibitors being investigated as potential new anti-tubercular drugs. In this thesis work, we are interested in the biosynthesis of L-serine and more specifically in the third and last enzyme of the pathway.

Focus on the phosphorylated L-serine pathway

In humans, L-serine (L-Ser) is considered a non-essential amino acid due to the fact that it can be synthesized *de novo* by the organism in parallel to its intake from diet, protein and phospholipids turnover, and conversion from glycine. L-Ser is a central nutrient for cell proliferation, growth, differentiation and function. It is the precursor of sphingolipids and phosphatidylserine (structural lipids of cell membranes and nervous tissues), methylenetetrahydrofolate (key metabolite source of one-carbon units for methylation reactions and nucleotide synthesis) and the neurotransmitter molecules glycine and D-serine. It plays an essential role in the development and functioning of the central nervous system. Patients who are deficient in this amino acid suffer from severe neurological disorders [53, 54].

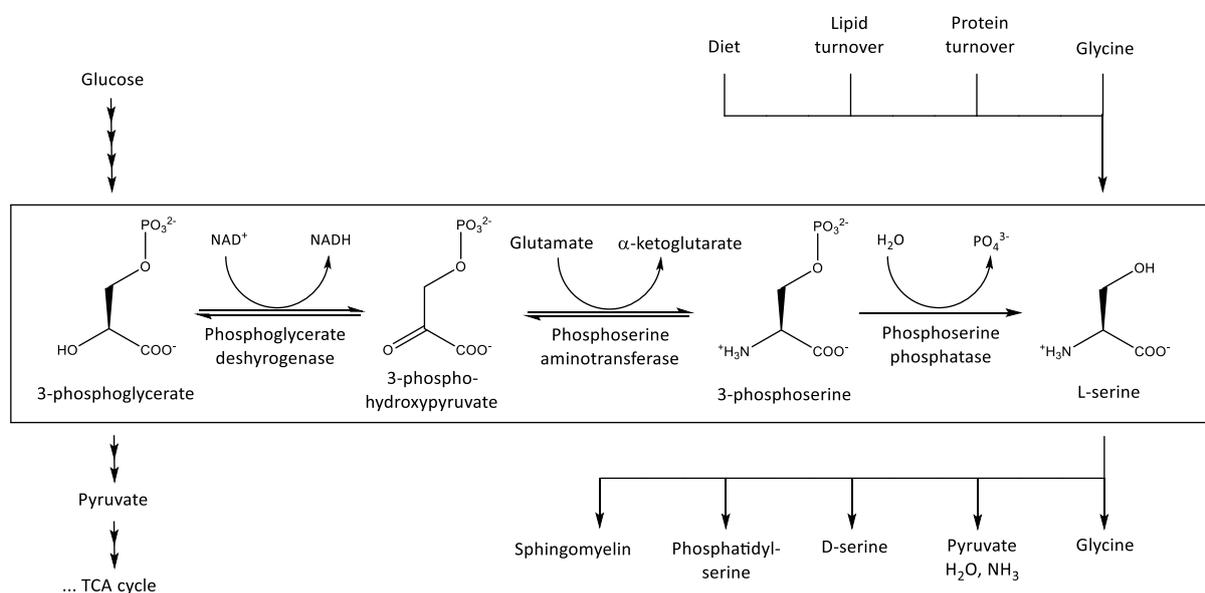


FIGURE 1.14: The phosphorylated pathway of L-serine biosynthesis

The major source of L-Ser is a 3-step biosynthetic route called the "phosphorylated pathway", a cytosolic pathway that starts from the 3-phosphoglycerate generated by glycolysis (Figure 1.14). This glycolytic intermediate is first converted to 3-phosphohydroxypyruvate by a phosphoglycerate dehydrogenase (PGDH, EC 1.1.1.95)

through a reversible NAD^+ -coupled oxidation. Then, a glutamate-linked transamination catalyzed by a phosphoserine aminotransferase (PSAT, EC 2.6.1.52) reversibly transforms the 3-phosphohydroxypyruvate into 3-phosphoserine. The latter is finally dephosphorylated to L-serine by a phosphoserine phosphatase (PSP, EC 3.1.3.3) in an irreversible Mg^{2+} -dependent step.

The phosphorylated pathway is conserved among animals, plants, and bacteria [55], among which the TB bacteria *M. tuberculosis* where the three enzymes are respectively encoded by the genes *serA1* (Rv2996c) and *serA2* (Rv0728c) for PGDH, *serC* (Rv0884c) for PSAT and *serB1* (Rv0505c) and *serB2* (Rv3042c) for PSP [43]. Among these, the genes encoding *MtSerA1*, *MtSerC* and *MtSerB2* have been identified as required for the growth of the pathogen *in vitro*. In addition, they are also crucial for its intracellular survival as it has been reported that *M. tuberculosis* does not uptake L-Ser from the host macrophages. [44, 56]. These observations make the phosphorylated L-serine pathway an interesting target for the design of new anti-tuberculosis compounds. Furthermore, subtle or more substantial structural differences between the human and mycobacterial enzymes play in favor of the possibility for a selective inhibition. The features of *MtSerA1*, *MtSerC* and *MtSerB2* and compounds inhibiting their activity are reviewed in reference [57]. Hereafter we briefly summarize the features aligned with the scope of this thesis for the first two enzymes and focus in detail on *MtSerB2*, our main subject, in the next section.

The phosphoglycerate dehydrogenase *MtSerA1* - The first step of the phosphorylated pathway is catalyzed by *MtSerA1*, a key enzyme in the regulation of L-Ser biosynthesis in *M. tuberculosis*. Three co-crystal structures of the enzyme in tetrameric form with various ligands are available in the PDB (1YGY, 3DC2, 3DDN) [58–60]. Each subunit is composed of a catalytic domain, an ASB noncatalytic substrate binding regulatory domain and an ACT amino acid binding regulatory domain. The particularity of *MtSerA1* crystallographic tetramer (Figure 1.15A) is its global asymmetry: no symmetry operation relates the 4 subunits together. This unusual feature originates from the fact that the asymmetric unit contains a dimer, reconstituting the tetramer by C2 symmetry, but that this dimer is itself asymmetric. The two subunits have indeed different conformations and hence do not superimpose as the ASB-ACT part is rotated 160° with respect to the catalytic domain (Figure 1.15B).

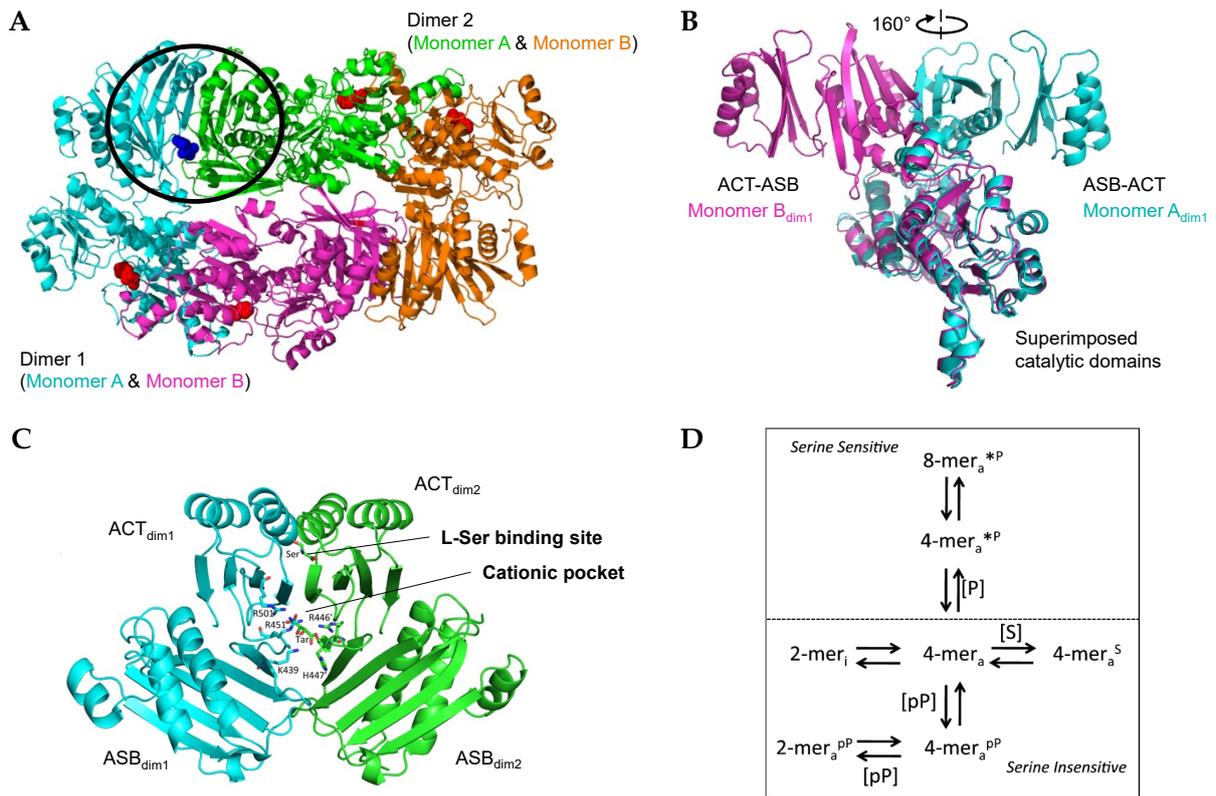


FIGURE 1.15: (A) Structure of *MtSerA1* entry 1YGY with each subunit displayed in a different color. Highlighted by the oval is one of the two ASB-ACT domain interface (shown in panel C). (B) Superimposition of the two subunits composing the dimeric asymmetric unit. (C) ASB-ACT domain interface highlighted in panel A and positive residues that bind tartrate (Tar). (D) Equilibrium model for *MtSerA1*. P = phosphate, pP = polyphosphate, S = substrate. Superscripts = bound ligands. Asterisk = serine sensitive alternative conformation. a = active. i = inactive. Figures from panels A, C and D are from reference [61].

The two C2 related dimers interact through their ACT domains, at the interface of which an L-Ser binding site is present and enables the allosteric feedback regulation of *MtSerA1* catalytic activity (Figure 1.15C). Another effector binding site is made up of cationic residues at the interface of the ACT and ASB domains of two dimers. The substrate, 3-phosphoglycerate, can interact with this pocket to elicit substrate inhibition but it has also been suggested that phosphate and short polyphosphates bind there. Thanks to kinetic studies and size exclusion chromatography experiments, it was shown that these negatively charged ions could make *MtSerA1* more or less sensitive to L-Ser feedback inhibition, probably through a mechanism of allosteric quaternary structure dynamics in line with the *morpheein* model (Figure 1.15D). Moreover, this feature makes *MtSerA1* very special because other PGDHs with the same primary architecture (CAT-ASB-ACT) are not inhibited by L-Ser even in the presence of phosphate. As hypothesized by Grant and its coworkers, these unusual properties could confer a unique metabolic advantage to intracellular pathogenic *Mycobacteria* [61–65].

The phosphoserine aminotransferase *MtSerC* - The second enzyme of the phosphorylated pathway, *MtSerC*, has not been described in such detail in the literature. A recent paper showed that *M. tuberculosis* mutant lacking *MtSerC* ($\Delta serC$) was indeed not able to grow without L-Ser supplementation and that this serine auxotroph was strongly attenuated for intracellular survival in THP-1 macrophages, thereby confirming the potential of the enzyme as a drug target [56]. Two structures were determined by Coulibaly *et al.* (PDB 2FYF and 3VOM), including one at 1.5 Å resolution co-crystallized with the pyridoxal 5'-phosphate (PLP) cofactor [67, 68]. *MtSerC* is composed of two α/β domains and, like human and *E. coli* PSATs, is an obligate symmetric (C2) homodimer in which the two active sites are located at the dimer interface (Figure 1.16A). However, if the three enzymes show the typical fold of the aspartate amino-transferase family of PLP-dependent enzymes and highly conserved PLP binding modes, additional contacts across the dimer interface of *MtSerC* restrict the access to the active sites (Figure 1.16B). The substrate binding region also differs in the residues that bind the anionic phosphate or γ -carboxylate moieties [66].

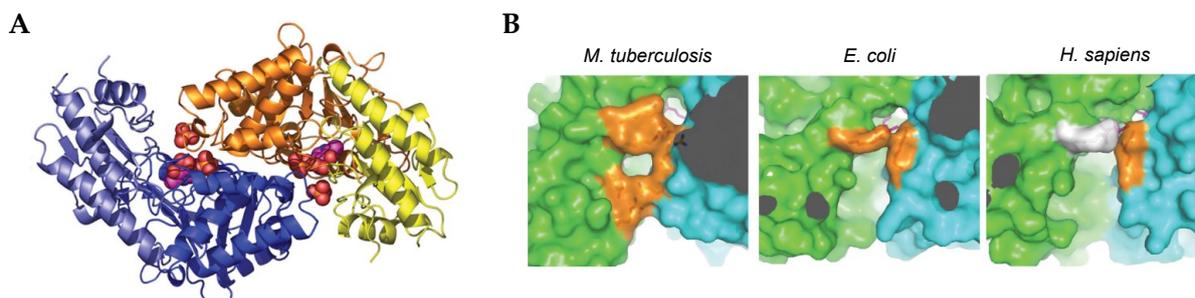


FIGURE 1.16: (A) Structure of *MtSerC* dimer entry 2FYF. Large domains: darker shades, small domains: lighter shades. The active sites are indicated by spheres representing the PLP cofactor and sulfate ion. (B) Molecular surface defining the active site groove of the PSAT enzymes of *M. tuberculosis*, *E. coli* and *H. sapiens*. Figures are from reference [66].

1.3 The phosphoserine phosphatase *MtSerB2*

1.3.1 A suspected secreted virulence factor

In a 2016 study, Shree *et al.* suggested that the essential metabolic enzyme *MtSerB2* could be also used by *M. tuberculosis* as a secreted effector protein to promote its survival and evade the host immune system [69]. A few years earlier, similar properties had been highlighted for the phosphoserine phosphatase SerB653 from *Porphyromonas gingivalis*, a periodontal pathogen infecting gingival epithelial cells [70, 71].

Through western blot analysis, they first showed that *MtSerB2* was detected in the cytosol of THP-1 macrophages infected with *M.tuberculosis* but also of macrophages obtained from bronchoalveolar lavage of TB patients. Then, the results of immunofluorescence experiments on infected THP-1 macrophages and on THP-1 macrophages challenged with exogenously added *MtSerB2* supported the fact that the enzyme was probably secreted upon infection and demonstrated that it elicited cytoskeleton rearrangements by colocalizing with host tubulin (Figure 1.17).

The authors then investigated the interaction of *MtSerB2* with components of the cytoskeleton by studying protein-protein interactions via pull-down assays. Cofilin, actin and α/β tubulin were identified as interaction partners. Moreover, *MtSerB2* was shown to mediate the dephosphorylation of phospho-cofilin both in *MtSerB2*-treated THP1-macrophages and *in vitro*. A previously characterized inactive mutant D341N/D345N failed to reproduce the same effect thereby confirming the result.

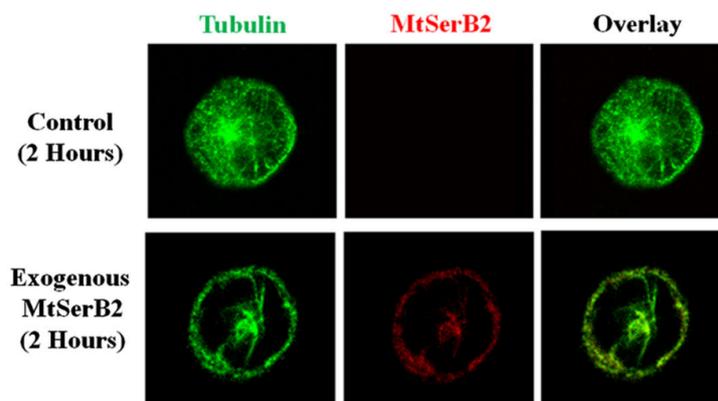


FIGURE 1.17: Confocal experiments showing the localization of exogenously added *MtSerB2* and alteration in tubulin dynamics in THP-1 macrophages at 2h time point. Cells were stained for α/β tubulin with Alexa fluor-488 (green) and *MtSerB2* with Texas Red (red). The figure is from reference [69].

MtSerB2 was also shown to affect the expression of genes that regulate actin dynamics by quantitative RT-PCR, to physically interact with anti-apoptotic proteins involved in the modulation of actin dynamics and to suppress the expression of the pro-inflammatory cytokine interleukin 8 (IL-8) probably via the dephosphorylation of nuclear factor $\text{NF-}\kappa\text{B}$ p65 and p38 mitogen activated protein kinase (MAPK p38), while the catalytically inactive mutant failed to do so.

Finally, the authors demonstrated the inhibitory properties of clofazimine (CFZ), an antibiotic used against MDR and XDR-TB, on *MtSerB2* via docking, dephosphorylation kinetics studies, isothermal titration calorimetry (ITC) and comparative inhibition assays involving other phosphatases. CFZ was then shown to prevent the effects of *MtSerB2* on host cell and proteins: in the presence of the inhibitor, the enzyme failed to dephosphorylate MAPK p38 and cofilin, to elicit cytoskeletal rearrangements in THP-1 cells and lost the ability to down-regulate IL-8 expression.

Based on these results, Shree *et al.* suggested that *MtSerB2* had gained new functions apart from its metabolic role and could be thought of as a *modulin*. As summarized in Figure 1.18, the enzyme would be secreted into the macrophage cytosol by *M. tuberculosis* where it would directly interact with host proteins and even dephosphorylate some of them. On one hand, the dephosphorylation-mediated activation of cofilin and the modulation of genes involved in actin dynamics would induce cytoskeletal

rearrangements, a strategy known to be employed by intracellular pathogens to promote their survival, replication and dissemination [72]. On the other hand, the dephosphorylation of MAPK p38 and NF- κ B p65 would result in the down-regulation of the chemokine IL-8, a chemoattractant and activator of neutrophils granulocytes. In the absence of this inflammatory mediator ("*chemokine paralysis*"), the host may no longer be able to detect the presence of the pathogen and direct the immune cells for its elimination [73].

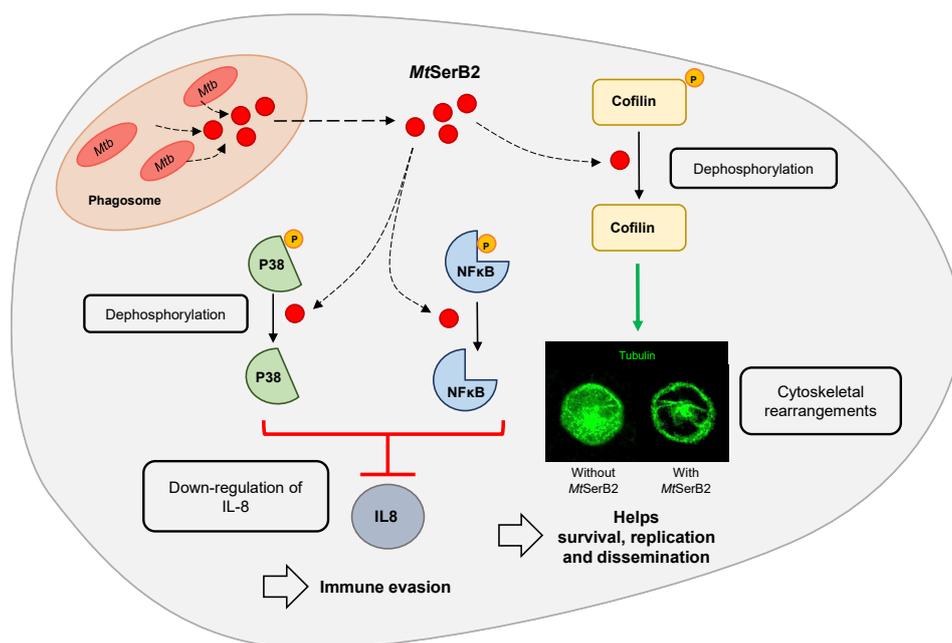


FIGURE 1.18: Schematic representation of the action of *MtSerB2* as a virulence factor of *M. tuberculosis* in macrophages. The figure is adapted from references [69] and [74].

... and therefore a potential moonlighting protein

The work of Shree *et al.* indicates that *MtSerB2* appears to have multiple functions: an enzymatic metabolic function in *M. tuberculosis* cytoplasm (the dephosphorylation of phosphoserine in L-Ser biosynthesis) and effector functions in the host cytoplasm (the dephosphorylation of host phospho-proteins, the interaction with cytoskeleton components and anti-apoptotic proteins, and the modulation of the expression of various genes). By definition, this suggests that *MtSerB2* could therefore be classified as a "**moonlighting protein**", a protein in which one polypeptide chain performs more than one physiologically relevant biochemical or biophysical function [75]. These

functions include the catalysis of metabolic reactions, acting as a cell surface receptor, mRNA binding, the regulation of DNA transcription, binding to other proteins, acting as a chaperone... The shift in function is related to more or less important structural changes, from a conformation change to an alteration of the oligomerization state, exposing hidden residues or structural motifs requested for the alternative function. These structure/function modifications can be triggered by environmental stress, changes in cellular concentration of substrates or specific ligands, or change in cellular localization upon secretion for example.

A recently described example of moonlighting protein used as an effector by *M. tuberculosis* to disrupt host immunity is the enoyl-CoA hydratase A1 (EchA1, Rv0222) [76]. In the bacterial cytoplasm, EchA1 plays a key role in lipid metabolism, catalyzing the hydration of *cis*-unsaturated fatty acid. Through extensive biochemical analyses, Wang *et al.* showed that EchA1 was also found in the host cytoplasm, and that its ubiquitylated form was able to inhibit TRAF6 signaling, thus preventing downstream pro-inflammatory responses through the impairment of NF- κ B activation [77]. As for *MtSerB2*, how EchA1 is transported from the bacterial cytoplasm into the host cytoplasm is yet to be determined.

1.3.2 A structurally unusual haloacid dehalogenase phosphatase

The PSP catalytic domain

Phosphoserine phosphatases belong to the haloacid dehalogenase (HAD) superfamily, a superfamily of hydrolases catalyzing a variety of functions with diverse substrate specificity. The most represented enzymes in this superfamily catalyze phosphoryl group transfer reactions, including about 20% ATPases and about 79% phosphatases [78, 79].

HAD phosphatases are readily identifiable via alignment of the amino acid sequence thanks to four highly conserved signature motifs where h is an hydrophobic residue and x is any residue [80]:

- **Motif I:** hhhDxDx[T/V][L/I/V]h
- **Motif II:** [S/T]
- **Motif III and IV:** K₁₈₋₃₀[G/S][D/S]_{x3-4}[D/E]

The identification of *MtSerB2* as a HAD phosphatase in this way was the first step carried out by Arora, Yadav and their coworkers in their respective characterizations of the enzyme in 2014 [81, 82].

The motifs described above contain the residues of the catalytic machinery, which are precisely positioned in space through a structural arrangement that is adopted by all members of the HAD phosphatase family. This typical fold is a Rossmann-like (or *Rossmannoid*) fold: a 3-layer $\alpha/\beta/\alpha$ sandwich with a central parallel β -sheet consisting of at least five strands in 3-2-1-4-5 order. The central sheet orients the loops and helices that bear the core residues (Figure 1.19A) [80].

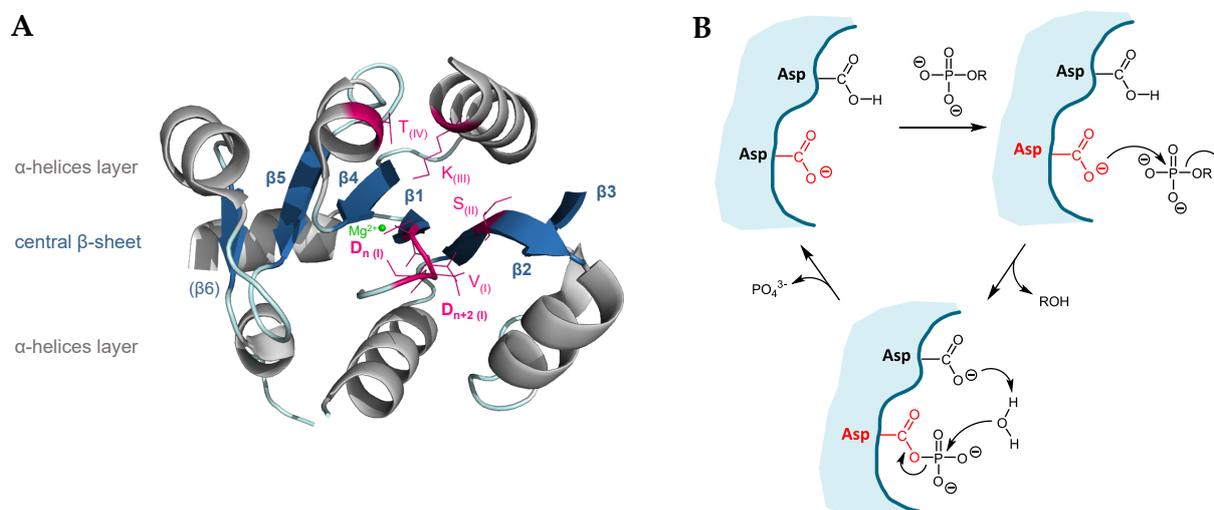


FIGURE 1.19: (A) The typical Rossmannoid fold exhibited by HAD phosphatases (here in human PSP, PDB: 6HYY). (B) General catalytic mechanism of HAD phosphatases. The nucleophilic aspartate residue (D_n) is depicted in red. The figure is adapted from reference [80].

These core residues are involved in positioning the substrate, the cofactor and the catalytic groups required to perform a two-step phosphotransfer reaction. The general reaction mechanism (Figure 1.19B) is well described and starts with the nucleophilic attack of the **first aspartate residue (D_n)** of **motif I** on the phosphoryl group of the substrate. This results in the formation of a phosphoaspartyl-enzyme intermediate and the displacement of the substrate leaving group, immediately protonated by the **second Asp residue (D_{n+2})** of **motif I**. The latter then deprotonates a water molecule that attacks the phosphoryl group to hydrolyze the phosphoester intermediate and release the phosphate ion.

A thorough structural characterization of human phosphoserine phosphatase performed in our laboratory highlights the role of conserved residues in stabilizing the cofactor and reaction intermediates [83]. In addition to their catalytic role, the two **Asp** residues of **motif I** also stabilize the **Mg²⁺ cofactor** on which HAD phosphatases depend. The presence of the Mg²⁺ ion helps the orientation of the substrate (3-phosphoserine in the case of PSPs) relative to the catalytic residues and provides some electrostatic stability by neutralizing the negative charges carried by the phosphoryl group and the Asp residues. The negative charges of the reaction intermediates are stabilized throughout the reaction. The residues **Val of motif I**, present between the two catalytic Asp, **Ser of motif II** and **Lys of motif III** interact mainly with the phosphoryl moiety while the carboxylate moiety establishes H-bonds with a **Thr** residue located before the last conserved Asp of **motif IV**.

Through their respective biochemical characterizations of *MtSerB2*, Arora, Yadav and colleagues proved that the enzyme did possess the functional characteristics of a HAD phosphatase. They verified that point mutants altered at the conserved residues of the active site indeed lost their catalytic activity. The preference for *O*-phospho-L-serine as a substrate was also established, as well as the dependence on the Mg²⁺ cofactor [81, 82].

MtSerB2 sequence and functional characteristics identify it as a HAD phosphatase, so by definition the enzyme should exhibit the typical Rossmannoid fold. However, there is no direct experimental evidence for this as the atomic structure of the enzyme remains undetermined to this day. To still be able to represent *MtSerB2* structure and rationalize the results of its characterization, Arora and Yadav employed homology modeling, an *in silico* modeling approach based on the concept that two homologous sequences form similar protein structures. The best template was identified as *Mycobacterium avium* phosphoserine phosphatase SerB (*MaSerB*), the closest crystallized homolog of *MtSerB2* with 83% sequence identity. This high percentage strongly suggests that the two enzymes adopt very similar folds and that *MtSerB2* structure can be described on the basis of the tertiary structure of *MaSerB* with a high degree of confidence.

The crystallographic structure of *MaSerB* is shown in Figure 1.20 (PDB: 3P96) [84, 85]. Although the asymmetric unit is composed of a monomer, the biological assembly is an intertwined dimer, retrieved in the crystal packing by applying a C2 symmetry

operation (Figure 1.20A and D). As shown in Figure 1.20B, *MaSerB* indeed bears the characteristic Rossmann-like fold of HAD phosphatases, along with other structural features.

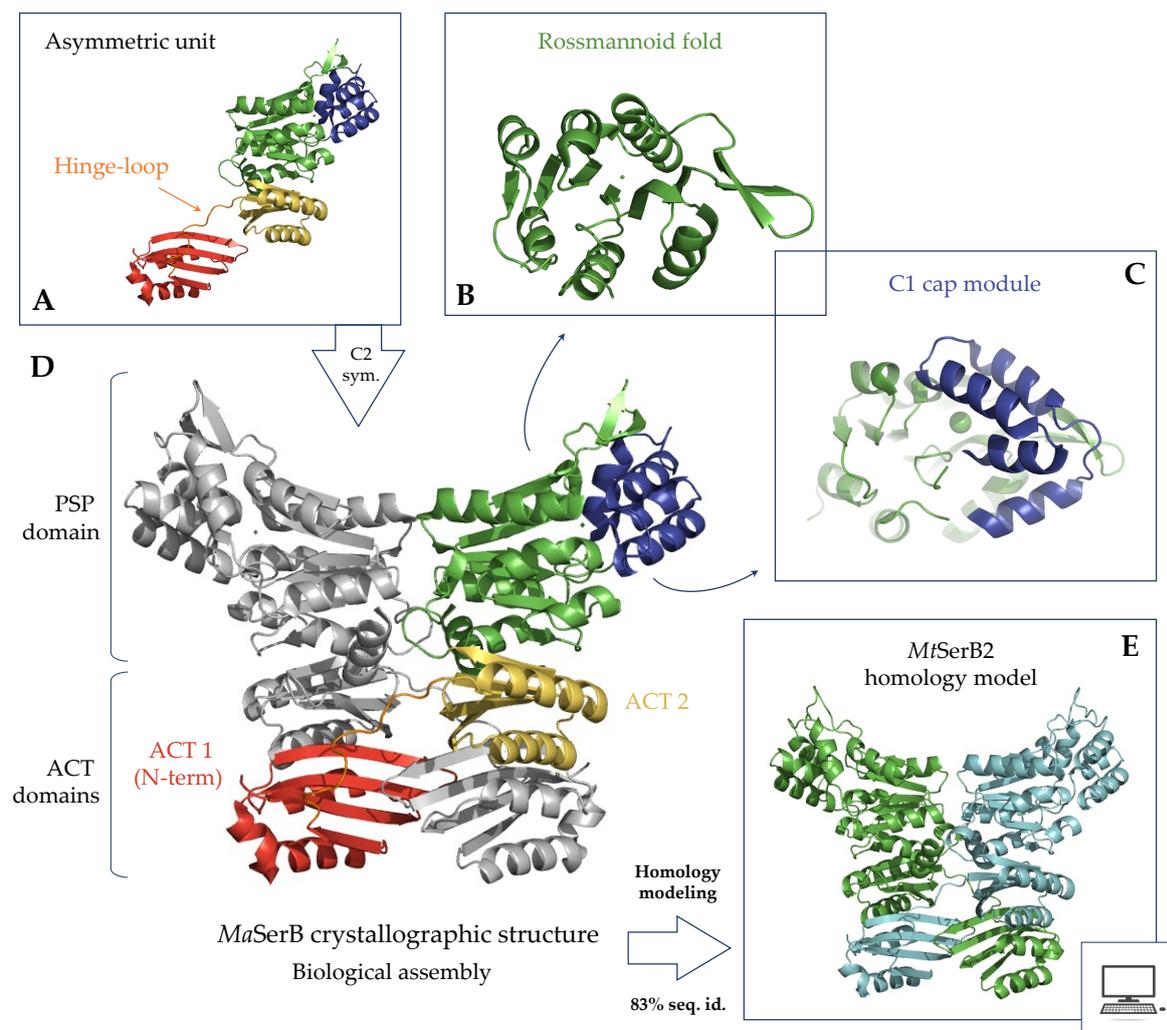


FIGURE 1.20: Architecture of *MaSerB* and *MtSerB2*. (A) Asymmetric unit of *MaSerB* crystallographic structure (PDB: 3P96). (B) *MaSerB* Rossmannoid fold. (C) *MaSerB* C1 cap module. (D) Biological assembly of *MaSerB* crystallographic structure (PDB: 3P96). (E) *MtSerB2* homology model built from *MaSerB* structure.

One of these additional structural features is a tetrahelical bundle inserted between strands β_1 and helix α_1 of the Rossmannoid fold, shielding the active site (Figure 1.20C). This insert is a mobile cap module of type C1, which role is to mediate the access of solvent, substrate and products to the active site during the catalytic cycle. Other types of cap modules, like the loop C0 cap or the more elaborated $\alpha+\beta$ C2 cap are also observed in HAD phosphatases. The type of cap determines the size and the nature of the substrate that can be catalyzed by the enzyme [80]. Throughout this manuscript, we refer to the combined Rossmannoid fold (HAD domain) and C1 cap module as the "PSP catalytic domain".

Prokaryotic HAD phosphatases are mainly small proteins consisting of a single hydrolase domain [80]. This is also the case for 77% of the sequences bearing an HAD domain in the Pfam protein family database. The rest of the enzymes carry additional domains, probably acquired by gene duplication events or horizontal gene transfer during evolution, that provide them with a diversity of specialized functions. *MaSerB*, and *MtSerB2* by structural homology, are part of them, carrying two additional folds called "ACT domains" in N-term of the HAD domain (Figure 1.20D). The crystallographic structure shows that the two monomers swap their N-terminal ACT1 domain to form an intertwined dimer, thanks to a 12 residue-long hinge-loop. This architecture makes *MaSerB* a *candidate* for domain-swapping, since no monomeric structure is available. In addition to being involved in the domain-swapping phenomenon, *MaSerB* ACT domains exhibit extensive ACT1-ACT2 and ACT1-ACT1 intermolecular interactions. The formation of such interfaces is an integral part of the nature of ACT domains, as detailed just below.

Further described in Chapter 3, the presence of the ACT domains also diametrically differentiates *MaSerB* and *MtSerB2* from their human counterpart, *Homo sapiens* PSP. The structure of the latter, known since 2002, shows that it simply consists of an HAD domain that dimerizes through a β -hairpin fold inserted between the β_3 -strand and α_3 -helix of the Rossmannoid core [86].

The ACT regulatory domains

The ACT domains are regulatory units of 70-80 residues recognized in a range of enzymes involved in amino acid and purine metabolism, as well as in some transcriptional regulators. They are named after three enzymes in which they are found: Aspartate kinase, Chorismate mutase and TyrA (prephenate dehydrogenase) [87].

Proposed by Aravind and Koonin as "a conserved, evolutionary mobile module", independently fused to a variety of enzymes to make them susceptible to the regulation of their catalytic activity [87], the ACT domain is identifiable by defining characteristics but exhibits significant functional and structural diversity depending on the enzyme to which it belongs. These features are thoroughly reviewed and exemplified in references [88], [89], and [90]. Hereafter, we provide a brief overview aimed at defining what an ACT domain is.

The $\beta\alpha\beta\beta\alpha\beta$ fold - The archetypal ACT domain is that of *E. coli* PGDH (*Ec*PGDH, Figure 1.21B). Its structure is composed of a $\beta\alpha\beta\beta\alpha\beta$ fold exhibiting a $\beta_4\beta_1\beta_3\beta_2$ antiparallel β -sheet on one face and a pair of antiparallel α -helices on the other, with the second helix twisted about 30° anticlockwise with respect to the first (Figure 1.21A). This topology is crucial for identifying ACT domains. Even if 6 distinct ACT domain families are recognized by Pfam, sequence-based identification alone is not enough as the domain shows significant sequence variation.

The self-association - In many structures, ACT domains are observed interacting with each other (inter- or intramolecularly), mediating oligomerization. As shown in Figure 1.22, the modes of interaction are diverse. In *Ec*PGDH homotetramer, the ACT domains of two different subunits form a dimer, interacting in a side-by-side orientation with a 180° rotational symmetry creating an eight-stranded antiparallel β -sheet with the four helices on the same side (Figure 1.21C). This kind of architecture is also seen in other enzymes, including *Mt*SerA1 (Figure 1.15C) and *Ma*SerB (Figure 1.20D). In the latter, the dimeric interaction is established between the ACT2 domain of one monomer and the exchanged N-terminal ACT1 domain from the other monomer, with both domains oriented in the same direction. Moreover, the two ACT1 domains interact with each other, in a staggered face-to-face orientation, via a portion of their β -sheets. The face-to-face β -sheet interaction is also observed in other enzymes and even leads to the formation of higher order oligomers (trimer, tetramers), alone or in combination with other modes of interaction.

The ability to bind small molecules - Allosteric ligand-binding sites are found at the interfaces between ACT domains. The diversity of sequences and spatial arrangements allows the recognition of many different effectors, usually amino acids, depending on the enzyme. The two side-to-side ACT domain dimers in *Ec*PGDH tetramer can each bind two L-serine molecules at the dimer interface. A L-Ser molecule is also

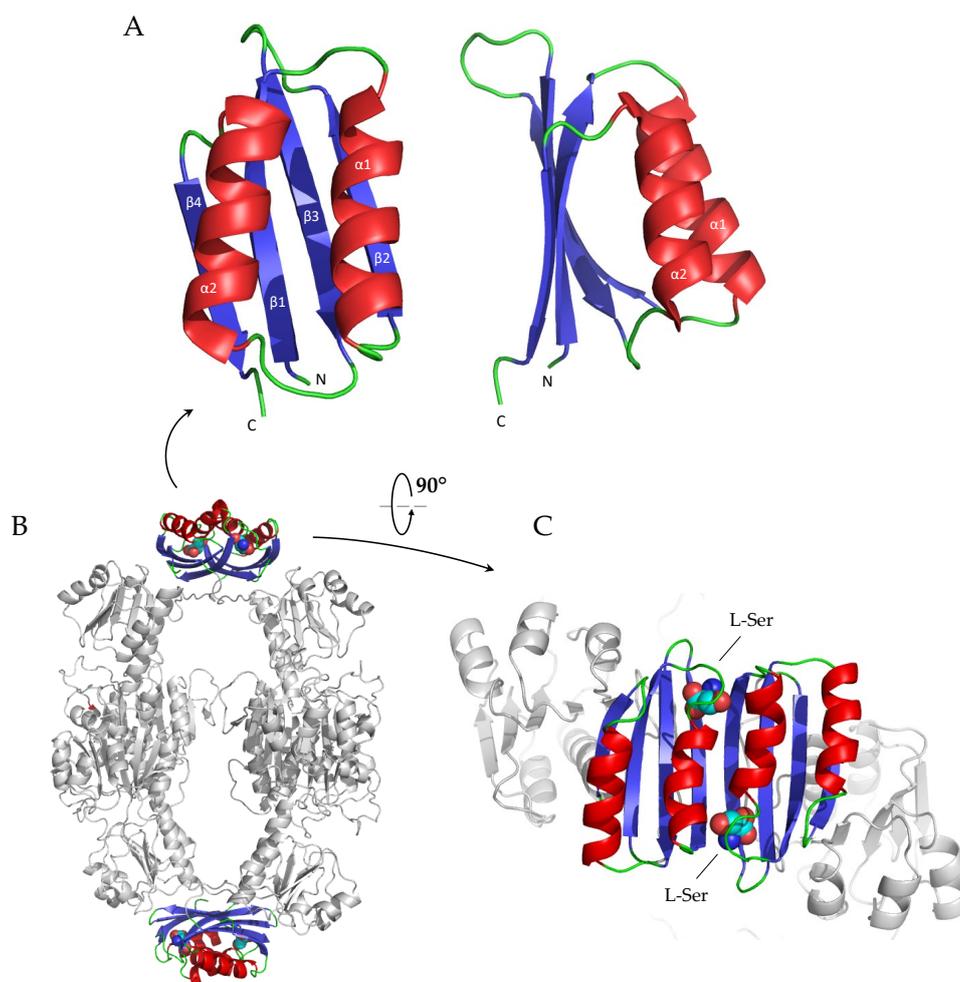


FIGURE 1.21: The archetypal ACT domains of *E. coli* PGDH. (A) The ACT signature fold. (B) *E. coli* PGDH crystallographic structure (PDB: 1PSD). (C) The dimeric self-association of the ACT domains observed in *EcPGDH* and the resulting L-Ser binding sites. L-Ser molecules are shown as spheres.

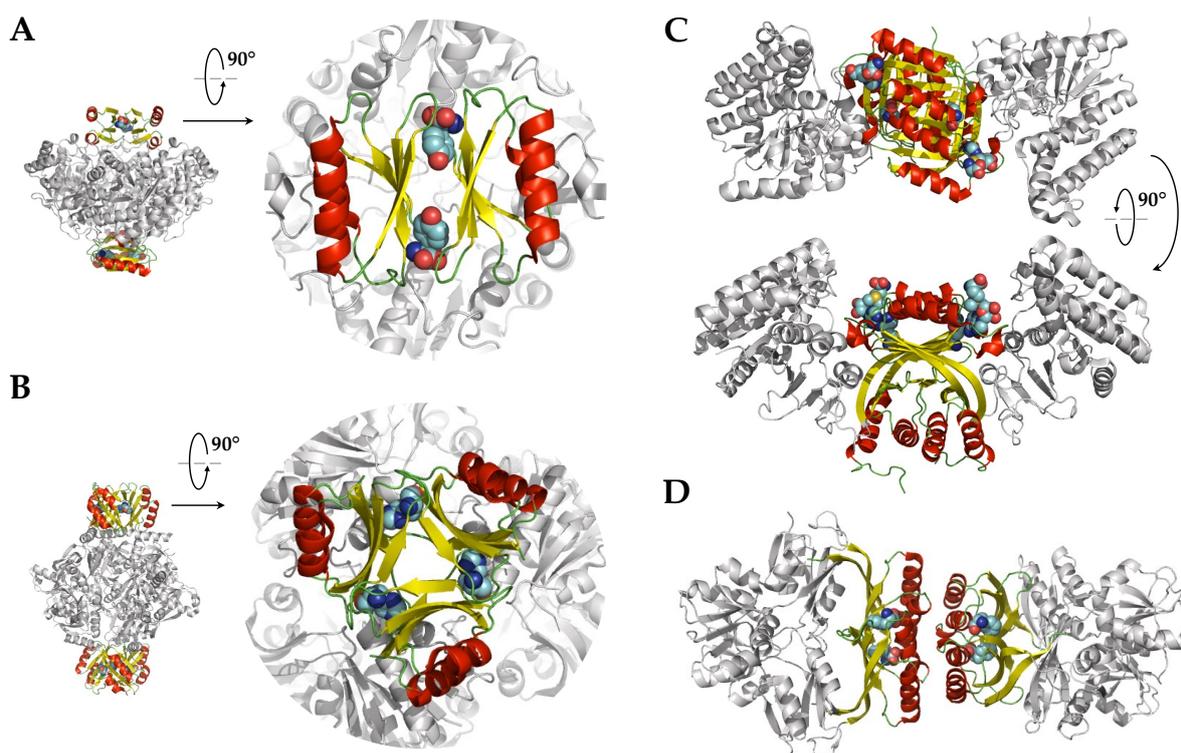


FIGURE 1.22: Non-exhaustive overview of the variety of interactions observed between ACT domains. The ligands are represented as spheres. (A) The face-to-face β -sheet dimeric arrangement of *Thermotoga maritima* DAH7PS (PDB:3PG9) ACT domains in the presence of L-tyrosine. (B) The face-to-face β -sheet trimeric arrangement of *M. tuberculosis* ATP-PRT (1NH8) ACT domains with L-histidine bound at the interface. (C) The tetrameric arrangement adopted by the successive ACT domains of *Arabidopsis thaliana* aspartate kinase I (2CDQ) with S-adenosylmethionine and L-lysine bound at the interface. (D) The unusual tetrameric assembly of *Chlorobium tepidum* prephenate dehydratase (2QMX) with L-phenylalanine bound at the interface.

observed at the two interfaces of ACT1 and ACT2 domains in co-crystal structures of *MaSerB* (PDB: 5JLR and 5JLP). Figure 1.22 shows the ligand binding mode in enzymes in which the ACT domains adopt other quaternary configurations. In some enzymes, like aspartate kinases, the particular tandem arrangement of the ACT domains allows the binding of two different compounds (Figure 1.22C). Sometimes, as is the case of L-tyrosine in *Thermotoga maritima* DAH7PS (Figure 1.22A), ligand binding is required to stabilize the interaction between ACT domains. In such proteins, the domains are not observed to interact in the absence of the ligand.

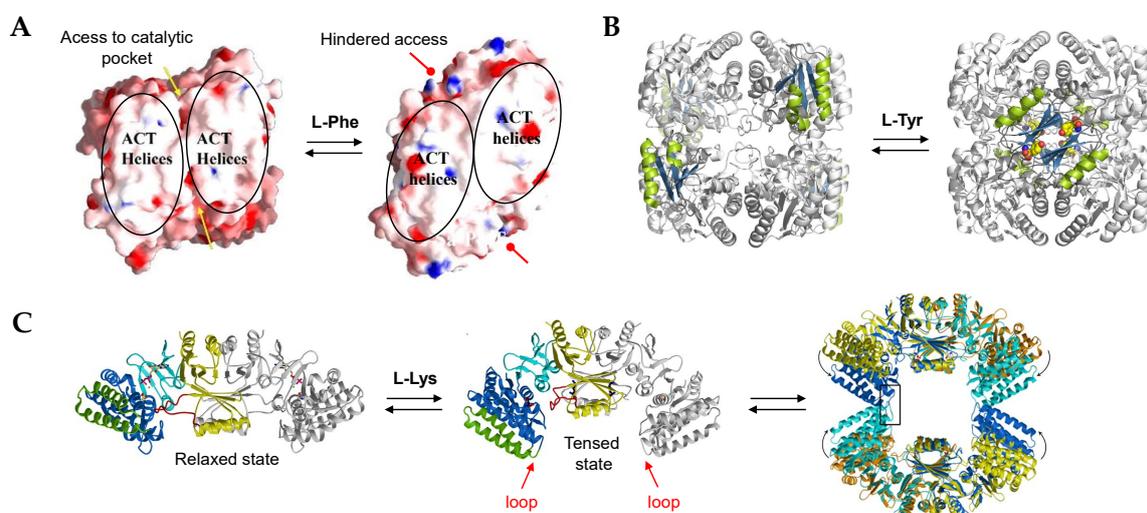


FIGURE 1.23: Examples of conformational changes in ACT containing enzymes upon effector binding. (A) Conformational change in prephenate dehydratase upon L-Phe binding [91]. (B) Conformational change in DAH7PS upon L-Tyr binding [90]. (C) Conformational change and tetramerization of *E. coli* aspartate kinase III upon L-Lys binding [92].

The transmission of allosteric signals - Ligand binding to the allosteric sites at the interface of the ACT domains leads to the transmission of a signal to regulate the catalytic activity of the enzyme. The signal is transmitted to the active site in the form of a conformational change caused by the binding of the effector. The regulation mechanism is dependent on the enzyme and involves conformational rearrangements of varying magnitude. In some enzymes, effector binding induces rather subtle rearrangements. This is the case of prephenate dehydratase, in which the ACT domain interface is tightened by L-Phe binding and causes a displacement of the adjacent catalytic domains relative to each other, resulting in partial closure of the active sites (Figure 1.23A). The elicited conformational change can also be of greater extent. In *T. maritima* DAH7PS, regulation occurs through the transition from an active open form

of the enzyme, where the ACT domains do not interact, to an inactive closed form in which the ACT interaction is stabilized by L-Tyr (Figure 1.23B). The ACT mediated allosteric regulation can even involve changes in oligomerization states, as with the example of *E. coli* aspartate kinase III. Upon L-Lys binding, the equilibrium is shifted from a relaxed to a tensed conformation of the enzyme. The transition involves a rotation of the catalytic domains and the loosening of a loop creating a new interface for the interaction between two tensed dimers, thus leading to the formation of an inactive tetramer (Figure 1.23C).

To summarize, ACT domains are evolutionary mobile regulatory units of defined secondary and tertiary structures but poorly conserved primary structure. They feature highly varied modes of self-association enabling the binding of specific ligands at their interface. They allow the transmission of allosteric signals for the control of enzymatic activity via conformational changes and/or the manipulation of the oligomeric equilibrium.

The structural characteristics of ACT domains, namely the topology and ligand-binding ability (L-serine), are indeed recognized in *MaSerB*. By homology, given the high sequence identity, their presence can be inferred in *MtSerB2*. Yadav, Shree and Grant then turned their attention to studying the last defining feature of ACT domains in *MtSerB2*: their functional role in regulation.

1.3.3 The regulation of *MtSerB2* activity: state of the art

In 2014 and 2017, Yadav and Shree, and Grant respectively published the results of experiments showing that the phosphatase activity of *MtSerB2* was influenced by the presence of L-serine, probably through the interaction with the ACT domains leading to a change in oligomerization state. Four major observations leading to that conclusion were made [82, 93].

1. *MtSerB2* activity is modulated by its ACT domains - Yadav and Shree showed that the presence of ACT domains influenced the activity of *MtSerB2* through steady-state kinetics experiments with truncated mutants. They noticed that the PSP domain alone was able to hydrolyze *O*-phospho-L-serine but with a catalytic constant (k_{cat}) 3-fold lower and an affinity constant (K_M) 6-fold higher than for the native enzyme. The loss in activity and affinity could be partially reversed by an equimolar addition of the purified ACT domains alone.

2. *MtSerB2* is inhibited by L-serine - On the basis of a detailed inhibition kinetics analysis, Grant demonstrated that *MtSerB2* was inhibited by L-Ser. The results of steady-state kinetics experiments showed that L-Ser acted as a partial competitive inhibitor of the enzyme (inhibitory constant $K_i = 19 \mu\text{M}$), binding the free enzyme with greater affinity than the enzyme-substrate complex but unable to drive the velocity to zero. On the other hand, phosphate was found to be a simple dead-end competitive inhibitor of *MtSerB2*, blocking the active site at high concentration. This was consistent with the irreversible nature of the reaction, in which the product cannot be converted back to the substrate. Grant demonstrated that the binding sites for L-Ser and phosphate were not mutually exclusive, meaning that L-Ser could bind to an allosteric site. On their side, Yadav and Shree also reported the inhibition of *MtSerB2* by L-Ser with an IC_{50} value of $0.78 \mu\text{M}$.

3. L-serine binds to the ACT domains of *MtSerB2* - Both teams highlighted the interaction of L-Ser with ACT domains through the analysis of the behavior of truncated and point *MtSerB2* mutants. Yadav and Shree first saw that L-Ser exhibited a drastic drop in its capacity to inhibit the PSP domain alone ($\text{IC}_{50} = 823.7 \mu\text{M}$) compared to the full-length enzyme ($\text{IC}_{50} = 0.78 \mu\text{M}$). Grant studied the L-Ser inhibition mechanism of two double *MtSerB2* point-mutants, G18A/G108A and D15A/E33A (ACT domain residues). Those mutants were designed on the basis of homology to *E. coli* PGDH ACT domains, in which the two homologous glycine residues are required for efficient ligand binding, and aspartate and glutamate residues are seen to interact with L-Ser. Both mutants were much less sensitive to inhibition by L-Ser, indicating that the latter indeed interacts with the ACT domains. Based on the kinetic behavior, Grant proposed that the G18A/G108A mutation made ligand-binding more difficult and that the D15A/E33A mutation abolished it completely.

4. L-serine mediates a specific oligomeric transition in *MtSerB2* - Lastly, Yadav and Shree observed that L-Ser mediated the transition of *MtSerB2* to a higher-order oligomer (Figure 1.24). Based on size-exclusion chromatography analyses in the presence and absence of L-Ser and electrophoresis under non-denaturing conditions, they proposed that "the dimeric population of *MtSerB2* shifted to a tetramer in the presence of a 0.8 molar ratio of L-Ser to *MtSerB2*". The authors also noted that mutants of ACT domains residues G18A and G108A did not undergo the transition. Phosphatase activity assays with the newly formed higher-oligomer showed that it was inactive.

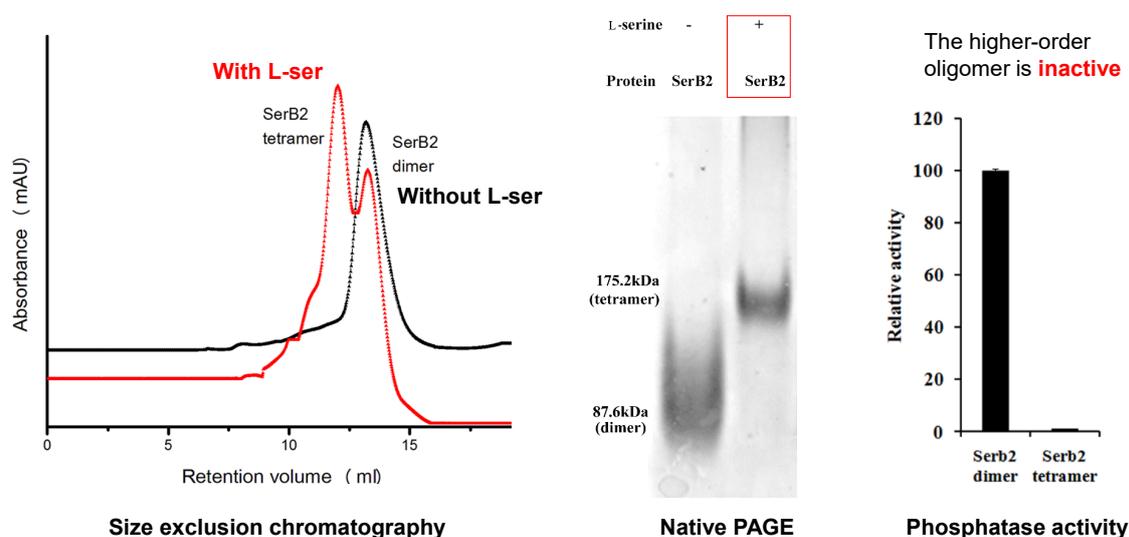


FIGURE 1.24: Yadav and Shree's experiments highlighting a specific transition to a higher-order *MtSerB2* oligomer in the presence of L-serine [82].

Taken together, all these results are consistent with the usual behavior observed in enzymes carrying ACT domains: the regulation of their activity via conformational changes induced by the binding to the ACT domains of a ligand belonging to the catalyzed metabolic pathway. For several enzymes, the structural mechanisms that result in this phenomenon have been elucidated [94]. In this thesis, using a structural biology approach, we attempted to answer the following question: "How does the binding of L-serine to *MtSerB2* ACT domains get translated to the regulation of its catalytic activity?".

Bibliography

- [1] K. Hashimoto, H. Nishi, S. Bryant, and A.R. Panchenko. Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. *Physical Biology*, 8(3):035007, 2011.
- [2] L. Thabault, M. Liberelle, and R. Frédérick. Targeting protein self-association in drug design. *Drug Discovery Today*, 26(5):1148–1163, 2021.
- [3] D.S. Goodsell and A.J. Olson. Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):105–153, 2000.
- [4] R. Koike, A. Kidera, and M. Ota. Alteration of oligomeric state and domain architecture is essential for functional transformation between transferase and hydrolase with the same scaffold. *Protein Science*, 18(10):2060–2066, 2009.
- [5] D.E. Koshland Jr, G. Némethy, and D. Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, 1966.
- [6] Changeux J.P. Monod J., Wyman J. On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology*, 12:88–118, 1965.
- [7] E.K. Jaffe. Morpheeins—a new structural paradigm for allosteric regulation. *Trends in Biochemical Sciences*, 30(9):490–497, 2005.
- [8] S.H. Lawrence and E.K. Jaffe. Expanding the concepts in protein structure-function relationships and enzyme kinetics: Teaching using morpheeins. *Biochemistry and Molecular Biology Education*, 36(4):274–283, 2008.
- [9] E.K. Jaffe and S.H. Lawrence. The morpheein model of allostery: evaluating proteins as potential morpheeins. In *Allostery*, pages 217–231. Springer, 2012.
- [10] W.SJ. Valdar and J.M. Thornton. Conservation helps to identify biologically relevant crystal contacts. *Journal of Molecular Biology*, 313(2):399–416, 2001.
- [11] Y. Tsuchiya, H. Nakamura, and K. Kinoshita. Discrimination between biological interfaces and crystal-packing contacts. *Advances and Applications in Bioinformatics and Chemistry: AABC*, 1:99, 2008.

- [12] S. Dasgupta, G.H. Iyer, S.H. Bryant, C.E. Lawrence, and J.A. Bell. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins: Structure, Function, and Bioinformatics*, 28(4):494–514, 1997.
- [13] M.H. Ali and B. Imperiali. Protein oligomerization: how and why. *Bioorganic & Medicinal Chemistry*, 13(17):5013–5020, 2005.
- [14] G.B. Irvine, O.M. El-Agnaf, G.M. Shankar, and D.M. Walsh. Protein aggregation in the brain: the molecular basis for Alzheimer’s and Parkinson’s diseases. *Molecular Medicine*, 14(7):451–464, 2008.
- [15] A. Gaber and M. Pavšič. Modeling and structure determination of homo-oligomeric proteins: An overview of challenges and current approaches. *International Journal of Molecular Sciences*, 22(16):9081, 2021.
- [16] L.S. Swapna, K. Srikeerthana, and N. Srinivasan. Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. *PloS One*, 7(5):e36688, 2012.
- [17] K. Hashimoto, T. Madej, S.H. Bryant, and A.R. Panchenko. Functional states of homooligomers: insights from the evolution of glycosyltransferases. *Journal of Molecular Biology*, 399(1):196–206, 2010.
- [18] G. D’Alessio. The evolutionary transition from monomeric to oligomeric proteins: tools, the environment, hypotheses. *Progress in Biophysics and Molecular Biology*, 72(3):271–298, 1999.
- [19] T. Perica, Y. Kondo, S.P. Tiwari, S.H. McLaughlin, K.R. Kemplen, X. Zhang, A. Steward, N. Reuter, J. Clarke, and S.A. Teichmann. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science*, 346(6216):1254346, 2014.
- [20] K. Hashimoto and A.R. Panchenko. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proceedings of the National Academy of Sciences*, 107(47):20352–20357, 2010.
- [21] P.B. Harbury, T. Zhang, P.S. Kim, and T. Alber. A switch between two-, three- and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, 262(5138):1401–1407, 1993.

- [22] J. Engel and R.A. Kammerer. What are oligomerization domains good for? *Matrix Biology*, 19(4):283–288, 2000.
- [23] M.J. Bennett, S. Choe, and D. Eisenberg. Domain swapping: entangling alliances between proteins. *Proceedings of the National Academy of Sciences*, 91(8):3127–3131, 1994.
- [24] F. Rousseau, J.W.H Schymkowitz, and L.S. Itzhaki. The unfolding story of three-dimensional domain swapping. *Structure*, 11(3):243–251, 2003.
- [25] F. Rousseau, J. Schymkowitz, and L.S. Itzhaki. Implications of 3D domain swapping for protein folding, misfolding and function. *Protein Dimerization and Oligomerization in biology*, pages 137–152, 2012.
- [26] M.V. Hayes, R.B. Sessions, R.L. Brady, and A.R. Clarke. Engineered assembly of intertwined oligomers of an immunoglobulin chain. *Journal of Molecular Biology*, 285(4):1857–1867, 1999.
- [27] N. Nandwani, P. Surana, H. Negi, N.M. Mascarenhas, J.B. Udgaonkar, R. Das, and S. Gosavi. A five-residue motif for the design of domain swapping in proteins. *Nature Communications*, 10(1):1–13, 2019.
- [28] V.L. Terse and S. Gosavi. The Molecular Mechanism of Domain Swapping of the C-Terminal Domain of the SARS-Coronavirus Main Protease. *Biophysical Journal*, 120(3):504–516, 2021.
- [29] I. Zegers, J. Deswarte, and L. Wyns. Trimeric domain-swapped barnase. *Proceedings of the National Academy of Sciences*, 96(3):818–822, 1999.
- [30] N.M. Mascarenhas and S. Gosavi. Understanding protein domain-swapping using structure-based models of protein folding. *Progress in Biophysics and Molecular Biology*, 128:113–120, 2017.
- [31] N. Koon, C.J. Squire, and E.N. Baker. Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proceedings of the National Academy of Sciences*, 101(22):8295–8300, 2004.
- [32] D. Cardinale, O. M.H. Salo-Ahen, S. Ferrari, G. Ponterini, and E. Cruciani. Homodimeric Enzymes as Drug Targets. *Current Medicinal Chemistry*, 17(9):826–846, 2010.

- [33] T.M. Bakheet and A.J. Doig. Properties and identification of antibiotic drug targets. *BMC Bioinformatics*, 11(1):1–10, 2010.
- [34] L.L. Silver. Appropriate targets for antibacterial drugs. *Cold Spring Harbor Perspectives in Medicine*, 6(12):a030239, 2016.
- [35] A. Cohen, V.D. Mathiasen, T. Schön, and C. Wejse. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal*, 54(3), 2019.
- [36] WHO Guidelines Review Committee. *Guidelines on the management of latent tuberculosis infection*. 2015.
- [37] CDC/ M.S. Ray Butler. Image en ligne: <https://phil.cdc.gov//PHIL-Images/9997/9997-lores.jpg>, 2006.
- [38] I. Barberis, N. L. Bragazzi, L. Galluzzo, and M. Martini. The history of tuberculosis: from the first historical records to the isolation of Koch’s bacillus. *Journal of Preventive Medicine and Hygiene*, 58(1):E9, 2017.
- [39] R. Koch. Die atologie der tuberkulose. *Berliner Klinische Wochenschrift*, 15(221-30), 1882.
- [40] S. Tiberi, M. Petersen, E. and Maeurer, F. Ntoumi, D. Yeboa-Manu, P. Mwaba, C. Vilaplana, O. Dar, M. Bates, T. Corrah, et al. Taking forward the stop TB partnership and world health organization joint theme for world TB day march 24th 2018—“Wanted: leaders for a TB-free world. You can make history. End TB”. *International Journal of Infectious Diseases*, 68:122–124, 2018.
- [41] WHO Global Tuberculosis Programme. *Global Tuberculosis Report 2021*. Global Tuberculosis Programme, 2021.
- [42] WHO Global Tuberculosis Programme. *WHO consolidated guidelines on tuberculosis, Module 4: Treatment - Drug-resistant tuberculosis treatment*. 2020.
- [43] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. 3rd Barry, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 396(6707):190–190, 1998.

- [44] C.M. Sasseti, D.H. Boyd, and E.J. Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1):77–84, 2003.
- [45] M.A. DeJesus, E.R. Gerrick, W. Xu, S.W. Park, J.E. Long, C.C. Boutte, E.J. Rubin, D. Schnappinger, S. Ehrt, S.M. Fortune, et al. Comprehensive essentiality analysis of the *Mycobacterium tuberculosis* genome via saturating transposon mutagenesis. *MBio*, 8(1):e02133–16, 2017.
- [46] J. Rengarajan, B.R. Bloom, and E.J. Rubin. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences*, 102(23):8327–8332, 2005.
- [47] G. Xu, Z. Ni, Y. Shi, X. Sun, H. Wang, C. Wei, G. Wang, and F. Li. Screening essential genes of *Mycobacterium tuberculosis* with the pathway enrichment method. *Molecular Biology Reports*, 41(11):7639–7644, 2014.
- [48] W. Ren, R. Rajendran, Y. Zhao, B. Tan, G. Wu, F.W. Bazer, G. Zhu, Y. Peng, X. Huang, J. Deng, et al. Amino acids as mediators of metabolic cross talk between host and pathogen. *Frontiers in Immunology*, 9:319, 2018.
- [49] Y.J. Zhang and E.J. Rubin. Feast or famine: the host–pathogen battle over amino acids. *Cellular Microbiology*, 15(7):1079–1087, 2013.
- [50] A. Dwivedy, A. Ashraf, B. Jha, D. Kumar, N. Agarwal, and B.K. Biswal. De novo histidine biosynthesis protects *Mycobacterium tuberculosis* from host IFN- γ mediated histidine starvation. *Communications Biology*, 4(1):1–15, 2021.
- [51] J.L. Gallant, A.J. Viljoen, P.D. Van Helden, and I. J.F. Wiid. Glutamate dehydrogenase is required by *Mycobacterium bovis* BCG for resistance to cellular stress. *PloS One*, 11(1):e0147706, 2016.
- [52] S.D. Yelamanchi and A. Surolia. Targeting amino acid metabolism of *Mycobacterium tuberculosis* for developing inhibitors to curtail its survival. *IUBMB life*, 73(4):643–658, 2021.
- [53] A.W. El-Hattab. Serine biosynthesis and transport defects. *Molecular Genetics and Metabolism*, 118(3):153–159, 2016.

- [54] G. Murtas, G.L. Marcone, S. Sacchi, and L. Pollegioni. L-serine synthesis via the phosphorylated pathway in humans. *Cellular and Molecular Life Sciences*, 77(24):5131–5148, 2020.
- [55] R. Caspi, R. Altman, T. and Billington, K. Dreher, H. Foerster, C.A. Fulcher, T.A. Holland, I.M. Keseler, A. Kothari, A. Kubo, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 42(D1):D459–D471, 2014.
- [56] K. Borah, M. Beyß, A. Theorell, H. Wu, P. Basu, T.A. Mendum, K. Noh, D.J.V. Beste, and J. McFadden. Intracellular *Mycobacterium tuberculosis* exploits multiple host nitrogen sources during growth in human macrophages. *Cell Reports*, 29(11):3580–3591, 2019.
- [57] M. Haufroid and J. Wouters. Targeting the serine pathway: A promising approach against tuberculosis? *Pharmaceuticals*, 12(2):66, 2019.
- [58] S. Dey, G.A. Grant, J.C. Sacchettini, and TB Structural Genomics Consortium (TBSGC). Crystal Structure of D-3-Phosphoglycerate dehydrogenase From *Mycobacterium tuberculosis*. doi:10.2210/pdb1YGY/pdb. 2005.
- [59] S. Dey and J.C. Sacchettini. Crystal structure of serine bound D-3-phosphoglycerate dehydrogenase from *Mycobacterium tuberculosis*. doi:10.2210/pdb3DC2/pdb. 2008.
- [60] S. Dey and J.C. Sacchettini. Crystal structure of hydroxypyruvic acid phosphate bound D-3-phosphoglycerate dehydrogenase in *Mycobacterium tuberculosis*. doi:10.2210/pdb3DDN/pdb. 2008.
- [61] X.L. Xu and G.A. Grant. Regulation of *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase by phosphate-modulated quaternary structure dynamics and a potential role for polyphosphate in enzyme regulation. *Biochemistry*, 53(26):4239–4249, 2014.
- [62] S. Dey, G.A. Grant, and J.C. Sacchettini. Crystal structure of *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase: extreme asymmetry in a tetramer of identical subunits. *Journal of Biological Chemistry*, 280(15):14892–14899, 2005.

- [63] S. Dey, R.L. Burton, G.A. Grant, and J.C. Sacchettini. Structural analysis of substrate and effector binding in *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase. *Biochemistry*, 47(32):8271–8282, 2008.
- [64] R.L. Burton, X.L. Chen, S. and Xu, and G.A. Grant. Role of the anion-binding site in catalysis and regulation of *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase. *Biochemistry*, 48(22):4808–4815, 2009.
- [65] X.L. Xu, S. Chen, N.D. Salinas, N.H. Tolia, and G.A. Grant. Comparison of Type 1 D-3-phosphoglycerate dehydrogenases reveals unique regulation in pathogenic *Mycobacteria*. *Archives of Biochemistry and Biophysics*, 570:32–39, 2015.
- [66] F. Coulibaly, E. Lassalle, H.M. Baker, and E.N. Baker. Structure of phosphoserine aminotransferase from *Mycobacterium tuberculosis*. *Acta Crystallographica Section D: Biological Crystallography*, 68(5):553–563, 2012.
- [67] F. Coulibaly, E. Lassalle, E.N. Baker, and *Mycobacterium Tuberculosis* Structural Proteomics Project (XMTB). Structure of a putative phosphoserine aminotransferase from *Mycobacterium tuberculosis*. doi:10.2210/pdb2FYF/pdb. 2007.
- [68] F. Coulibaly, E. Lassalle, H.M. Baker, and E.N. Baker. Structure of a putative phosphoserine aminotransferase from *Mycobacterium tuberculosis*. doi:10.2210/pdb3VOM/pdb. 2012.
- [69] S. Shree, A.K. Singh, R. Saxena, H. Kumar, A. Agarwal, V.K. Sharma, K. Srivastava, K.K. Srivastava, S. Sanyal, and R. Ramachandran. The *M. tuberculosis* HAD phosphatase (Rv3042c) interacts with host proteins and is inhibited by Clofazimine. *Cellular and Molecular Life Sciences*, 73(17):3401–3417, 2016.
- [70] G.D. Tribble, S. Mao, C.E. James, and R.J. Lamont. A *Porphyromonas gingivalis* haloacid dehalogenase family phosphatase interacts with human phosphoproteins and is important for invasion. *Proceedings of the National Academy of Sciences*, 103(29):11027–11032, 2006.
- [71] Y. Hasegawa, G.D. Tribble, H.V. Baker, J.J. Mans, M. Handfield, and R.J. Lamont. Role of *Porphyromonas gingivalis* SerB in gingival epithelial cell cytoskeletal remodeling and cytokine production. *Infection and Immunity*, 76(6):2420–2427, 2008.

- [72] C.M. Haglund and M.D. Welch. Pathogens and polymers: microbe–host interactions illuminate the cytoskeleton. *Journal of Cell Biology*, 195(1):7–17, 2011.
- [73] R.P. Darveau, C.M. Belton, R.A. Reife, and R.J. Lamont. Local chemokine paralysis, a novel pathogenic mechanism for *Porphyromonas gingivalis*. *Infection and Immunity*, 66(4):1660–1665, 1998.
- [74] A.K. Sharma, N. Dhasmana, N. Dubey, N. Kumar, A. Gangwal, M. Gupta, and Y. Singh. Bacterial virulence factors: secreted for survival. *Indian Journal of Microbiology*, 57(1):1–10, 2017.
- [75] H. Liu and C.J. Jeffery. Moonlighting proteins in the fuzzy logic of cellular metabolism. *Molecules*, 25(15):3440, 2020.
- [76] T. Truong and B.H. Penn. An *M. tuberculosis* metabolic enzyme moonlights as an anti-inflammatory effector protein. *Cell Host & Microbe*, 27(3):310–312, 2020.
- [77] L. Wang, J. Wu, J. Li, H. Yang, T. Tang, H. Liang, M. Zuo, J. Wang, H. Liu, F. Liu, et al. Host-mediated ubiquitination of a mycobacterial protein suppresses immunity. *Nature*, 577(7792):682–688, 2020.
- [78] K.N. Allen and D. Dunaway-Mariano. Markers of fitness in a successful enzyme superfamily. *Current Opinion in Structural Biology*, 19(6):658–665, 2009.
- [79] E.V. Koonin and R.L. Tatusov. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity: application of an iterative approach to database search. *Journal of Molecular Biology*, 244(1):125–132, 1994.
- [80] A. Seifried, J. Schultz, and A. Gohla. Human HAD phosphatases: structure, mechanism, and roles in health and disease. *The FEBS journal*, 280(2):549–571, 2013.
- [81] G. Arora, P. Tiwari, R.S. Mandal, A. Gupta, S. Sharma, D. and Saha, and R. Singh. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *Journal of Biological Chemistry*, 289(36):25149–25165, 2014.

- [82] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [83] M. Haufroid, M. Mirgaux, L. Leherte, and J. Wouters. Crystal structures and snapshots along the reaction pathway of human phosphoserine phosphatase. *Acta Crystallographica Section D: Structural Biology*, 75(6):592–604, 2019.
- [84] J. Abendroth, A.S. Gardberg, J.I. Robinson, J.S. Christensen, B. L Staker, P.J. Myler, L.J. Stewart, and T.E. Edwards. SAD phasing using iodide ions in a high-throughput structural genomics environment. *Journal of Structural and Functional Genomics*, 12(2):83–95, 2011.
- [85] Seattle Structural Genomics Center for Infectious Disease (SSGCID). Crystal structure of Phosphoserine phosphatase SerB from *Mycobacterium avium*, native form. doi:10.2210/pdb3P96/pdb. 2016.
- [86] H-Y. Kim, Y-S. Heo, J.H. Kim, M.H. Park, J. Moon, E. Kim, D. Kwon, J. Yoon, D. Shin, E-j. Jeong, et al. Molecular basis for the local conformational rearrangement of human phosphoserine phosphatase. *Journal of Biological Chemistry*, 277(48):46651–46658, 2002.
- [87] L. Aravind and E.V. Koonin. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *Journal of Molecular Biology*, 287(5):1023–1040, 1999.
- [88] D.M. Chipman and B. Shaanan. The ACT domain family. *Current Opinion in Structural Biology*, 11(6):694–700, 2001.
- [89] G.A. Grant. The ACT domain: a small molecule binding domain and its role as a common regulatory element. *Journal of Biological Chemistry*, 281(45):33825–33829, 2006.
- [90] E.J.M. Lang, P.J. Cross, G. Mittelstädt, G.B. Jameson, and E.J. Parker. Allosteric ACTION: the varied ACT domains regulating enzymes of amino-acid metabolism. *Current Opinion in Structural Biology*, 29:102–111, 2014.
- [91] K. Tan, H. Li, R. Zhang, M. Gu, S.T. Clancy, and A. Joachimiak. Structures of open (R) and close (T) states of prephenate dehydratase (PDT)—implication of

- allosteric regulation by L-phenylalanine. *Journal of Structural Biology*, 162(1):94–107, 2008.
- [92] M. Kotaka, J. Ren, M. Lockyer, A. R. Hawkins, and D.K. Stammers. Structures of R-and T-state *Escherichia coli* aspartokinase III: Mechanisms of the allosteric transition and inhibition by lysine. *Journal of Biological Chemistry*, 281(42):31544–31552, 2006.
- [93] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.
- [94] J.S. Liberles, M. Thórólfsson, and A. Martinez. Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. *Amino Acids*, 28(1):1–12, 2005.

Part II

Objectives and strategy

Chapter 2

Studying *MtSerB2* self-assembly and regulation: why and how?

This thesis aims at the structural and functional characterization of the phosphoserine phosphatase SerB2 from *Mycobacterium tuberculosis* (*MtSerB2*). Involved in L-serine biosynthesis, this enzyme is essential for the survival of the tuberculous pathogen but is also thought to play a role of virulence factor in the infected macrophages, facilitating immune invasion and evasion [1, 2]. Such properties make *MtSerB2* an interesting target for the development of new antibiotics against multi-drug resistant tuberculosis.

The identification of small molecule inhibitors of *MtSerB2* phosphatase activity by various groups [2–4], including ours [5], has shown that *MtSerB2* is a druggable target. In addition to the classical drug-design strategy aiming at the occupation of the active site, *MtSerB2* offers the possibility of an approach by allosteric inhibition through its ability to self-assemble.

As inferred from the structure of its closest crystallized homolog, SerB from *M. avium*, the enzyme would exist in the form of an active intertwined dimer, swapping regulatory "ACT" domains. The disruption of this homodimer could be considered as a new therapeutic strategy but is not the only option available: *MtSerB2* also appears to be capable of forming an inactive higher-order oligomer when interacting with its endogenous feedback inhibitor, L-serine [3, 6]. Shifting the equilibrium towards the

formation of this species could also provide a basis for the design of new allosteric inhibitors. Furthermore, targeting *MtSerB2* quaternary structure offers a selective approach avoiding inhibition of the smaller human phosphoserine phosphatase, which does not possess ACT domains and exhibits a very different dimerization mode [7].

To be most effective, the design of inhibitors is based on accurate structural information about the targeted macromolecule. The problem with *MtSerB2* system is that no direct, experimental structural information is available. The structure used to rationalize the results presented in the literature is a homology model and there are no data shedding light on the architecture or stoichiometry of the higher-order oligomer formed in the presence of L-serine.

We aim to address this gap with the research work proposed here (Figure 2.1). The goal is to provide a deeper understanding of self-assembly and regulation in *MtSerB2* as a basis for the development of allosteric inhibitors. To that end, after overexpression and purification, the system will be studied in the absence and presence of L-serine using orthogonal techniques and following a three step workflow:

1. Qualitative and quantitative identification of species in solution by size exclusion chromatography (SEC), electrophoresis under non-denaturing conditions (Native PAGE) and multi-angle light scattering (MALS)
2. Structural characterization of the identified species via X-ray crystallography (XRD) and small-angle X-Ray scattering (SAXS), coupled with molecular modeling
3. Functional characterization via steady-state enzyme kinetics analyses

The particularity of this work also lies in the parallel characterization of other phosphoserine phosphatases, either homologous natural variants present in other organisms or mutants designed *in vitro* by site-directed mutagenesis. Their behavior combined with the difference between the primary sequences can be exploited to highlight the role of key residues and domains in *MtSerB2* quaternary states.

The results of this work are presented in three separate chapters. Chapter 3 includes studies on *MtSerB2* homologs, performed upstream or in parallel of our work on *MtSerB2*. Its content helps to rationalize the behavior observed for *MtSerB2* by

comparison. Then, Chapter 4 focuses on the characterization of *MtSerB2* in the absence of L-serine and starts with the presentation of an unexpected result with respect to the state of the art presented in the introduction. Finally, Chapter 5 presents the behavior of *MtSerB2* in the presence of L-serine and aims to advance the knowledge regarding the formation of the higher order oligomer. The content of the three chapters will then be discussed in the final section of the manuscript, in order to provide an overall picture of self-assembly and regulation in *MtSerB2*.

These three chapters present results obtained through the use of the biophysical and biochemical techniques mentioned above. A brief reminder of how these work and the type of information they provide can be found in Appendix A, with a particular focus on MALS and SAXS.¹ Detailed experimental protocols are available in Appendix B.

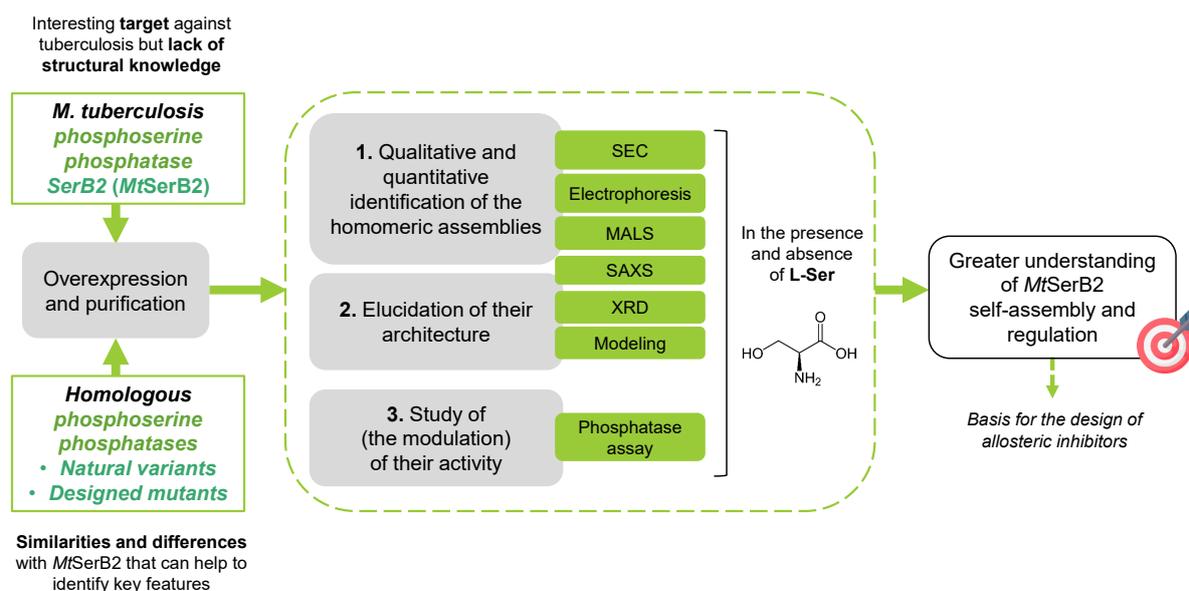


FIGURE 2.1: Strategy adopted in this thesis to better characterize *MtSerB2* self-assembly and regulation.

¹Access to the SEC-MALS equipment was kindly provided by Professor Marianne Fillet from The Laboratory for the Analysis of Medicines (LAM), CIRM, ULiège, and SEC-SAXS experiments were performed on the SWING beamline at SOLEIL Synchrotron (Paris), access to which was granted by Doctor Javier Pérez and supported by iNEXT-Discovery, grant number 871037, funded by the Horizon 2020 program of the European Commission.

Bibliography

- [1] C.M. Sassetti, D.H. Boyd, and E.J. Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1):77–84, 2003.
- [2] S. Shree, A.K. Singh, R. Saxena, H. Kumar, A. Agarwal, V.K. Sharma, K. Srivastava, K.K. Srivastava, S. Sanyal, and R. Ramachandran. The *M. tuberculosis* HAD phosphatase (Rv3042c) interacts with host proteins and is inhibited by Clofazimine. *Cellular and Molecular Life Sciences*, 73(17):3401–3417, 2016.
- [3] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [4] G. Arora, P. Tiwari, R.S. Mandal, A. Gupta, S. Sharma, D. and Saha, and R. Singh. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *Journal of Biological Chemistry*, 289(36):25149–25165, 2014.
- [5] E. Pierson, M. Haufroid, T.P. Gosain, P. Chopra, R. Singh, and J. Wouters. Identification and repurposing of trisubstituted harmine derivatives as novel inhibitors of *Mycobacterium tuberculosis* phosphoserine phosphatase. *Molecules*, 25(2):415, 2020.
- [6] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.
- [7] H-Y. Kim, Y-S. Heo, J.H. Kim, M.H. Park, J. Moon, E. Kim, D. Kwon, J. Yoon, D. Shin, E-j. Jeong, et al. Molecular basis for the local conformational rearrangement of human phosphoserine phosphatase. *Journal of Biological Chemistry*, 277(48):46651–46658, 2002.

Part III

Results and discussions

Chapter 3

Getting to know *MtSerB2* orthologs

3.1 Homologs selection

At the beginning of this project, we were in possession of plasmids encoding the human and *M. avium* phosphoserine phosphatases (respectively *HsPSP* and *MaSerB*). The plasmid allowing the overexpression of *MaSerB* was generously provided by the SSGCID. Their construction (AVA0421) endowed with an N-terminal cleavable His-Tag being very convenient for downstream purifications, and for the sake of reproducibility between constructions, we also requested the plasmid encoding *MtSerB2* from them [1]. Clones for orthologous PSPs were also available, among which those of the PSPs from *B. melitensis* (*BmSerB*) and *M. marinum* (*MmSerB2*). These orthologs were selected for further comparison with *MtSerB2* on the basis of their sequence, as detailed in the next section. Table 3.1 summarizes the source of the constructions used in this thesis and their sequence parameters.

TABLE 3.1: Sequence parameters of the enzymes studied in this thesis. The parameters were calculated using ProtParam on the basis of the exact sequences of the characterized constructs (after cleavage, or with His-tag for *HsPSP*). pI: isoelectric point. The sequences were confirmed by plasmid sequencing. Mismatches observed with the UniProt reference sequences are as follows: (1) V inserted between M1 and N2 + mutations G31R and G152E (2) L inserted between M1 and L2.

Enzyme	UniProt	Length	Molar mass (Da)	pI	Ext. coef. ($M^{-1} \text{cm}^{-1}$)	Rem
<i>MtSerB2</i>	O53289	413	43 357	4.78	11460	-
<i>MmSerB2</i>	B2HHH0	416	43 840	4.67	11460	-
<i>MaSerB</i>	A0QJI1	416	43 907	4.77	7450	(1)
<i>BmSerB</i>	Q8YI30	307	32 514	5.28	11460	(2)
<i>HsPSP</i>	P78330	245	27 171	6.26	11460	-

3.2 Sequence conservation analysis

3.2.1 Global alignment

The primary sequences of *MtSerB2* and the orthologous PSPs characterized in this thesis are aligned in Figure 3.1. The sequence alignment was divided into ACT and PSP domains on the basis of *MaSerB* (PDB: 3P96) secondary structure elements (depicted on the first line).

The first observation that can be made is that the aligned PSPs have different lengths. The mycobacterial enzymes *MtSerB2*, *MmSerB2* and *MaSerB* are 409, 412 and 412 residues long, respectively. *BmSerB* is significantly shorter, with 303 residues, and *HsPSP* is the shortest of the studied orthologs, with 225 residues. Based on the correspondence with the structure of *MaSerB*, and as depicted in Figure 3.1, this difference in length could be related to the presence of one (in *BmSerB*) or two (in mycobacterial enzymes) additional N-terminal ACT regulatory domains.

To further investigate the presence of these domains, the sequences were searched against the content of the Pfam protein family database (<http://pfam.xfam.org/>). The results of this search are shown in Table 3.2. A C-term HAD-like hydrolase domain is identified for all enzymes. In *HsPSP*, it corresponds to the whole sequence. For all three mycobacterial enzymes, an N-terminal ACT domain (of type *ACT6*) is found and a second ACT domain (of type *ACT*), located between the N-term ACT

domain and the HAD-like domain, is proposed as an insignificant Pfam match. In *BmSerB*, the 91 residues-sequence located upstream of the HAD-like domain is left unassigned. Although Pfam recognizes 6 distinct types of ACT domains, their identification remains complicated because of the substantial sequence variation within the family [6].

TABLE 3.2: Pfam matches for the five PSPs studied in this thesis. CL0137: HAD superfamily clan, CL0070: ACT-like domain clan, S = significant, I = insignificant

Enzyme	Domain	Description	Clan	Start-end (residue)	E-value	Pfam match
<i>MtSerB2</i>	HAD	HAD-like hydrolase	CL0137	186-354	3.30E-29	S
	ACT_6	ACT domain	CL0070	10-86	8.00E-19	S
	ACT	ACT domain	CL0070	100-168	0.006	I
<i>MmSerB2</i>	HAD	HAD-like hydrolase	CL0137	185-353	1.40E-27	S
	ACT_6	ACT domain	CL0070	9-85	6.70E-19	S
	ACT	ACT domain	CL0070	99-164	0.16	I
<i>MaSerB</i>	HAD	HAD-like hydrolase	CL0137	185-353	1.70E-28	S
	ACT_6	ACT domain	CL0070	9-85	1.40E-17	S
	ACT	ACT domain	CL0070	112-167	1.2	I
<i>BmSerB</i>	HAD	HAD-like hydrolase	CL0137	92-259	1.10E-16	S
<i>HsPSP</i>	Hydrolase	HAD-like hydrolase	CL0137	14-191	5.10E-17	S

Percentages of sequence identity and similarity were also calculated based on the alignment shown in Figure 3.1. The data is shown in Table 3.3.

TABLE 3.3: Identity and similarity percentages between the five PSP studied in this thesis. Values marked with an asterisk* were calculated on overlapping regions only.

%id (%sim)	<i>MtSerB2</i>	<i>MmSerB2</i>	<i>MaSerB</i>	<i>BmSerB</i>
<i>MmSerB2</i>	85.7 (90.8)	-	-	-
<i>MaSerB</i>	83.3 (88.1)	83.0 (89.8)	-	-
<i>BmSerB</i>	23.4 (38.8) *29.5 (48.9)	23.7 (39.0) *30.2 (49.5)	24.5 (40.4) *31.1 (51.4)	-
<i>HsPSP</i>	14.3 (24.6) *23.7 (41.0)	14.6 (24.9) *24.5 (41.8)	14.4 (24.2) *24.1 (40.6)	19.1 (30.4) *25.3 (40.3)

Sharing over 83% identity and about 90% sequence similarity, mycobacterial SerBs are closely related. It is worth noticing that *MtSerB2* is even more similar to *MmSerB2*

than to its crystallized ortholog *MaSerB*. When these are compared to *BmSerB* and *HsPSP*, the percentages decrease to around 24% (40%) and 14% (24%) respectively. Because the sequences are of different lengths and the identity calculation is performed on the total length of the alignment, these percentages are slightly underestimated compared to those obtained by comparing the even-sized mycobacterial SerBs. Therefore, when the calculation is performed on the overlapping regions only, the percentages rise to around 30% (50%) for Mycobacteriae vs. *Brucella*, and 24% (41%) for Bacteriae vs. Human. To further visualize the similarities, separate alignments within each domain region can be made.

3.2.2 Alignment of PSP domains

The sequences corresponding to the PSP domains of the different orthologs are aligned in Figure 3.2 and the results of the identity calculations based on this alignment are shown in Table 3.4. As expected, the PSP domains of mycobacterial SerBs are highly similar (>90%). When these are compared to *BmSerB*, there are still about 50% of similar residues and 35% of identical residues. The similarity is a little lower when comparing the PSP domains of bacterial enzymes to the human enzyme (40%). The four signature motifs I (hhhDxDx[T/V][L/I/V]h), II ([S/T]), and III-IV (K18-30[G/S][D/S]x3-4[D/E]) of HAD phosphatases are found in all five enzymes [7]. Besides these motifs, 28 residues are strictly conserved among their PSP domains, of which 13 are in the C1 cap region and 8 are glycines. It is also interesting to note the presence of gaps in the sequence alignment at the dimerization region of *HsPSP*. The human enzyme contains four more residues than the other enzymes in this region.

TABLE 3.4: Identity and similarity percentages between the PSP domains of the five PSP studied in this thesis.

%id (%sim)	<i>MtSerB2_{PSP}</i>	<i>MmSerB2_{PSP}</i>	<i>MaSerB_{PSP}</i>	<i>BmSerB_{PSP}</i>
<i>MmSerB2_{PSP}</i>	87.8 (92.4)	-	-	-
<i>MaSerB_{PSP}</i>	86.6 (92.0)	86.1 (91.6)	-	-
<i>BmSerB_{PSP}</i>	34.7 (51.1)	36.0 (52.3)	36.8 (53.6)	-
<i>HsPSP</i>	23.3 (40.2)	24.1 (41.0)	23.7 (39.8)	24.5 (39.5)

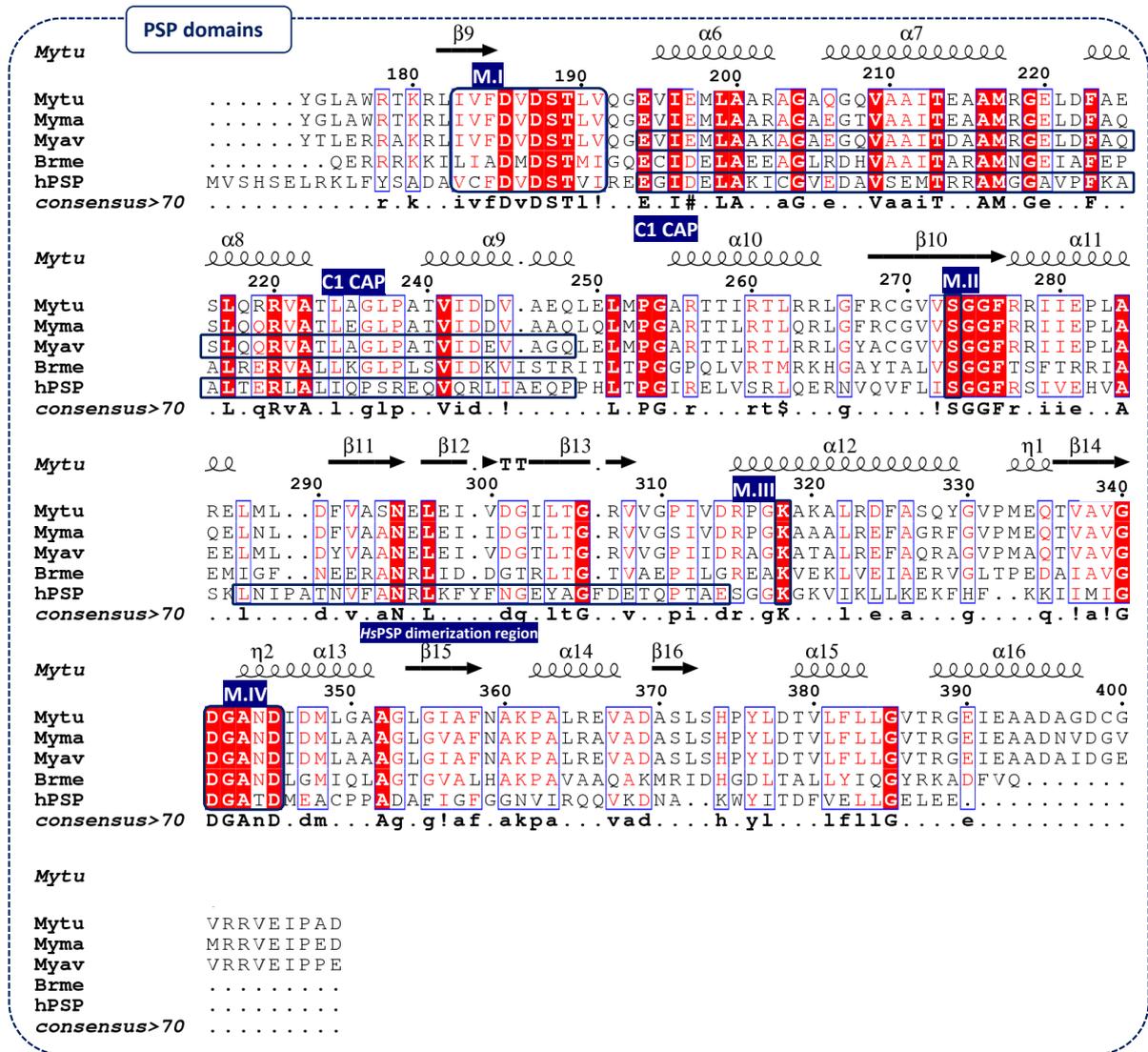


FIGURE 3.2: Multiple sequence alignment of the PSP domains of the five PSPs studied in this thesis.

3.2.3 Alignment of (putative) ACT domains

The sequences corresponding to the ACT domains regions of the studied orthologs are aligned in Figure 3.3 and the results of identity calculations based on the alignment are shown in Table 3.5.

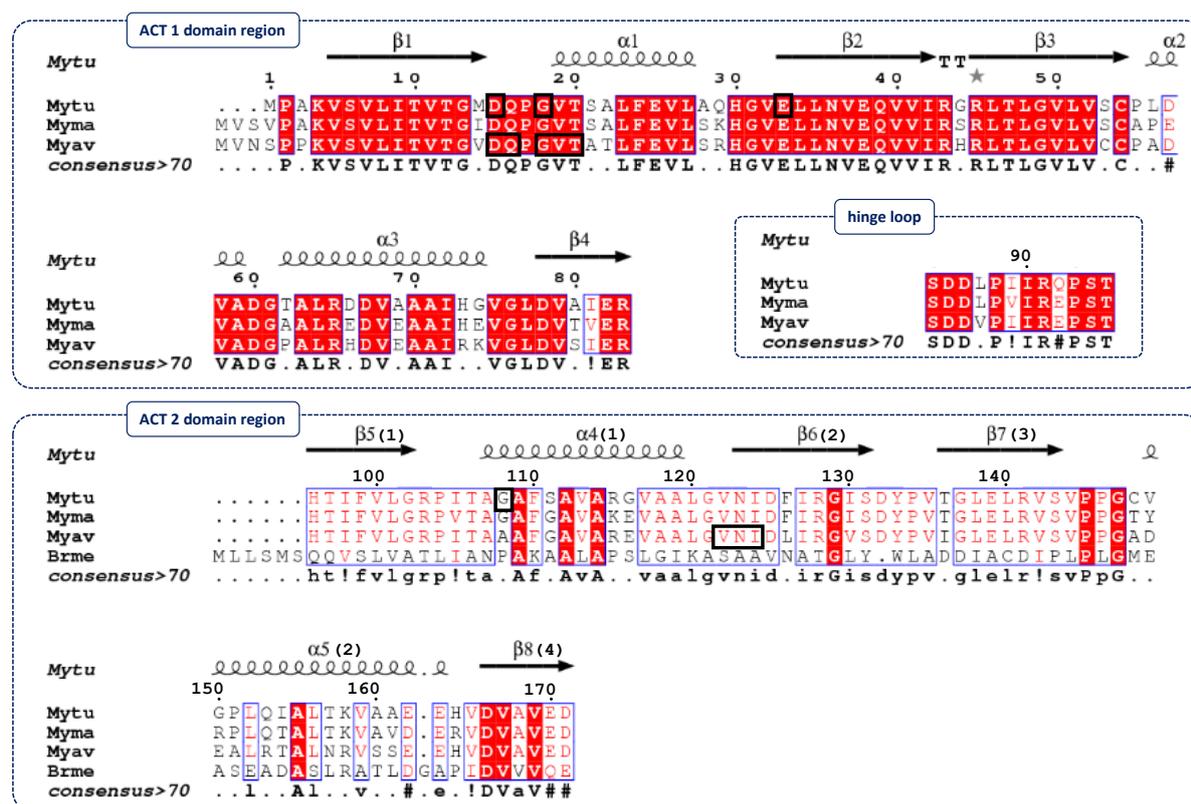


FIGURE 3.3: Multiple sequence alignment of the ACT domains regions of the five PSPs studied in this thesis. Residues identified as important for L-serine binding in *MtSerB2* [2, 3] and *MaSerB* [4, 5] are boxed in black.

Again, the ACT1 domain regions of *MtSerB2*, *MmSerB2* and *MaSerB* are highly similar (>80%, **in bold** in Table 3.5). The non-conserved residues are evenly distributed over the alignment. The similarity is even a little higher over the ACT2 domain area, with *MtSerB2* and *MmSerB2* sharing 92% sequence similarity in this area. Remarkably, however, the alpha helix areas of the ACT2 domain are the least conserved parts of the alignment, with α_4 helix including 4 of the 17 non-conserved residues, and α_5 helix including 10 of the 17 non-conserved residues. It should also be noted that the residues (boxed in Figure 3.3) that were identified as important for the interaction of *MtSerB2* and *MaSerB* with L-Ser in references [2–5] are retained

TABLE 3.5: Identity and similarity percentages between the ACT domains regions of the five PSP studied in this thesis.

%id (%sim)	<i>MtSerB2_{ACT1}</i>	<i>MmSerB2_{ACT1}</i>	<i>MaSerB_{ACT1}</i>		<i>BmSerB_{ACT2}</i>	<i>MtSerB2_{ACT2}</i>	<i>MmSerB2_{ACT2}</i>
<i>MmSerB2_{ACT1}</i>	80.2 (83.7)	-	-		-	-	-
<i>MaSerB_{ACT1}</i>	77.9 (80.2)	79.1 (86.1)	-		-	-	-
<i>BmSerB_{ACT2}</i>	<u>13.8 (47.1)</u>	<u>12.6 (41.4)</u>	<u>12.6 (42.5)</u>		-	-	-
<i>MtSerB2_{ACT2}</i>	18.1 (38.6)	17.4 (36.1)	11.6 (31.4)		12.8 (34.9)	-	-
<i>MmSerB2_{ACT2}</i>	16.9 (34.9)	15.1 (32.6)	11.6 (30.2)		12.8 (33.7)	85.5 (92.1)	-
<i>MaSerB_{ACT2}</i>	18.1 (39.8)	17.4 (38.4)	12.8 (36.1)		15.1 (36.1)	79.0 (82.9)	77.6 (86.8)

among the 3 enzymes, with the exception of A112 in *MaSerB*. The hinge loop area, where there are 3 variations, is also very well conserved (83% identity, 100 % similarity). Lastly, the sequence conservation between the ACT1 and ACT2 domain regions is relatively low, with percentages of identity and similarity between 11.6 and 18.1%, and 30.2 and 39.7% respectively (alignment not shown). Based on these observations, the Pfam analysis and by analogy with the experimental structure of *MaSerB*, the existence of 2 consecutive N-term ACT1 and ACT2 domains in *MtSerB2* and *MmSerB2* can be confirmed despite the absence of structural evidence.

This conclusion cannot be drawn unambiguously for the presence of an ACT domain in *BmSerB*. Indeed, the sequence identity and similarity percentages are quite low when comparing the ACT2 region to that of its mycobacterial counterparts (12-15% id, 34-36% sim, *in italics* in Table 3.5). However, interestingly, when the region is aligned with the ACT1 domains of *MaSerB*, *MtSerB2*, and *MmSerB2* (underlined in Table 3.5) the identity percentages remain comparable (12-14%) but the similarity percentages are significantly higher (41-47%). The highest sequence similarity is found with the ACT1 domain of *MtSerB2* (47.1%). Still, as described in section 3.2.1, Pfam failed to associate this region with a domain of the ACT family. A structure-based approach should be used to further investigate the nature of the fold associated with the unassigned sequence. Such analyses are described in section 3.3.5.

3.2.4 Conclusion of the sequence conservation analysis

In addition to the human and *M. avium* PSPs, already available in the laboratory at the beginning of this work, we chose to study the PSPs of *M. marinum* and *B. melitensis* in parallel with *MtSerB2*. Following the alignment of the sequences of these enzymes, we noticed two key elements:

- *MmSerB2* is an even closer homolog of *MtSerB2* than *MaSerB* in terms of sequence identity.
- *BmSerB* is 92 residues shorter than *MtSerB2*. The enzyme would consist of a C-term PSP domain quite similar (51.0%) to that of *MtSerB2*, preceded at the N-term by a 90 amino acid sequence that could potentially fold into an ACT domain.

Given the interesting association of their similarities and differences with *MtSerB2*, we decided to further characterize *MmSerB2* and *BmSerB*. Their biophysical and biochemical parameters are described in the following section. The characterization of *HsPSP* and *MaSerB* has been mainly the subject of previous work from the lab or other groups but key information will also be described alongside the results obtained during this thesis.

3.3 Biophysical and biochemical features of the orthologs

3.3.1 A word about enzyme production and analysis

Overexpression and purification – The bacterial enzymes analyzed in this thesis are all encoded by plasmids AVA0421 (pAVA0421) received from the SSGCID and were all produced following the same overexpression and purification protocol (see Appendix B). Each enzyme could be successfully overexpressed by IPTG induction in transfected *E.coli* bacteria and then purified with yields of at least 20 mg/L culture. The pAVA0421 allows the production of enzymes carrying a N-terminal hexahistidine tag cleavable by the Human Rhinovirus 3C (HRV3C) protease. The implemented purification strategy therefore included the following 3 steps:

1. **IMAC:** a first step of immobilized metal affinity chromatography (IMAC) to separate the recombinant protein from the rest of the cell lysate
2. **His₆-tag cleavage** using a His-tagged HRV3C protease
3. **Reverse IMAC:** a second IMAC step to separate the cleaved protein from nickel-affine contaminants and His-tagged protease

This workflow allowed to recover highly pure recombinant enzymes, as proven by the presence of a single intense band on SDS-PAGE gel. Enzymes produced with this protocol are also very close to the native form since only 4 residual amino acids remain at the N-term after tag cleavage (Gly-Pro-Gly-Ser).

The SDS-PAGE gel shown in Figure 3.4 monitors the expression and purification of *MtSerB2* (43 kDa) and illustrates the aforementioned production protocol. The efficiency of the overexpression can be observed by the appearance of an intense band between 37 kDa and 50 kDa in the post-induction lane (AI) compared to the pre-induction lane (BI). This band, also present in lane S shows that the protein is soluble in the bacterial lysis supernatant. As indicated by a very intense band in lane IM, the first IMAC allows for the recovery of the His₆-tagged protein, although other bands reveal the presence of contaminants. The change in electrophoretic mobility of the band between IM lane and C lane is the result of the mass loss (2 kDa) due to the removal of the hexahistidine tag (MAHHHHHHMGTLEAQTQ) following cleavage

by HRV3C protease. Finally, the intense band of the rIM lane reflects the high purity of the protein after the final reverse IMAC step. Similar gels were obtained for all bacterial orthologs purified in the laboratory.

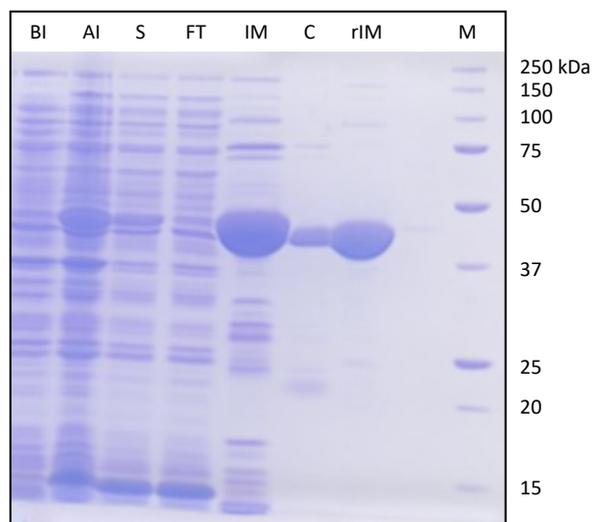


FIGURE 3.4: SDS-PAGE analysis monitoring the overexpression and purification of recombinant *MtSerB2*. BI = before induction, AI = after induction, S = soluble fraction, FT = flow through, IM = after the first IMAC step, C = after tag cleavage, rIM = after the reverse IMAC step, M = protein standards.

The plasmid encoding the human enzyme *HsPSP* was kindly provided by collaborators from UCLouvain for a previous project. Pure frozen His-tagged protein used for crystallization experiments was already available at the beginning of this work and the characterizations were performed directly on this stock.

Characterization by SEC-MALS and SEC-SAXS – Unless otherwise specified, the elution buffer used for each SEC-MALS and SEC-SAXS analysis was composed of 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM TCEP. The chromatograms shown in this chapter were obtained following the analysis of the samples on a BioResolveSEC mAb 200Å 2.5 µm 7.8 x 300 mm (Waters) column at a flow rate of 0.5 mL min⁻¹.

3.3.2 PSP from *H. sapiens*, the human PSP (*HsPSP*)

The human phosphoserine phosphatase is already well characterized in the literature. Expressed in all human tissues, it is a Mg^{2+} -dependent dimeric enzyme that specifically dephosphorylates both D and L enantiomers of *O*-phospho-serine at pH optima between 5.6 and 6.6 [8, 9]. Thanks to mutagenesis and the solving of its crystallographic structure with various ligands and inhibitors (6 structures deposited in the PDB, see Table 3.6), its catalytic mechanism involving the formation of a phosphoenzyme intermediate has been extensively described by various research groups, including ours. It is now well documented that the enzyme exists in closed and open conformations between which it switches through the unfolding of the α_2 -helix of the C1-cap (residues 40 to 56) [10–14]. Structure based knowledge also helped to understand how *HsPSP* is uncompetitively feedback regulated by L-serine and inhibited when Ca^{2+} replaces Mg^{2+} in the active site [14, 15]. In addition to its structural and biochemical properties, its involvement in L-serine synthesis defects, neurodevelopmental diseases and various cancers has been widely reported [16–21].

TABLE 3.6: The 6 structures of *HsPSP* deposited in the PDB. D-AP3 = D-2-amino-3-phosphono-propionic acid

PDB entry	Resolution (Å)	Nature	Ligands	Authors	Year
1L8L	2.51	wt	D-AP3	Kim <i>et al.</i>	2002
1L8O	2.8	wt	L-Ser, PO_4^{3-}	Kim <i>et al.</i>	2002
1NNL	1.53	wt	Ca^{2+} , Cl^-	Peeraer <i>et al.</i>	2003
6HYJ	1.929	wt	Ca^{2+} , L-phosphoserine, L-Ser	Wouters <i>et al.</i>	2019
6HYY	1.566	wt	Ca^{2+} , Mg^{2+} , PO_4^{3-} , L-Ser	Wouters <i>et al.</i>	2019
6Q6J	1.985	wt	Ca^{2+} , Cl^- , L-homocysteic acid	Wouters <i>et al.</i>	2019

HsPSP forms a particularly elongated dimer (Figure 3.5). The narrow dimer interface is located in the β -hairpin fold (residues 130-140) that is inserted between the β_3 -strand and α_3 -helix of the Rossmannoid core. The β -hairpins of two protomers interact together to form a dimer. A detailed description of this interface in structure 1L8L has been reported by Kim *et al* [14]. The residues F139, L135, K136, A144, Y138, R134, E142, F137, Y143, F146 are essential for the stabilization of the latter via hydrogen bonding, salt bridges and the formation of hydrophobic clusters.

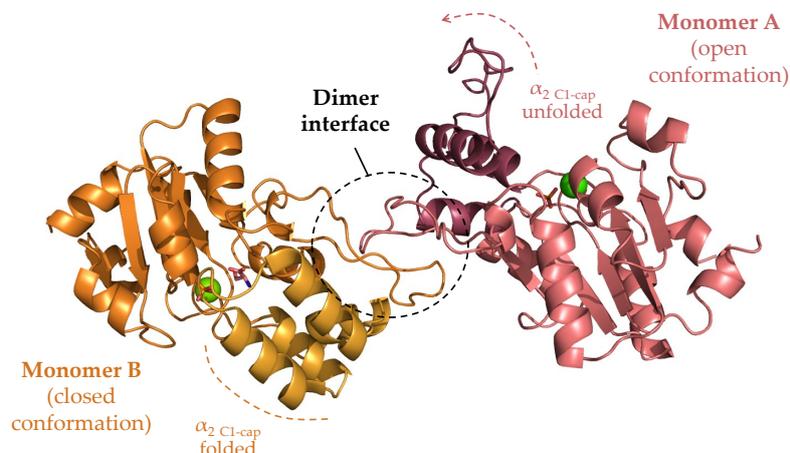


FIGURE 3.5: Crystallographic structure of *HsPSP* dimer (PDB: 6HYY)

Investigation of *HsPSP* structure in solution

It is briefly reported in the aforementioned references that *HsPSP* is also dimeric in solution, as assessed by SEC and dynamic light scattering. However, only reference [9] reported a chromatogram. In order to supplement the available data about its behavior in solution, we further characterized *HsPSP* by SEC-MALS and SEC-SAXS.

The use of SEC-MALS allows to determine the molar mass of a biomolecule in solution in an absolute way. To enable the analysis of a monodisperse solution, the substance of interest is separated from any other constituents of the solution by SEC. Its elution is monitored by a UV detector which also permits a measurement of its concentration. It then passes through a MALS detector, where the intensity of the light scattered by the molecule is measured and converted into a molar mass value via physical equations¹.

As can be seen on the SEC-UV chromatogram in Figure 3.6, *HsPSP* elutes as a single but heterogeneous peak. The peak is slightly tailing and two shoulders are present at its front. The molar mass measured by MALS over the entire peak is 49.3 kDa and effectively corresponds to the mass of *HsPSP* dimer (MW_{monomer} : 27 kDa). The observed deformations are indicative of conformational polydispersity. The hydrodynamic radius of the enzyme likely varies depending on the conformation in which each of its protomers is found (fully open to fully closed) or the relative spatial

¹See Appendix A for a more detailed explanation of the technique

arrangement of the two protomers in the dimer. *HsPSP* could therefore exist as several forms that are retained for shorter or longer retention times within the column.

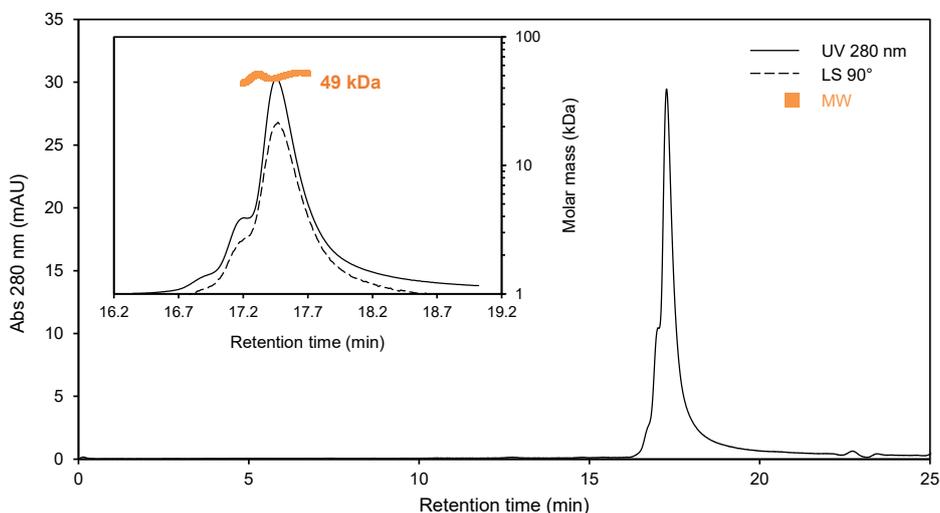


FIGURE 3.6: SEC-UV-MALS analysis of *HsPSP*.

HsPSP solution structure was also assessed by SEC-SAXS. In SAXS, the intensity of X-rays scattered by a sample made monodisperse by SEC is measured as a function of the scattering angle. The resulting 2D isotropic scattering signal is integrated to yield a scattering curve $I(q)$ vs q from which many shape parameters can be determined through replots. Information about folding, flexibility, shape, conformation and oligomerization state among others can be accessed through this technique².

The parameters derived from the analysis of the experimental scattering curve (Figure 3.7A, gray dots) are presented in Table 3.7. The R_g value is high compared to that of globular proteins of similar molar mass and reflects the elongated shape of the enzyme. For example, ovalbumin and BSA have indeed R_g values around 23 to 28 Å [22] and the empirical value calculated for a globular protein of 54.2 kDa using Equation 3.1 [23] is 23.9 Å.

$$R_g = \left(\frac{3}{5}\right)^{\frac{1}{2}} \left(\frac{M_W(Da)}{0.44(Da/\text{Å}^3)} \frac{3}{4\pi}\right)^{\frac{1}{3}} \quad (3.1)$$

²See Appendix A for a more detailed explanation of the technique.

The deviation from globularity is also indicated by the high D_{\max} value (139 Å). Browsing the Small Angle Scattering Biological Data Bank (SASBDB), it can be seen that 50 kDa globular proteins have maximal dimensions around 80 Å. The molecular weight values calculated from Bayesian inference, 53.1 kDa (CI: 49.2 to 53.8), and from Porod volume, 63.5 kDa, confirmed the dimeric state of the enzyme in solution.

TABLE 3.7: SAXS parameters of *HsPSP* (1) derived from the Guinier plot (2) calculated from the $P(r)$ function. R_g and $I(0)$ were evaluated with a Guinier fit to the data up to $q_{\max}R_g < 0.8$, the linear range for elongated samples.

Radius of gyration	R_g (Å)	34.28 ± 0.19 ⁽¹⁾ ; 34.45 ± 0.09 ⁽²⁾
Forward scattering intensity	$I(0)$ (cm ⁻¹)	$0.04 \pm 5.71 \cdot 10^{-5}$ ⁽¹⁾ ; $0.04 \pm 4.36 \cdot 10^{-5}$ ⁽²⁾
Maximum dimension	D_{\max} (Å)	139
Porod volume	V_p (nm ³)	76.5

Another sign of the deviation from globularity is the shape of the dimensionless Kratky plot (Figure 3.7B). For a globular (compact) protein, such a plot exhibits a bell-shape with a maximum value of 1.104 for $qR_g = \sqrt{3}$. Here, the maximum value is shifted towards the right and exceeds 1.104. The shape of the plot corresponds well to that of an extended protein composed of several folded domains including flexible linkers of rather compact conformations [24]. The early peak of the $P(r)$ function as well as the long tail appearing at large r (Figure 3.7C) is also indicative of an elongated molecular shape. These observations are in full agreement with the crystallographic dimer structure and the well-documented dynamic behavior of the enzyme, oscillating between open and closed conformations.

Finally, the experimental curve could be compared with the theoretical scattering of *HsPSP* atomic structure. To this end, CRY SOL was used to evaluate and fit the theoretical scattering curve of *HsPSP* dimer to the experimental data. Figure 3.7A shows the fit for the dimer formed by monomer A of 6HYY and its crystallographic neighbor but all available crystallographic dimers were tested and all produced similar results, with χ^2 values comprised between 14.4 and 40.4. The theoretical and experimental curves show the same overall shape but the theoretical data deviate from the experimental data at q values from 0.15 to 0.275 and from 0.35 to 0.5 Å⁻¹. Again, these discrepancies can be explained by the conformational polydispersity of *HsPSP* as the medium angle and high angle scattering provide information about domain motions

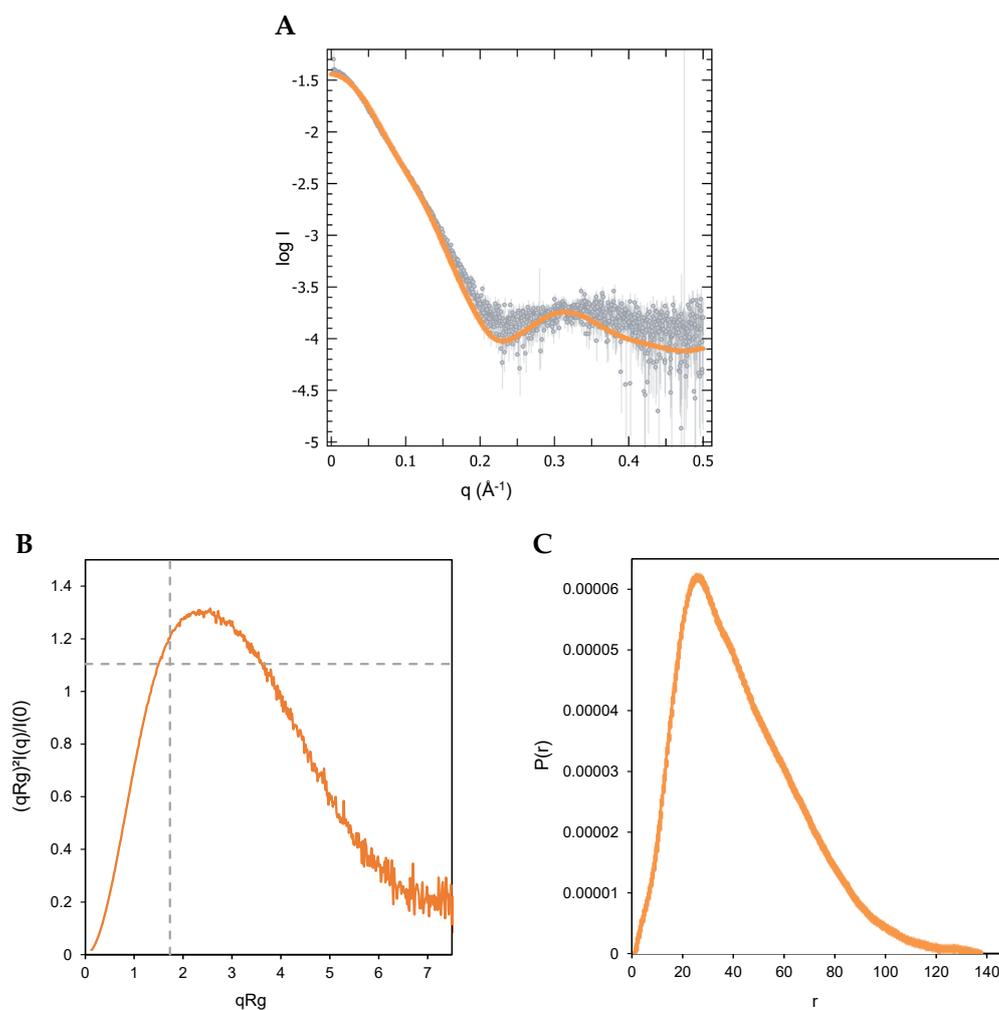


FIGURE 3.7: Analysis of *HsPSP* solution structure from SAXS data. (A) Grey: experimental SAXS profile ($\log I(q)$ versus q) of *HsPSP*. Orange: fit of *HsPSP* crystal dimer structure 6HYY to the experimental data ($\chi^2 = 22.4$). (B) Dimensionless Kratky plot. (C) Calculated $P(r)$ function.

and small structural fluctuations. It can therefore be concluded that the crystallographic structure of *HsPSP* does not perfectly represent its structure in solution. As shown by the SAXS data, the overall shape of the enzyme is conserved. But, in solution, the protomers forming the dimer probably adopt different relative positions and intermediate conformations not sampled in the solid state.

3.3.3 SerB from *M. avium* (*MaSerB*)

SerB from *M. avium* is the only phosphoserine phosphatase with two ACT domains whose atomic structure has been solved. Nine structures are available in the PDB (Table 3.8), the first of which was solved by the Seattle Structural Genomics Center for Infectious Disease (SSGCID) [25]. The biological assembly features a domain-swapped homodimer in which two identical monomeric subunits are associated through C2 symmetry and interchange their N-terminal ACT1 domains. All the structures contain a magnesium ion within the active site and 3 of the entries of Shree *et al.* were co-crystallized with L-serine which is found either in the catalytic site or at the interface of the ACT 1 and ACT 2 domains (Figure 3.8). The 9 structures superimpose very well: the alignment of each of them to structure 3P96 produces RMSD values on C α atoms between 0.23 and 0.47 Å. The differences come exclusively from the conformation of some side chains exposed to the solvent. The positions of the residues composing the active site and interacting with L-Ser at ACT domains are well conserved (see Appendix C). Since the two enzymes share 83.3% sequence identity, *MaSerB* has been used as a template homolog to model *MtSerB2*.

TABLE 3.8: The 9 structures of *MaSerB* deposited in the PDB.

PDB entry	Resolution (Å)	Nature	Ligands	Authors	Year
3P96	2.05	wt	-	SSGCID	2010
5JJB	2.31	D343G	-	Shree <i>et al.</i>	2016
5T41	3.15	S275A/R279A	-	Shree <i>et al.</i>	2016
5IT0	1.97	D343N/D347G	-	Shree <i>et al.</i>	2016
5JMA	2.03	wt	L-Ser (PSP dom.)	Shree <i>et al.</i>	2016
5IT4	2.10	D343N	-	Shree <i>et al.</i>	2016
5IS2	1.88	wt	-	Shree <i>et al.</i>	2016
5JLR	2.26	wt	L-Ser (ACT dom.)	Shree <i>et al.</i>	2016
5JLP	2.50	wt	L-Ser (ACT dom.)	Shree <i>et al.</i>	2016

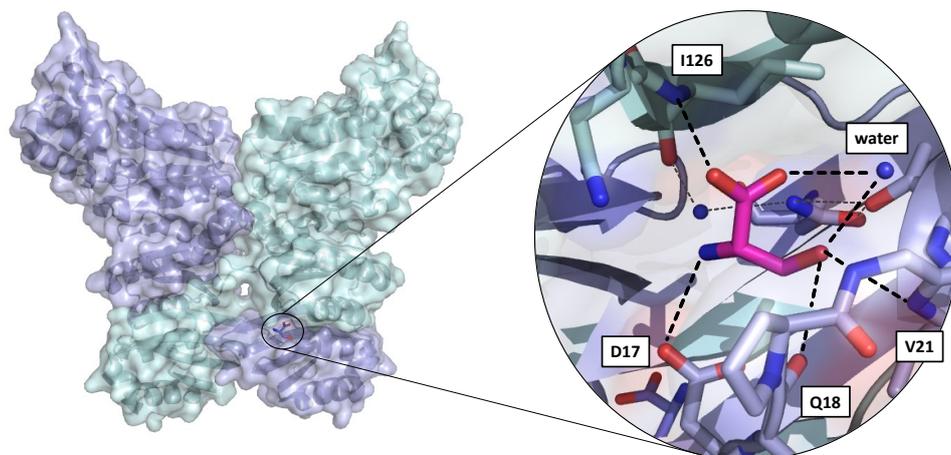


FIGURE 3.8: Crystallographic structure of *MaSerB* (PDB: 5JLR) and focus on the L-Ser binding site located at the ACT1-ACT2 interface. The monomers that form the dimer are colored in different shades of blue to highlight ACT1 domain-swapping. As the dimer is symmetric, another L-Ser molecule is found at the same site on the other side of the enzyme. Hydrogen bonds are depicted as dotted lines.

Investigation of *MaSerB* structure in solution

To complete the knowledge brought by the crystalline state of *MaSerB*, we also investigated its form in solution. First, SEC-MALS allowed to identify the oligomeric state of the enzyme. As can be seen on the SEC-UV chromatogram in Figure 3.9, the enzyme elutes as a major peak around 15 minutes (94% of area under the curve, AUC) followed by a small bump (6% AUC). Using the MALS detector, the main peak could be associated with the elution of *MaSerB* dimer, with a measured molar mass of 85.5 kDa (MW_{monomer} : 43.9 kDa). Measurement of the molar mass associated with the second peak resulted in a value of 69.8 kDa. However, it should be kept in mind that the very low concentration of this species makes the estimate less accurate. It cannot be decided unambiguously whether this small peak corresponds to the elution of a monomer or of a truncated form of the enzyme.

The solution structure of *MaSerB* was also investigated by SEC-SAXS. Size parameters were determined from the experimental curve of Figure 3.10A and are listed in Table 3.9. The experimental R_g value is slightly higher than the theoretical value of 28.1 Å calculated by Equation 3.1 for a perfectly globular 88 kDa protein. The molecular weight values calculated from Bayesian inference and from Porod volume, respectively 91.2 kDa (CI: 84.3 to 95.8) and 104.7 kDa, are in line with the SEC-MALS

results and confirm the dimeric state of the enzyme in solution.

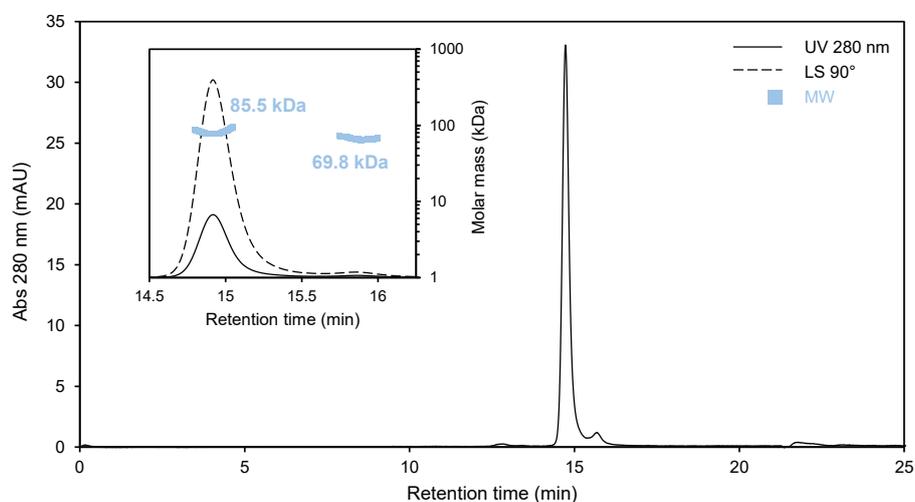


FIGURE 3.9: SEC-UV-MALS analysis of *MaSerB*.

TABLE 3.9: SAXS parameters of *MaSerB*. (1) derived from the Guinier plot
(2) calculated from the $P(r)$ function

Radius of gyration	R_g (Å)	32.62 ± 0.03 ⁽¹⁾ ; 32.84 ± 0.02 ⁽²⁾
Forward scattering intensity	$I(0)$ (cm ⁻¹)	$0.11 \pm 5.99 \cdot 10^{-5}$ ⁽¹⁾ ; $0.11 \pm 5.0 \cdot 10^{-5}$ ⁽²⁾
Maximum dimension	D_{\max} (Å)	101
Porod volume	V_p (nm ³)	126

Using CRY SOL, we were also able to calculate the theoretical scattering curve from *MaSerB* crystallographic structure and compare it to the SAXS data. The calculated curve is depicted in light blue in Figure 3.10 and although its global profile is similar to that of the experimental data, a systematic deviation is observed from $q = 0.125 \text{ \AA}^{-1}$. This discrepancy can be explained by the adoption of slightly different conformations in solution and in the crystalline state, a phenomenon often encountered in multi-domain proteins. Refinement of the crystallographic structure against SAXS data allowed us to visualize the extent of such conformational change. *MaSerB* 3P96 structure was submitted to the program DADIMODO and four conformers better fitting the experimental data were obtained. One of them, shown as an example in Figure 3.10B, superimposes with a RMSD on C_α atoms of 3.11 \AA over 760 residues to structure 3P96. As can be seen in Figure 3.10A, its theoretical scattering curve (dark blue) aligns better with the experimental data ($\chi^2 = 1.3$) than that of structure 3P96

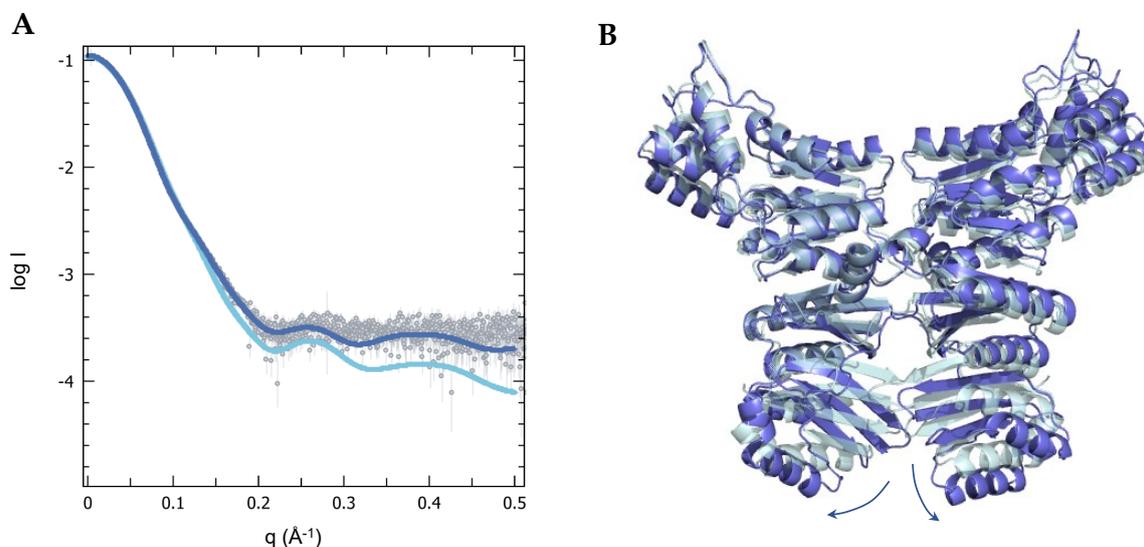


FIGURE 3.10: Analysis of *MaSerB* solution structure from SAXS data. (A) Grey: experimental SAXS profile ($\log I(q)$ versus q) of *MaSerB*. Light blue: fit of *MaSerB* crystal dimer structure 3P96 to the experimental data ($\chi^2 = 21.8$). Dark blue: fit of a *MaSerB* model generated by DADIMODO to the experimental data ($\chi^2 = 1.3$) (B) Superimposition of *MaSerB* crystal structure and model generated by DADIMODO.

($\chi^2 = 21.8$). This conformer illustrates that a slight displacement of the ACT1 domains with respect to the rest of the structure can explain the difference observed between the theoretical and experimental curves from $q = 0.125 \text{ \AA}^{-1}$.

3.3.4 SerB2 from *M. marinum* (*MmSerB2*)

Enzymological characterization of *MmSerB2*

Among the orthologs studied in this work, *MmSerB2* has the closest sequence to *MtSerB2*. This observation is not surprising when considering that *M. marinum* and *M. tuberculosis* share over 3000 orthologous genes with an average amino acid identity of 85% [26]. We published an enzymological characterization of *MmSerB2* and comparison with *MtSerB2* in *Biochemical and Biophysical Research Communications* in 2020 [27]. The paper is presented in Appendix D.

In that publication, we showed that *MmSerB2* has an enzymological behavior almost identical to that reported for *MtSerB2* in references [2, 3, 28]. Using *O*-phospho-L-serine (PS) as the substrate, the highest phosphatase activity was observed at 42 °C

and pH 7.6 in the presence of a Mg^{2+} cofactor. The enzyme was also active, although to a lesser extent, in the presence of Mn^{2+} and inactive in the apo form or with other divalent cations (Ca^{2+} , Cu^{2+} , Ni^{2+} , Zn^{2+}).

Next, steady-state kinetics experiments demonstrated that *MmSerB2* and *MtSerB2* catalyzed the hydrolysis of PS to L-serine at comparable rates of about 12 s^{-1} (k_{cat}). However, in terms of K_M , *MmSerB2* was at least twofold less affine for PS but was inhibited by the latter at lower concentrations than *MtSerB2*. We found out that this kinetic behavior correlated well with *MmSerB2* being a homodimeric enzyme most active when only one of its two active sites is occupied by the substrate. The model was also coherent with the possible interaction of PS (or newly synthesized L-serine) at an allosteric inhibitory site. We hypothesized that it could be the site at the ACT1/ACT2 domain interface where L-serine is known to bind in *MaSerB* (PDB: 5JLP and 5JLR) and exert a regulatory effect in *MtPGDH*. The behavior of *MtSerB2* non-substrate inhibited mutant G18A/G108A reported by Grant [3] and docking studies of PS in that pocket supported our hypothesis (Figure 3.11).

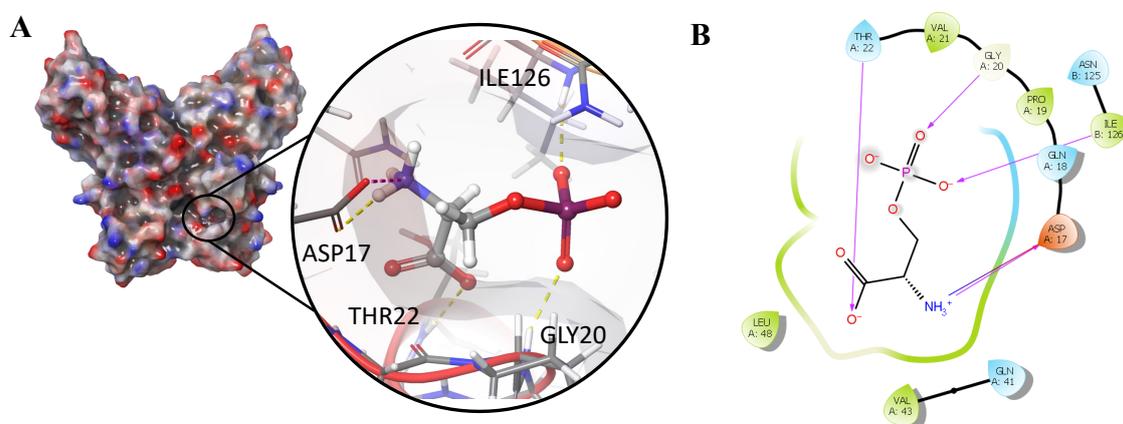


FIGURE 3.11: Docking of *O*-phospho-L-serine (PS) at pH 7.4 in the L-serine binding site of *MmSerB2* homology model. (A) Structure of docked PS in the L-serine binding site at the interface of ACT1 (of monomer 1) and ACT2 (of monomer 2) domains. (B) 2D diagram of the interactions between docked PS and the residues of the L-serine binding site. H-bonds are depicted in pink. The salt bridge is depicted in a gradient from blue to red.

Investigation of *MmSerB2* structure in solution

Like the other homologs, we had the opportunity to further characterize *MmSerB2* by light and X-ray scattering techniques. The use of SEC-MALS first gave us insight into the oligomeric state of *MmSerB2*. The SEC-UV presented in Figure 3.12 shows that the enzyme has an elution profile very similar to that of *MaSerB*. Two distinct peaks appear: a major peak eluting just before 15 min (95% AUC) followed by a small bump (5% AUC). Again, MALS allowed the association of the main peak with a dimeric form of the enzyme (measured molar mass: 88.2 kDa, theoretical MW_{monomer}: 43.8 kDa). The small peak, associated to a measured molar mass of 66.0 kDa, could also correspond to the elution of a monomer or of a truncated form of the enzyme.

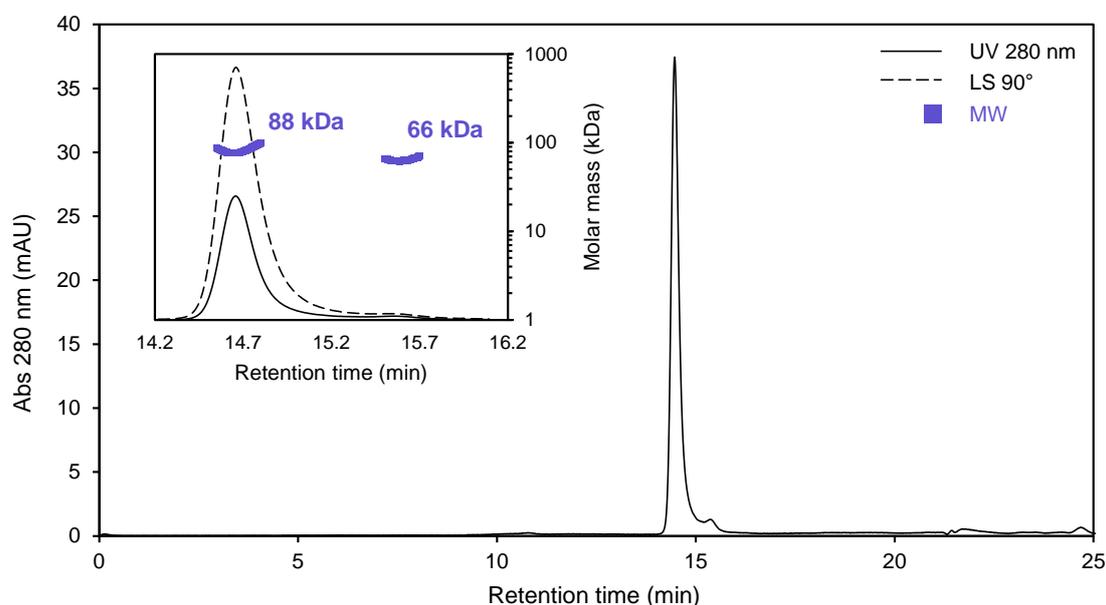


FIGURE 3.12: SEC-UV-MALS analysis of *MmSerB*.

MmSerB2 was also analyzed by SEC-SAXS. The experimental scattering curve is presented in Figure 3.13A and the derived parameters are shown in Table 3.10.

No experimental atomic structure of *MmSerB2* is available at this time. In our publication, a homology model has been constructed based on the structure of *MaSerB* (PDB: 3P96) to illustrate and discuss the results of the enzymological characterization. The very high percentage of sequence identity between *MmSerB2* and *MaSerB* (83.0 %) is already a good argument for the reliability of the model, but the comparison of SEC-SAXS data provides further experimental evidence that the two enzymes fold

TABLE 3.10: SAXS parameters of *MmSerB*. (1) derived from the Guinier plot (2) calculated from the $P(r)$ function

Radius of gyration	R_g (Å)	32.32 ± 0.03 ⁽¹⁾ ; 32.54 ± 0.02 ⁽²⁾
Forward scattering intensity	$I(0)$ (cm^{-1})	$0.1 \pm 5.74 \cdot 10^{-5}$ ⁽¹⁾ ; $0.1 \pm 4.84 \cdot 10^{-5}$ ⁽²⁾
Maximum dimension	D_{max} (Å)	99
Porod volume	V_p (nm^3)	125

almost identically in solution. In Figure 3.13A, we can see that the SAXS curve of *MmSerB2* overlaps precisely with that of *MaSerB2*. Consequently, the parameters determined for the two enzymes are almost identical. The extent of shape similarity is even better represented by the superimposition of the $P(r)$ functions (Figure 3.13C). The dimensionless Kratky plots are also identical and show the typical bell-shaped allure obtained for a well-folded globular protein, with a maximum value of 1.104 at $qRg = \sqrt{3}$.

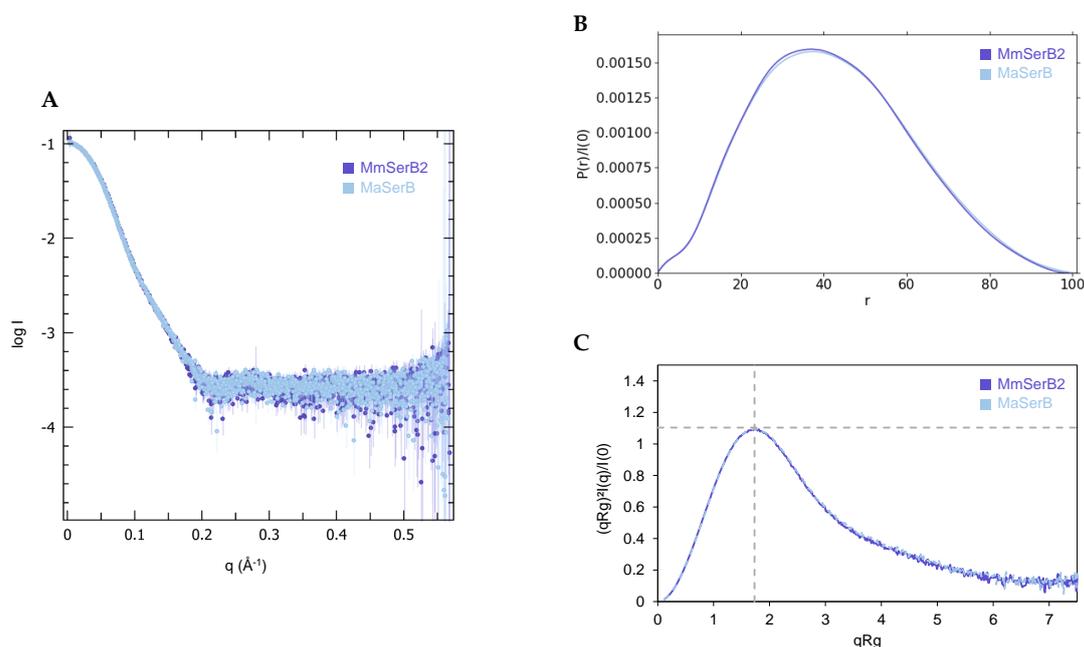


FIGURE 3.13: Analysis of *MmSerB2* solution structure from SAXS data and comparison with *MaSerB*. (A) Experimental SAXS profiles ($\log I(q)$ versus q) of *MmSerB2* and *MaSerB*. (B) Calculated $P(r)$ functions. (C) Dimensionless Kratky plots.

3.3.5 SerB from *B. melitensis* (*BmSerB*)

Bacteria of the genus *Brucella* are intracellular pathogens that cause a chronic infection leading to sterility and abortion in animals, and debilitating disorders in humans. To the best of our knowledge, there is no specific mention of SerB from *Brucella melitensis* in the literature. However, its ortholog in *B. abortus* (99% sequence identity) has been demonstrated to be important to intracellular proliferation and virulence. Revora *et al.* indeed showed that *B. abortus serB* deletion mutant was auxotrophic for L-Ser, unable to replicate intracellularly and was attenuated in mice in the acute and chronic phases of infection [29]. As for *MtSerB2*, the determination of the structure of *Brucella* SerB is of therapeutic interest.

Determination of the structure of *BmSerB* (on-hold PDB entry 7QPL)

BmSerB possesses a 90 residues sequence upstream of its PSP domain. The nature of the domain family to which this sequence corresponds is difficult to predict via sequence-based only methods. Since the structure of this enzyme had never been solved, crystallization assays were undertaken to shed light on the identity of the fold.

The enzyme was produced as explained in section 3.3.1, then concentrated at 25.7 mg mL^{-1} . Sitting drop vapor diffusion crystallogensis assays were performed under the 96 conditions of three different crystallization screens from *Molecular Dimensions*: PACT Premier, JSCG Plus and BCS. On day 5 after the set up of the experiment, dodecahedral crystals (Figure 3.14) were observed in condition 1-40 from the BCS screen. The solution consisted of 0.2 M sodium formate, 0.1 M sodium phosphate pH 6.2, 20% w/v PEG smear medium and 10% glycerol.

Crystals of the same morphology were easily reproduced in 24-well hanging drop vapor diffusion experiments using conditions close to those of the hit (pH 6.2-6.8, 14-24% w/v PEG smear medium). Among the crystals analyzed by X-ray diffraction, the best crystal grew in 0.2 M sodium formate, 0.1 M sodium phosphate pH 6.8, 22% w/v PEG smear medium and 10% glycerol and diffracted to a resolution of 1.77 \AA . It was found to belong to the cubic space group $I2_13$, with unit cell parameters $a = b = c = 143.21 \text{ \AA}$. Initial phases were calculated by molecular replacement using two parts of a *BmSerB* model predicted by Alphafold2 as search models. The structure was then built and refined to best fit the electron density to R_{work} and R_{free} values of 15.8% and

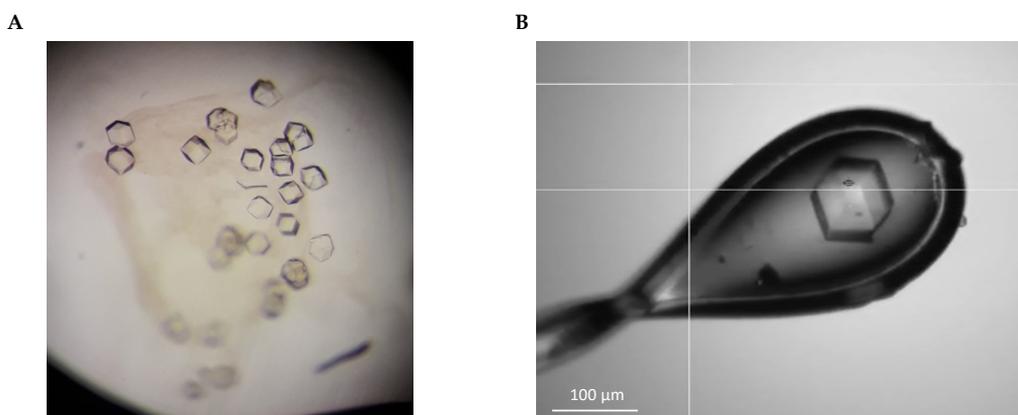


FIGURE 3.14: *BmSerB* crystals. (A) As obtained in condition 1-40 of the BCS Screen. (B) Mounted in a 0.3 mm CryoLoop placed in the cryostream for data collection. The crystal measures about 85 μm.

18.3% respectively. The data collection and refinement statistics are given in Appendix E.

The crystal structure (Figure 3.15) shows only one molecule in the asymmetric unit. As expected from sequence alignments with the orthologous PSPs, it is readily noticed that the enzyme is composed of two distinct domains linked by a short 7-residues loop:

- A large domain spanning from residues Ile90 to Gln303 and corresponding to the **PSP domain** of the enzyme. As predicted by the Pfam search, this domain presents the typical Rossmann-like fold of the HAD phosphatases: a 3-layer $\alpha/\beta/\alpha$ sandwich with a central 6-strands parallel β -sheet in 3-2-1-4-5-6 order. Access to the active site, where Mg^{2+} and a phosphate ion are found, is shielded by a tetrahelical C1 cap module similar to those found in *HsPSP* and *MaSerB*. The cap is located between the strand β_1 and helix α_1 of the Rossmannoid fold. A loop containing two anti-parallel beta strands is also inserted between strand β_3 and helix α_3 of the Rossmannoid fold.
- A small domain spanning from residues Val9 to Glu82 and presenting the characteristic architecture of an **ACT domain**: a $\beta\alpha\beta\beta\alpha\beta$ fold featuring a $\beta_4\beta_1\beta_3\beta_2$ antiparallel β -sheet flanked by a pair of antiparallel α -helices presenting a twist angle of 37.5° with respect to each other.

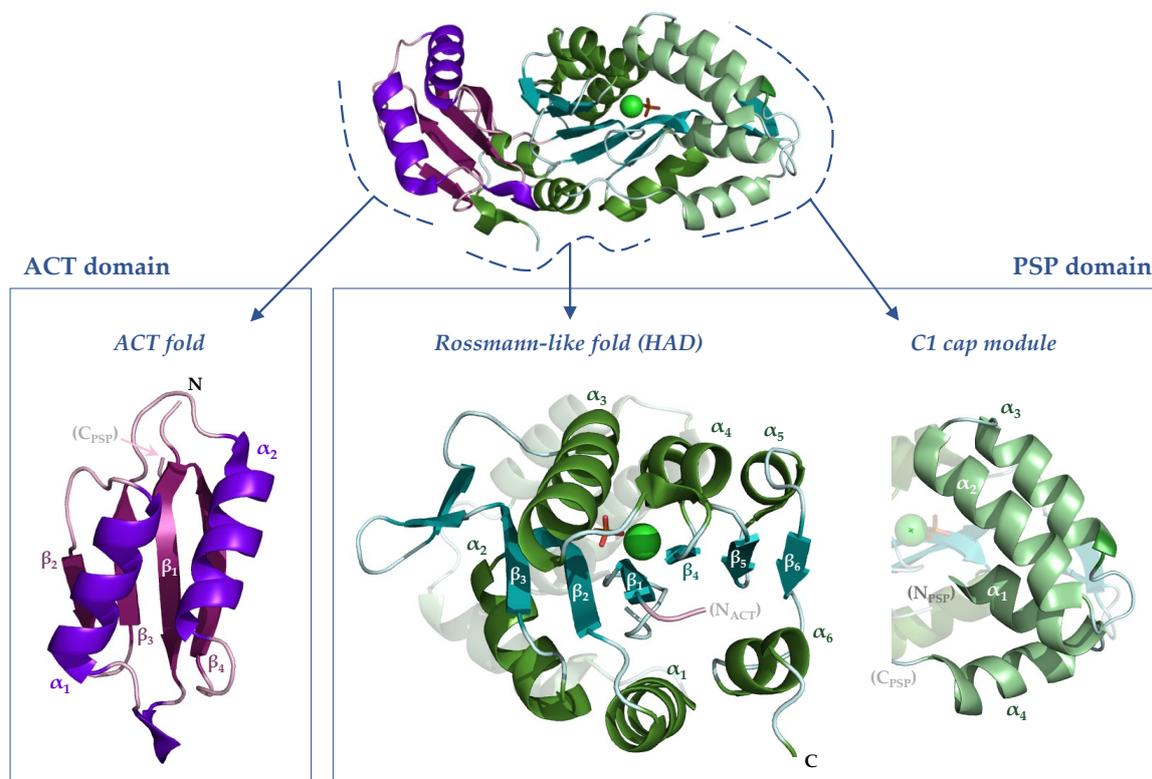


FIGURE 3.15: Crystallographic structure of *BmSerB* and focus on its domains and folds. N = N-terminus, C = C-terminus, (N_x) and (C_x) = direction of the N- or C- terminus and adjacent fold.

BmSerB crystallized with a phosphate ion and a divalent magnesium ion in its active site (Figure 3.16). Both interact with the conserved residues of the typical HAD phosphatase motifs I, II, III and IV highlighted in Figure 3.2. The Mg^{2+} ion is hexacoordinated by two water molecules, two oxygen atoms from the side chains of D251 (motif IV) and D94 (motif I), the oxygen atom from the backbone carbonyl of D96 (motif I) and an oxygen atom from the phosphate ion. The latter interacts mainly by hydrogen bonding with the sides chains of S183 (motif II), D94 and D96 (motif I), N254 (motif IV), the oxygen atom from the backbone carbonyl of D96 and three water molecules. It also forms a salt bridge with K228 (motif III) side chain.

The structure superimposes quite well with that of the other two crystallized orthologs studied in this work (Figure 3.17, left). RMSD values on C_α atoms for the alignments with the whole structures and domains of *MaSerB* (PDB: 3P96) and *HsPSP* (PDB: 6HYY) are reported in Table 3.11. Key residues from conserved HAD motifs

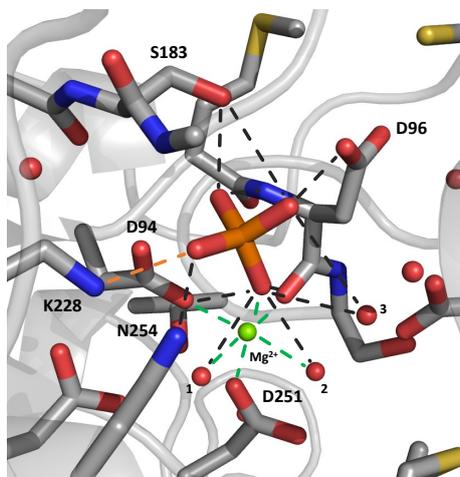


FIGURE 3.16: Structure of the active site of *BmSerB*. Interactions of the phosphate ion (orange) and the magnesium ion (green) with the active sites residues are depicted in dotted lines: hydrogen bonds (black), salt bridge (orange) and coordinate bonds (green).

I-IV overlap well (Figure 3.17, right). The magnesium ions (or calcium ion for *HsPSP* monomer B) are found at close positions in all structures, and the phosphate ion of *BmSerB* is superimposed with that of *HsPSP* monomer B and the chlorine ion of *MaSerB*. In *HsPSP* monomer A, the phosphate ion occupies a different position because the enzyme is in an alternative conformation [12].

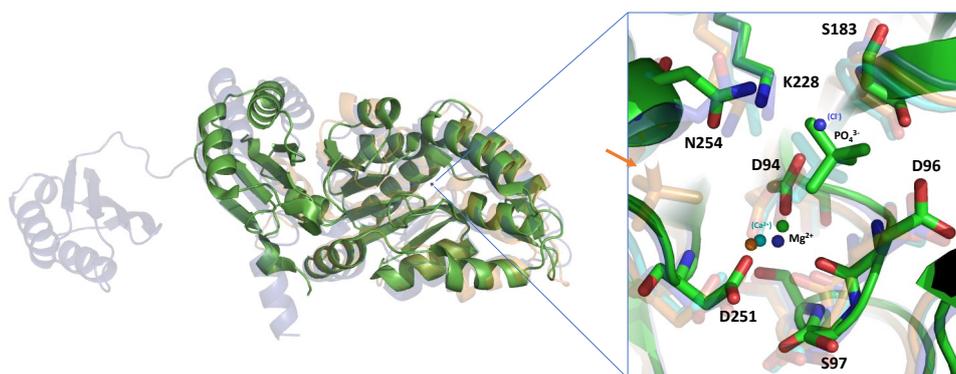


FIGURE 3.17: Structure alignment of *BmSerB*, *MaSerB* (PDB: 3P96, dark blue) and *HsPSP* (PDB: 6HYY, orange). Left: global alignment. Right: focus on the active site. *HsPSP* monomers A and B are depicted in orange and cyan respectively. The position of phosphate ion in monomer A is indicated by an arrow.

TABLE 3.11: RMSD values for the structure alignment of *BmSerB* with *MaSerB* (PDB: 3P96) and *HsPSP* (PDB: 6HYY)

<i>BmSerB</i>	aligned with	RMSD on C_{α} atoms (Å)
Whole structure	<i>MaSerB</i> (whole)	2.61 (over 288 residues)
	<i>HsPSP</i> (whole)	2.59 (over 208 residues)
PSP domain	<i>MaSerB</i> PSP	1.33 (over 208 residues)
	<i>HsPSP</i> monomer A	2.51 (over 208 residues)
	<i>HsPSP</i> monomer B	3.36 (over 176 residues)
ACT domain	<i>MaSerB</i> ACT1	3.42 (over 64 residues)
	<i>MaSerB</i> ACT2	3.08 (over 64 residues)

Upon examination of other PSP structures deposited on the PDB, we found two that were similar to *BmSerB*: the PSP from *Vibrio cholerae* (PDB: 3N28) and the HAD hydrolase t0658 from *Salmonella enterica* (PDB: 4EZE). The three structures are superimposed in Figure 3.18. These two enzymes also feature an additional N-term domain with a resembling ACT topology in addition to their PSP domain. Structure alignment of those domains with the ACT domain of *BmSerB* respectively give RMSD values on C_{α} atoms of 3.16 Å over 64 residues and 3.47 Å over 56 residues for 4EZE and 3N28.

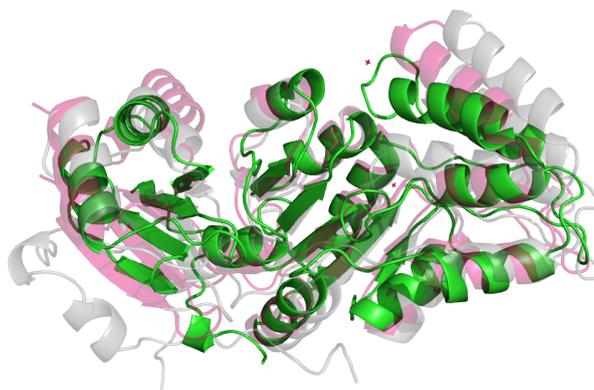


FIGURE 3.18: Structure alignment of *BmSerB* with the PSP from *V. cholerae* (PDB: 3N28, grey) and the HAD hydrolase t0658 from *S. enterica* (PDB: 4EZE, magenta).

Investigation of *BmSerB* structure in solution

Although *BmSerB* crystallized with a single molecule in the asymmetric unit, it does not mean that it is necessarily monomeric. However, biological interactions with the crystallographic neighbors are difficult to discriminate from crystalline contacts. SEC-SAXS and SEC-MALS allowed us to unambiguously determine the oligomeric state of the enzyme and characterize its structure in solution.

SEC-MALS was used to measure the molecular weight of the *BmSerB*. As shown on the SEC-UV chromatogram (Figure 3.19), the enzyme elutes in a single peak corresponding to a molecular weight of 32.1 kDa as measured by MALS. Considering that the theoretical weight for a monomer is 32.5 kDa, this indicates that the enzyme exists only as a monomer in solution.

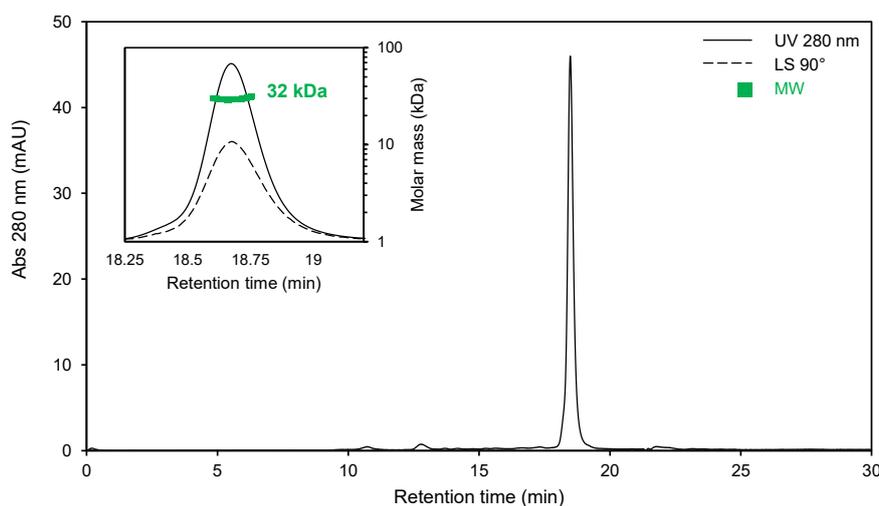


FIGURE 3.19: SEC-UV-MALS analysis of *BmSerB*.

BmSerB might have been expected to form a dimer like *HsPSP* in solution as it also possess the β -hairpin loop that allows the latter to dimerize. However, based on the sequence alignment of Figure 3.2, it can be seen that *BmSerB* loop is four residues shorter and contains only two of the ten residues known to be involved in *HsPSP* dimeric interface. This deletion/substitution combination is probably the reason why *BmSerB* does not dimerize.

SEC-SAXS allowed us to further characterize *BmSerB* solution structure. SAXS parameters (Table 3.12) were determined from the experimental scattering curve (Figure 3.20A). The R_g value is well in line with the theoretical value for a 32 kDa globular

protein (20.2 Å) calculated with Equation 3.1. The bell-shaped dimensionless Kratky plot and P(r) function (Figure 3.20C and D) also show that the enzyme is well folded and rather globular in solution. Molecular weight calculated from Bayesian inference and from Porod volume were of 28.9 kDa (confidence interval: 27.2 to 29.2 kDa) and 29.9 kDa respectively. These results confirm the monomeric state identified by SEC-MALS.

TABLE 3.12: SAXS parameters of *BmSerB*. (1) derived from the Guinier plot (2) calculated from the P(r) function

Radius of gyration	R_g (Å)	21.36 ± 0.03 ⁽¹⁾ ; 21.52 ± 0.05 ⁽²⁾
Forward scattering intensity	$I(0)$ (cm ⁻¹)	$0.04 \pm 2.77 \cdot 10^{-5}$ ⁽¹⁾ ; $0.04 \pm 3.28 \cdot 10^{-5}$ ⁽²⁾
Maximum dimension	D_{\max} (Å)	80
Porod volume	V_p (nm ³)	55.2

In addition to these parameters, the SAXS curve provided the opportunity to confront the crystallographic structure described above with the shape of *BmSerB* in solution. To this end, the scattering profile of *BmSerB* crystallographic structure was evaluated (light green curve) and fitted to the experimental scattering curve (gray dots) using CRY SOL (Figure 3.20A). At first sight, the allure of the simulated curve corresponds well to that of the experimental curve. We notice however a slight systematic shift starting around $q = 0.175 \text{ \AA}^{-1}$ which can be explained by a conformational difference between the crystal structure and the enzyme in solution. To reflect this change, the crystallographic structure was submitted to SREFLEX. The program outputted a set of 10 models representing possible conformational changes that improved the agreement with the experimental curve. The improvement is reflected in the χ^2 value which went from 3.91 for the crystal structure to values between 1.02 and 1.13 for the generated conformers. The model corresponding to the best fit is superimposed on *BmSerB* crystallographic structure in Figure 3.20B (RMSD on C_α atoms: 4.39 Å over 288 residues). It can be seen that the agreement with the experimental curve can be improved by simply moving loops and changing the relative position of the ACT domain with respect to the PSP domain.

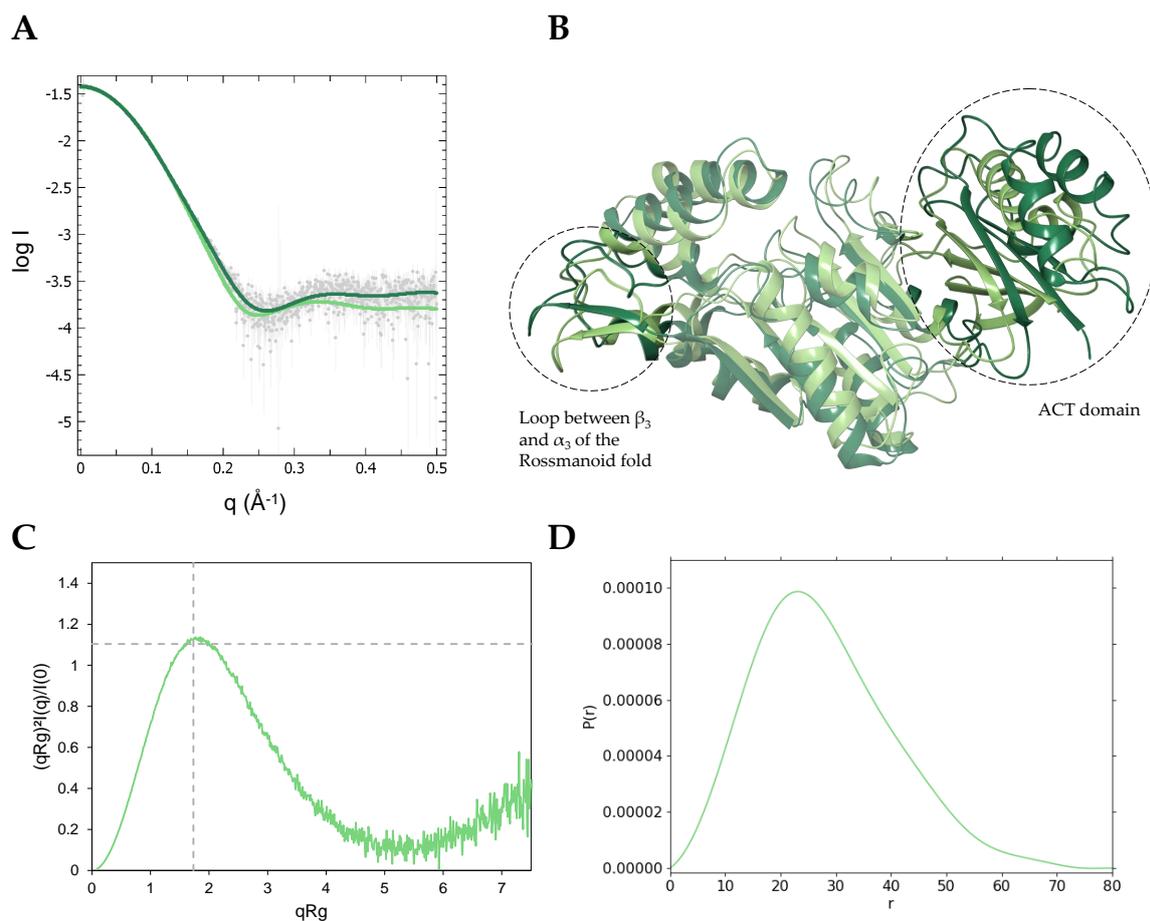


FIGURE 3.20: (A) Grey: experimental SAXS profile ($\log I(q)$ versus q) of *BmSerB*. Light green: fit of *BmSerB* crystal structure to the experimental data ($\chi^2 = 3.91$). Dark green: fit of a *BmSerB* model generated by SREFLEX to the experimental data ($\chi^2 = 1.02$) (B) Superimposition of *BmSerB* crystal structure and model generated by SREFLEX. (C) Dimensionless Kratky plot. (D) Calculated $P(r)$ function.

Enzymological characterization of *BmSerB* and behavior in the presence of L-serine

Solving *BmSerB* structure allowed us to identify the presence of a distinct N-terminal domain possessing the topology of an ACT domain. However, in addition to their characteristic architecture, the ACT domains are also defined by a functional feature: they elicit an allosteric response in the fused catalytic domain upon binding a small molecule. The associated binding site is generally found at the interface between ACT domains, as the latter work in concert [6].

To verify the effector function of the ACT domain identified in *BmSerB*, we examined its oligomeric and catalytic behavior in the presence of L-serine.

SEC chromatograms showed that *BmSerB* was eluted around at the same retention time in the presence and absence of 2.5 mM L-Ser. The molecular weights associated with these peaks were determined by MALS and values of 32.2 kDa without L-serine and 34.5 kDa value with L-serine were obtained. These results show that, under these buffer conditions, L-Ser does not affect the monomeric state of *BmSerB*.

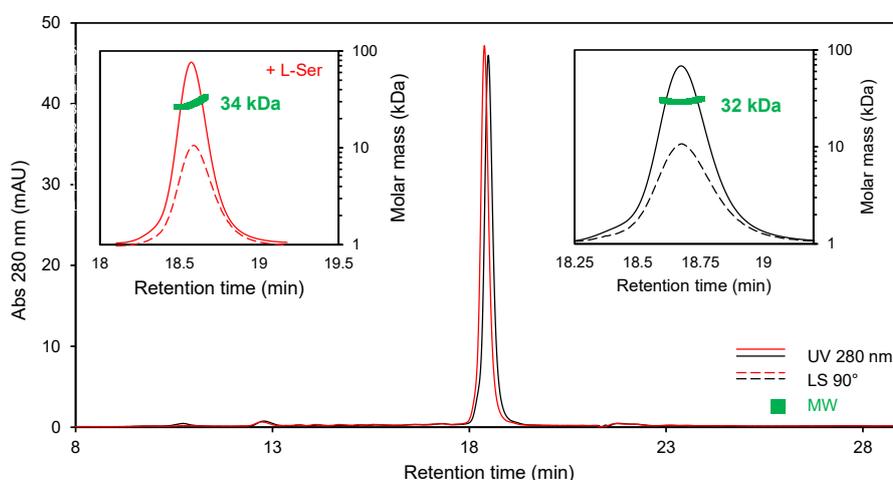


FIGURE 3.21: SEC-UV-MALS analysis of *BmSerB* in the presence and absence of 2.5 mM L-Ser in the mobile phase (molar ratio 1:100).

The kinetic behavior of *BmSerB* was also assessed at different L-Ser concentrations using the phosphatase assay. First, the steady-state kinetics of the enzyme in the absence of L-Ser were studied (Figure 3.22A). As can be seen, the velocity curve rises to a maximum and then decreases as the concentration of the substrate (*O*-phospho-L-serine, PS) is increased. This deviation from typical Michaelis-Menten behavior

is characteristic of the inhibition of the enzyme by its substrate. The equation that accounts for total uncompetitive substrate inhibition (Equation 3.2) could be fitted to the data.

$$v = \frac{V_{max}[S]}{K_M + [S](1 + \frac{[S]}{K_{iS}})} \quad (3.2)$$

In Equation 3.2, V_{max} is the maximum velocity, $[S]$ is the free substrate (*O*-phospho-L-serine, PS) concentration, K_M is the Michaelis constant (substrate dissociation constant) and K_{iS} is the dissociation constant of the substrate for the inhibitory site. The value of the catalytic constant k_{cat} is obtained by dividing V_{max} by the total enzyme concentration $[E]_t$.

It indicates that a second molecule of PS probably binds to *BmSerB*-PS complex at a non-catalytic inhibitory site and leads to the formation of a non-productive PS-*BmSerB*-PS complex³. The dissociation constants of PS for both sites as well as the reaction rate of *BmSerB*-PS complex were determined by non-linear regression and are reported in Table 3.13. Figure 3.22B shows the velocity curve obtained in the presence of increasing L-Ser concentrations. As reflected by the superposition of the 6 curves, L-Ser does not seem to modulate the catalytic activity of *BmSerB*.

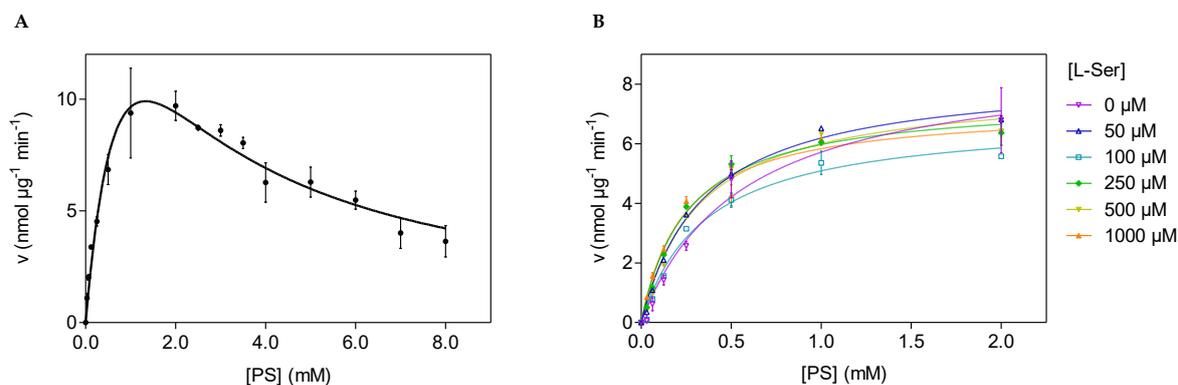


FIGURE 3.22: Velocity curves for the dephosphorylation of *O*-phospho-L-serine (PS) by *BmSerB* (A) in the absence of L-Ser (B) in the presence of increasing L-Ser concentrations

³See section 4.3.1 of Chapter 4 for more details about substrate inhibition

TABLE 3.13: Kinetic constants for the dephosphorylation of *O*-phospho-L-serine by *BmSerB* in the absence of L-Ser

Dissociation constant of PS for catalytic site	K_M (mM)	1.23 ± 0.44
Dissociation constant of PS for inhibitory site	K_{iS} (mM)	1.44 ± 0.49
Maximum reaction velocity	V_{\max} (nmol μg^{-1} min $^{-1}$)	28.24 ± 7.10
Catalytic constant	k_{cat} (s $^{-1}$)	15.30 ± 3.85

Since L-Ser neither seems to modulate the catalytic activity nor to participate in a change in the oligomerization state of *BmSerB*, we can only refer to its ACT domain as "putative" [6]. However, it remains possible that other yet untested ligands or a pH variation could highlight a functional role for this putative ACT domain.

3.4 Summary of the findings

To better understand the behavior of *MtSerB2*, the biophysical and biochemical properties of four orthologous enzymes were studied. The orthologs were selected on the basis of their sequence. These are either very similar with only a few point differences (*MaSerB*, *MmSerB2*) or are shorter by one (*BmSerB*) or more domains (*HsPSP*). These characteristics will allow us to highlight the role of domains or residues in the biochemical and biophysical behavior of *MtSerB2*.

The use of SEC-MALS and SEC-SAXS allowed us to shed light on the quaternary architecture and conformation of each ortholog in solution. In the case of *BmSerB*, X-ray crystallography has also given us access to an atomic level of detail. Comparing its crystallographic structure, as well as those of *MaSerB* and *HsPSP* to their solution structures illustrated that small conformational changes such as the displacement of domains or loops was common between the solution state and the solid state. The diversity in the morphologies of the studied homologs also highlighted that SAXS, already through the shape of the diffusion curve, the Kratky plot, the $P(r)$ function and the values of R_g and D_{max} , is a rich source of information regarding the enzyme shape.

Data on the hydrolysis kinetics of *O*-phospho-L-serine by *BmSerB* were also presented and showed that its putative ACT domain did not appear to be involved in the regulation of phosphatase activity in the presence of L-Ser. Similar experiments were also performed for *MmSerB2* and *MaSerB* and will be presented and discussed later, in parallel to *MtSerB2*.

The key information described in this chapter and useful for further discussion is summarized in Figure 3.23.

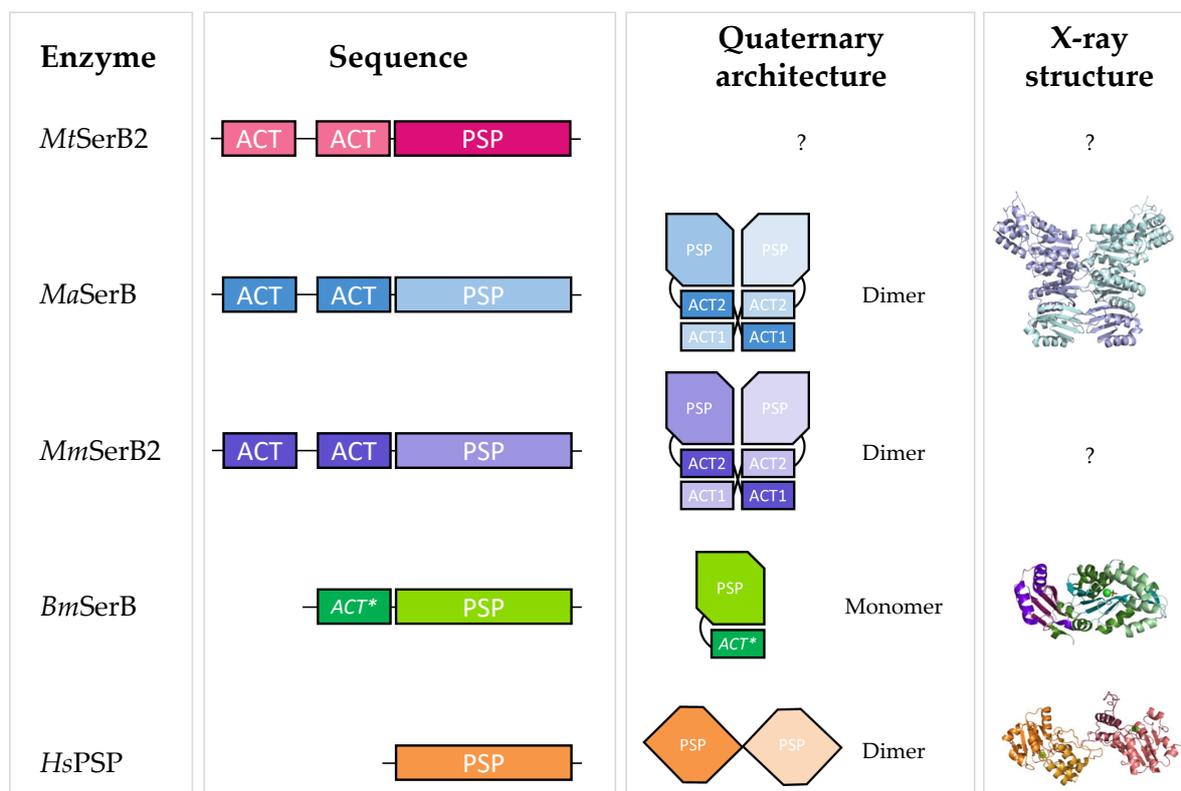


FIGURE 3.23: Graphical summary of structural differences studied in *MtSerB2* orthologs in this chapter. The presence of an asterisk refers to a putative ACT domain.

Bibliography

- [1] R. Choi, A. Kelley, D. Leibly, S. Nakazawa Hewitt, A. Napuli, and W. Van Voorhis. Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 67(9):998–1005, 2011.
- [2] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [3] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.

- [4] S. Shree and R. Ramachandran. Crystal structure of *Mycobacterium avium* SerB2 in complex with serine at ACT domain. doi:10.2210/pdb5JLP/pdb. 2016.
- [5] S. Shree and R. Ramachandran. Crystal structure of *Mycobacterium avium* SerB2 with serine present at slightly different position near ACT domain. doi:10.2210/pdb5JLR/pdb. 2016.
- [6] E.J.M. Lang, P.J. Cross, G. Mittelstädt, G.B. Jameson, and E.J. Parker. Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. *Current Opinion in Structural Biology*, 29:102–111, 2014.
- [7] A. Seifried, J. Schultz, and A. Gohla. Human HAD phosphatases: structure, mechanism, and roles in health and disease. *The FEBS journal*, 280(2):549–571, 2013.
- [8] A.M. Moro-Furlani, V.S. Turner, and D.A. Hopkinson. Genetical and biochemical studies on human phosphoserine phosphatase. *Annals of Human Genetics*, 43(4):323–333, 1980.
- [9] K. Taranath Shetty et al. Phosphoserine phosphatase of human brain: partial purification, characterization, regional distribution, and effect of certain modulators including psychoactive drugs. *Neurochemical Research*, 15(12):1203–1210, 1990.
- [10] Y. Peeraer, A. Rabijns, C. Verboven, J-F. Collet, E. Van Schaftingen, and C. De Ranter. High-resolution structure of human phosphoserine phosphatase in open conformation. *Acta Crystallographica Section D: Biological Crystallography*, 59(6):971–977, 2003.
- [11] J-F. Collet, M.H. Gerin, I. and Rider, M. Veiga-da Cunha, and E. Van Schaftingen. Human L-3-phosphoserine phosphatase: sequence, expression and evidence for a phosphoenzyme intermediate. *FEBS Letters*, 408(3):281–284, 1997.
- [12] M. Haufroid, M. Mirgaux, L. Leherte, and J. Wouters. Crystal structures and snapshots along the reaction pathway of human phosphoserine phosphatase. *Acta Crystallographica Section D: Structural Biology*, 75(6):592–604, 2019.
- [13] L. Leherte, M. Haufroid, M. Mirgaux, and J. Wouters. Investigation of bound and unbound phosphoserine phosphatase conformations through Elastic Network

- Models and Molecular Dynamics simulations. *Journal of Biomolecular Structure and Dynamics*, 39(11):3958–3974, 2021.
- [14] H-Y. Kim, Y-S. Heo, J.H. Kim, M.H. Park, J. Moon, E. Kim, D. Kwon, J. Yoon, D. Shin, E-j. Jeong, et al. Molecular basis for the local conformational rearrangement of human phosphoserine phosphatase. *Journal of Biological Chemistry*, 277(48):46651–46658, 2002.
- [15] Y. Peeraer, A. Rabbijns, J-F. Collet, E. Van Schaftingen, and C. De Ranter. How calcium inhibits the magnesium-dependent enzyme human phosphoserine phosphatase. *European Journal of Biochemistry*, 271(16):3421–3427, 2004.
- [16] M. Veiga-da Cunha, J-F. Collet, B. Prieur, J. Jaeken, Y. Peeraer, A. Rabbijns, and E. Van Schaftingen. Mutations responsible for 3-phosphoserine phosphatase deficiency. *European Journal of Human Genetics*, 12(2):163–166, 2004.
- [17] J. Jaeken, M. Detheux, J-P. Fryns, J-F. Collet, P. Alliet, and E. Van Schaftingen. Phosphoserine phosphatase deficiency in a patient with williams syndrome. *Journal of Medical Genetics*, 34(7):594–596, 1997.
- [18] V. Rawat, P. Malvi, E.S. Della Manna, D.and Yang, S. Bugide, X. Zhang, and N. Gupta, R.and Wajapeyee. PSPH promotes melanoma growth and metastasis by metabolic deregulation-mediated transcriptional activation of NR4A1. *Oncogene*, 40(13):2448–2462, 2021.
- [19] L. Liao, M. Ge, Q. Zhan, R. Huang, X. Ji, X. Liang, and X. Zhou. PSPH mediates the metastasis and proliferation of non-small cell lung cancer through MAPK signaling pathways. *International Journal of Biological Sciences*, 15(1):183, 2019.
- [20] M.A. Bachelor, Y. Lu, and D.M. Owens. L-3-Phosphoserine phosphatase (PSPH) regulates cutaneous squamous cell carcinoma proliferation independent of L-serine biosynthesis. *Journal of Dermatological Science*, 63(3):164–172, 2011.
- [21] K. Sato, T. Masuda, Q. Hu, T. Tobo, S. Kidogami, Y. Ogawa, T. Saito, S. Nambara, H. Komatsu, H. Hirata, et al. Phosphoserine phosphatase is a novel prognostic biomarker on chromosome 7 in colorectal cancer. *Anticancer Research*, 37(5):2365–2371, 2017.

- [22] D-M. Smilgies and E. Folta-Stogniew. Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. *Journal of Applied Crystallography*, 48(5):1604–1606, 2015.
- [23] C.D. Putnam, M. Hammel, G.L. Hura, and J.A. Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, 40(3):191–285, 2007.
- [24] How random are intrinsically disordered proteins? A small angle scattering perspective, author=Receveur-Bréchet, V. and Durand, D. *Current Protein and Peptide Science*, 13(1):55–75, 2012.
- [25] J. Abendroth, A.S. Gardberg, J.I. Robinson, J.S. Christensen, B. L Staker, P.J. Myler, L.J. Stewart, and T.E. Edwards. SAD phasing using iodide ions in a high-throughput structural genomics environment. *Journal of Structural and Functional Genomics*, 12(2):83–95, 2011.
- [26] T.P. Stinear, T. Seemann, P.F. Harrison, G.A. Jenkin, J.K. Davies, P.D.R. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, et al. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Research*, 18(5):729–741, 2008.
- [27] Elise Pierson and Johan Wouters. Biochemical characterization of phosphoserine phosphatase serb2 from *Mycobacterium marinum*. *Biochemical and Biophysical Research Communications*, 530(4):739–744, 2020.
- [28] G. Arora, P. Tiwari, R.S. Mandal, A. Gupta, S. Sharma, D. and Saha, and R. Singh. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *Journal of Biological Chemistry*, 289(36):25149–25165, 2014.
- [29] V. Révora, M.I. Marchesini, and D.J. Comerici. *Brucella abortus* depends on L-serine biosynthesis for intracellular proliferation. *Infection and Immunity*, 88(2):e00840–19, 2020.

Chapter 4

Characterization of *MtSerB2* oligomeric states in the absence of L-serine

4.1 Observation and identification of the oligomeric states

4.1.1 *MtSerB2* elutes and migrates as two oligomeric species

When a milligram amount of pure *MtSerB2* obtained after IMAC, His-tag cleavage and reverse IMAC (as described in section 3.3.1) is applied to a semi-preparative SEC column in a Tris-NaCl pH 7.4 buffer, the resulting UV chromatogram shows two main absorbance peaks (Figure 4.1A).

Two elements ensure that each of these two peaks actually correspond to the detection of *MtSerB2* by the 280 nanometer UV monitor and not to that of a contaminating protein:

- The *MtSerB2* sample applied to the column is pure, as proven by the presence of a single intense band on SDS-PAGE gel (Figure 3.4 of Chapter 3). This band lies between the 37 kDa and 50 kDa references, which is consistent with

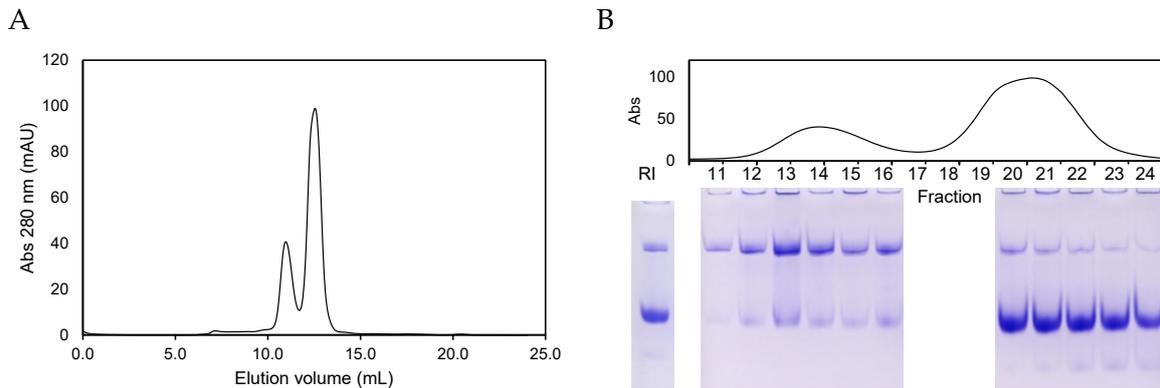


FIGURE 4.1: (A) SEC-UV analysis of *MtSerB2* obtained after reverse IMAC and (B) native PAGE analysis of *MtSerB2* before SEC (RI) and of the eluted fractions. Column: Superdex 200 10/300 GL.

the molecular weight of *MtSerB2* monomer (43.4 kDa). Given the His-tagged protein specific 3-step purification protocol applied, it can be considered that this band corresponds to *MtSerB2* only and not to a mixture of *MtSerB2* with a contaminant of the same monomeric molecular weight.

- When the same sample is analyzed by native PAGE, two distinct bands appear on the gel after coloration by Coomassie brilliant blue (Figure 4.1B, lane RI). The analysis of each fraction eluted during the SEC experiment by this technique allows to assign each of these bands to a peak of the chromatogram. By doing so, the upper band, corresponding to the species with the lowest electrophoretic mobility, can be associated with the first peak. The lower band, corresponding to the species with the highest electrophoretic mobility, matches the second peak.

These observations and considerations thus indicate that *MtSerB2* exists as two major forms in solution:

- A larger species, which elutes first from the SEC column (first peak) and migrates less far on the native PAGE gel (upper band)
- A smaller species, which elutes last from the SEC column (second peak) and migrates further on the native PAGE gel (lower band)

These two species are two distinct oligomeric states of *MtSerB2*. It can be ruled out that they correspond to two different conformations (e.g, extended and more compact) of the same oligomeric state. If this were the case, it would not be possible

to differentiate them in view of the limited resolution of the semi-preparative column and the gel.

This observation differs from the SEC results published by Yadav and Shree in 2014 [1]. Although the mobile phase was similar (Tris-NaCl buffer pH 8.0, versus pH 7.4 in this thesis) and the experiment was performed with the same column model, the authors report that *MtSerB2* elutes in a single peak in the absence of L-serine. However, their purification protocol involved differential precipitation of *MtSerB2* using ammonium sulfate after a first IMAC step, followed by preparative SEC as the final purification step before analysis. This alternative experimental design could explain the discrepancy in results. One hypothesis could be that one of the species remained in solution during the differential precipitation and was therefore excluded from the rest of the workflow. Another explanation, although less likely, could be that the authors overlooked the smallest peak during the pooling of preparative SEC fractions, resulting in a *MtSerB2* batch containing a single oligomeric species. The elution volume at peak maximum reported in reference [1] and the results presented further in this thesis indicate that the species observed by Yadav and Shree under L-Ser-free conditions matches the smaller oligomer (second peak, lower band) that we observed by SEC and native PAGE.

4.1.2 *MtSerB2* can form a dimer and a tetramer

The subsequent logical step was to identify the two oligomeric states of *MtSerB2* by determining their stoichiometry. To achieve this, two methods of molecular weight (M_W) determination were used and confronted:

- The relative estimation of M_W by calibration of the SEC column
- The absolute measure of M_W by SEC coupled to multi-angle light scattering (SEC-MALS)

Relative estimation of M_W by calibration of the SEC column

One of the methods used for molecular weight estimation relies on the calibration of a SEC column using globular protein standards. A sigmoidal relationship exists between the elution parameter (volume, time) of a series of homologous compounds analyzed by SEC and the logarithm of their molecular weights. The M_W of any similar compound can therefore be estimated by comparing its elution parameter with

the values obtained for the standards via a calibration curve in the range where the relationship between $\log(M_W)$ and the elution parameter is virtually linear.

This method was used for a first estimation of the M_W associated with the two oligomeric *MtSerB2* species. The orthologs presented in Chapter 3 were also analyzed in this way. The operation was performed on the two SEC columns used in this work, a semi-preparative one and an analytical one. The chromatograms and associated calibration curves are shown in Figure 4.2. The elution parameters and estimated M_W values for *MtSerB2* and its orthologs are listed in Table 4.1.

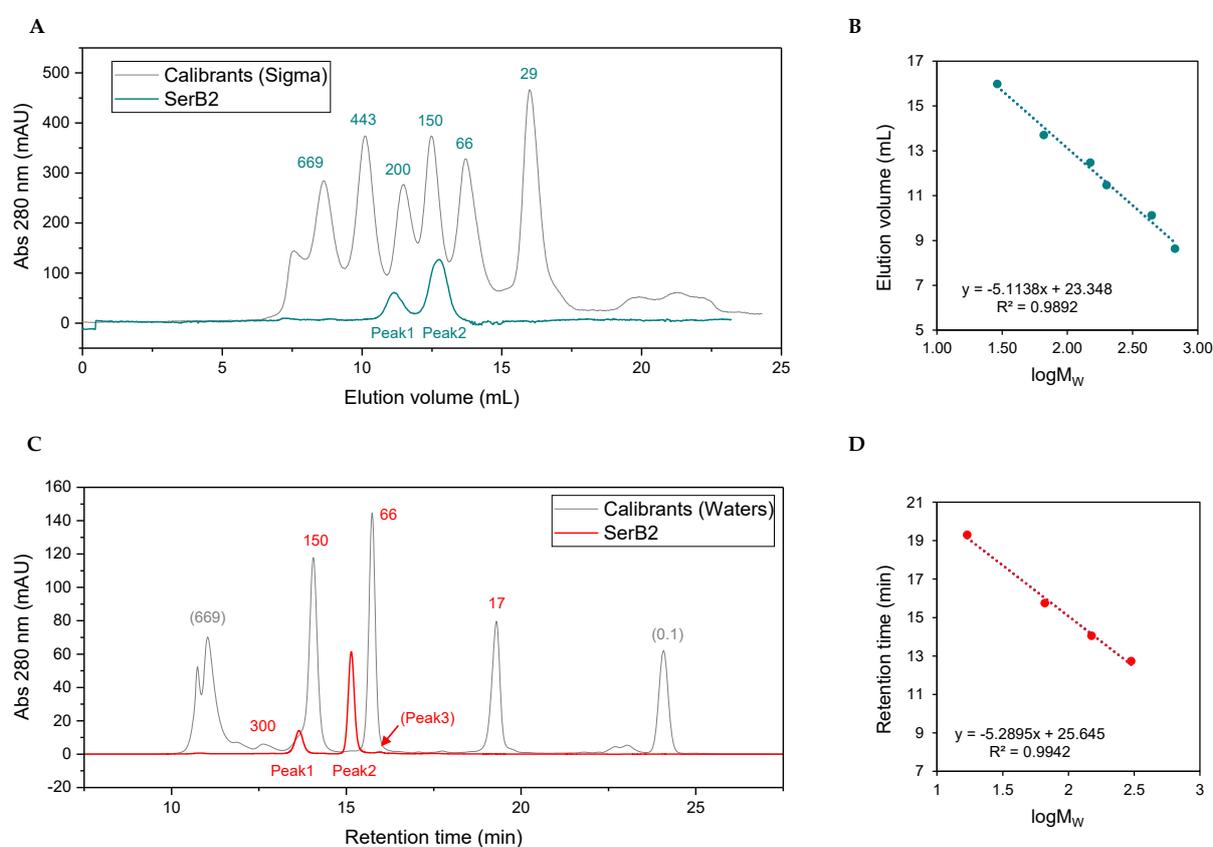


FIGURE 4.2: SEC column calibration using globular commercial calibrants. (A) Superimposition of the chromatograms obtained for *MtSerB2* and a mix of calibrants on a semi-preparative Superdex 200 10/300 GL column in Tris-NaCl pH 7.4 buffer, 0.2 mL min^{-1} . (B) Calibration curve corresponding to the data shown in chromatogram A. (C) Superimposition of the chromatograms obtained for *MtSerB2* and a mix of calibrants on an analytical BioResolveSEC mAb 200 \AA 2.5 \mu m $7.8 \times 300 \text{ mm}$ column in Tris-NaCl pH 7.4 buffer, 0.5 mL min^{-1} . (D) Calibration curve corresponding to the data shown in chromatogram C.

TABLE 4.1: Elution parameters for the analysis of *MtSerB2* and orthologous proteins by SEC and M_W values estimated by the calibration of two different columns.

Enzyme	Expected oligomeric state	Semi-prep. (Superdex)		Analytical (BioResolve)		
		Elution volume (mL)	Est. MW (kDa)	Retention time (min)	Est. MW (kDa)	Est. stoichiometry
<i>MtSerB2</i> - Peak 1	?	11.14	244	13.65	181	6/5/4
<i>MtSerB2</i> - Peak 2	?	12.76	118	15.14	95	3/2
<i>MtSerB2</i> - (Peak 3)	?	-	-	15.96	67	2/1
<i>MaSerB</i>	Dimer (88 kDa)	12.52	131	14.96	103	3/2
<i>MmSerB2</i>	Dimer (88 kDa)	n.d.	n.d.	14.79	111	3/2
<i>BmSerB</i>	Monomer (33 kDa)	14.95	44	18.35	24	2/1
<i>HsPSP</i>	Dimer (54 kDa)	n.d.	n.d.	16.12	63	2

As can be seen from the estimated oligomer stoichiometries, this method does not offer an unambiguous determination of the oligomerization states. Using the known oligomerization state of *MtSerB2* orthologs as a benchmark, we find that the molecular weights are systematically overestimated, except in the case of *BmSerB* on the analytical column. The effect is even more pronounced for the semi-preparative column. This phenomenon could be explained by a difference in shape between the analytes and the globular standards. This is illustrated by the overestimation of the M_W of *HsPSP*, which is known to have a particularly elongated dimeric quaternary structure. Deviations from ideal SEC behavior can also explain inaccuracies in M_W estimation. For example, ionic or hydrophobic interactions between the column resin and the enzyme can take place. In that case, the retention time within the column depends on these interactions in addition to steric effects and deviation from the linear relationship between $\log(M_W)$ and the elution parameter occur.

Absolute measure of M_W by SEC coupled to Multi-Angle Light Scattering (SEC-MALS)

To shed light on the oligomerization states of *MtSerB2*, we used a second molecular weight determination technique: MALS. The determination of M_W by MALS is more reliable because it does not require comparison with standards. Indeed, it is an absolute technique that determines the molecular weight of an analyte in solution from physical equations¹. The use of a SEC column upstream of the MALS detector allows to separate the different species in solution so that they can be analyzed individually.

¹See Appendix B for a more detailed explanation of the technique

The SEC-UV-MALS chromatogram obtained for the analysis of *MtSerB2* on the analytical SEC column is shown in Figure 4.3. As can be seen from the UV signal, *MtSerB2* elutes in three distinct peaks. The 3 peaks make up respectively 26.7, 69.5 and 3.8 % of the total area under the curve. The molar mass associated to each of these peaks was measured by MALS: 165.3 kDa, 86.9 kDa and 55.5 kDa in order of elution. Considering a molar mass of 43.4 kDa for a monomer, the main peaks can be respectively associated to a tetramer and a dimer. The third and smaller peak could correspond to a monomer or a truncated form of the enzyme. It is interesting to note that such a peak seems to be characteristic of PSPs with two ACT domains, as it also appears in the chromatograms of *MmSerB2* and *MaSerB*. Analyses performed on the semi-preparative column do not show the presence of this species due to the lower separation resolution. The independent analysis of a *MtSerB2* sample, conducted by Wyatt Technology Corporation, corroborates our results. The enzyme eluted in two peaks on a Superdex 200 Increase 10/300 semi-preparative column. The experiment was performed with various injection volumes and resulted in the identification of a tetramer for the first peak (160-180 kDa) and a dimer for the second peak (80-88 kDa).

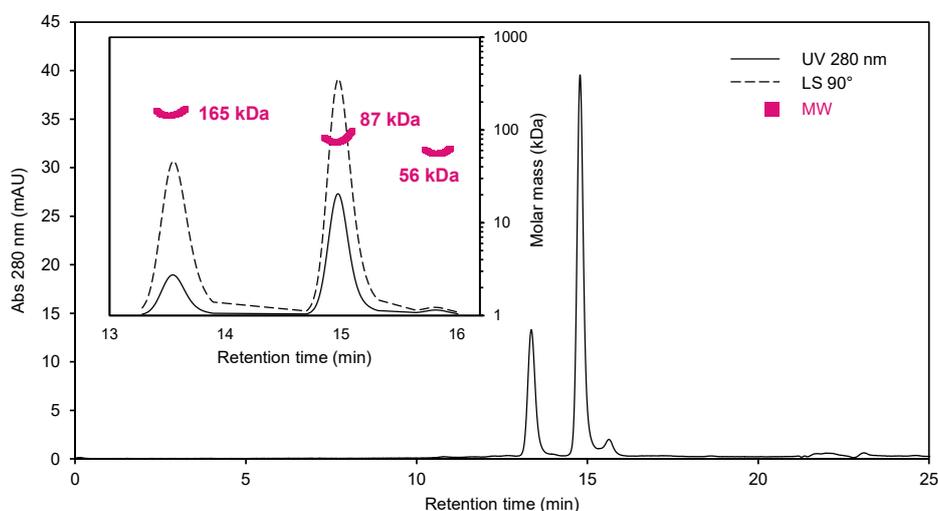


FIGURE 4.3: SEC-UV-MALS analysis of *MtSerB2*.

The SEC-MALS results, together with the calibration of the SEC columns, allow us to state that *MtSerB2* exists in two main oligomeric forms in solution: a dimer and a tetramer. The analytical SEC experiments also reveal the presence of a minority species that could be associated with the monomeric form. An examination of the base of the peaks would also reveal the presence in trace of a species of intermediate size

between those of the dimer and the tetramer and species of greater size than that of the tetramer. These suspected species are indicated by arrows on the chromatogram in Figure 4.4. The existence of a trimeric species, which could correspond to the intermediate peak, will be demonstrated in Chapter 5. As for the first peaks, they could indicate the presence of trace amounts of aggregates in the analyzed solution.

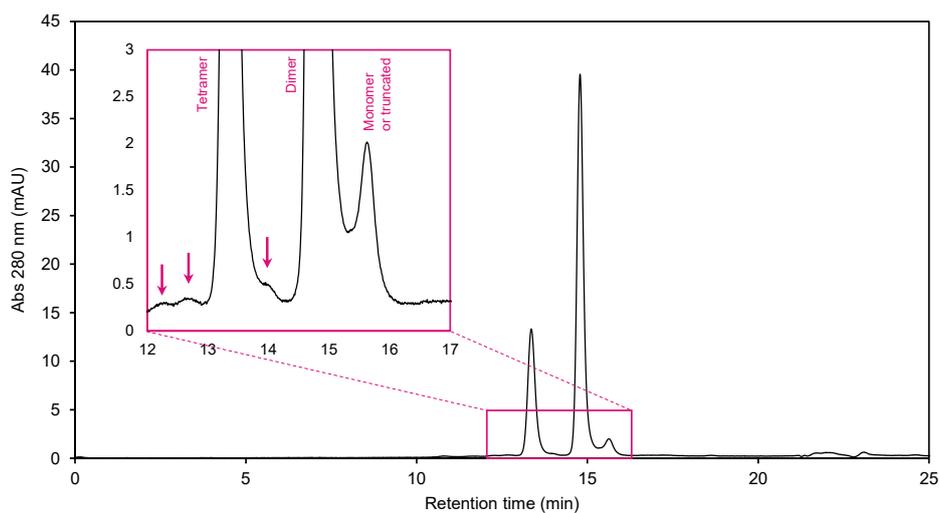


FIGURE 4.4: SEC-UV analysis of *MtSerB2* and zoom at the base of the peaks.

4.1.3 *MtSerB2* dimer and tetramer populations interconvert through a slow equilibrium independent of the total protein concentration

The equilibrium between *MtSerB2* tetramer and dimer forms was probed. To do so, native PAGE and SEC were mainly used but our observations could be completed by preliminary results of mass photometry (MP)². MP enables the quantification and the measure of the molecular mass of individual biomolecules in solution by measuring a light scattering signal proportional to their mass as they land on an illuminated measurement surface.

The dimer and the tetramer are stable species that can be separated and individually concentrated

The chromatography equipment we have in the lab allows us to fractionate the eluate at the exit of the SEC column. We were therefore able to recover separately the dimeric and tetrameric populations of *MtSerB2* following their separation in a semi-preparative SEC column. The populations could then be individually analyzed by native PAGE at their concentration at the column exit and after concentration up to 25 mg mL⁻¹ by centrifugation (Figure 4.5). As can be seen from the native gel, the dimeric population remains dimeric after eluting from the SEC column (lane 3) and the same is true for the tetrameric population (lane 1). Both also retain their oligomeric state after concentration (lane 4 and 2). This property is crucial for setting up the crystallogensis assays detailed in section 4.2.1.

The stability of the dimeric species could be shown by another SEC experiment. A sample of *MtSerB2* was fractionated by SEC, the dimeric population was recovered, concentrated, and then re-injected into the column under the same experimental conditions. The superimposition of the two chromatograms in Figure 4.6 shows the presence of the dimeric population alone (a single peak) after re-injection. The tetramer was not reformed in the meantime. Re-injection of the tetrameric population was undertaken with the same method but the experiment was not successful. The amount of tetramer recovered after the first SEC step is very small and the loss due

²Access to this technique was provided during the demonstration of a Refeyn TwoMP instrument at the Laboratory for the Analysis of Medicines (LAM, ULiège) on March 29, 2022

to the concentration step do not allow to obtain a sufficient amount of protein for a peak to be visible on the chromatogram after re-injection.

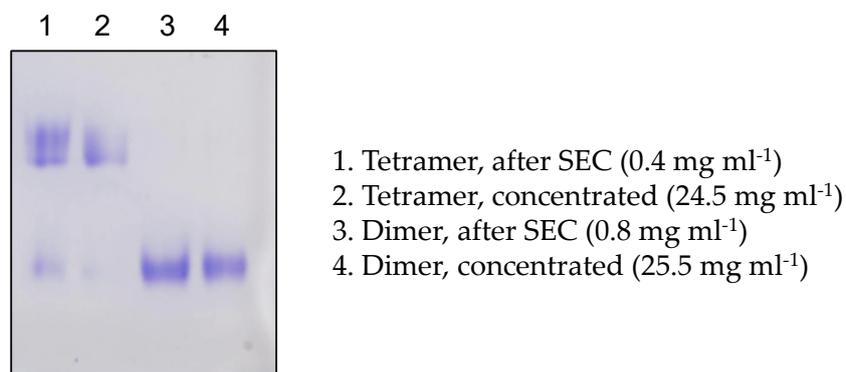


FIGURE 4.5: Native PAGE analysis of *MtSerB2* tetramer and dimer fractions after eluting from the SEC column and after centrifugation concentration.

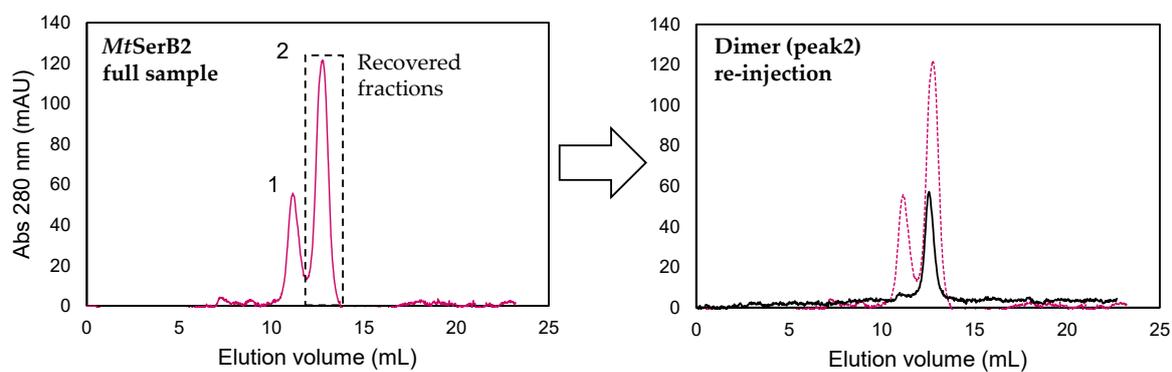


FIGURE 4.6: SEC-UV analysis of a full *MtSerB2* sample (left) and of the re-injected dimer peak (right).

The dimer and tetramer populations rearrange over time

During the analyses of separated dimer and tetramer batches by SEC and native PAGE, we noticed that the tetramer batches are often contaminated with a small quantity of dimer while the dimer batches are always (almost) pure. This observation led us to test whether this phenomenon was simply due to poor upstream separation or to population rearrangement over time. For this purpose, tetramer and dimer samples of different ages were analyzed by native PAGE and MP. The gel analysis presented in Figure 4.7 shows that a portion of the tetramer population of the tetramer batch already converted to dimer after 3 weeks at 4 °C. Regarding the dimer batch, the dimer population mostly retained its oligomeric state, as evidenced by the comparable intensities of the two dimer bands after thawing and after 3 weeks. However, a faint tetramer band could be observed for the thawed sample and the latter appeared slightly more intense after 3 weeks. It is therefore possible that a very small proportion of the dimer population had shifted to a tetrameric state. The same observations can be made for the samples that stayed for 2 months at 4 °C.

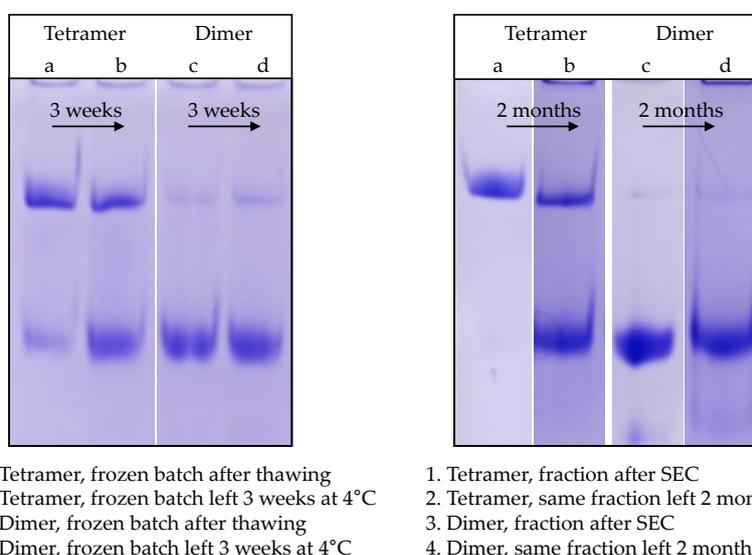


FIGURE 4.7: Native PAGE analysis of *MtSerB2* dimer and tetramer samples. The samples were fresh, freshly thawed or aged for 3 weeks or 2 months at 4 °C.

These observations were corroborated by MP analyses shown in Figure 4.8, where the proportions of dimer and tetramer in samples of different ages were quantified. The analyzed tetramer batch was initially contaminated with a non-negligible amount of dimer, but it can be seen that the proportion of tetramer decreased with sample age, while the amount of dimer increased. For the sample kept 3 months at 4 °C, a dimer to tetramer ratio of 80/20 was observed. These proportions approached the 85/15 ratio seen in the total unseparated sample (assumed to be at equilibrium) during the same MP analysis series. The behavior of the dimer batch was different. The dimer/tetramer ratio remained unchanged after two weeks at 4°C. After 3 months, a small amount of tetramer seemed to have reformed.

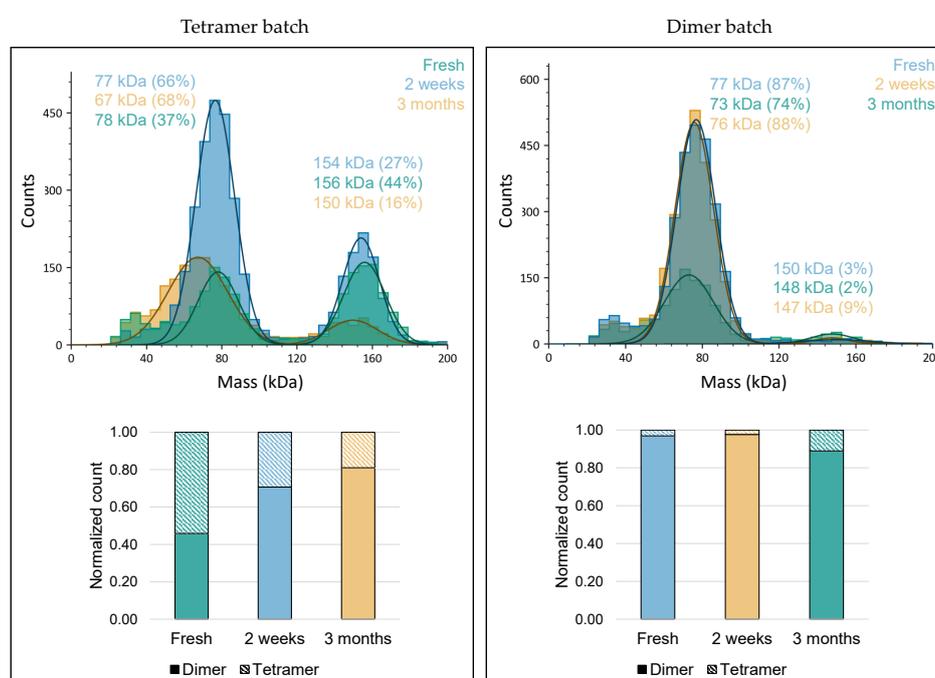


FIGURE 4.8: Mass photometry analysis of *MtSerB2* dimer and tetramer batches. The batches were freshly thawed or aged for 2 weeks and 3 months at 4 °C. Between brackets are indicated the percentage of the total number of counts. The height of histograms bars represent the total amount of dimer and tetramer.

Based on these observations, it can be concluded that it seems much easier for the dimer to be reformed from the tetramer than the other way around. The interconversion of species evolves on a scale of weeks. It appears that the dimeric and tetrameric populations of *MtSerB2* are in slow equilibrium.

The interconversion between dimer and tetramer is not affected by sample concentration

It could also be shown that the equilibrium between the two oligomeric forms is not affected by the total protein concentration. Several *MtSerB2* samples of different total concentrations were analyzed by native PAGE. The samples were prepared in two ways: (1) by gradually concentrating an initial 1.2 mg mL^{-1} solution, and (2) by gradually diluting an initial 14.9 mg mL^{-1} solution. A fixed volume was taken at each concentration or dilution step for electrophoretic analysis. The intensity of the bands on the gel, proportional to the amount of each species, was quantified by gel densitometry. As shown in the native gels and associated densitometry histograms (Figure 4.9), the proportions of dimer and tetramer do not vary with the concentration of the sample.

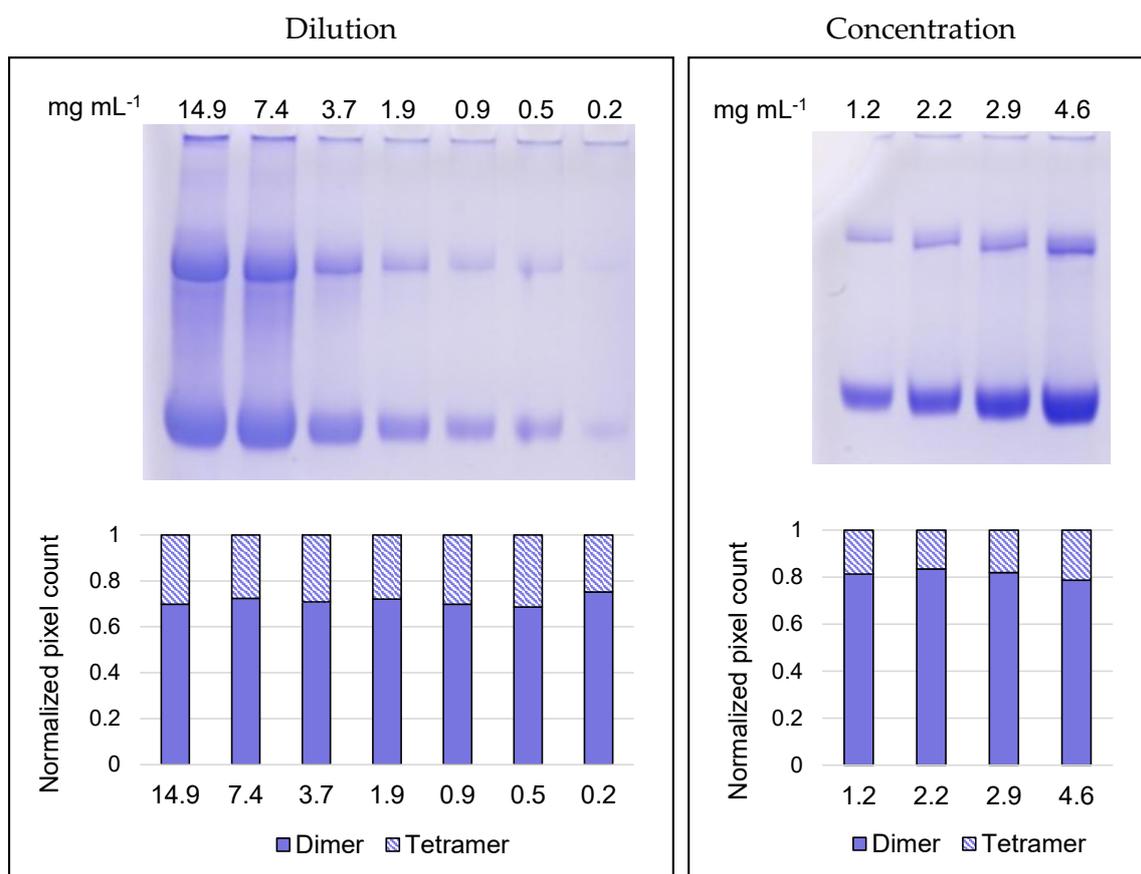


FIGURE 4.9: Native PAGE analysis of a *MtSerB2* sample during dilution and concentration, and corresponding densitometry histograms.

The same conclusion could be drawn from SEC experiments. A volume of 30 μL of a 1.5 mg mL^{-1} *MtSerB2* solution was injected into the analytical BioResolve SEC column and 250 μL of the same solution concentrated to 15 mg mL^{-1} were injected into the semi-preparative Superdex SEC column. Area under the curve measurements indicate that the proportions of dimer and tetramer do not change, despite the difference in concentration of the sample and within the column (Figure 4.10). The third peak, present during the analysis with the analytical column, does not appear on the chromatogram associated with the semi-preparative column. Given the lower resolution capacity of the latter, the species associated with this peak is probably co-eluting with the dimeric population. The area under the curve of the dimeric peak is therefore compared with the sum of the areas of peaks 2 and 3 of the BioResolve analysis.

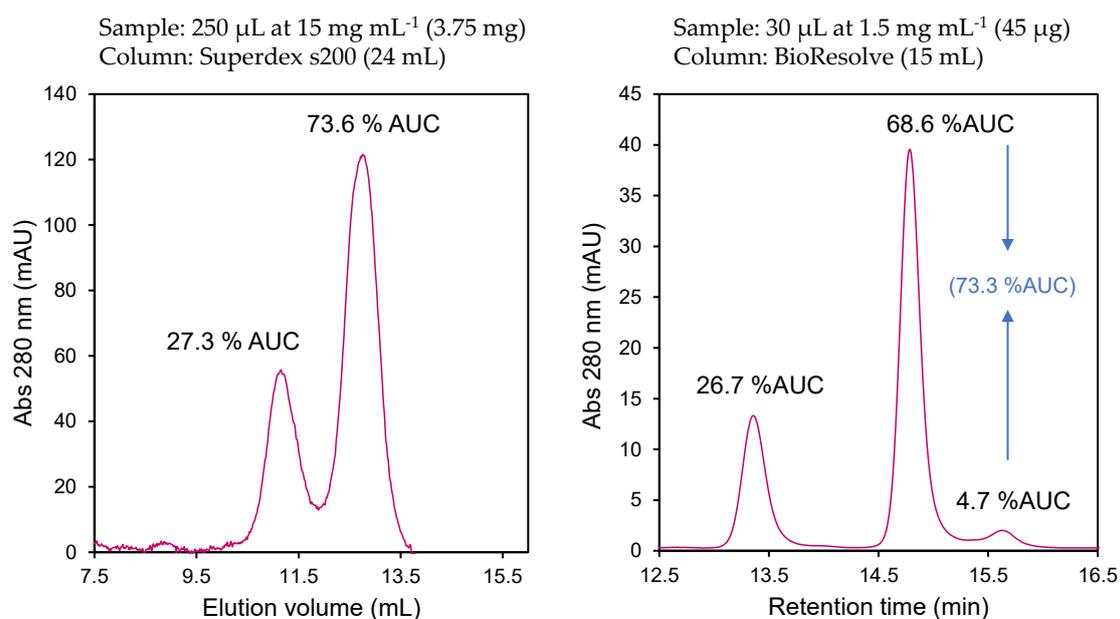


FIGURE 4.10: SEC-UV analysis of a *MtSerB2* solution at different concentrations on different columns.

Phosphate and pH do not seem to influence the dimer/tetramer equilibrium

The influence of phosphate, one of the reaction products of *MtSerB2*, was also assessed. SEC analyses were performed in a phosphate NaCl buffer that also enabled the oligomeric behavior to be studied over a range of two pH units (5.8 to 7.8). The results are shown in Figure 4.11. Area under the curve calculations showed that *MtSerB2* dimer and tetramer were still found in the same proportions in the phosphate buffer compared to the Tris buffer and that these quantities remained constant regardless of the pH.

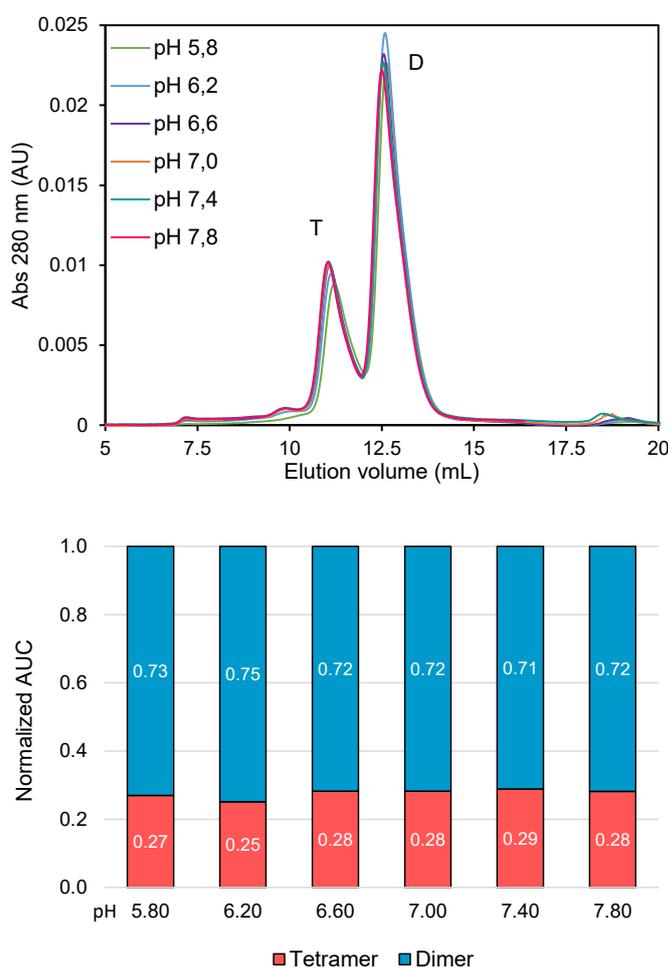


FIGURE 4.11: SEC-UV analysis of *MtSerB2* in a phosphate NaCl buffer at different pH and corresponding area under the curve histograms for the dimer and tetramer peaks.

4.2 Structural characterization of the oligomeric species

Our work confirmed that *MtSerB2* forms a dimer in solution and provided the first quantitative experimental evidence for the existence of a tetramer of *MtSerB2*. We then focused on studying the structures of these oligomers: does the dimer resemble that of *MaSerB* and what is the architecture adopted by *MtSerB2* tetramer?

4.2.1 *MtSerB2* dimer and tetramer are difficult to crystallize

We tried to crystallize *MtSerB2* many times during the project. Initial attempts were made with the full sample containing dimer and tetramer, at varying protein concentrations, using *Hampton Research's Crystal Screen 1 and 2* crystallization screens. Conditions around the crystallization conditions of *MaSerB* provided in the PDB were also tested. No results were obtained from these assays.

Given the polydispersity of the sample, prior separation of the two oligomeric populations seemed to be a good strategy to promote *MtSerB2* crystallization. The following tests were therefore performed after separation of dimer and tetramer populations by preparative SEC. As shown in Figure 4.5, both species are stable after separation and can be concentrated while maintaining their oligomerization state. This makes it possible to try to crystallize them separately.

Each oligomeric population was then set up for crystallization in the conditions of *Molecular Dimensions' JCSG-plus* and *PACT premier* screens. These two screens are complementary: JCSG screen is a general sparse-matrix screen covering previously successful crystallization space, while PACT screen is based on a systematic test of pH, anions and cations with the precipitant polyethylene glycol (PEG). Their combination has been demonstrated as an efficient strategy for the identification of new crystallization conditions [2]. A third screen, the *BCS Screen*, was also used. In this screen, the precipitating agent consists of mixtures of PEGs grouped by molecular weight (PEG smears), which allows broader chemical space to be covered while reducing the number of PEG variables [3]. Solutions with compositions close to the crystallization conditions of *MaSerB* were also tried.

A few crystals have been obtained from these experiments. Unfortunately, during their analysis by X-Ray diffraction at SOLEIL Synchrotron, they turned out to be salt

or did not diffract. A table summarizing the crystals tested and the conditions under which they were obtained is available in Appendix G.

Crystallization of *MtSerB2* oligomers has been an ongoing effort throughout this thesis project but no atomic structure could be revealed to date. Fortunately, other techniques, either indirect such as site-directed mutagenesis or direct, such as SAXS, have allowed us to learn more about the structures adopted by the enzyme.

4.2.2 Investigation of the major species: *MtSerB2* dimer

In the absence of crystallographic or other structural data, insight into the structure of *MtSerB2* is based on homology modeling. *MtSerB2* is usually modeled based on the structure of SerB from *M. avium*, its closest crystallized homolog (83.3 % identity). Homology models of *MtSerB2* based on the structure of *MaSerB* have been used for docking studies in references [1, 4, 5]. However, although the high percentage of sequence identity is a good indicator of model reliability, no experimental evidence of the resemblance between *MtSerB2* and *MaSerB* dimers is available. We performed this verification first through the design of a truncated mutant and then with the use of SAXS when the opportunity for such measurements occurred.

***MtSerB2* depends on its ACT1 domain to oligomerize**

As shown by its crystallographic structure, *MaSerB* dimer is very likely formed by domain-swapping. The swapped domain is the N-terminal ACT1 domain which, thanks to the hinge-loop, is positioned below the ACT2 domain of the other monomer to form an 8-stranded anti-parallel β -sheet.

An analysis of the dimeric interface allowed us to identify the residues involved in the interaction between the two monomers. Figure 4.12 schematizes the result of this scan, where for each residue of monomer 1, the number of residues of monomer 2 within 4 Å and the domain to which they belong are indicated.

The analysis quantifies the importance of the ACT1 domain in the dimeric interaction. ACT1 residues are involved in a large number of proximity pairs (44% of total pairs), particularly at the 2nd β -strand. Interactions with the other monomer at the ACT1 (18%), ACT2 (24%) and hinge loop (6%) regions could be identified. In contrast, ACT2 domain residues participate in fewer pairs (27%), mainly with the ACT1

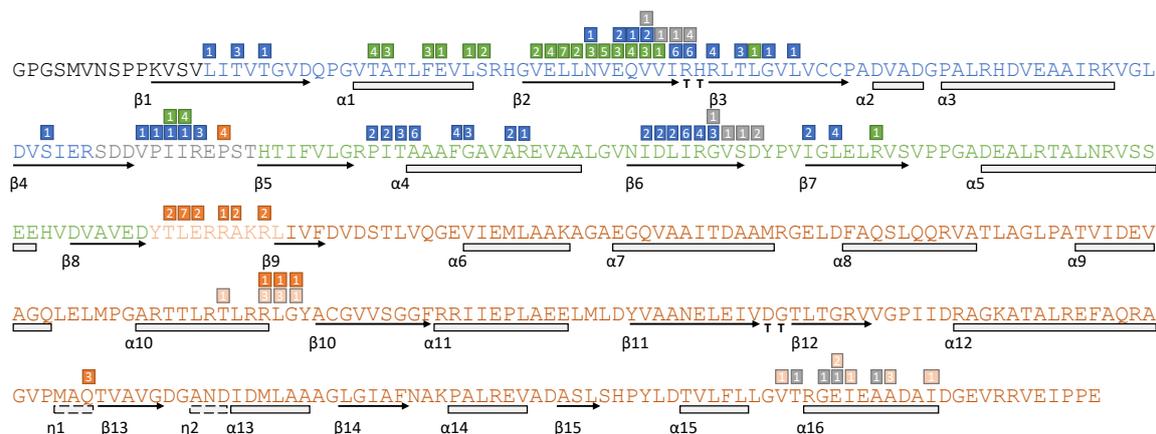


FIGURE 4.12: Residues involved in the dimeric interface of *MaSerB*. Monomer 1 sequence is depicted. Above each residue in this sequence, the number of residues of monomer 2 within 4 Å is indicated. The region to which each residue belongs is indicated by color: ACT1 (blue), ACT2 (green), PSP (orange), hinge-loop (gray) or ACT2-PSP loop (salmon). Secondary structure elements based on 5IS2 structure are shown below the sequence.

domain. PSP domain residues are less involved in the dimeric interface (13%) and interact mainly with the hinge loop and the ACT2-PSP flexible loop. The analysis also highlights that half of the hinge loop residues play a role in stabilizing the interface by interacting with residues in the ACT1, ACT2 and PSP domains.

The presence of the ACT1 domain thus seems to play a crucial role in the dimerization of *MaSerB*. Assuming that it adopts the same fold in solution, then this postulate should also hold for *MtSerB2* dimer. Based on this, we wanted to investigate the oligomeric behavior of a truncated form of *MtSerB2*, lacking the ACT1 domain. For this experiment, the truncated *MtSerB2* Δ ACT1 mutant was designed by site-directed deletion mutagenesis following the method of Liu *et al.* [6]. Its oligomeric behavior was then assessed by SEC-MALS (Figure 4.13) and compared with that of full *MtSerB2* and *BmSerB*.

As can be seen on the SEC-UV-MALS chromatogram, *MtSerB2* Δ ACT1 elutes in a major peak corresponding to a measured molar mass of 37.6 kDa. This result suggests that the mutant is mostly monomeric, as the calculated MW_{monomer} is 34.8 kDa. The conclusion is also supported by the similar elution volume to *BmSerB* (32.5 kDa, monomeric). However, an absorbance signal at and after the dead volume shows the presence of high molecular weight aggregates in the sample. Truncation, preventing

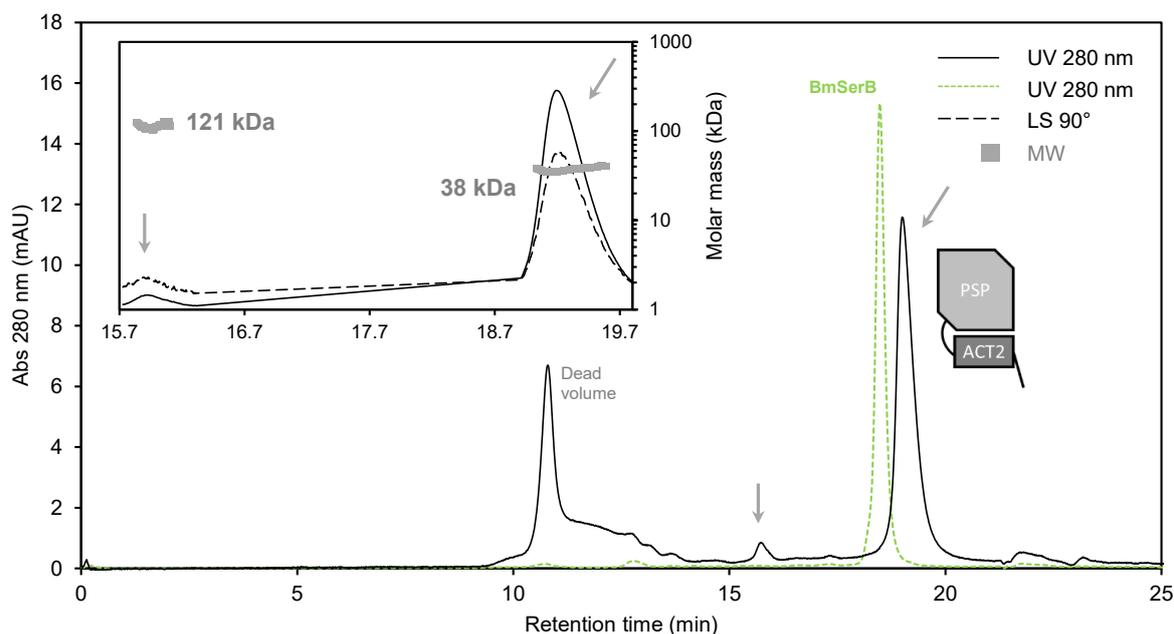


FIGURE 4.13: SEC-UV-MALS analysis of *MtSerB2*ΔACT1 mutant

dimerization, probably leads to the exposure of hydrophobic areas to the solvent. This explains the formation of aggregates but also why *MtSerB2*ΔACT1 monomer elutes after *BmSerB*: non-specific hydrophobic interactions with the matrix could increase the residence time in the SEC column. Another, smaller peak is present between the aggregates and the monomeric peak. The MALS results for this peak indicate a molar mass of 121.2 kDa, which could correspond to that of a trimer ($121.2/37.6 = 3.2$).

The oligomeric behavior of *MtSerB2*ΔACT1 mutant, along with that of *BmSerB*, thereby highlights the crucial role of the ACT1 domain in *MtSerB2* self-assembly. ACT1 absence prevents dimer formation, which is a good indication that *MtSerB2* also adopts a domain-swapped butterfly fold similar to that of *MaSerB*. It seems to also prevent tetramer formation but would lead to the formation of a small trimeric population.

MtSerB2 forms a dimer similar to that of *MaSerB*

Access to SAXS allowed us to verify that *MtSerB2* dimer is indeed folded like *MaSerB* in solution. As shown by the superposition of their scattering curves (Figure 4.14A), the comparable shape of their $P(r)$ functions (Figure 4.14B) and the very close values of their shape parameters (Table 4.2), the two enzymes likely adopt almost identical global shapes in solution. Minor local conformational variations may exist between the two enzymes, as reflected by the slight deviation of the SAXS curves between q values of 0.150 and 0.250 \AA^{-1} , and the local differences observed between the $P(r)$ functions along a somewhat higher D_{max} value for *MtSerB2*. The dimensionless Kratky plot (Figure 4.14C) shows that *MtSerB2* dimer is also globular and well folded. By extension, in view of the identical SAXS profiles between *MaSerB* and *MmSerB2* (see Chapter 3, page 81), it can be concluded that the shape of *MtSerB2* dimer is also very similar to that of *MmSerB2*.

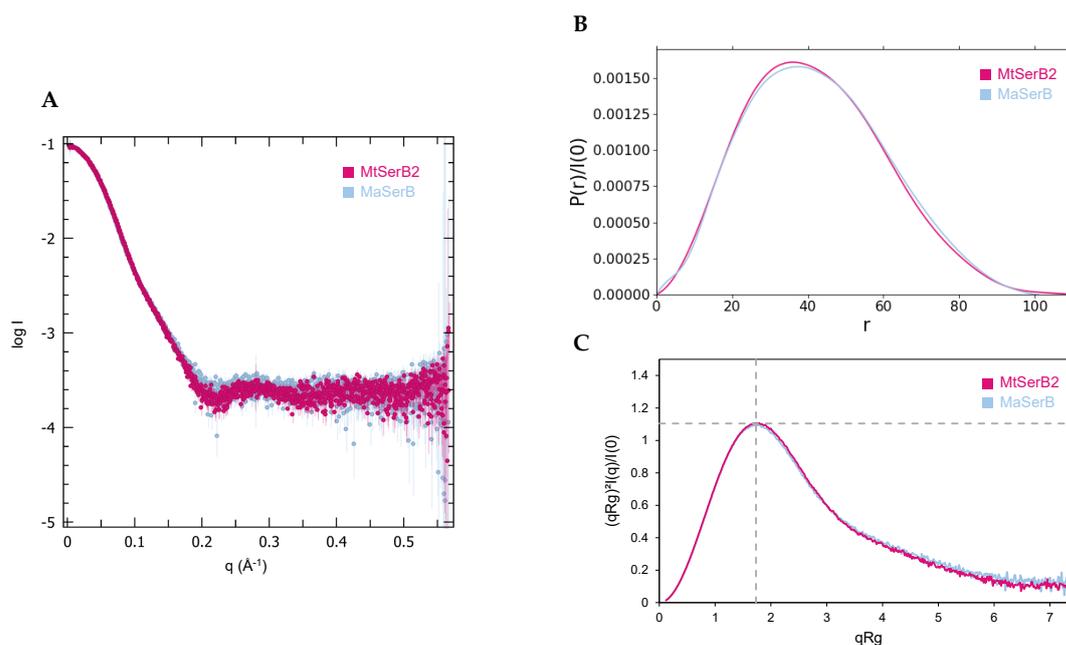


FIGURE 4.14: Analysis of *MtSerB2* solution structure from SAXS data and comparison with *MaSerB*. (A) Experimental SAXS profiles ($\log I(q)$ versus q) of *MtSerB2* and *MaSerB*. (B) Calculated $P(r)$ functions. (C) Dimensionless Kratky plots.

TABLE 4.2: SAXS parameters of *MtSerB2* dimer compared to those of *MaSerB*. (1) derived from the Guinier plot (2) calculated from the $P(r)$ function

	<i>MtSerB2</i> dimer	<i>MaSerB</i> dimer
R_g (Å)	32.47 ± 0.02 ⁽¹⁾ ; 32.55 ± 0.02 ⁽²⁾	32.62 ± 0.03 ⁽¹⁾ ; 32.84 ± 0.02 ⁽²⁾
$I(0)$ (cm ⁻¹)	$0.09 \pm 3.32 \cdot 10^{-5}$ ⁽¹⁾ ; $0.09 \pm 3.64 \cdot 10^{-5}$ ⁽²⁾	$0.11 \pm 5.99 \cdot 10^{-5}$ ⁽¹⁾ ; $0.11 \pm 5.0 \cdot 10^{-5}$ ⁽²⁾
D_{max} (Å)	111	101
V_p (nm ³)	127	126
M_W Bayesian inference (kDa)	88.3 (84.3 to 95.5)	91.2 (84.3 to 95.8)
M_W Porod volume (kDa)	105.5	104.7

4.2.3 Investigation of the minor species: *MtSerB2* tetramer

To shed light on *MtSerB2* tetramer structure, several previously investigated properties of the enzyme and its orthologs can be exploited. First, its two closest homologs *MaSerB* and *MmSerB2* are dimeric and unable to form tetramer under the conditions studied in this thesis. The residues responsible for the tetramerization of *MtSerB2* can therefore be identified on the basis of the sequence difference between these enzymes. Next, the ACT1 domain appears to be essential for dimer formation, as highlighted by the oligomeric behavior of *MtSerB2* Δ ACT1 mutant and supported by that of *BmSerB*. Finally, on this basis and by comparison with the SAXS profiles of *MaSerB* and *MmSerB2*, it has been argued that *MtSerB2* dimer is also formed through ACT1 domain-swapping. The hypothesis that *MtSerB2* forms a domain-swapped tetramer as well cannot be excluded. In this context, it is worth looking at the determinant region of domain-swapping: the hinge loop.

Gln92 is involved in *MtSerB2* propensity to form a tetramer

Figure 4.15 depicts the sequence alignment of *MtSerB2*, *MaSerB* and *MmSerB2* focused on the hinge loop region. The sequences are particularly conserved but 3 point variations distinguish the loops of the three enzymes:

- Leucine L87 in *MtSerB2* is a valine (V) in *MaSerB*
- Isoleucine I89 in *MtSerB2* is a valine (V) in *MmSerB2*
- Glutamine Q92 in *MtSerB2* is a glutamate (E) in both *MmSerB2* and *MaSerB*

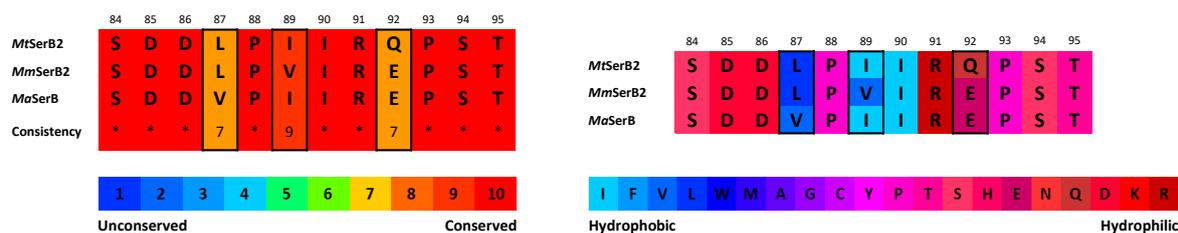


FIGURE 4.15: Alignment of the sequences of *MtSerB2*, *MaSerB* and *MmSerB2* focused on the hinge loop region performed with PRALINE tool [7]. Residues are colored according to their conservation (left) and hydrophilicity (right).

None of these mutations look particularly drastic. Due to their size properties and hydrophobic character, valine, leucine and isoleucine residues are frequently substituted by each other. The same is true for the glutamate/glutamine variation, two polar residues with very similar carboxylate and amide functional groups [8]. However, a notable difference exists between these two residues: glutamate bears a net negative charge at pH 7.4 while glutamine is neutral. This mutation is thus the only one that really differentiates *MtSerB2* from its two exclusively dimeric counterparts at the hinge loop.

We wanted to probe the potential involvement of Gln92 in the propensity of *MtSerB2* to form tetramer. For this purpose, a ***MtSerB2* Q92E** mutant was designed following the method of Liu *et al.* [6] and produced according to the usual protocol. Its oligomeric behavior was analyzed by native PAGE and analytical SEC (Figure 4.16).

The native PAGE first provided us with a qualitative response. As shown by the gel in Figure 4.16A, *MaSerB* only forms dimer (lower band) while *MtSerB2* and *MtSerB2* Q92E form dimer and tetramer (upper band). However, the band corresponding to the tetramer of *MtSerB2* Q92E sample appears less dense than that of *MtSerB2*. The mutant would therefore form less tetramer.

This observation could be verified and quantified by SEC-UV and integration of the area under the curve for each peak. As shown in Figure 4.16B, *MtSerB2* Q92E sample contains less tetramer (13.7 %AUC) than *MtSerB2* sample (26.7 %AUC). Normalization of the chromatograms to the dimeric peak provides a clearer picture of the lower proportion of the tetrameric population in *MtSerB2* Q92E. AUC ratios (tetramer

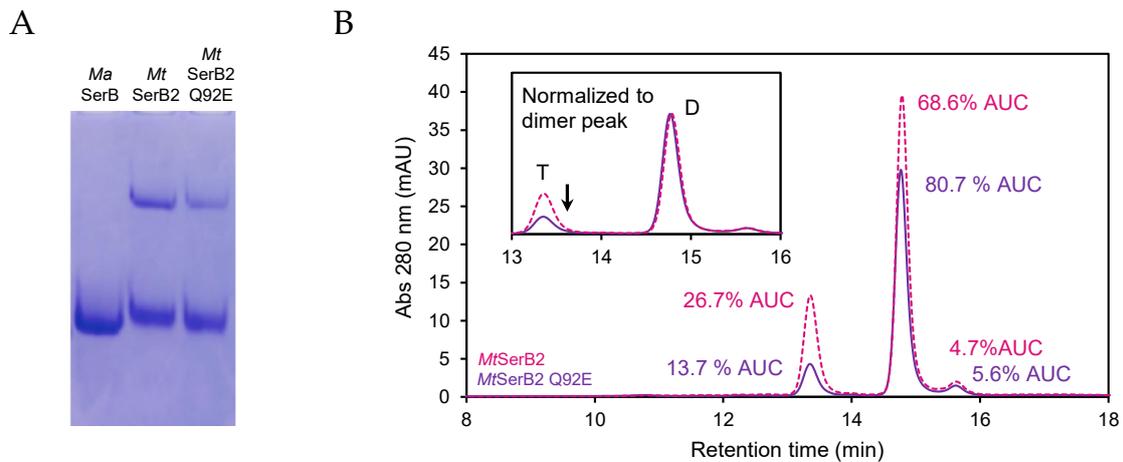


FIGURE 4.16: Analysis of the oligomeric behavior of *MtSerB2* Q92E. (A) Native PAGE analysis of *MaSerB*, *MtSerB2* and *MtSerB2* Q92E. (B) Superimposition of the SEC-UV chromatograms of *MtSerB2* (pink dotted line) and *MtSerB2* Q92E (purple). Insert: data normalized to the dimer peak.

to dimer) for *MtSerB2* Q92E and *MtSerB2* are 0.17 and 0.40 respectively.

These experiments thus show that mutation of the Gln92 residue present in the hinge loop of *MtSerB2* to glutamate has an effect on the self-assembly state of the enzyme: we observed a decrease in the tetrameric population and an increase in the dimeric population. This glutamine could therefore play a role in the tetramerization propensity of the enzyme. A first hypothesis could be that the Q92E mutation would disrupt the oligomeric equilibrium by influencing the conformation of the hinge-loop. Indeed, a conformational restriction could take place through the interaction of the net negative charge of glutamate and the positive charge carried by the upstream arginine. Another possibility would be that the glutamate destabilizes the tetrameric architecture through the effect of unfavorable interactions at an interface due to its negative charge.

The N-term part of the second α -helix of ACT2 domain is a hotspot for tetramer formation

Residue Q92 is involved in the tetramerization ability of *MtSerB2* but its presence is not necessary for the phenomenon to occur since the enzyme can still form tetramer when it is mutated to a glutamate like in the exclusively dimeric SerBs. Other residues, specific to *MtSerB2* should therefore participate in the stabilization of the

tetramer. As briefly mentioned in section 3.2.3 (Chapter 3), the sequence alignment of *MtSerB2*, *MmSerB2* and *MaSerB* shows that the second alpha helix of the ACT2 domain ($\alpha 5$ in Figure 4.17) is particularly poorly conserved between the three enzymes, especially at the N-terminal part. Comparing *MtSerB2* to its two exclusively dimeric orthologs, it can be seen that the nature of the residues changes significantly in this region. While *MtSerB2* bears rather small, neutral, hydrophobic and apolar residues (C148, V149, G150 and I154), the residues are larger, charged and more polar in *MmSerB2* (T, Y, R, T) and *MaSerB* (A, D, E, T). The alignment in the right box at the bottom of Figure 4.17 highlights the polarity shift of the residues at these four positions.

To test whether these residues are responsible for *MtSerB2* tetramerization, a new mutant was generated: ***MtSerB2 α*** . This mutant, in addition to the Q92E mutation, bears the mutations C148T, V149Y, G150R, and I154T, where the residues found in *MtSerB2* are replaced by those found in *MmSerB2*, the latter being the closest homolog of *MtSerB2* and existing only in dimeric form under the studied conditions. After production according to the usual protocol, the oligomeric behavior of this mutant was assessed by SEC and compared with that of native *MtSerB2*.

The superimposition of the SEC-UV chromatograms (Figure 4.18) clearly shows that *MtSerB2 α* exists almost essentially in the dimeric state in the analyzed sample (94.9 %AUC). The area under the peak that should correspond to the elution of a tetrameric species is only 1.1 %AUC. In comparison, for a similar amount of enzyme injected into the column as a control, the native *MtSerB2* sample contains tetramer (27.2 %AUC) and dimer (72.5 %AUC). It appears that the mutant also forms some larger aggregates, eluting in the dead volume of the column (3.0 %AUC). Finally, an interesting feature to notice is the presence of a shoulder at the back of the dimer peak, which could be due to the co-elution of a smaller species. This analysis alone cannot give a more precise answer as to the identity of this species, but a plausible hypothesis would be the presence of a monomeric population in solution, greater than in the case of native *MtSerB2*.

The combination of mutations C148T, V149Y, G150R, I154T and Q92E impedes the tetramerization of *MtSerB2*. To go further and identify the residue(s) with the strongest impact, it would be necessary to test the effect of each mutation separately on the oligomeric behavior of *MtSerB2*. So far, we can only suggest that the

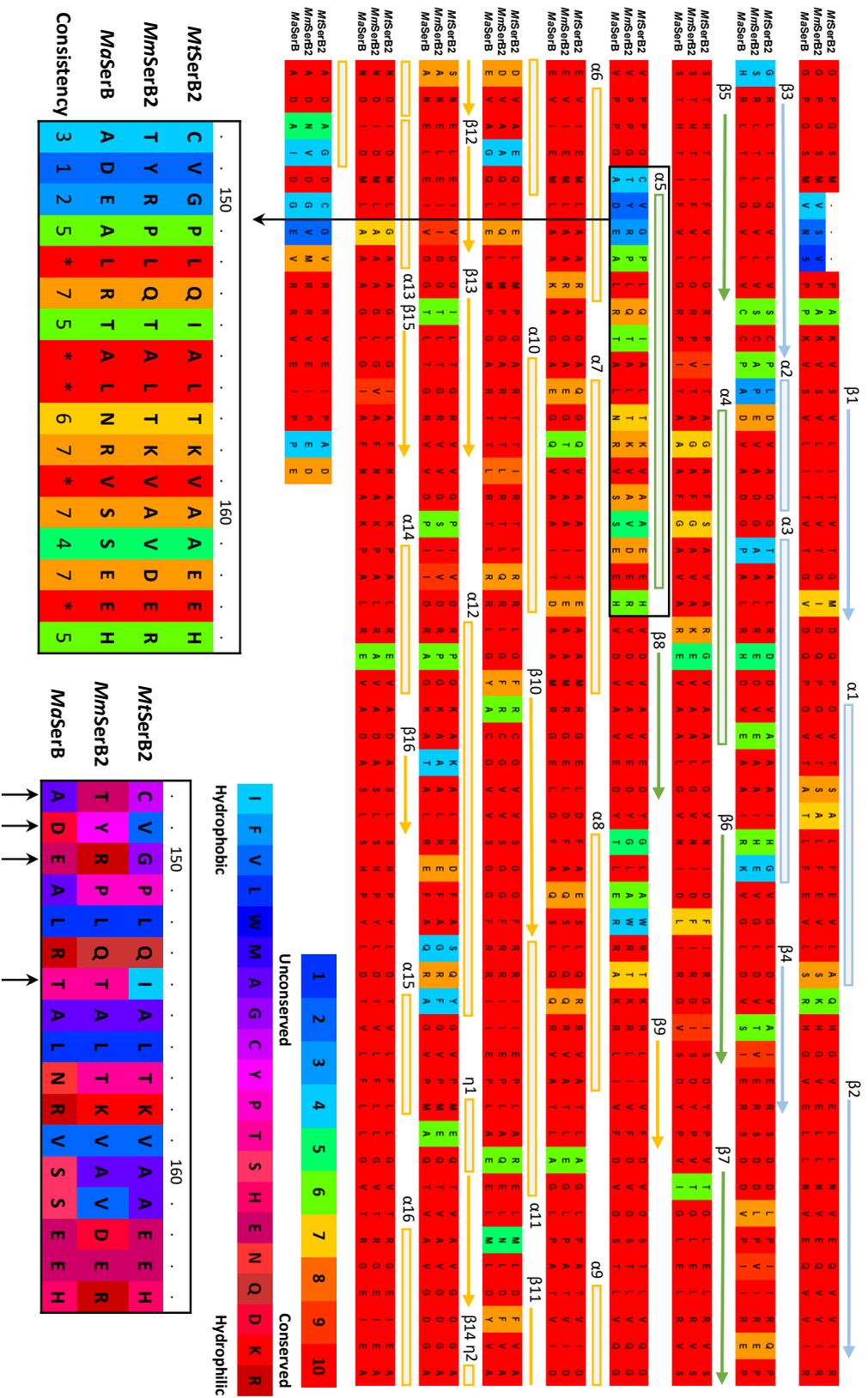


Figure 4.17: Alignment of the sequences of *MtSerB2*, *MaSerB* and *MmSerB2* performed with PRALINE tool [7]. The secondary structure elements based on *MaSerB* structure are depicted above the corresponding residues and colored according to the domain they belong to: ACT1 (blue), ACT2 (green) or PSP (orange). Boxes at the bottom: alignment focused on the second alpha helix of the ACT2 domain ($\alpha 5$). Residues are colored according to their conservation (left) and hydrophobicity (right).

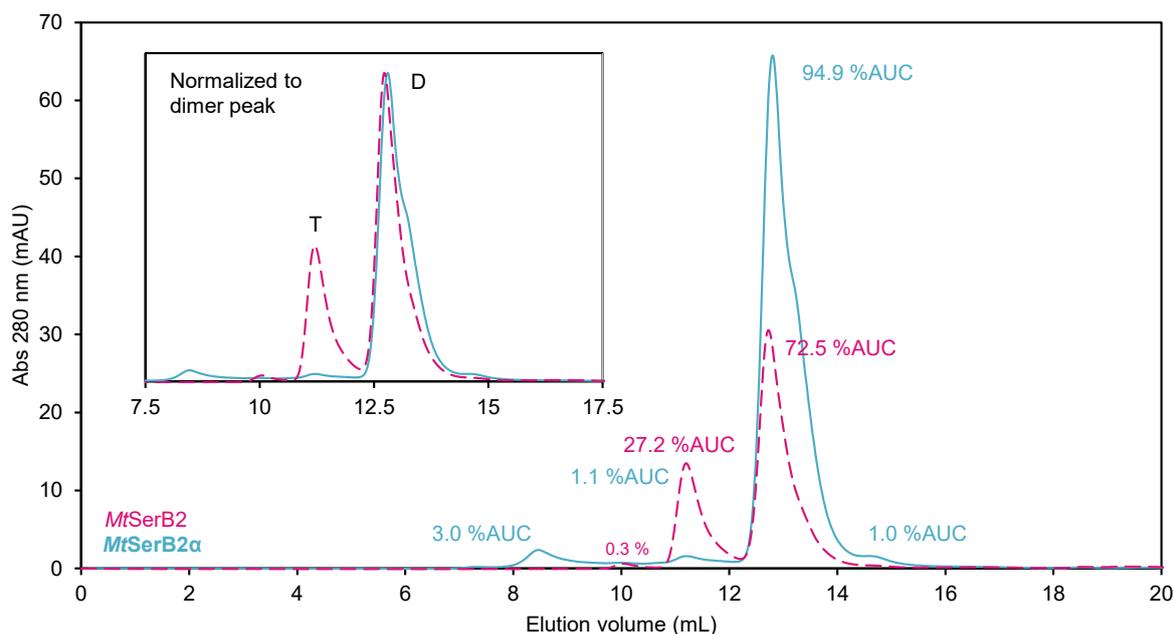


FIGURE 4.18: Superimposition of the SEC-UV chromatograms of *MtSerB2* (pink dotted line) and *MtSerB2 α* mutant (blue). Insert: data normalized to the dimer peak.

N-terminal part of the second alpha helix of the ACT2 domain is involved in the tetramerization of the enzyme. The crystallographic structure of *MaSerB* dimer shows that the residues mutated in this experiment point to the solvent (Figure 4.19).

The side chains of these residues are not involved in the stabilization of the dimer by intramolecular interactions. In a hypothetical tetramer, it is possible that these solvated, larger, charged and polar surface residues found in *MmSerB2* and *MaSerB* would destabilize the interface in which they would be located. It could be more favorable for apolar and smaller residues such as C, V, G and I found in *MtSerB2* to be buried within it. Besides this interfacial hypothesis, it cannot be excluded that the mutations made in this experiment rather induce a conformational change preventing tetramer formation. These so-called *allosteric mutations* therefore act indirectly from a distance to destabilize the interface. The effect that such mutations can have on the conformational dynamics of a protein is well documented in the case of the dimer-tetramer equilibrium of the PyrR family of pyrimidine operon attenuators [9].

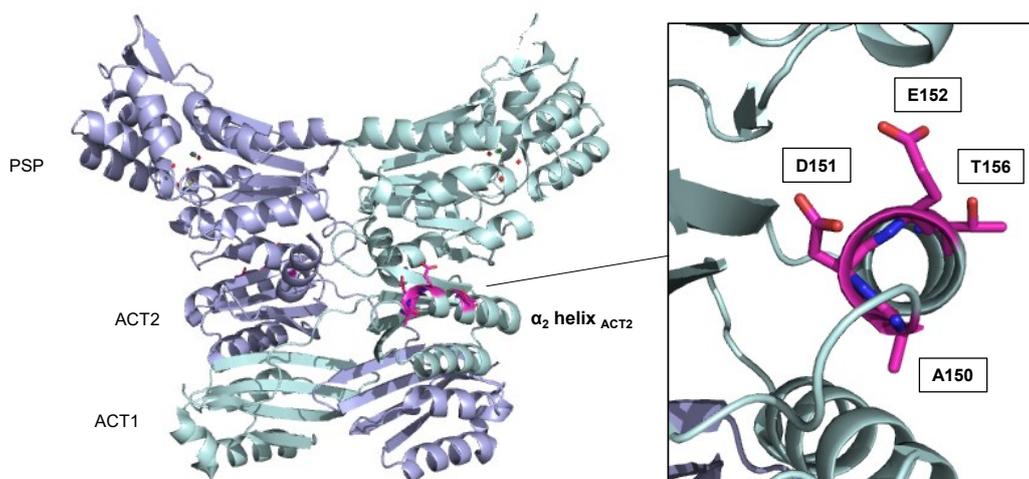


FIGURE 4.19: Localization of the α_2 -helix of the ACT2 domain in *MaSerB* structure 3P96. The residues shown in magenta colored sticks are the residues homologous to those shown to be involved in the tetramerization of *MtSerB2* by SEC.

***MtSerB2* tetramer is a rather globular, distinct oligomeric species**

SAXS provided the opportunity to obtain direct structural information about the tetramer formed by *MtSerB2*. SEC-SAXS experiments with *MtSerB2* tetramer also yielded high quality scattering curves (Figure 4.20A) giving access to its size parameters (Table 4.3). First, the molar mass values calculated based on the SAXS data corroborate the SEC-MALS results and confirm the tetrameric state of the enzyme. Obtaining these data provided confirmation that the tetramer is a well-folded species that behaves like a globular protein in solution, as evidenced by the shape and maximum position of the dimensionless Kratky plot (Figure 4.20C). As for the dimer, the experimental R_g value is slightly higher than the theoretical value of 35.2 Å calculated for a perfectly globular 173 kDa protein. The differences observed in the shapes of the scattering curves and related $P(r)$ functions (Figure 4.20B) also indicates that the tetramer is structurally distinct from the dimer. Its larger size is well reflected in the greater R_g and D_{max} values.

TABLE 4.3: SAXS parameters of *MtSerB2* tetramer compared to those of *MtSerB2* dimer. (1) derived from the Guinier plot (2) calculated from the $P(r)$ function

	<i>MtSerB2</i> tetramer	<i>MtSerB2</i> dimer
R_g (Å)	39.89 ± 0.06 ⁽¹⁾ ; 39.62 ± 0.04 ⁽²⁾	32.47 ± 0.02 ⁽¹⁾ ; 32.55 ± 0.02 ⁽²⁾
$I(0)$ (cm ⁻¹)	$0.06 \pm 5.30 \cdot 10^{-5}$ ⁽¹⁾ ; $0.06 \pm 5.26 \cdot 10^{-5}$ ⁽²⁾	$0.09 \pm 3.32 \cdot 10^{-5}$ ⁽¹⁾ ; $0.09 \pm 3.64 \cdot 10^{-5}$ ⁽²⁾
D_{max} (Å)	127	111
V_p (nm ³)	254	127
M_W Bayesian inference (kDa)	185.8 (162.7 to 221.1)	88.3 (84.3 to 95.5)
M_W Porod volume (kDa)	210.7	105.5

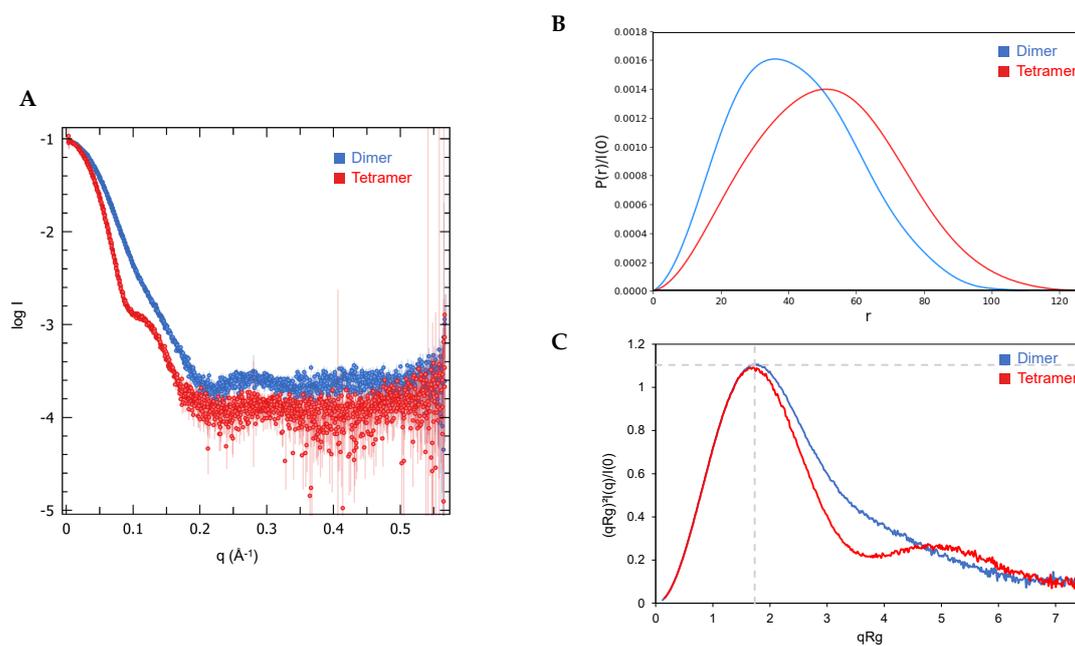


FIGURE 4.20: Analysis of *MtSerB2* tetramer solution structure from SAXS data and comparison with *MtSerB2* dimer. (A) Experimental SAXS profiles ($\log I(q)$ versus q) of *MtSerB2* tetramer and dimer. (B) Calculated $P(r)$ functions. (C) Dimensionless Kratky plots.

***MtSerB2* tetramer could be an assembly of monomers with global D2 symmetry**

The subsequent step was to investigate the architecture of *MtSerB2* tetramer by combining the SAXS data with the structural knowledge acquired on the enzyme directly and through the study of the orthologs. Our strategy was to model plausible tetrameric assemblies using protein-protein docking and to confront them by comparing their calculated SAXS profile to the experimental curve. Three types of potential tetrameric assemblies were probed:

1. A **dimer of domain-swapped dimers (DD)**, resulting from the interaction between two dimers possessing the known *butterfly-like* domain-swapped architecture (Figure 4.21).

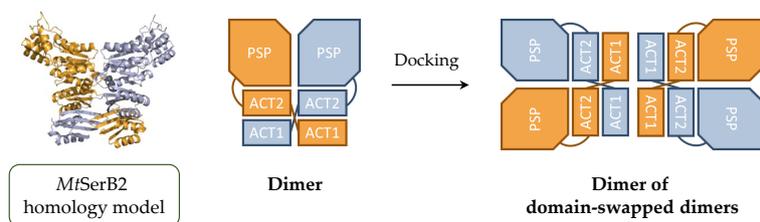


FIGURE 4.21: Modeling strategy for the dimers of domain-swapped dimers.

2. A **tetramer composed of 4 closed monomers (T4)**, based on the underlying principles of domain-swapping (Figure 4.22). In theory, the monomer associated with *MtSerB2* domain-swapped dimer would exist in a form where the ACT1 domain is located underneath its own ACT2 domain so that the interactions established between these two domains (i.e. the *primary interface*) are identical to those established in the dimer. The relative spatial arrangement of the three domains (PSP-ACT2-ACT1) would thus be identical to that observed for a dimer half, with the only part whose conformation differs between the dimer and the monomer being the hinge-loop. A closed monomer could therefore be modeled from *MtSerB2* dimer homology model. The positions of PSP and ACT2 domains of protomer 1 and ACT1 domain of protomer 2 were conserved, and then the hinge loop from the ACT1 domain was manually connected to the ACT2 domain. The reassembled structure was subjected to a 200 ns Molecular Dynamics simulation at 310 K and 1 bar to better represent the conformation that a closed monomer might adopt in solution.

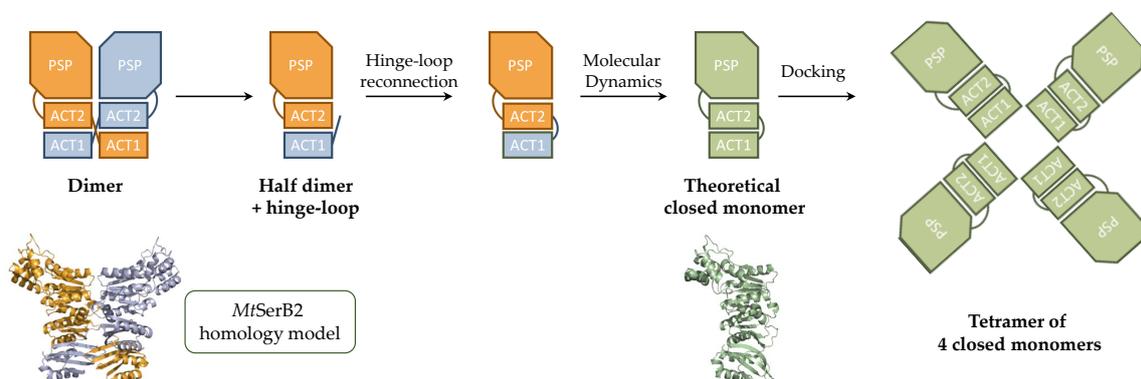


FIGURE 4.22: Modeling strategy for the tetramers composed of 4 closed monomers.

3. An **ACT1 domain-swapped tetramer (DST)**, in which the monomers exchange their ACT1 domains with each other (Figure 4.23). As in the case of the dimer, the tetramer could also be formed through domain-swapping. Instead of being back-swapped as in the dimer, ACT1 domain of protomer 1 would interact with ACT2 domain of protomer 2, whose ACT1 domain would interact with ACT2 domain of protomer 3 and so on until a finite assembly of 4 subunits is built. Again, according to the principles of domain-swapping, the relative positions of PSP, ACT2 and ACT1 domains would be conserved in each subunit and only the conformation of the hinge loop would have changed from that adopted in the dimer. To represent this, the subunit introduced in the docking process consists of a dimer half lacking the hinge-loop region to be reconnected subsequently.

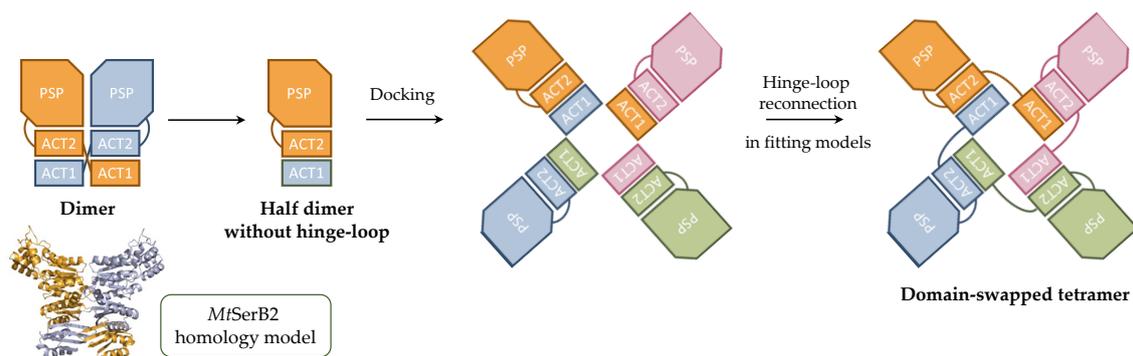


FIGURE 4.23: Modeling strategy for the ACT1 domain-swapped tetramers.

It is well established that the majority of homo-oligomeric proteins exhibit a globally symmetric structure since it provides the oligomeric complex with advantages in terms of stability and function. [10] Assemblies with asymmetrically arranged subunits exist but are very rare (less than 5% of homo-oligomeric structures) [11, 12]. Accordingly, our working hypothesis implied that docking should provide symmetric tetramer models. Two symmetry groups are possible for a tetramer: the **cyclic point group C4** and the **dihedral point group D2**.

- The **cyclic group C4** contains a single axis of rotational symmetry and allows for the assembly of ring-shaped tetramers where the 4 subunits are symmetrically arranged around the axis.
- The **dihedral group D2** contains an axis of second-order rotational symmetry perpendicular to a second axis of the same type.

Symmetry-aware protein docking algorithms were therefore needed to construct reasonable tetrameric structures. The software used are presented in Table 4.4 with their symmetry constraints and the types of tetramer they allowed us to model.

TABLE 4.4: Software allowing for the modeling of symmetric homomers used in this work and types of tetramer that could be modeled using them. DD: dimer of domain-swapped dimers, T4: tetramer of 4 closed monomers, DST: domain-swapped tetramer.

Software	Symmetry constraints	Tetramer types	Webserver	Ref.
M-ZDOCK	C2 to C24	DD, T4, DST	https://zdock.umassmed.edu/m-zdock/	[13]
ClusPro	C2 and C3	DD	https://cluspro.bu.edu/login.php	[14–17]
GalaxyTongDock	C _n and D _n (up to 12)	DD, T4, DST	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TONGDOCK_INTRO	[18]

A total of 197 models (137 DD, 30 T4, and 30 DST) were generated by these docking algorithms. The theoretical scattering curve of each model was calculated and fitted

to the experimental SAXS curve using CRY SOL. Models yielding a χ^2 value over 15.00 were instantly discarded, leading to 6 models remaining for further examination.

Among these 6 models, 5 were generated via ClusPro by docking the dimeric structure under C2 symmetry constraint and yielded χ^2 values between 8.05 and 14.01 (Figure 4.24). However, since the dimer was used as the asymmetric unit in the docking, none of these models is globally symmetric with respect to the positioning of the monomer in the structure. Therefore, we did not consider these asymmetric dimers of dimers as plausible tetramer candidates. Moreover, the fact that the tetrameric population is not readily reformed upon concentration of the dimeric population, as shown in Section 4.1.3 (page 110), also supports the exclusion of tetramers formed by dimers of domain-swapped dimers without prior conformational change.

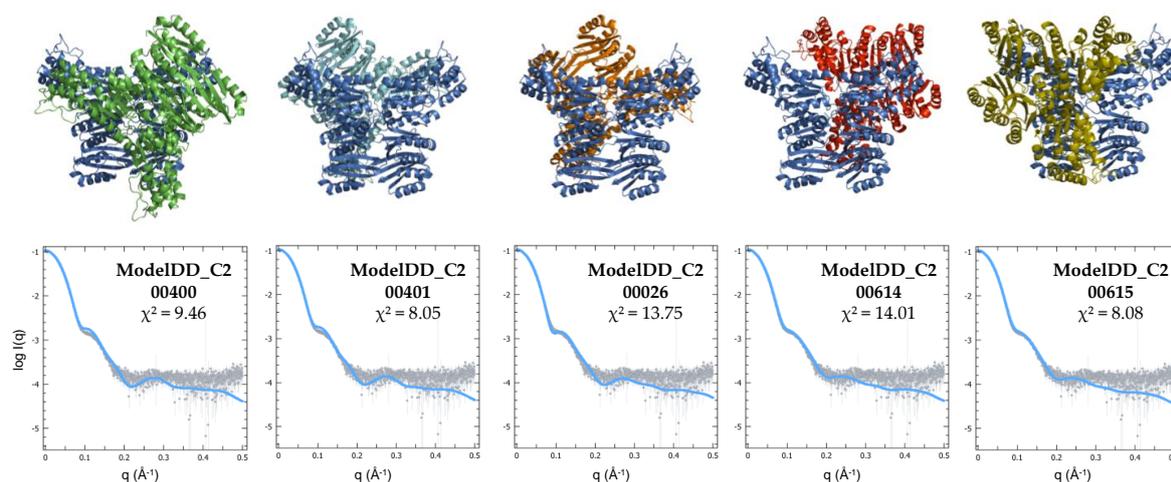


FIGURE 4.24: Representation of the asymmetric dimers of domain-swapped dimer models obtained following the protein-protein docking procedure and corresponding fits to the experimental SAXS data.

The only remaining model was obtained via GalaxyTongDock and is also the one that produced the best fit to the experimental SAXS data, with a χ^2 of 7.69. It represents a tetramer with a global D2 symmetry formed by 4 closed monomers (Figure 4.25). As can be seen, the fit to the data could be improved. The model was therefore refined with SREFLEX and the resulting generated conformers showed that the SAXS profile could be better fitted through a slight adjustment of the relative position of the monomers, as could happen in solution.

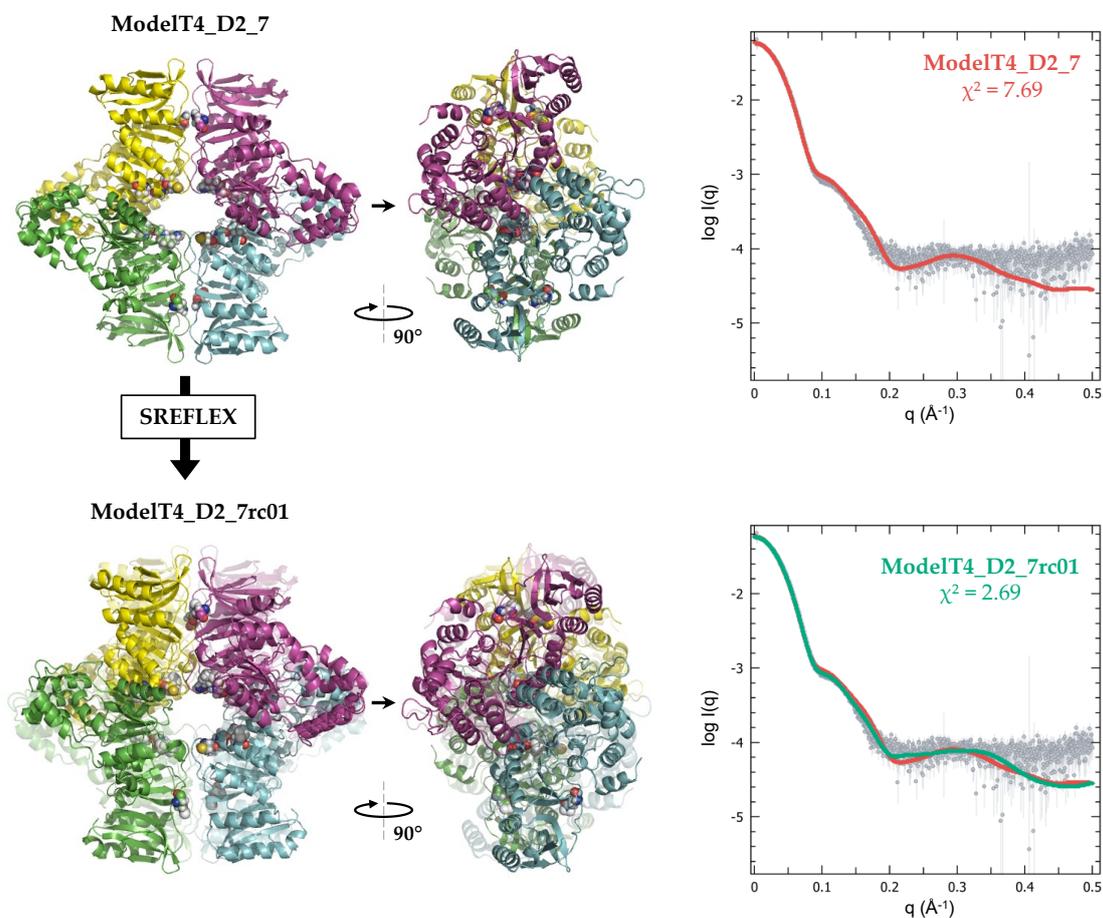


FIGURE 4.25: Representation of the only symmetric tetramer model retained following the protein-protein docking procedure. Top panel: architecture obtained as the docking result and corresponding fit to the experimental SAXS data. Bottom panel: architecture obtained following structure refinement using SREFLEX (plain) superimposed with the docking result and superimposition of the corresponding fits to the experimental SAXS data. Residues C148, V149, G150, I254 and Q92 are depicted as spheres.

This model is supported by the non-trivial argument that dihedral symmetry prevails among soluble cytoplasmic enzymes forming higher-order oligomers. D2 tetramers are much more common than C4 tetramers as the existence of different types of interface along the two-fold axes provides these structures not only with greater stability but also with the ability to build allosteric control [10]. Moreover, in line with our site-directed mutagenesis experiments, residues C148, V149, G150, and I154, involved in *MtSerB2* tetramerization, are located at the intersection of the

two-fold axes and core of the model. Residue Q92, in the hinge-loop, is also present at the interface where one of the axes is located. These residues are shown as spheres in the model of Figure 4.25. Another interesting observation is the interface established at the α_2 -helices of the C1 cap, known in *HsPSP* to unfold to accept substrates and release products of the dephosphorylation reaction. The enzyme kinetics experiments that will be presented in section 4.3 make this structural feature very significant.

Although supported by biological facts and our experimental observations, this model remains a hypothesis to be validated. To generate it, we used a protein-protein docking approach. However, one of major challenges of the method include dealing with flexibility and conformational changes. With rigid docking, the probability of obtaining correct solutions critically depends on the starting protomer structure, whose conformation must match the one adopted within the oligomer. This strategy was adopted in our work, considering that the PSP-ACT2-ACT1 structural motif found in the domain-swapped dimer would also be found in the tetramer. Yet, it is not impossible that the protomer forming *MtSerB2* tetramer adopts a totally different conformation from our hypothesis given the highly flexible nature of the hinge-loop.

This possibility has been explored and illustrated using the CORAL algorithm of the ATSAS suite. CORAL performs SAXS-based rigid body modeling of multi-domain protein complexes that miss some fragments such as termini portions or interdomain linkers [19]. Briefly, CORAL works by randomly moving domains from one another. When a relative position is found such that a flexible linker of defined length can be attached between the N- and C-terminal portions of subsequent domains, it is kept and the missing flexible linker is modeled using dummy atoms. To find the optimal positions and orientations of the domains, and the approximate conformations of the linkers, CORAL employs a simulated annealing protocol. The later allows for the construction of an assembly of subunits without steric clashes that minimizes the discrepancy between the experimental scattering data and the curves calculated from the appropriate subunits ensembles. The model resulting in the best χ^2 value of fit is proposed as the result.

To run CORAL, the coordinate file of *MtSerB2* homology model was split in two separate files. The first one contains the PSP-ACT2 part and the second one contains the ACT1 domain. A flexible linker of 12 residues, corresponding to the hinge-loop,

was imposed between the C-terminal portion of PSP-ACT2 and the N-terminal portion of ACT1 domain. C4 and D2 symmetries were also imposed in order to ensure the modeling of a symmetrical tetramer. Ten independent runs were performed for each symmetry to generate a pool of possible models.

These calculations resulted in 20 structurally distinct tetramer models, all of which corresponded well to the experimental curve: ten C4 symmetry models with χ^2 values between 2.21 and 3.19, and ten D2 symmetry models with χ^2 values between 2.00 and 2.68. The three C4 and D2 models with the best fits are shown in Figure 4.26 as an example. Even though they are all very different, it is not surprising that all these models fit the experimental data well, since CORAL is dedicated to provide results meeting this condition. However, it should be kept in mind that CORAL, in contrast to protein-protein docking algorithms, does not use equations describing steric and physicochemical complementarity at the protein-protein interface to define the best relative orientation of the protomers in the complex. The proposed results are therefore not necessarily valid from a biological point of view. The conformations of the dummy residues hinge-loops are also not all adoptable by the actual sequence of residues composing the hinge-loop of *MtSerB2*. All in all, this makes the models proposed by CORAL difficult to discriminate from each other. Nevertheless, these results remind us that other protomer configurations, where the two ACT domains are separated for example, could as well explain the shape of the SAXS curve.

Whether it is the D2 tetramer isolated from our protein-protein docking strategy, or the tetramers obtained by CORAL, a common feature confirmed by the Kratky plot emerges from the generated models: *MtSerB2* tetramer is a rather globular species. The assemblies are also composed of four interacting monomers, and would be a clue that the monomeric form of *MtSerB2* could be the common denominator between the dimeric and tetrameric forms of the protein. It would appeal for a much more complex mechanism than a direct equilibrium between the dimer and the tetramer by involving dissociation of the oligomeric species and re-association of the monomeric species. We will see in Chapter 5 that this hypothesis will be strengthened by other experimental evidence.

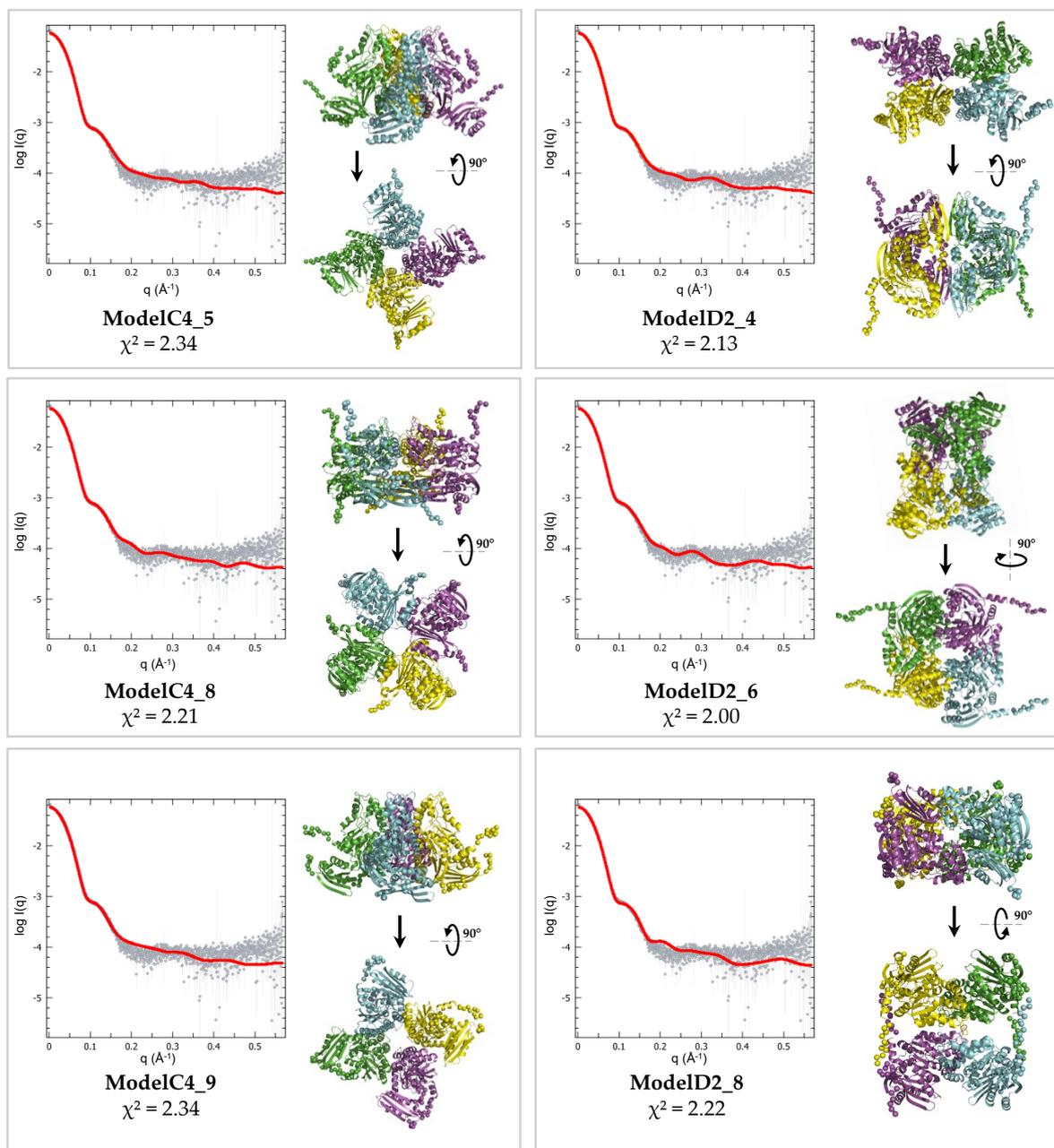


FIGURE 4.26: Best-fitting *MtSerB2* tetramers models generated by CORAL with C4 and D2 symmetry constraints.

4.3 Enzymological characterization of the oligomeric species

The works of Yadav, Shree, Arora, Grant and co-workers report assays of *MtSerB2* phosphatase activity [1, 4, 5, 20]. However, unbeknownst to the experimenter, it is likely that the samples used for these assays contained both dimer and tetramer of *MtSerB2*. With this in mind, we aimed to characterize the individual activity of the two native oligomeric states of the enzyme.

To this end, steady-state kinetics studies of *O*-phospho-L-serine (PS) dephosphorylation by *MtSerB2* have been performed. Enzyme activity was followed by quantifying the inorganic phosphate produced under the initial velocity conditions at different PS concentrations.

4.3.1 *MtSerB2* dimer is active and exhibits catalytic properties similar to those of *MmSerB2* and *MaSerB*

The plot of initial velocity versus PS concentration for *MtSerB2* dimer is shown in Figure 4.27A. A native-PAGE analysis (Figure 4.27B) of the studied protein batch attests to the purity of the dimeric sample. The shape of the curve, where the velocity rises to a maximum before decreasing as PS concentration increases, is indicative of substrate inhibition. The kinetic parameters were therefore determined by fitting Equation 4.1, accounting for total uncompetitive substrate inhibition, to the experimental data. As shown in the first two rows of Table 4.5, the results are in line with the parameters determined by Grant in reference [20], with however a lower PS inhibition constant (K_{iS}).

$$v = \frac{V_{max}[S]}{K_M + [S](1 + \frac{[S]}{K_{iS}})} \quad (4.1)$$

In Equation 4.1, derived from the scheme presented in Figure 4.28, V_{max} is the maximum velocity, $[S]$ is the free substrate (*O*-phospho-L-serine, PS) concentration, K_M is the Michaelis constant (substrate dissociation constant) and K_{iS} is the dissociation constant of the substrate for the inhibitory site. The value of the catalytic constant k_{cat} is obtained by dividing V_{max} by the total enzyme concentration $[E]_t$.

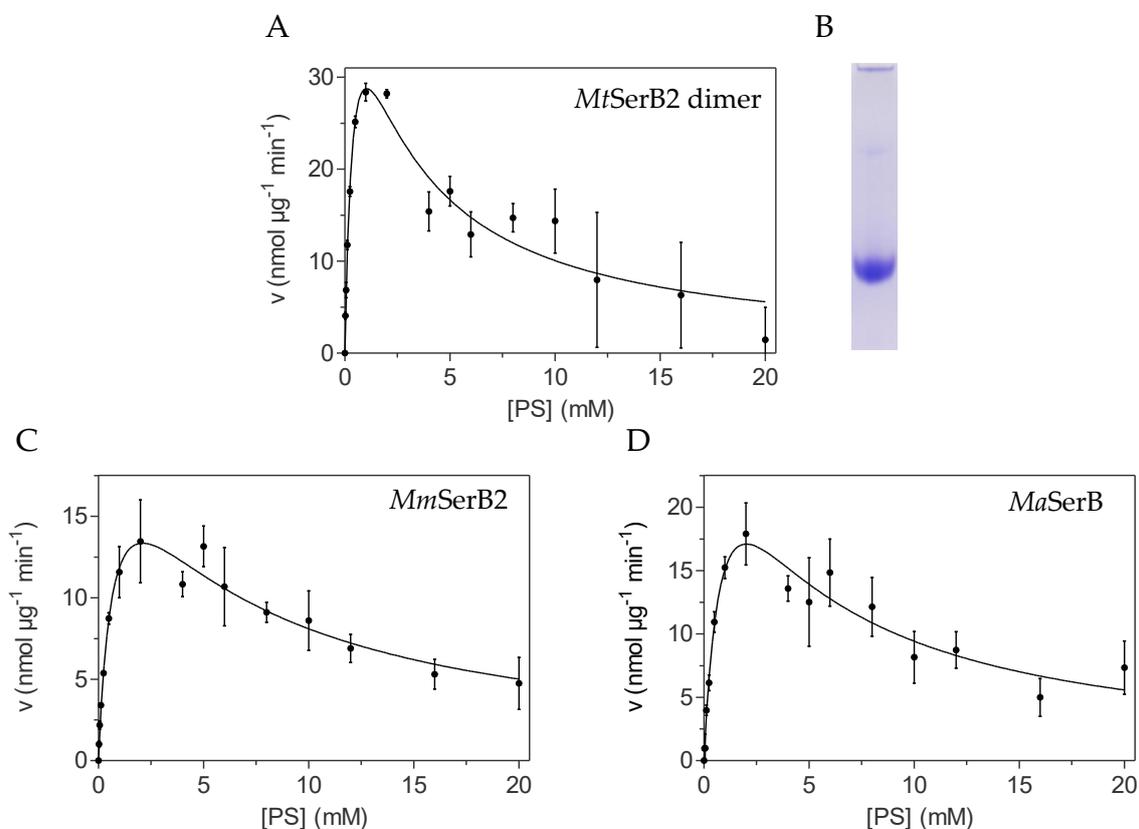


FIGURE 4.27: Initial velocity vs substrate concentration curves for the dephosphorylation of *O*-phospho-L-serine (PS) by (A) *MtSerB2* (C) *MmSerB2* (D) *MaSerB*. (B) Native PAGE analysis of *MtSerB2* dimer sample used for the kinetics analysis.

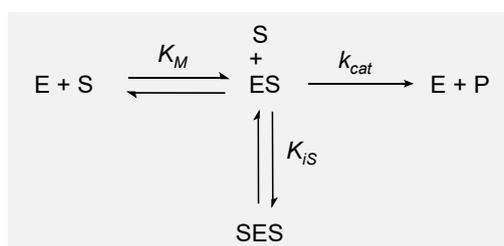


FIGURE 4.28: Total uncompetitive substrate inhibition mechanism scheme. E = enzyme, S = substrate, P = product, k_{cat} = catalytic constant, K_M = substrate dissociation constant, K_{IS} = dissociation constant of the substrate at the inhibitory site.

For comparison purposes, the experiment was repeated with *MaSerB* and *MmSerB2*. The velocity curves obtained for both enzymes show a similar shape to that obtained for *MtSerB2* (Figures 4.27C and D). The equation for substrate inhibition also fitted the experimental data well and was used to determine the associated kinetic parameters. It can be seen that the latter are essentially of the same order of magnitude for all three orthologs and that no drastic variations are observed.

TABLE 4.5: Kinetic parameters for the dephosphorylation of *O*-phospho-L-serine by *MtSerB2* dimer, *MmSerB2* and *MaSerB*.

Enzyme	K_M (mM)	V_{\max} (nmol μg^{-1} min $^{-1}$)	k_{cat} (s $^{-1}$)	K_{IS} (mM)
<i>MtSerB2</i> dimer	0.48 ± 0.17	55.23 ± 10.90	39.91 ± 7.87	2.26 ± 0.70
<i>MtSerB2</i> (Grant [20])	0.38 ± 0.05	not shown	18.6 ± 1.1	23 ± 5
<i>MmSerB2</i>	0.81 ± 0.18	23.69 ± 2.77	17.31 ± 2.02	5.42 ± 1.12
<i>MaSerB</i>	1.07 ± 0.37	35.21 ± 7.32	25.77 ± 5.36	3.81 ± 1.24

These results indicate that *MtSerB2* dimer, like many enzymes [21], is inhibited by its substrate. Its kinetic characteristics resemble those of *MmSerB2* and *MaSerB2*, which is consistent with their structural similarity. The model that could be successfully fitted to the experimental data of the three enzymes involves the binding of PS to an inhibitory site distinct from the catalytic site. Binding to this allosteric site leads to the formation of a non-productive PS-*MtSerB2*-PS complex (Figure 4.28). As shown by the structures of *MaSerB* co-crystallized with L-serine and Grant's site-directed mutagenesis experiments, dual-ACT PSPs possess a regulatory allosteric site at the interface of ACT1 and ACT2 domains [20, 22, 23]. However, while it is possible that PS partly binds to this site, we do not believe the phenomenon is responsible for the total observed substrate inhibition. As a matter of fact, PS inhibits *BmSerB* in a similar way while the enzyme has only one ACT domain and thus lacks the ACT1-ACT2 interface found in *MtSerB2*, *MaSerB* and *MmSerB2* (see section 3.3.5 of Chapter 3). The mechanism of substrate inhibition observed in PSPs is therefore probably independent of the presence of the ACT1 domain. Nevertheless, it should not be excluded that the two types of constructions exhibit different substrate inhibition mechanisms.

4.3.2 *MtSerB2* tetramer is inactive

Next, the activity of *MtSerB2* tetramer was evaluated using the same method. The plot of initial velocity versus PS concentration is shown in Figure 4.29A and its shape is also indicative of substrate inhibition. The kinetic parameters for the sample were determined using Equation 4.1 and are outlined in Table 4.6. While K_M and K_{IS} values are in line with those shown by the dimer, the k_{cat} value is significantly lower as it went from 39.91 to 8.65 s⁻¹.

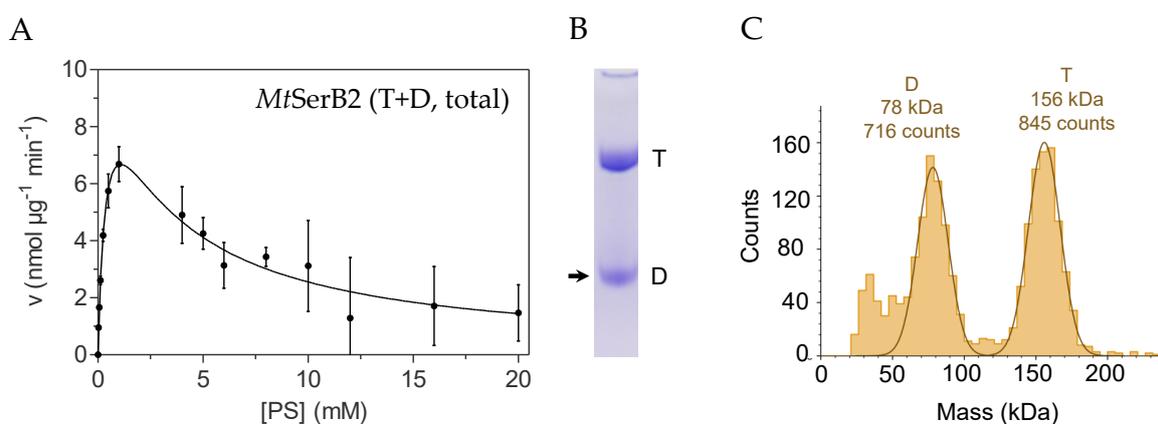


FIGURE 4.29: (A) Initial velocity vs substrate concentration curves for the dephosphorylation of *O*-phospho-L-serine (PS) by *MtSerB2* enriched tetramer batch (T+D) (B) Native PAGE analysis of *MtSerB2* enriched tetramer batch (T+D) used for the kinetics analysis. (C) Mass photometry analysis of *MtSerB2* enriched tetramer batch (T+D).

It is important to note that the sample was already contaminated with some amount of dimer, as shown on the native PAGE gel in Figure 4.29B. The low activity observed for this enriched tetramer batch could therefore be solely due to the presence of the dimer fraction. To verify this, we re-calculated the activity of the sample considering that the dimer was the only active species.

The first step was to determine the concentration of the dimer in the sample. The single-molecule detection properties of mass photometry allowed us to accurately determine the ratio of populations as shown in Figure 4.29C. From this ratio and from the amount of total subunit measured in the sample by absorbance measurement at 280 nm, we were able to recalculate the assumed amount of active subunit, i.e. belonging to the dimer, in the sample. This corrected quantity could be substituted

for the total subunit amount in the calculation of the activity per microgram of enzyme per minute ³.

TABLE 4.6: Kinetic parameters for the dephosphorylation of *O*-phospho-L-serine by the batches containing *MtSerB2* tetramer and dimer (T+D, determined using the total monomeric amount and only the corrected amount of active species) and *MmSerB2* dimer (pure).

Enzyme	K_M (mM)	V_{\max} (nmol μg^{-1} min $^{-1}$)	k_{cat} (s $^{-1}$)	K_{iS} (mM)
<i>MtSerB2</i> T+D (total)	0.43 ± 0.12	11.97 ± 1.74	8.65 ± 1.26	2.75 ± 0.62
<i>MtSerB2</i> T+D (corrected)	0.43 ± 0.12	54.46 ± 7.90	39.36 ± 5.71	2.75 ± 0.62
<i>MtSerB2</i> dimer (pure)	0.48 ± 0.17	55.23 ± 10.90	39.91 ± 7.87	2.26 ± 0.70

The recalculated kinetic parameters, considering that the dimer is the only active species in solution, are presented in Table 4.6 under the name "*MtSerB2* T+D (corrected)". The value of k_{cat} is now equivalent to that obtained for the pure dimer sample (39.36 and 39.91 s $^{-1}$), indicating that the tetramer present in the analyzed sample is likely inactive. The fact that the K_M and K_{iS} values match between the pure dimer batch and the contaminated tetramer batch also suggests that the tetramer does not bind PS. Otherwise, an increase in these parameters would have been observed since a greater substrate concentration would have been needed to overcome the competition between binding to the active dimer and binding to the inactive tetramer.

The apparent inactivity of *MtSerB2* tetramer and its inability to bind phosphoserine support the D2 symmetric model (*ModelT4-D2-7*) isolated as a result of the protein-protein docking process. Indeed, in this model, the four α_2 -helices of the C1 caps are involved in protein-protein interfaces. As demonstrated in the crystallographic structures of *HsPSP* and by Molecular Dynamics simulations [24], this particular helix unfolds to let in and out substrates and products of the dephosphorylation reaction and is therefore deeply involved in the catalytic activity of the enzyme. The architecture of the proposed *MtSerB2* tetramer model suggests that these helices would be constrained by the interactions they establish at the oligomeric interface, loosing the dynamic behavior necessary for enzyme activity. Access to the active site would thus be prevented in the tetramer, which would result in the inactivity of this species.

³Details of the calculation are presented in Appendix H.

4.3.3 *MtSerB2* in the monomeric form could be almost inactive

Finally, the activity of the monomeric *MtSerB2* Δ ACT1 mutant was assayed. The plot of initial velocity versus PS concentration is shown in Figure 4.30. The kinetic parameters are reported in Table 4.7 and were also determined by fitting Equation 4.1 to the data, although the error associated with higher PS concentrations could also have allowed the fit of the standard Michaelis-Menten equation.

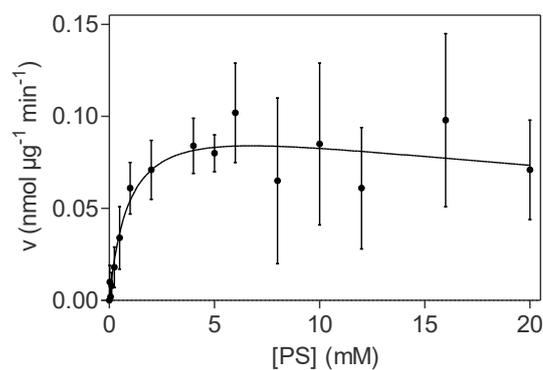


FIGURE 4.30: Initial velocity vs substrate concentration curves for the dephosphorylation of *O*-phospho-*L*-serine (PS) by monomeric *MtSerB2* Δ ACT1 mutant.

TABLE 4.7: Kinetic parameters for the dephosphorylation of *O*-phospho-*L*-serine by monomeric *MtSerB2* Δ ACT1 mutant.

Enzyme	K_M (mM)	V_{max} (nmol μg^{-1} min $^{-1}$)	k_{cat} (s $^{-1}$)	K_{iS} (mM)
<i>MtSerB2</i> Δ ACT1	0.96 ± 0.45	0.11 ± 0.02	0.08 ± 0.01	47.37 ± 40.50

The shape of the curve shows that the mutant has non-zero enzymatic activity, but its velocity is drastically decreased compared to the native *MtSerB2* dimer, as evidenced by the decrease in the k_{cat} value from 39.91 to 0.08 s $^{-1}$. In comparison, *BmSerB* which has a similar PSP-ACT2 architecture has a turnover number of 15.30 s $^{-1}$. The loss of activity could therefore be associated with a specific conformation adopted by *MtSerB2* in the absence of its ACT1 domain. This result provides a first indication that a native monomeric form of *MtSerB2* could be (almost) inactive. The fact that *MtSerB2* appears to have to dimerize in order to be active, supported by the fact that little or no monomer is observed in solution, also supports this hypothesis.

4.4 Summary of the findings

In the literature, *MtSerB2* is described as an ACT1 domain-swapped dimer forming tetramer in the presence of L-serine. Early in this work, we made an unexpected observation against this description following size exclusion chromatography and non-denaturing electrophoresis experiments. We found that even in the absence of L-Ser, *MtSerB2* existed as two distinct oligomeric species. Using MALS, these species could be identified as dimeric (majority species) and tetrameric (minority species) forms of the enzyme. The chromatograms also showed the possible presence of a small amount of monomer, and intermediate and larger trace size species.

The interconversion between the dimer and the tetramer has been probed. Both species are stable and can be separated and analyzed individually. Their respective concentrations in solution seem to be governed by a slow equilibrium in favor of the dimer and independent of the total protein concentration of the sample.

Crystallization of the oligomers proved unsuccessful, but insights into their architecture were obtained via SAXS and study of variants designed by site-directed mutagenesis.

First, it was confirmed that *MtSerB2* dimer indeed adopts a fold similar to that of *MaSerB* and *MmSerB2* homologs in solution: a butterfly-like architecture in which two monomers swap their N-terminal ACT1 domain. As demonstrated by the behavior of a truncated mutant, this small domain is essential for the dimerization of the enzyme since it establishes the majority of stabilizing interactions at the oligomeric interface. Its truncation results in the formation of a virtually inactive monomer and a small amount of trimer in solution. In addition to their structural similarity, the catalytic behavior of *MtSerB2*, *MaSerB* and *MmSerB2* towards *O*-phospho-L-serine, their substrate, is comparable as demonstrated by the study of their dephosphorylation kinetics.

Next, SAXS experiments showed that *MtSerB2* tetramer is a globular species, structurally distinct from the dimer in solution. Combined with protein-protein docking and rigid body modeling, the experimental data also allowed to model plausible tetramer structures. A promising model, composed of 4 monomers folded on themselves according to the principles of domain-swapping and assembled following a global D2 dihedral symmetry has been isolated. This model is interesting because

residues involved in the tetramerization propensity of the enzyme (Q92, C148, V149, G150, I154), identified by mutagenesis and SEC, are located at the protein-protein interfaces. The model also explains the inactivity of the tetrameric form, revealed during kinetic analyses, by the shielding of the dynamic helix that governs the access to the active site. Still, this model is a hypothesis that needs to be challenged by orthogonal experimental methods. The assumption of a dimer-of-dimers like tetramer having been ruled out by the total concentration independence of the dimer/tetramer equilibrium and the symmetry constraint, a more complex mechanism involving the dissociation of oligomers into monomers could govern the equilibrium between the two major species. The following chapter provides an argument supporting the hypothesis of such a mechanism.

Bibliography

- [1] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [2] Janet Newman, David Egan, Thomas S Walter, Ran Meged, Ian Berry, Marouane Ben Jelloul, Joel L Sussman, David I Stuart, and Anastassis Perrakis. Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*, 61(10):1426–1431, 2005.
- [3] A. Chaikuad, S. Knapp, and F. von Delft. Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens. *Acta Crystallographica Section D: Biological Crystallography*, 71(8):1627–1639, 2015.
- [4] S. Shree, A.K. Singh, R. Saxena, H. Kumar, A. Agarwal, V.K. Sharma, K. Srivastava, K.K. Srivastava, S. Sanyal, and R. Ramachandran. The *M. tuberculosis* HAD phosphatase (Rv3042c) interacts with host proteins and is inhibited by Clofazimine. *Cellular and Molecular Life Sciences*, 73(17):3401–3417, 2016.
- [5] G. Arora, P. Tiwari, R.S. Mandal, A. Gupta, S. Sharma, D. and Saha, and R. Singh. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *Journal of Biological Chemistry*,

289(36):25149–25165, 2014.

- [6] Huanting Liu and James H Naismith. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, 8(1):1–10, 2008.
- [7] P. Bawono and J. Heringa. PRALINE: a versatile multiple sequence alignment toolkit. In *Multiple Sequence Alignment Methods*, pages 245–262. Springer, 2014.
- [8] M.J. Betts and R.B. Russell. Amino acid properties and consequences of substitutions. *Bioinformatics for Geneticists*, 317(289):10–1002, 2003.
- [9] T. Perica, Y. Kondo, S.P. Tiwari, S.H. McLaughlin, K.R. Kemplen, X. Zhang, A. Steward, N. Reuter, J. Clarke, and S.A. Teichmann. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science*, 346(6216):1254346, 2014.
- [10] D.S Goodsell and A.J. Olson. Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):105–153, 2000.
- [11] L.S. Swapna, K. Srikeerthana, and N. Srinivasan. Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. *PloS One*, 7(5):e36688, 2012.
- [12] A. Gaber and M. Pavšič. Modeling and structure determination of homo-oligomeric proteins: An overview of challenges and current approaches. *International Journal of Molecular Sciences*, 22(16):9081, 2021.
- [13] B. Pierce, W. Tong, and Z. Weng. M-ZDOCK: a grid-based approach for C_n symmetric multimer docking. *Bioinformatics*, 21(8):1472–1478, 2005.
- [14] I.T. Desta, K.A. Porter, B. Xia, D. Kozakov, and S. Vajda. Performance and its limits in rigid body protein-protein docking. *Structure*, 28(9):1071–1081, 2020.
- [15] C. Vajda, S. and Yueh, D. Beglov, T. Bohnuud, S. E Mottarella, B. Xia, D.R. Hall, and D. Kozakov. New additions to the Clus Pro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*, 85(3):435–444, 2017.
- [16] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, D. Beglov, and S. Vajda. The ClusPro web server for protein–protein docking. *Nature Protocols*, 12(2):255–278, 2017.

- [17] D. Kozakov, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, and S. Vajda. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*, 81(12):2159–2166, 2013.
- [18] T. Park, M. Baek, H. Lee, and C. Seok. GalaxyTongDock: Symmetric and asymmetric ab initio protein–protein docking web server with improved energy parameters. *Journal of Computational Chemistry*, 40(27):2413–2417, 2019.
- [19] M.V. Petoukhov, D. Franke, A.V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P.V. Konarev, and D.I. Svergun. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(2):342–350, 2012.
- [20] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.
- [21] M.C. Reed, A. Lieb, and H. F. Nijhout. The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays*, 32(5):422–429, 2010.
- [22] S. Shree and R. Ramachandran. Crystal structure of *Mycobacterium avium* SerB2 in complex with serine at ACT domain. doi:10.2210/pdb5JLP/pdb. 2016.
- [23] S. Shree and R. Ramachandran. Crystal structure of *Mycobacterium avium* SerB2 with serine present at slightly different position near ACT domain. doi:10.2210/pdb5JLR/pdb. 2016.
- [24] M. Haufroid, M. Mirgaux, L. Leherte, and J. Wouters. Crystal structures and snapshots along the reaction pathway of human phosphoserine phosphatase. *Acta Crystallographica Section D: Structural Biology*, 75(6):592–604, 2019.

Chapter 5

Characterization of *MtSerB2* oligomeric states in the presence of L-serine

5.1 Observation and identification of the oligomeric states

5.1.1 *MtSerB2* forms a new species in the presence of L-Ser

L-Serine (L-Ser) is the product of the reaction catalyzed by *MtSerB2* and has been shown to act as a feedback inhibitor of the enzyme [1, 2]. Yadav and Shree (2014) reported SEC experiments where the dimeric population of *MtSerB2* shifts to a higher order oligomeric species they called *tetramer* in the presence of L-Ser in the mobile phase [2]. We reproduced their experiment by analyzing *MtSerB2* by SEC in the usual Tris-NaCl mobile phase supplemented with 10 mM L-Ser.

Figure 5.1A shows the superimposition of the SEC-UV chromatograms obtained for the analysis of *MtSerB2* in the serine-free mobile phase (black trace) and in the 10 mM L-Ser mobile phase (red trace). In the absence of L-Ser, the two usual peaks corresponding to the tetramer and the dimer are seen. In the presence of L-Ser, the

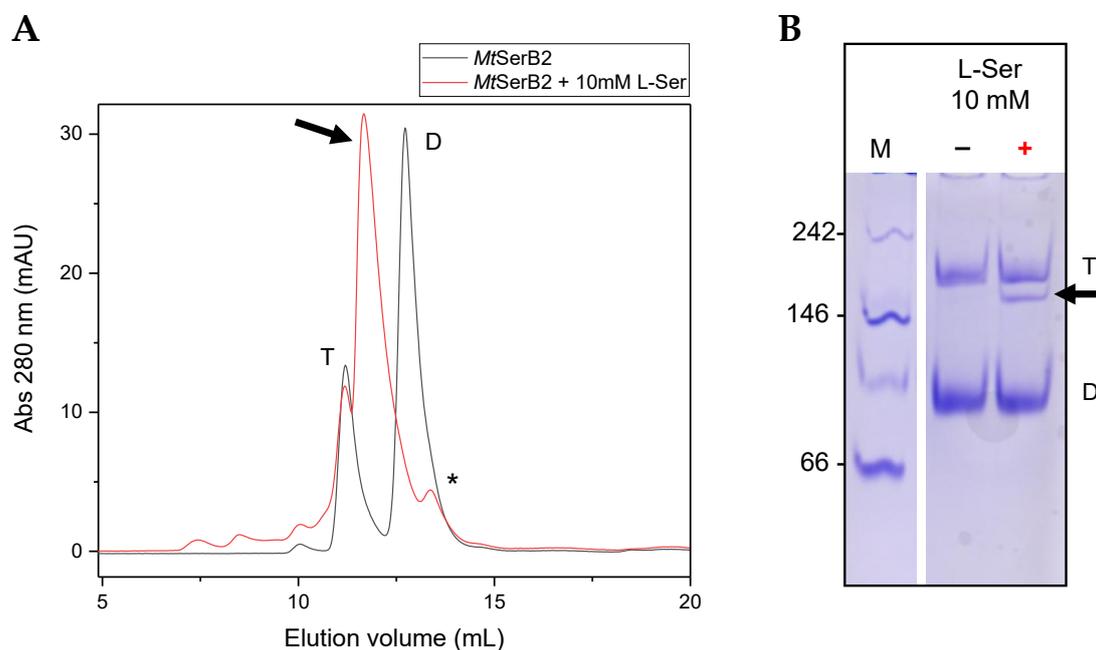


FIGURE 5.1: Observation of a new *MtSerB2* species in the presence of L-Ser. (A) SEC-UV analysis of *MtSerB2* (0.75 mg) in the presence and absence of 10 mM L-Ser in the mobile phase (molar ratio 1:60). Column: Superdex 200 Increase 10/300 GL (B) Native PAGE analysis of *MtSerB2* with and without prior incubation with 10 mM L-Ser.

elution profile is significantly different: a new peak (indicated by the arrow) appears at an elution volume between those of the tetrameric and dimeric species. This peak could therefore correspond to the elution of a species of intermediate size. The native PAGE gel shown in Figure 5.1B corroborates this observation with the appearance of a new band of intermediate electrophoretic mobility when *MtSerB2* sample is incubated with 10 mM L-Ser. The three small peaks eluting before the tetramer on the chromatogram indicate that L-Ser also appears to encourage the formation of a small amount of higher molecular weight species. Another interesting peak (*) also appears just at the elution volume of the dimeric peak shoulder observed in the serine-free experiment. In comparison with the SEC analyses performed on the analytical column where a third peak appears (*, Figure 5.2A), and as discussed in section 4.1.2 of Chapter 4, we believe that this peak may correspond to the elution of a monomer or a truncated form of *MtSerB2*.

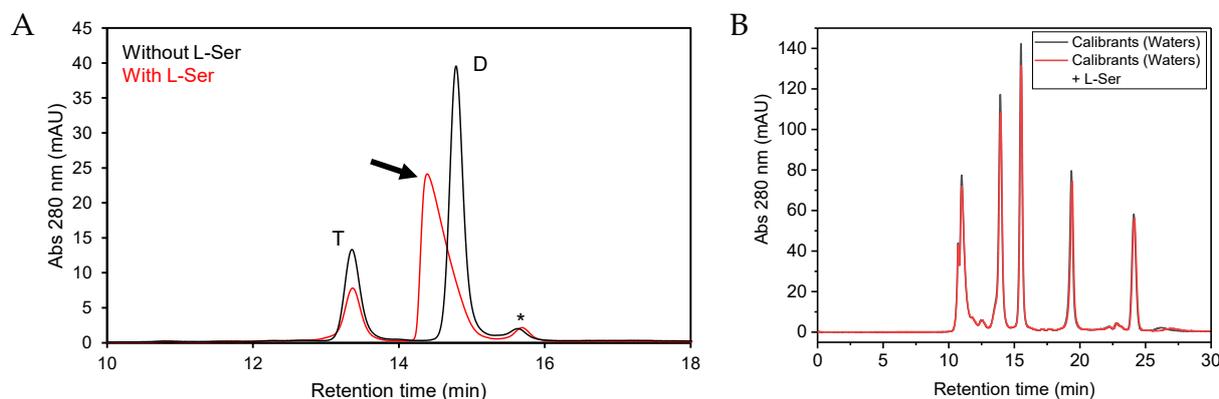


FIGURE 5.2: SEC-UV analysis using the analytical column BioResolveSEC mAb 200Å 2.5 µm 7.8 × 300 mm. (A) SEC-UV analysis of *MtSerB2* (45 µg) in the presence and absence of 2.5 mM L-Ser in the mobile phase (molar ratio 1:100). (B) SEC-UV analysis of BEH200 SEC protein standard mix (Waters) in the presence and absence of 2.5 mM L-Ser in the mobile phase.

To rule out a buffer effect, we verified that the presence of L-Ser did not change the retention times of the standards used to calibrate the SEC column in section 4.1.2 of Chapter 4. The chromatogram shown in Figure 5.2B demonstrates that the retention time of none of the calibrants changes in the presence of L-Ser. The shape of the peaks also remains the same.

5.1.2 *MtSerB2* L-Ser induced form is a trimer

We first attempted to identify the stoichiometry of the new species by MALS (Figure 5.3). However, as indicated by the molar mass measurement along the elution, the peak is not monodisperse. Its tailing shape indicates that there appears to be a species of larger hydrodynamic radius than the dimer that elutes at the beginning of the peak and then a distribution of species of smaller and smaller sizes. The co-elution makes the MALS measurement difficult. A molar mass of 98.9 kDa was measured for this peak.

As exemplified several times throughout this document, molecular weight can also be calculated from SAXS data. Figure 5.4A shows the superimposition of the SEC-UV chromatograms obtained for *MtSerB2* in the presence (red trace) and absence (black trace) of L-Ser during SAXS analysis. After the passage in the UV cell, the samples are analyzed in the SAXS cell where a scattering pattern is acquired every second

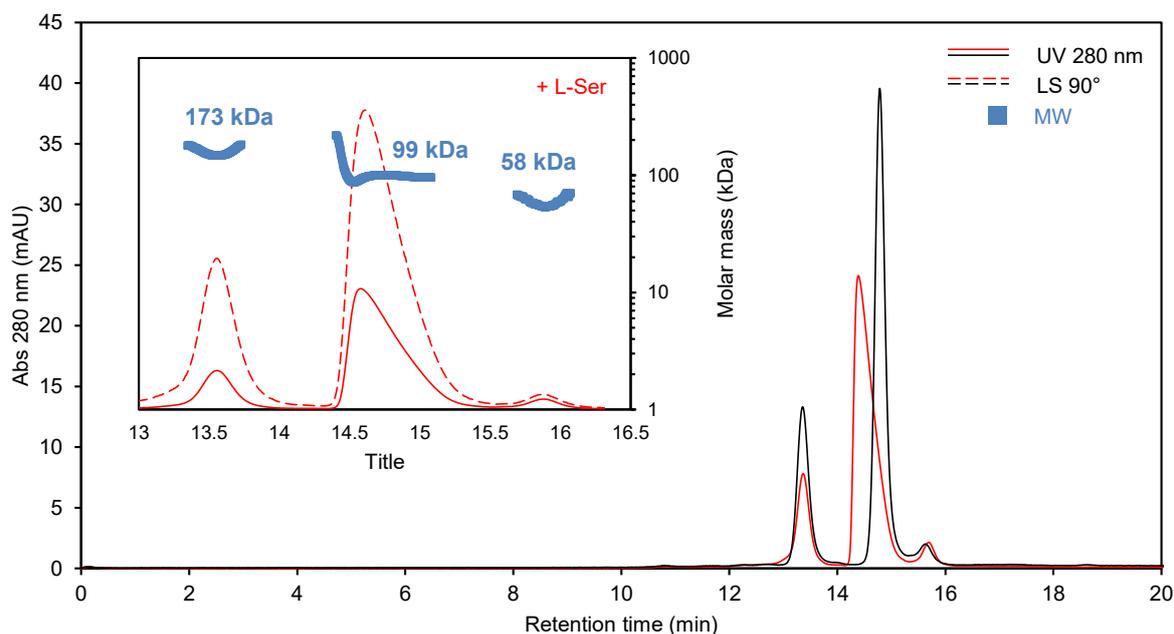


FIGURE 5.3: SEC-UV-MALS analysis of *MtSerB2* in the presence and absence of 2.5 mM L-Ser in the mobile phase (molar ratio 1:100). Column: BioResolveSEC mAb 200Å 2.5 μ m 7.8 x 300 mm.

(frame). To each frame corresponds a value of $I(0)$ and a value of R_g , which makes it possible to identify the frames corresponding to the same species. For *MtSerB2* in the presence of L-Ser, frames 180 to 192 of the front end of the new peak were identified as associated with the same species based on the fact that the standard deviation on the corresponding R_g values did not exceed 0.3 Å (Figure 5.4B). These frames were then averaged and the buffer scattering signal was subtracted to generate the SAXS curve shown in Figure 5.5, thus corresponding to the species whose formation is induced by L-Ser. The size parameters corresponding to the new species (reported in Table 5.1 of section 5.2.2) were then calculated from this curve and used for the determination of the associated molecular weight. The Bayesian inference approach yielded a value of 130.9 kDa with a confidence interval of 121.5 to 142.2 kDa while a value of 151.4 kDa was obtained from the Porod volume. Considering a molar mass of 43.4 kDa for *MtSerB2* monomer, the new species is thus thought to be a trimer. This estimation is consistent with the position of the peak on the chromatogram shown in Figure 5.4A: the peak is located midway between the tetramer and dimer peaks, where would logically elute a trimeric species.

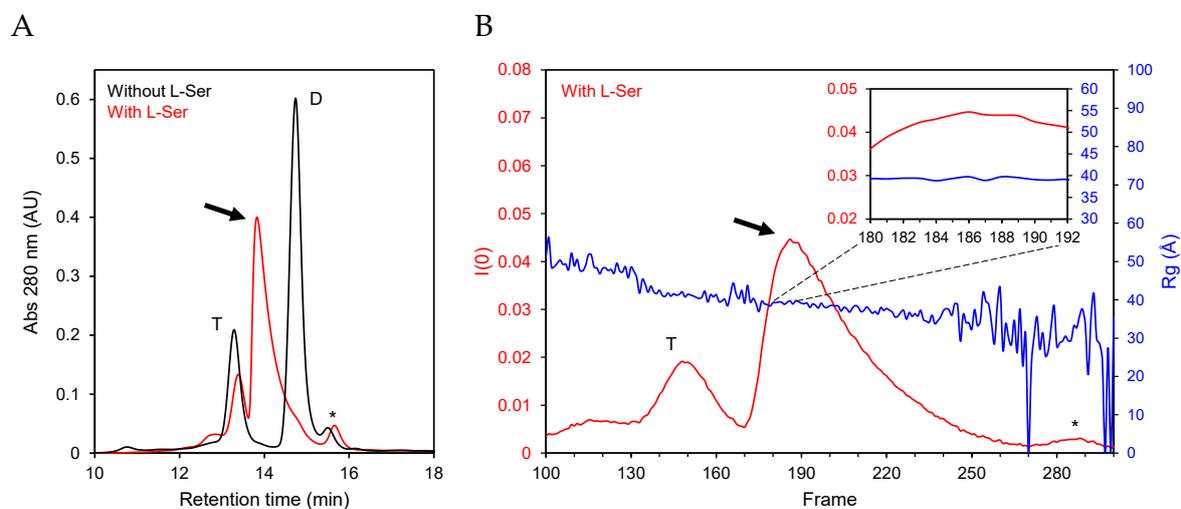


FIGURE 5.4: SEC-UV-SAXS analysis of *MtSerB2* in the presence and absence of 10 mM L-Ser in the mobile phase (molar ratio 1:30). Column: BioResolveSEC mAb 200Å 2.5 μm 7.8 \times 300 mm. (A) SEC-UV chromatogram. (B) SEC-SAXS analysis of *MtSerB2* elution in the presence of L-Ser: measure of $I(0)$ and R_g in the SAXS cell with one frame per second.

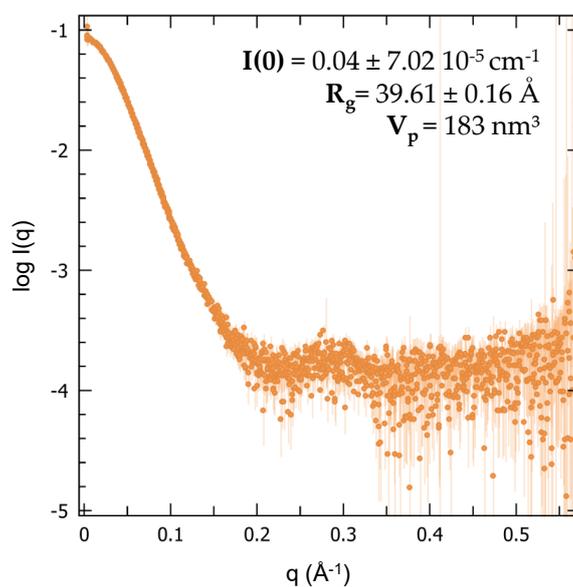


FIGURE 5.5: Experimental SAXS profile ($\log I(q)$ versus q) of L-Ser induced *MtSerB2* species and associated size parameters obtained from Guinier approximation.

5.1.3 L-Ser-induced trimer formation is specific to *MtSerB2*

MtSerB2 orthologs and mutants presented in previous chapters were also analyzed by SEC in the presence of a 1:100 enzyme to L-Ser molar ratio. Superimpositions of chromatograms for the analyses in the presence and absence of L-Ser are shown in Figure 5.6. For *MmSerB2* and *HsPSP*, the shape and position of the peak is not modified by the presence of L-Ser in the mobile phase. It can be confidently concluded that the amino acid does not alter their oligomerization state.

MaSerB, *BmSerB* and *MtSerB2* Δ ACT1 peaks undergo a slight leftwards shift. However, this effect does not appear to be due to a change in oligomerization state. First, similar SEC experiments performed on the same column during SEC-SAXS analyses at 10-fold concentrations of *MaSerB* and *BmSerB* show slight shifts to the right instead. Then, the R_g values measured by SAXS for these two enzymes vary very little: 32.6 Å and 31.2 Å for *MaSerB* in the absence and presence of L-Ser respectively, and 21.4 Å and 21.5 Å for *BmSerB*. The molecular weights calculated by the Bayesian approach also remain identical in the presence of L-Ser: 91.2 kDa for *MaSerB* and 28.9 kDa for *BmSerB*. We find that the R_g of *MaSerB* varies more markedly than that of *BmSerB* in the presence of L-Ser and this could be explained by a local conformational change of the dimer but likely not by a transition to a higher oligomerization state. These controls show that the change in the position of the peak could rather be due to the experimental conditions or possibly to a modification of the interaction of the protein with the column resin.

Finally, from the shape of their chromatograms, it is found that mutants *MtSerB2* Q92E and *MtSerB2* α undergo the same transition to a trimeric form as native *MtSerB2*. Residues C148, V149, G150, I154 and Q92 would therefore not prevent the interaction between the enzyme and L-Ser, nor the formation of the trimer. All in all, these results strongly suggest that the transition to a trimeric state mediated by L-Ser is specific to *MtSerB2*. Orthologous enzymes studied in this thesis do not appear to undergo this change.

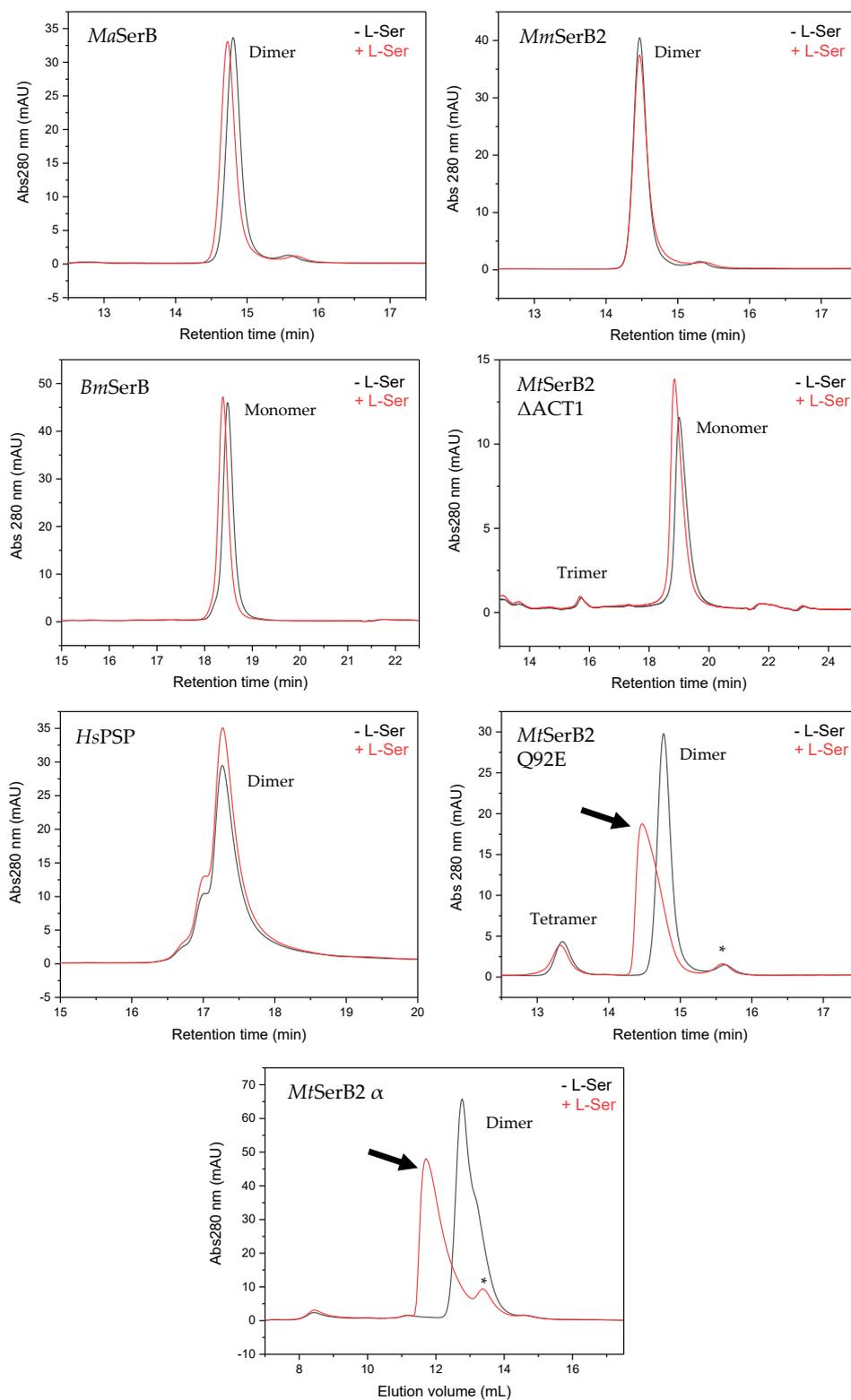


FIGURE 5.6: SEC-UV analysis of *MtSerB2* orthologs and mutants in the absence and presence of L-Ser in the mobile phase (molar ratio 1:100). Column: BioResolveSEC mAb 200Å 2.5 μ m 7.8 x 300 mm, except for *MtSerB2 α* analyzed with Superdex S200 Increase 10/300 GL.

5.1.4 *MtSerB2* trimer is formed from the dimer

The analytical SEC experiments were repeated at several *MtSerB2* to L-Ser molar ratios. The superimposition of the related chromatograms is shown in Figure 5.7 along with a histogram displaying how the percentage of the area under the curve (AUC) of each peak varies with the molar ratio. The tetramer peak is not displaced nor changes shape whatever the amount of L-Ser present in the mobile phase. The same holds for the third small peak. However, a dose-response effect is observed for the second peak. The peak tails and shifts to the left already under the effect of L-Ser in a 1:1 molar ratio in the mobile phase. At 1:10 ratio, the peak is further shifted to the left and tails even more. This displacement and shape are retained at 1:100 and 1:1000 ratios.

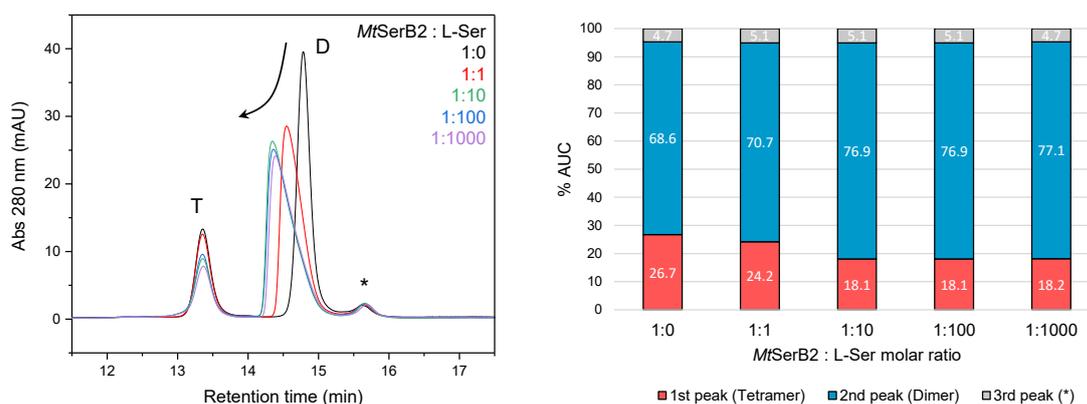


FIGURE 5.7: SEC-UV analysis of *MtSerB2* in the presence of increasing enzyme to L-Ser molar ratios and histogram of the percent of area under the curve corresponding to the first, second and third peaks for each ratio.

As shown by the histogram, the AUC variation is consistent with this observation: the proportion of tetramer progressively decreases from 26.7 %AUC to 24.2 %AUC as the molar ratio increases from 1:0 to 1:1, then to 18.1 %AUC for the 1:10 ratio and remains constant for the ratios 1:100 and 1:1000. Since the area below the third peak does not appear to vary, a small amount of tetramer could be converted to a species eluting below the second peak upon the addition of L-Ser in 1:1 and 1:10 proportions. Meanwhile, the only peak undergoing a shift and shape change under the effect of L-Ser is the second peak. The species from which the trimer is formed would therefore be the dimer.

5.2 Structural characterization of *MtSerB2* L-Ser induced trimer

5.2.1 *MtSerB2* trimer has proven difficult to crystallize

Crystallization assays were performed to try to unveil the atomic structure of *MtSerB2* trimer. To this end, *MtSerB2* dimer (27.6 mg mL^{-1}) was incubated with L-Ser (1:16 molar ratio) and set up in vapor diffusion experiments against the conditions of *JCSG-plus*, *PACT premier* and *BCS* screens (described in section 4.2.1 of Chapter 4).

A few crystals were obtained in nine of the 288 screened conditions (Appendix G) but were identified as salt or did not diffract upon their analysis by X-ray diffraction. However, an interesting result was obtained in condition 2-23 of the *PACT premier* screen (0.2 M sodium citrate tribasic dihydrate 0.1 M Bis-Tris propane pH 6.5, 20% PEG 3350). As shown in Figure 5.8, the drop contained a mix of spherulites and needle shaped crystals growing in what seemed to be the dispersed phase of a liquid-liquid phase separation.

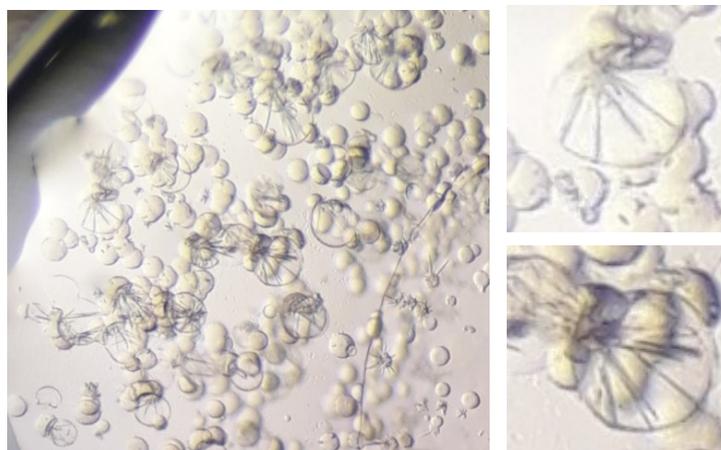


FIGURE 5.8: Needle shaped crystals ($<0.1 \text{ mm}$) obtained in condition 2-23 of the *PACT premier* screen: 0.2 M sodium citrate tribasic dihydrate 0.1 M Bis-Tris propane pH 6.5, 20% PEG 3350.

We first tried to reproduce the result on a larger scale and to optimize the shape of the crystals by varying the percentage of polyethylene glycol (PEG), the pH of the solution and *MtSerB2* concentration but only spherulites could be obtained (Figure

5.9). Nevertheless, these spherulites were crushed to be used for *seeding*, an optimization method where an ordered, solid phase nucleant is introduced in a fresh drop to act as a growth surface for new crystals, removing the need for *de novo* nucleation [3, 4]. Sea urchin like crystals were obtained after the first round of seeding, and a second round of seeding using these crystals, crushed, gave thin needles clusters and 3D crystals clusters in conditions containing lower *MtSerB2* concentrations. A piece of 3D crystal cluster could be harvested and analyzed by X-ray diffraction at the synchrotron. Unfortunately, it did not diffract. However, these crystallization conditions seem promising and it would be interesting to investigate them further, in combination with seeding, by methodically varying each component and by trying various additives.

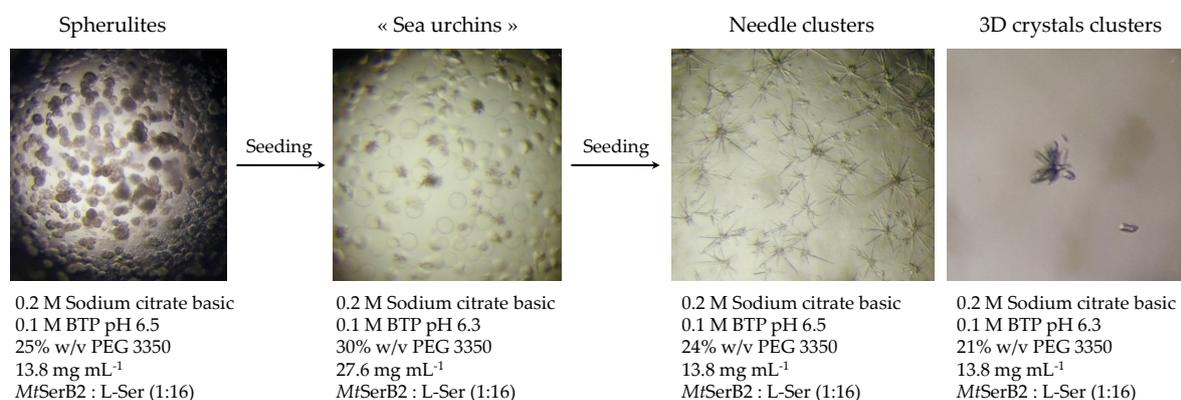


FIGURE 5.9: Results of the crystallization assays obtained during the optimization of condition 2-23 of the *PACT premier* screen. BTP: Bis-Tris propane.

5.2.2 *MtSerB2* trimer is an extended assembly

The SAXS data acquired for *MtSerB2* trimer (Figure 5.10 and Table 5.1) allowed us to reveal some of its structural features. First, an important characteristic to note is its tendency to aggregate. For the trimer, four SEC-SAXS experiments were performed at different protein concentrations. Only the measurement made at the lowest concentration was free of aggregation and could be used for the determination of the size parameters. For the three other measurements, the Guinier plot showed a characteristic upturn at low q values, indicating the presence of aggregates in those samples. Next, the superimpositions of the SAXS curve of *MtSerB2* oligomeric species and of their $P(r)$ functions (Figure 5.10A and B) show that the trimer is indeed a structurally

distinct species. Moreover, the data indicate that *MtSerB2* trimer is rather extended. Several observations support this statement:

- The linear range of the Guinier plot does not extend above qR_g values of 1.0, which is a characteristic of extended scatterers. The values of R_g and $I(0)$ were therefore determined by fitting the data in this restricted range.
- The dimensionless Kratky plot (Figure 5.10C) exhibits a maximum that exceeds the values of 1.104 for $(qR_g)^2 I(q)/I(0)$ and $\sqrt{3}$ for qR_g . The bell shape is also less defined than for *MtSerB2* dimer and tetramer, which could reflect a higher degree of flexibility in the trimer.
- Despite the difference in stoichiometry, the trimer has a R_g (39.6 Å) very close to that of the tetramer (39.9 Å, Table 5.1). If the protein was globular, it would likely have exhibited a R_g value between those of the dimer and the tetramer.
- The $P(r)$ function (Figure 5.10B) displays a longer tail and the D_{\max} value determined from the function is larger (143 Å) than that calculated for the tetramer (127 Å, Table 5.1). Once again, a globular species would have shown a regular bell-shaped $P(r)$ function and the D_{\max} would have been comprised between those of the dimer and the tetramer.

TABLE 5.1: SAXS parameters of *MtSerB2* trimer compared to those of *MtSerB2* dimer and tetramer. (1) derived from the Guinier plot (2) calculated from the $P(r)$ function. For the trimer, R_g and $I(0)$ were evaluated with a Guinier fit to the data up to $q_{\max}R_g < 1.0$ in the linear range for an extended particle.

	<i>MtSerB2</i> tetramer	<i>MtSerB2</i> trimer	<i>MtSerB2</i> dimer
R_g (Å)	39.89 ± 0.06 ⁽¹⁾ ; 39.62 ± 0.04 ⁽²⁾	39.61 ± 0.16 ⁽¹⁾ ; 39.94 ± 0.1 ⁽²⁾	32.47 ± 0.02 ⁽¹⁾ ; 32.55 ± 0.02 ⁽²⁾
$I(0)$ (cm ⁻¹)	$0.06 \pm 5.30 \cdot 10^{-5}$ ⁽¹⁾ ; $0.06 \pm 5.26 \cdot 10^{-5}$ ⁽²⁾	$0.04 \pm 7.02 \cdot 10^{-5}$ ⁽¹⁾ ; $0.04 \pm 5.95 \cdot 10^{-5}$ ⁽²⁾	$0.09 \pm 3.32 \cdot 10^{-5}$ ⁽¹⁾ ; $0.09 \pm 3.64 \cdot 10^{-5}$ ⁽²⁾
D_{\max} (Å)	127	143	111
V_p (nm ³)	254	183	127
M_W Bayesian inference (kDa)	185.8 (162.7 to 221.1)	130.9 (121.5 to 142.2)	88.3 (84.3 to 95.5)
M_W Porod volume (kDa)	210.7	152.2	105.5

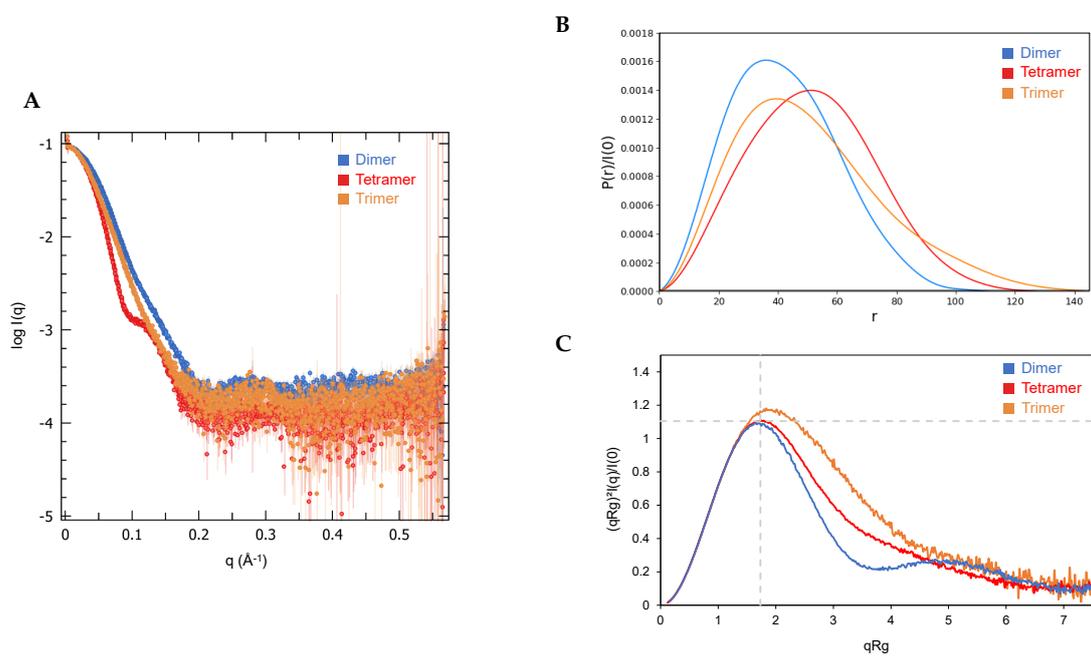


FIGURE 5.10: Analysis of *MtSerB2* trimer solution structure from SAXS data and comparison with *MtSerB2* dimer and tetramer. (A) Experimental SAXS profiles ($\log I(q)$ versus q) of *MtSerB2* trimer, tetramer and dimer. (B) Calculated $P(r)$ functions. (C) Dimensionless Kratky plots.

5.2.3 The trimeric interface of *MtSerB2* is probably formed by the PSP and ACT2 domains

Molecular modeling was undertaken to see what *MtSerB2* trimer might look like. As was done for the tetramer, we hypothesized that the trimer was also a symmetric species. The only possible symmetry for this stoichiometry was the **C3 symmetry**, where a single 3-fold axis of rotational symmetry allows for the assembly of a ring-shaped trimer where the 3 subunits are symmetrically distributed around the axis.

Protein-protein docking using the algorithms of M-ZDOCK, ClusPro and Galaxy-TongDock with the closed monomer model structure as the input resulted in 104 trimer models of C3 symmetry. When comparing their simulated SAXS profile with the experimental data, only eight of these models gave fits with χ^2 values below 15.00. Among them, three ClusPro models were redundant, which left 6 structurally distinct models for further examination (Figure 5.11).

These six models exhibit a common feature in line with the above SAXS data analysis: they possess a flattened shape, extending mainly in one spatial direction. However, the R_g associated to these models is systematically underestimated by more than 1 Å compared to the measured experimental value of 39.6 Å. This observation suggest that the protein may adopt a less compact structure in solution than those represented by the docking models. The monomers composing the assembly could adopt a more extended conformation, for example via extension of the hinge loop and separation of ACT1 and ACT2 domains. In four of the six models, including the two with χ^2 values below 10.00, the monomers interact with each other primarily through the PSP and ACT2 domains, with the ACT1 domain pointing away from the structure and left free of any interaction with the other monomers. This kind of architecture makes our hypothesis of such a conformational extension plausible.

Trimer modeling attempts have also been made using CORAL and the results are in agreement with this proposal. Of the 10 models generated, 7 models with χ^2 values between 1.81 and 3.40 have an architecture in which the core of the trimer is composed of the PSP and ACT2 domains and the ACT1 domain points to the outside of the structure, separated from the ACT2 domain by the hinge-loop in an extended conformation. The R_g values associated with these models are between 38.6 and 39.2 Å and do not deviate more than 1.0 Å from the experimental value. As an illustration, the structures of the three best fitting models and their fits to the experimental data

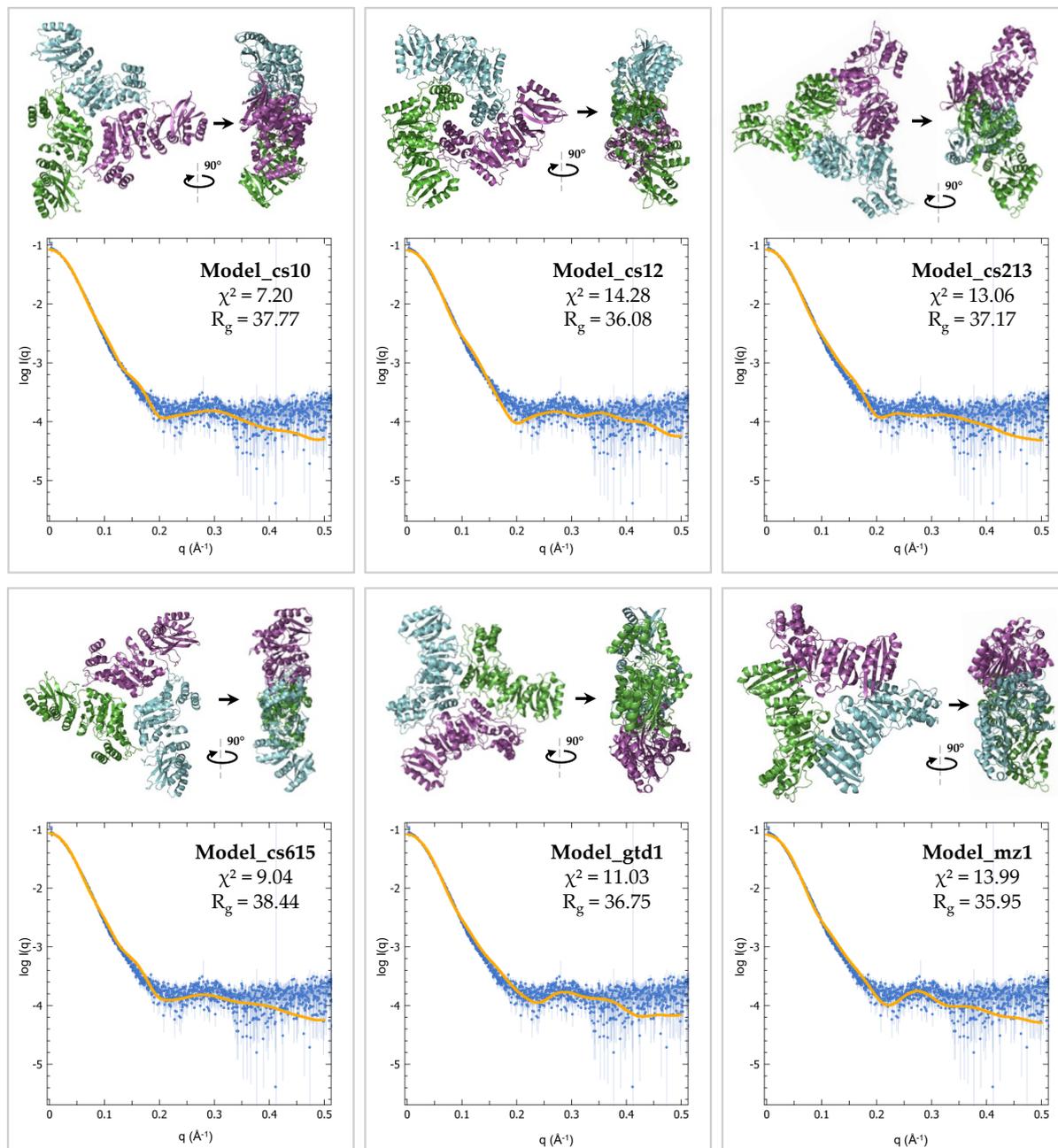


FIGURE 5.11: The 6 structurally distinct C3 *MtSerB2* trimer models selected for examination ($\chi^2 < 15.00$) following the protein-protein docking procedure.

are shown in Figure 5.12. On the other hand, the 3 remaining models featured the ACT1 domain in the center of the structure and fitted the experimental data less well ($4.17 < \chi^2 < 4.62$). They also exhibited too small R_g values comprised between 37.3 and 37.8 Å.

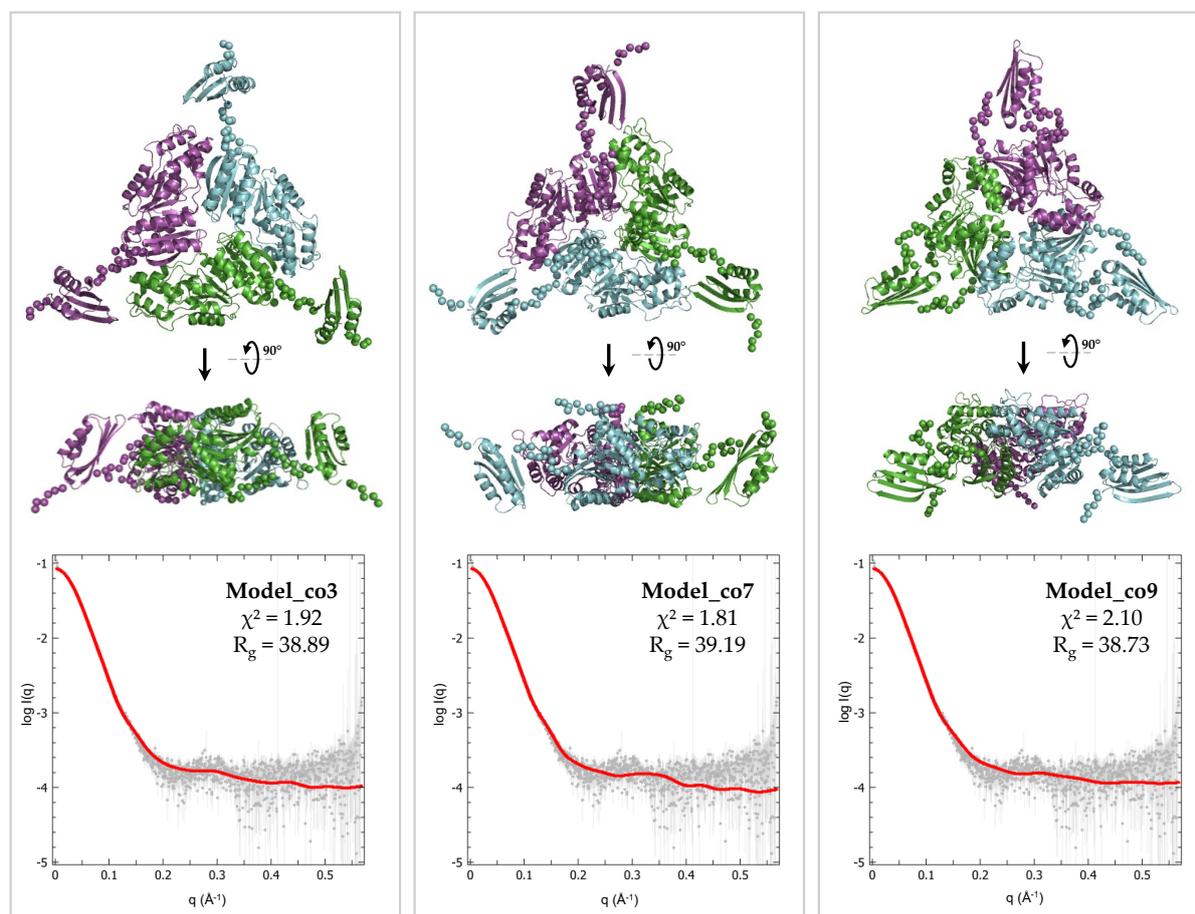


FIGURE 5.12: Best-fitting *MtSerB2* trimer models generated by CORAL with C3 symmetry constraints.

In addition to the fact that they constitute the majority of the generated models and best fit the experimental SAXS data, the trimer models interacting through the PSP and ACT2 domains seem to be better candidates thanks to a previous experimental observation. In section 4.2.2 of Chapter 4 (page 116), we described the oligomeric behavior of *MtSerB2* Δ ACT1 mutant. As shown by SEC-MALS, this mutant composed only of PSP and ACT2 domains is able to form a small amount of trimer. This result suggests that the ACT1 domain is not involved in the trimeric interface and that it is entirely possible that it is directed outwards from the assembly without interacting with the other monomers.

The hypothesis that the ACT1 and ACT2 domains would no longer interact together in the trimer is supported by the fact that the models in which they are separated match the experimental data both in terms of fit and R_g value. Two related arguments consolidate this hypothesis. First, the Kratky plot shown in Figure 5.10C indicates some degree of flexibility in the trimer. This flexibility could be due to the deployment of the hinge-loop separating the two ACT domains. Then, by definition, a flexible molecule can adopt multiple conformations, giving rise to species of slightly varying hydrodynamic radii. The tailing shape of the peak obtained in SEC in the presence of L-Ser thus fully supports the elution of a flexible *MtSerB2* species, where the extended hinge-loop would allow various positioning of the ACT1 domain with respect to the rest of the molecule. The next section provides insight into why the ACT1 and ACT2 domains may be separated in solution, despite the fact that the latter seem to form a stabilizing interface in the dimer.

5.3 Investigation of *MtSerB2* trimer formation mechanism

5.3.1 *MtSerB2* dimer possesses a second L-Ser binding pocket

As shown by SEC in section 5.1.3, L-Ser induces trimer formation in *MtSerB2* while the close homologs of the enzyme, *MaSerB* and *MmSerB2*, retain their dimeric state in the presence of this amino acid. We sought to identify the differences that lead to the formation of the trimer. To this end, the PrankWeb server was used to predict ligand binding sites in *MtSerB2*, *MaSerB* and *MmSerB2* dimers [5, 6]. The results are compared in Figure 5.13.

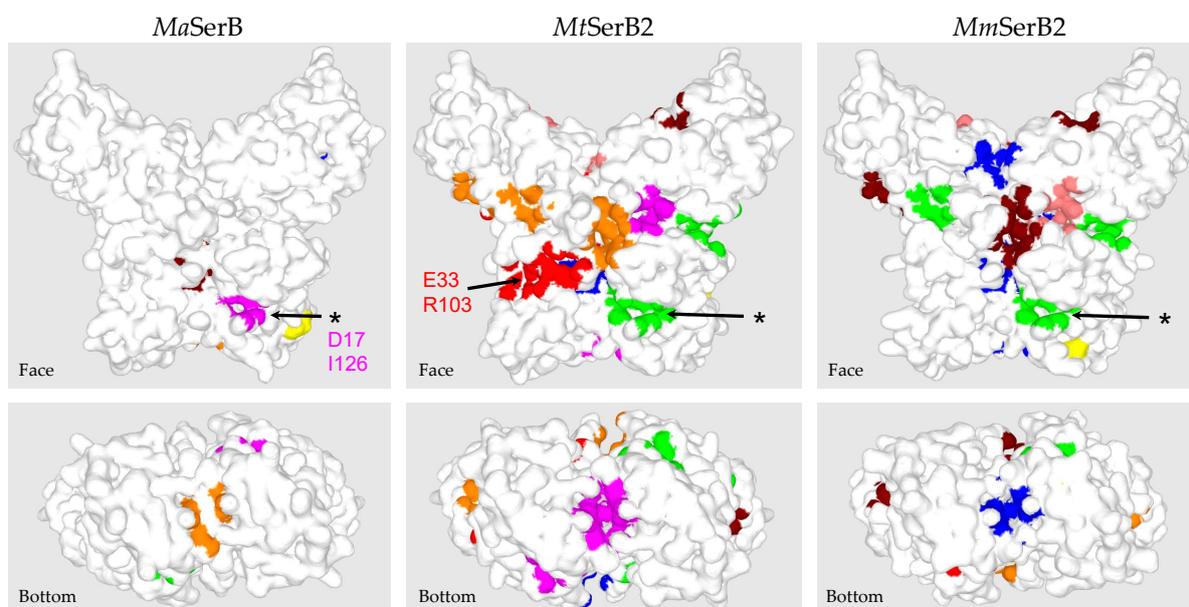


FIGURE 5.13: Ligand binding sites in *MaSerB*, *MtSerB2*, and *MmSerB2* predicted by PrankWeb server. The arrows indicate L-Ser binding pockets in the three orthologs. In addition to the main pocket evidenced in *MaSerB* structures 5JLR and 5JLP (*), another pocket only present in *MtSerB2* was identified (red).

Interestingly, the prediction highlights a pocket that would exist in *MtSerB2* but not in the two homologs (displayed in red in Figure 5.13). Moreover, this pocket is located at the interface of the ACT1 and ACT2 domains (from two distinct monomers due to domain-swapping) and contains the residues E33 and R103. As demonstrated in reference [1], these two residues are thought to be involved in L-Ser binding on

the basis of homology with the ACT domain of *E. coli* phosphoglycerate dehydrogenase (PGDH). It is important to note that this pocket is distinct from the known L-Ser-binding pocket (* in Figure 5.13, containing D17 and I126) evidenced in *MaSerB* structures 5JLR and 5JLP (see section 3.3.3 of Chapter 3). This known pocket is also predicted in *MtSerB2* and *MmSerB2*.

A docking experiment of L-Ser in the newly identified pocket was performed to further probe the interacting residues. Nineteen possible poses of L-Ser were generated by an induced fit docking protocol at pH 7.4 in *MtSerB2* dimer homology model. The top 5 poses in terms of docking score are detailed and superimposed in Figure 5.14A. It can be seen that all of them are very similar with respect to the positioning of L-Ser in the pocket and that they involve recurrent interactions (see Figure 5.14B and Table 5.2) with the side chains of residues E33 and T136, as well as with the backbone carbonyl oxygen of L34. Pose 2 also involves a salt bridge between the carboxylic acid moiety of L-Ser and R103 guanidinium moiety through a conformational change of the latter allowed by the induced fit.

TABLE 5.2: Established interactions between L-Ser and *MtSerB2* in the top 5 poses of the induced fit docking of L-Ser at pH 7.4 in a binding pocket of *MtSerB2* dimer homology model containing E33 and R103.

L-Ser group	Interaction	Residue group	Residue	Pose
OH side chain	H-bond	CO backbone	L34	2, 3, 4, 5, 7
NH ₃ ⁺	Salt bridge	COO ⁻ side chain	E33	2, 3, 5, 7
NH ₃ ⁺	H-bond	COO ⁻ side chain	E33	2, 3, 4, 7
OH side chain	H-bond	OH side chain	T136	2, 3, 4, 7
COO ⁻	H-bond	OH side chain	T136	5
COO ⁻	Salt bridge	NHC(NH ₂) ₂ ⁺ side chain	R103	2

5.3.2 Glu33, Arg103 and Thr136 are implicated in *MtSerB2* trimer formation

The next step was to verify that these residues are indeed involved in L-Ser binding *in vitro* and that the secondary L-Ser binding-site they constitute is responsible for *MtSerB2* trimer formation. To this end, the oligomeric behavior of mutant *MtSerB2* E33A R103A T136A (***MtSerB2* 3A**) was studied. In this mutant, E33, R103 and T136 are converted to alanine residues whose side chains lack the functional groups required for ligand binding.

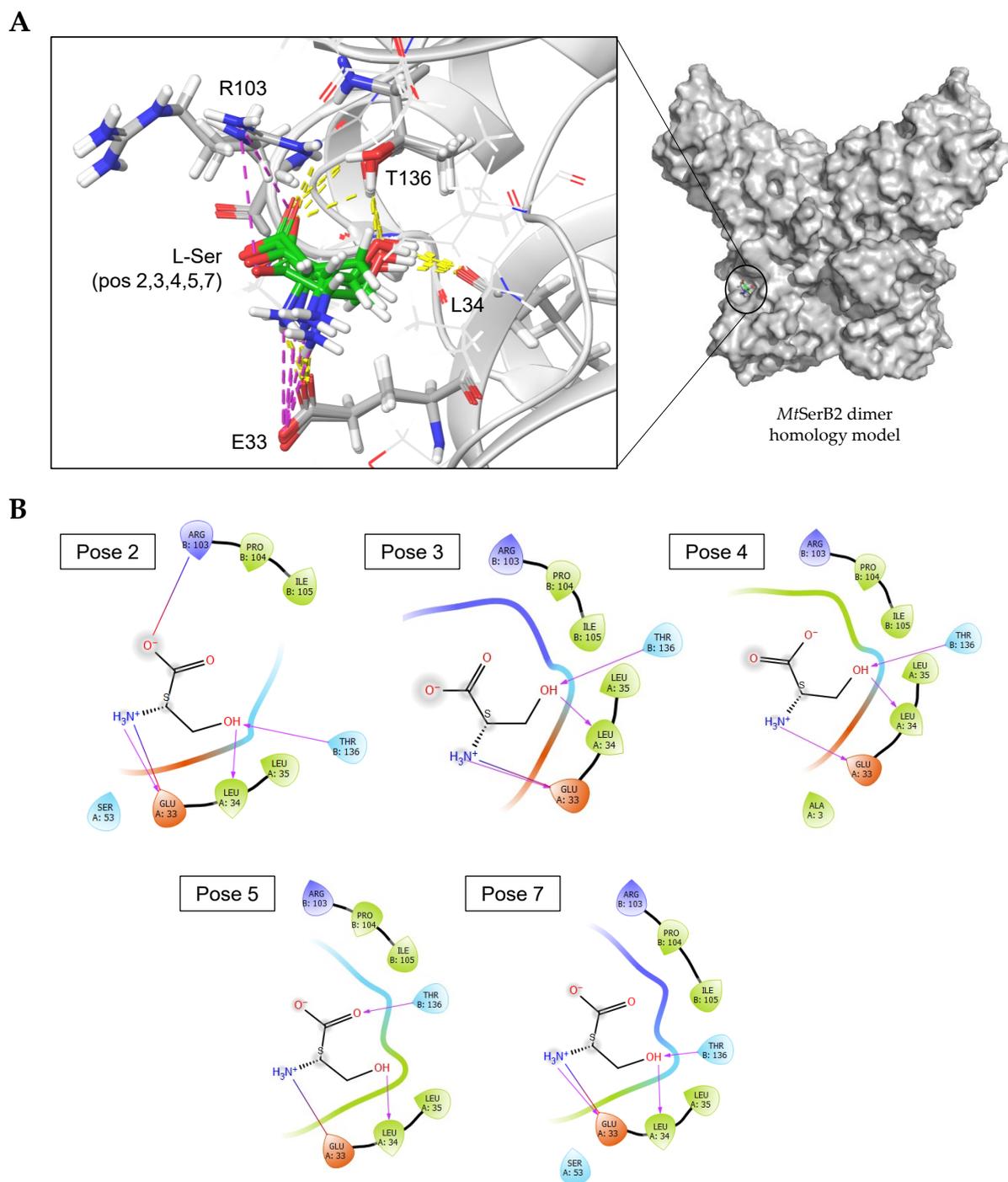


FIGURE 5.14: Induced-fit docking of L-Ser at pH 7.4 in a binding pocket of *MtSerB2* dimer homology model containing E33 and R103. (A) Superimposition of the top 5 poses of L-Ser in the binding pocket and interacting residues. Salt bridges are depicted in purple and H-bonds in yellow. (B) 2D ligand interaction maps corresponding to the top 5 poses. H-bonds are depicted in pink and salt bridges in a gradient from blue to red.

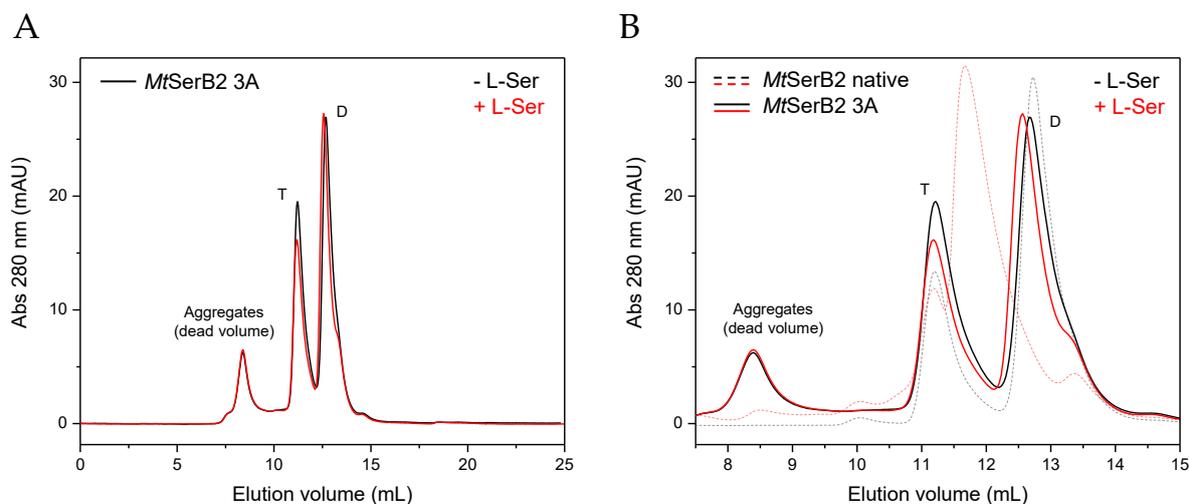


FIGURE 5.15: SEC-UV analysis of mutant *MtSerB2* E33A R103A T136A (3A) and native *MtSerB2* in the presence and absence of 10 mM L-Ser in the mobile phase (molar ratio 1:60). (A) Superimposition of SEC chromatograms for the analysis of mutant *MtSerB2* 3A with and without L-Ser. (B) Superimposition of SEC chromatograms for the analysis of mutant *MtSerB2* 3A and native *MtSerB2* with and without L-Ser.

Figure 5.15A shows the superimposition of the chromatograms obtained for the SEC analysis of *MtSerB2* 3A in the absence and presence of L-Ser (molar ratio 1:60). The two traces overlap almost perfectly and the dimer peak no longer undergoes a leftward shift in the presence of L-Ser, in contrast to what is observed for native *MtSerB2* (Figure 5.15B). An additional peak eluting around 8.5 mL (column dead volume) indicates that the mutant is more prone to aggregation, perhaps due to the destabilization of the ACT1-ACT2 interface. This experiment shows that the E33A, R103A and T136A mutations abolish L-Ser mediated trimer formation. At least one of these three residues is thus probably involved in a key interaction with L-Ser. It would be interesting to precisely determine which residue(s) is/are engaged in the mechanism by assessing the effect of single-residue mutations.

The binding of L-serine in this secondary pocket, located at the interface of the ACT1 and ACT2 domains of *MtSerB2* dimer, would thus be the origin of the trimer formation. The stoichiometry of this oligomeric transition implies a perturbation of the dimeric species that could lead to its dissociation into monomers in order to form the trimeric species. For this reason we believe that the interaction of L-Ser with

the residues of this pocket disrupts the interface between the ACT1 and ACT2 domains. Disruption of this essential interface for dimer stabilization (as evidenced by the monomeric mutant *MtSerB2* Δ ACT1) would induce a conformational change leading to dimer dissociation. As supported by the assumption that the ACT1 and ACT2 domains no longer interact in the trimeric species, L-Ser would remain in interaction with one of the two ACT domains, thus preventing the reassembly of the ACT1-ACT2 interface.

5.4 Enzymological characterization of the oligomeric species in the presence of L-Ser

5.4.1 L-Ser acts as a mixed-type predominantly competitive inhibitor of *MtSerB2* dimer

The last part of this work was to study the kinetic mechanism of *MtSerB2* dimer inhibition by L-Ser. Steady-state kinetics of *O*-phospho-L-serine (PS) dephosphorylation were assessed as described in section 4.3 (Chapter 5), in the presence of fixed L-Ser concentrations in the assay. The resulting initial velocity vs PS concentration curves are presented in Figure 5.16A. The appearance of the graph indicates that L-Ser effectively acts as an inhibitor of *MtSerB2* dimer, decreasing the velocity of the enzyme in a dose dependent manner.

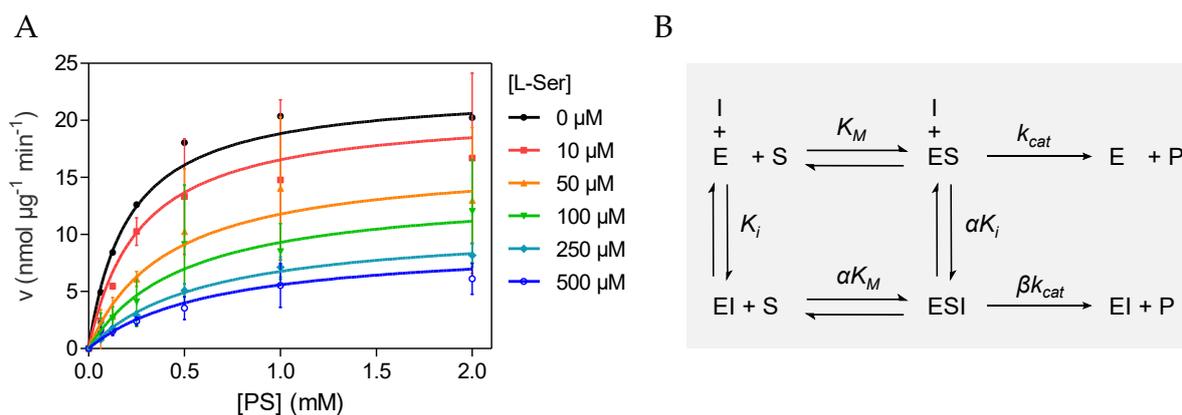


FIGURE 5.16: L-Ser inhibition kinetics of *MtSerB2*. (A) Initial velocity vs substrate concentration curves for the dephosphorylation of *O*-phospho-L-serine (PS) in the presence of fixed L-Ser concentrations. (B) General modifier mechanism scheme. E = enzyme, S = substrate, P = product, I = modifier (inhibitor), K_M = substrate dissociation constant, K_i = inhibitor dissociation constant, α = reciprocal allosteric coupling constant, β = factor by which the modifier affects the catalytic constant k_{cat} .

The methodology proposed by Antonio Baici was followed to unravel the kinetic mechanism of inhibition (<https://www.enzyme-modifier.ch>) [7]. The condensed form of the rate equation (Equation 5.1) derived from the general modifier mechanism (Figure 5.16B) was first fitted to the experimental data for each L-Ser concentration to determine the apparent $k_{cat,app}$ and $K_{M,app}$ values for each L-Ser concentration.

$$\frac{v}{[E]_t} = \frac{k_{cat,app}[S]}{K_{M,app} + [S]} = \frac{k_{cat} \left(1 + \beta \frac{[I]}{\alpha K_i}\right) [S]}{K_M \left(1 + \frac{[I]}{K_i}\right) + [S] \left(1 + \frac{[I]}{\alpha K_i}\right)} \quad (5.1)$$

In Equation 5.1, $[E]_t$ is the total enzyme concentration, $[S]$ is the free substrate (O-phospho-L-serine, PS) concentration, $[I]$ is the free inhibitor (L-Ser) concentration, k_{cat} is the catalytic constant, K_M is the Michaelis constant (substrate dissociation constant), K_i is the inhibitor dissociation constant, α is the reciprocal allosteric coupling constant and β is the factor by which the inhibitor affects k_{cat} .

Then, the dependence of $k_{cat,app}$, $1/k_{cat,app}$, $K_{M,app}$, $k_{cat,app}/K_{M,app}$, and $K_{M,app}/k_{cat,app}$ on L-Ser concentration was analyzed (Figure 5.17). The shape of the replots gave the characteristic pattern of a *hyperbolic mixed, predominantly specific inhibition* (HMx(Sp>Ca)I), according to the nomenclature used by Baici. In this mechanism, L-Ser binds the free form of MtSerB2 with greater affinity but can also interact with the substrate bound form ($\alpha > 1$). The resulting L-Ser-MtSerB2-PS complex is still productive but shows a decreased turnover rate ($0 < \beta < 1$). In generic terms, these results identify a *mixed predominantly competitive partial inhibition* mechanism of MtSerB2 by L-Ser.

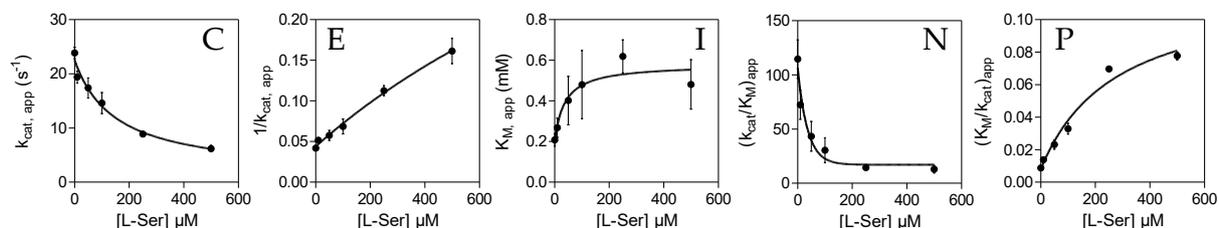


FIGURE 5.17: Dependence of the apparent kinetic parameters on L-Ser concentration. The letters C, E, I, N and P identify the shape of the replots according to Baici's methodology [7]. Plain lines are the fits of the dependency of kinetic parameters on $[L-Ser]$ to the experimental data based on the general modifier equation.

In view of the quasi-linear appearance of the replots of $1/k_{cat,app}$ and $K_{M,app}/k_{cat,app}$ vs L-Ser concentrations, it was difficult to distinguish between a complete/linear and a partial/hyperbolic inhibition mechanism. The partial/hyperbolic inhibition mechanism was identified from the shape of the v vs $[L-Ser]$ replot in Figure 5.18. It shows a plateau at finite velocity indicating the inability of an increasing L-Ser concentration to drive the velocity to zero.

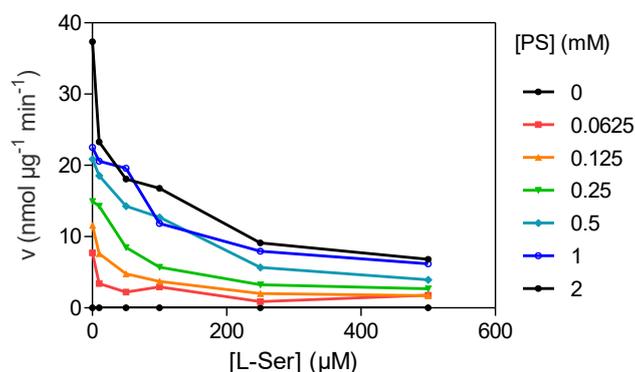


FIGURE 5.18: Plot of velocity vs L-Ser concentration at constant PS concentrations.

Our analysis matches the *partial competitive inhibition* mechanism identified by Grant [1]. Fitting the complete form of the general modifier equation (Equation 5.1) to the experimental data yielded a K_i value of $22.0 \pm 5.3 \mu\text{M}$ agreeing with the value of $19 \pm 2 \mu\text{M}$ presented in the paper. For the parameters α and β , values of 3.52 ± 1.20 and 0.0006 ± 0.0004 were respectively determined, effectively translating a 2 to 5 times lower affinity of L-Ser for the enzyme-substrate complex and an almost zero turnover rate of the L-Ser-enzyme-substrate complex ($\beta k_{\text{cat}} = 0.01 \text{ s}^{-1}$).

5.4.2 L-Ser acts differently on *MtSerB2* and its dual-ACT domain orthologs

The kinetic mechanisms of inhibition of *MaSerB* and *MmSerB* were evaluated using the same method. The initial velocity curves in the presence of fixed L-Ser concentrations and the replots of the kinetic parameters versus L-Ser concentration are shown in Figure 5.19.

The shape of the replots (Figure 5.20) and that of the v vs $[L\text{-Ser}]$ (Figure 5.21), where a saturating concentration of L-Ser brings the enzyme velocity to zero, allows to state that L-Ser is a *mixed* and *linear (complete)* inhibitor of both enzymes. The $K_{M,\text{app}}$ vs $[L\text{-Ser}]$ replots, however, show disparate $K_{M,\text{app}}$ values that do not clearly follow one of the four possible trends allowed by the general modifier mechanism. This makes the identification of the predominant character of the mixed inhibition

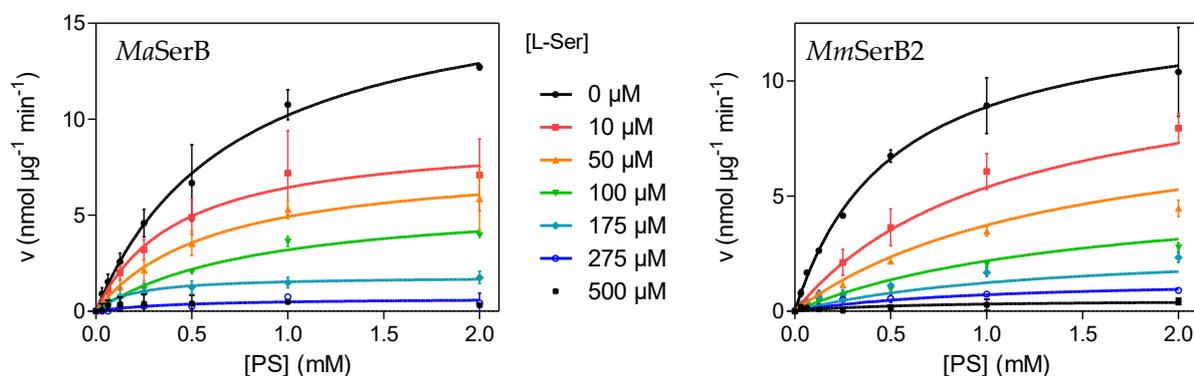


FIGURE 5.19: Initial velocity vs substrate concentration curves for the dephosphorylation of *O*-phospho-L-serine by (PS) by *MaSerB* (left) and *MmSerB2* (right) in the presence of fixed L-Ser concentrations.

ambiguous. From the decreasing replot pattern, *MaSerB* seems to show a predominant *catalytic/uncompetitive* character, where the inhibitor binds preferentially to the *MaSerB*-PS complex, but the trend is less clear for *MmSerB2*.

However, we noticed that the $1/k_{cat,app}$ vs $[L-Ser]$ replots are parabolic and therefore do not correspond to any pattern of the dichotomous determination key proposed by Baici. This observation suggests a more complex mechanism of inhibition, *parabolic inhibition*, in which the enzyme is able to bind two L-Ser molecules at distinct sites [8]. Different types of parabolic inhibition mechanism exist and can be identified by inspecting the dependence of $1/k_{cat,app}$ and $K_{M,app}/k_{cat,app}$ on L-Ser concentration [9].

Equation 5.2 is the double-reciprocal transformation of Equation 5.1 ($1/v$ vs $1/[S]$):

$$\frac{[E]_t}{v} = \text{intercept} + \text{slope} \frac{1}{[S]} = \frac{1}{k_{cat,app}} + \frac{K_{M,app}}{k_{cat,app}} \frac{1}{[S]} \quad (5.2)$$

The term $1/k_{cat,app}$ corresponds to the y-intercept (I) of the $1/v$ vs $1/[S]$ plot and $K_{M,app}/k_{cat,app}$ to its slope (S). In the case of the inhibition of *MaSerB* and *MmSerB2* by L-Ser, the slope replot is linear while the intercept replot is parabolic. This case is called *S-linear I-parabolic noncompetitive* inhibition according to Cleland's nomenclature [10]. In this mechanism, a molecule of L-Ser can bind the free enzyme as well as the enzyme-PS complex, at one or the other site, and a quaternary complex L-Ser-enzyme-PS-L-Ser is formed through the simultaneous interaction of L-Ser at both sites. The

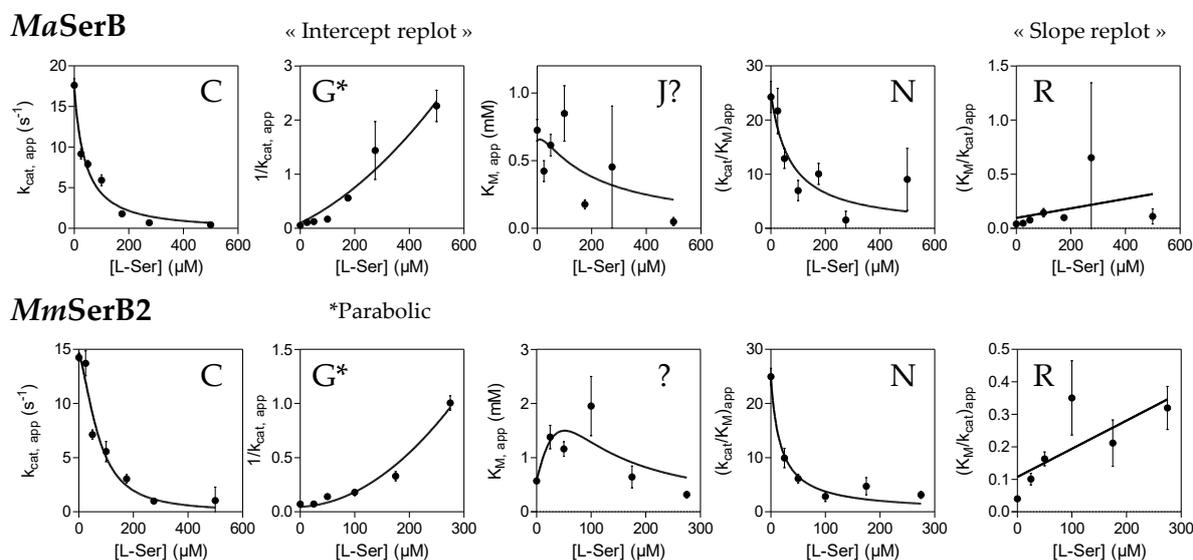


FIGURE 5.20: Dependence of the apparent kinetic parameters of *MaSerB* and *MmSerB2* on L-Ser concentration. The letters C, G, J, N and R identify the shape of the replots according to Baici's methodology [7]. Plain lines are the fits of the dependency of kinetic parameters on [L-Ser] to the experimental data based on the rate equation for S-linear I-parabolic noncompetitive inhibition.

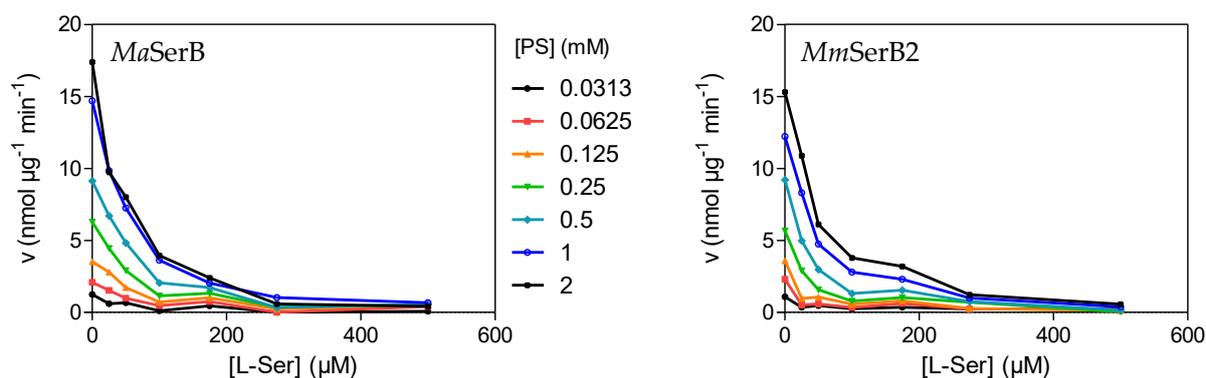


FIGURE 5.21: Plot of velocity vs L-Ser concentration at constant PS concentrations for *MaSerB* and *MmSerB2*.

ternary and quaternary complexes are non productive, which results in a complete inhibition. Such a mechanism is depicted in Figure 5.22 and is described by the general rate expression of Equation 5.3.

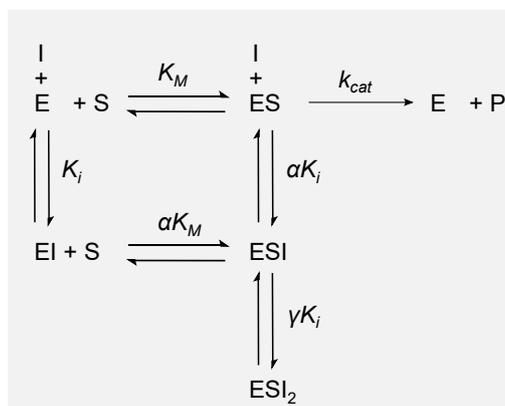


FIGURE 5.22: S-linear I-parabolic inhibition mechanism scheme. E = enzyme, S = substrate, P = product, I = modifier (inhibitor), k_{cat} = catalytic constant, K_M = substrate dissociation constant, K_i = inhibitor dissociation constant, α , γ = reciprocal allosteric coupling constant

$$\frac{v}{[E]_t} = \frac{k_{cat}[S]}{K_M(1 + a[I]) + [PS](1 + b[I] + c[I]^2)} \quad (5.3)$$

where $[I]$ is the free L-Ser concentration, and a , b and c are interaction factors containing the parameter K_i as well as the reciprocal allosteric coupling constant α and γ whose values depend on binding cooperativity. It can be seen that the full form of the general modifier equation (Equation 5.1) is no longer adequate to fit the experimental data since $[PS]$ in the denominator is now multiplied by a second-degree polynomial translating the parabolic character and no longer by a linear term. Therefore, the expressions for $k_{cat,app}$ and $K_{M,app}$ as a function of L-Ser concentration now contain this term and this explains why the replots of $K_{M,app}$ vs $[L-Ser]$ did not show one of the 4 expected shapes.

Since the precise mechanism is unknown (*can L-Ser bind to one site as well as the other in the free enzyme? ... in the enzyme-substrate complex? Is the affinity for the first site equal to that for the second site? Is there binding cooperativity?*), the component terms of the interaction factors are not known either and they cannot be constrained for

the non-linear regression. Therefore, determining accurate K_i , α and γ values for the inhibition of *MaSerB* and *MmSerB2* by L-Ser is an arduous task.

Comparison of these results with those obtained in the previous section reveals that *MtSerB2* is inhibited by L-Ser through a different mechanism than that shown by its orthologs *MaSerB* and *MmSerB2*.

Indeed, on the one hand, *MtSerB2* is partially inhibited in a predominantly competitive way with respect to substrate binding via the action of L-Ser on an inhibitory site. We suspect that the L-Ser binding site involved in the mechanism is the one containing E33, R103 and T136. As shown in this chapter, the interaction of L-Ser with this site induces a conformational change, leading eventually to the formation of a trimer. According to the identified mechanism, PS can still bind the enzyme after this conformational change, but with less affinity. The resulting ternary complex is still active, but catalyzes the dephosphorylation reaction very poorly (4000 times lower rate).

On the other hand, *MmSerB2* and *MaSerB* are totally inhibited in a mixed manner by interacting with two L-Ser molecules at two distinct sites. The first site, where L-Ser can bind in the free enzyme but also in the enzyme-PS complex could be the one containing D17, Q18, V21 and I126 in *MaSerB* (the site evidenced in the 5JLR and 5JLP crystallographic structures, marked by an asterisk in Figure 5.13) and the homologous site found in *MmSerB2*. When L-Ser is bound there and PS is bound in the active site, a second L-Ser molecule can interact with the enzyme. Potentially, it binds either to a new site created somewhere on the enzyme by a conformational change due to the binding of the PS and the first L-Ser molecule, or in the active site, near or interacting with PS. Further experiments would be required to identify the precise inhibitory sites and the order of interaction, but we know so far that the binding of two L-Ser molecules completely abolishes enzyme activity.

It would be very interesting to study the kinetic behavior of the *MtSerB2* 3A mutant to verify if the singular L-Ser binding site identified in *MtSerB2* is indeed responsible for the alternative inhibition mechanism. This hypothesis would be confirmed if the 3A mutant showed a kinetic behavior similar to *MmSerB2* and *MaSerB* in the presence of L-Ser.

5.5 Summary of the findings

In 2014, Yadav and Shree reported that *MtSerB2* underwent a change in its oligomerization state upon the interaction with L-serine, its reaction product and allosteric feedback inhibitor [2]. The authors described the new species as a tetramer, without providing further experimental evidence to this statement. In Chapter 4 of this thesis, we showed that *MtSerB2* could already form a tetramer in the absence of L-Ser. Following this observation, we aimed to reproduce Yadav and Shree's experiment and determine the stoichiometry of the new species formed.

Through SEC and native PAGE analyses, we demonstrated that a new *MtSerB2* oligomeric species was actually formed in the presence of L-Ser and SAXS allowed for the identification of this new oligomer as a trimer, not a tetramer. Crystallization assays have been undertaken but have not yet allowed us to reveal the atomic detail of the new assembly. Nevertheless, X-Ray scattering data gave valuable insights into its architecture as it could be seen that the trimer was extended and probably more flexible than *MtSerB2* dimer and tetramer.

Plausible trimer models have been constructed using protein-protein docking and rigid body modeling in combination with SAXS data. The best fits were obtained from models that shared a common architecture: a rather flattened assembly, whose oligomeric interface was built by contacts between the PSP-ACT2 domain part, leaving the ACT1 domains pointing to the outside of the assembly. Even better fits were obtained with models in which the ACT1 domains were separated from the core assembly by the hinge-loop in an extended conformation. These models are consistent with the flattened shape and flexibility suggested by the experimental data.

Additional SEC experiments indicated that the trimer was formed from the dimer. Indeed, on chromatograms, the dimer peak underwent a L-Ser dose-dependent displacement and deformation, whereas the tetramer peak was left unaffected. This experiment, replicated with each of the orthologs studied in this thesis, showed that trimer formation from the dimer was specific to *MtSerB2*. Remarkably, the oligomeric state of its two closest orthologs, *MaSerB* and *MmSerB2* did not change in the presence of L-Ser. By predicting the potential ligand-binding sites for the three enzymes, a L-Ser binding pocket also located at the ACT1-ACT2 interface but specific to *MtSerB2* could be identified. The implication of this special pocket in the mechanism of trimer

formation was confirmed using a triple alanine *MtSerB2* mutant. This mutant, designed to prevent the eventual binding of L-Ser in the newly identified pocket, was unable to form trimer in the presence of the amino acid.

Finally, steady-state inhibition kinetics experiments suggest that this pocket implies a different L-Ser regulation mechanism in *MtSerB2* compared to its close homologs *MaSerB* and *MmSerB2*. Interaction of one L-Ser molecule with this pocket would induce a partial, predominantly competitive inhibition of *MtSerB2* with respect to the substrate, while *MaSerB* and *MmSerB2* would be totally inhibited by the consecutive binding of two L-Ser molecules at two different sites in the enzyme-substrate complex.

We suggest that the interaction of L-Ser with the pocket disrupts the ACT1-ACT2 domain interface, which induces a conformational change within the enzyme, resulting in the separation of the ACT1 domain from the rest of the structure. The domain-swapped architecture of the dimer is disrupted and dissociation into monomers takes place. The monomers, unstable in solution, re-associate very quickly into trimers by interacting via the PSP-ACT2 part. This hypothesis is supported by the oligomeric behavior of a *MtSerB2* mutant lacking the ACT1 domain: it is mainly monomeric in solution but it is able to form a small amount of trimer. Consistent with the competitive inhibition kinetics mechanism, the trimer could still bind *O*-phospho-L-serine, but with much lower affinity than the dimer. It would still catalyze the dephosphorylation reaction but about 4000 times slower than the dimer.

Bibliography

- [1] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.
- [2] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [3] A. D’Arcy, T. Bergfors, S.W. Cowan-Jacob, and M. Marsh. Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallographica Section F: Structural Biology Communications*, 70(9):1117–1126, 2014.

- [4] T. Bergfors. Succeeding with seeding: some practical advice. In *Evolving Methods for Macromolecular Crystallography*, pages 1–10. Springer, 2007.
- [5] L. Jendele, R. Krivak, P. Skoda, M. Novotny, and D. Hoksza. FrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Research*, 47(W1):W345–W349, 2019.
- [6] R. Krivák and D. Hoksza. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):1–12, 2018.
- [7] A. Baici. *Kinetics of enzyme-modifier interactions*, volume 1. Springer, 2015.
- [8] I.H. Segel. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems*. Wiley New York, 1975.
- [9] V. Leskovac. *Comprehensive enzyme kinetics*. Springer Science & Business Media, 2003.
- [10] W.W. Cleland. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: II. Inhibition: Nomenclature and theory. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Enzymological Subjects*, 67:173–187, 1963.

Part IV

General discussion, conclusions and outlooks

Chapter 6

MtSerB2, a potential morpheein

6.1 General discussion

The phosphoserine phosphatase *MtSerB2* is a promising enzymatic target from *M. tuberculosis* for the development of new anti-tuberculosis agents. Studies identifying inhibitors demonstrated that the enzyme is druggable [1–4], but designing compounds targeting the active site turns out not to be an optimal strategy in terms of selectivity since the human homolog *HsPSP* possesses a highly conserved catalytic pocket. Nevertheless, studies prior to this thesis indicated that *MtSerB2* was likely present as an active homodimer in solution and underwent a quaternary structure change to an inactive homomeric species in the presence of its endogenous allosteric feedback inhibitor, L-serine [2, 5]. These observations laid the groundwork for an alternative therapeutic approach: the allosteric targeting of *MtSerB2* via a homomeric disruption strategy. The structural basis was nevertheless not very solid to afford a rational inhibitor design: what was known about the system before this work was limited to a homology model and an obscure L-Ser feedback regulation mechanism, involving a quaternary structure change to a higher-order oligomer of undefined stoichiometry (Figure 6.1A).

In this work, we aimed to deepen the structural knowledge regarding *MtSerB2* to offer a comprehensive basis for the design of allosteric inhibitors. In order to provide further structural insight into *MtSerB2* self-assembly and the mechanisms of regulation of its catalytic activity, we characterized the system in the absence and presence

of L-Ser via orthogonal analytical techniques (chromatography, electrophoresis, light scattering, crystallography) and enzyme kinetics studies. The workflow consisted in 1) the identification of the oligomeric species present in solution 2) the elucidation of their architecture 3) the study of their enzymatic activity. The characterization was also extended to more or less similar *MtSerB2* orthologs and site-specific mutants designed *in vitro*. The primary sequence differences allowed us to rationalize the involvement of key residues and domains in the structural mechanisms observed in *MtSerB2*.

The observations made throughout this work allow us to propose the following hypothesis: ***MtSerB2* is a morpheein**. By definition, a morpheein is a protein that can form two or more functionally and structurally distinct homo-oligomers, called *morpheein forms*, interconverting through a conformationally flexible dissociated state. For *MtSerB2*, we propose the morpheein equilibrium outlined in Figure 6.2. We will discuss each of its elements based on the results presented in the previous three chapters.

We can start the discussion with what is now obvious: **there is an active dimeric form of *MtSerB2* in solution**. Initially, this was assumed only on the basis of the structure of its closest crystallized homolog *MaSerB* (84% sequence identity). During this thesis, this could be verified on the basis of SEC-MALS and SEC-SAXS experiments confirming the stoichiometry of the species. Enzyme kinetics analysis, assaying the dephosphorylation of *O*-phospho-L-serine, also proved that it was active. The architecture of this homodimer could also be elucidated. To date, we have not been able to determine its atomic structure, the enzyme being recalcitrant to crystallization, but comparison of SAXS data obtained for *MaSerB* and *MtSerB2* dimer revealed that the two enzymes had identical shapes in solution. Moreover, thanks to the behavior of a truncated, monomeric mutant in solution, *MtSerB2* Δ ACT1, we were able to demonstrate that the ACT1 domain was necessary for the dimerization. On this basis, and given the very high sequence identity percentage with *MaSerB*, we are confident that *MtSerB2* dimer also exhibits a ACT1 domain-swapped *butterfly-like* architecture in solution (Figure 6.1B).

In our SEC experiments, we made an unexpected observation with respect to what was described for *MtSerB2* in the literature: in addition to the dimeric form, ***MtSerB2* can also exist as a tetramer in solution**. According to native PAGE gel quantification,

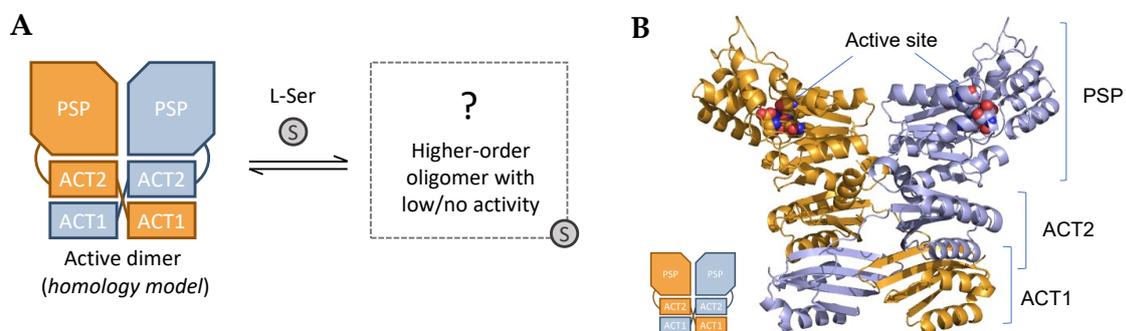


FIGURE 6.1: (A) *MtSerB2* equilibrium model prior to this work. (B) *MtSerB2* homology model based on *MaSerB* crystallographic structure (PDB: 3P96).

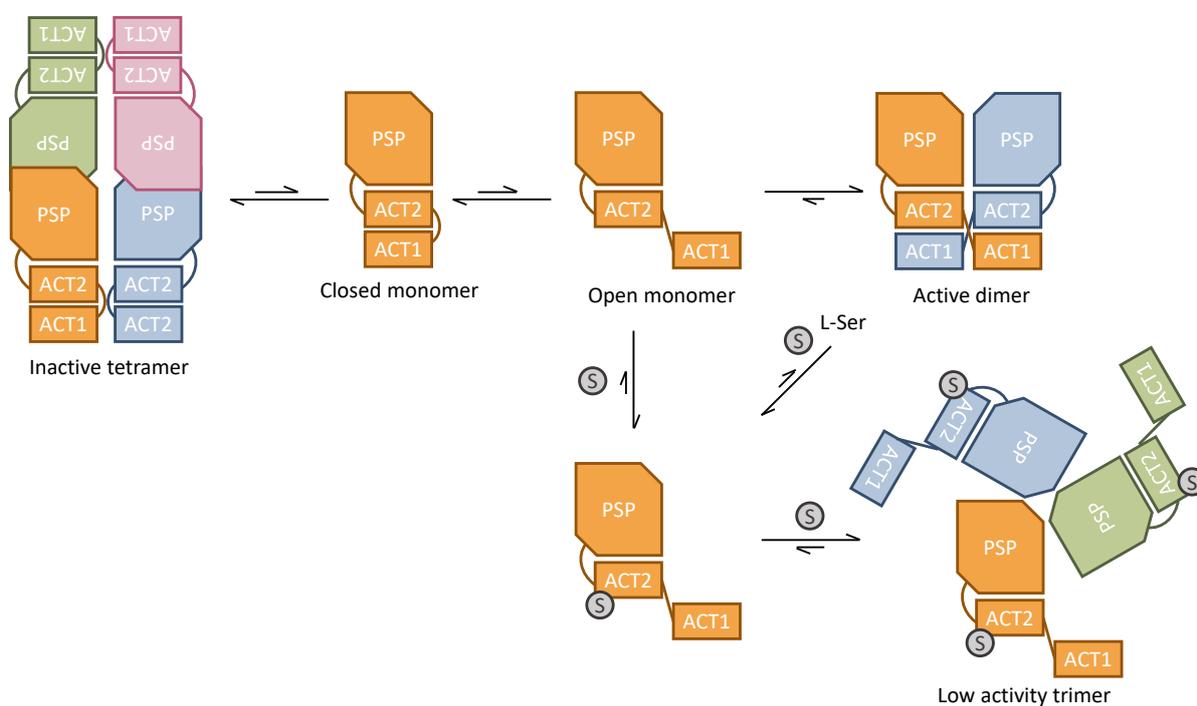


FIGURE 6.2: *MtSerB2* morpheein equilibrium model proposed on the basis of this work.

SEC-UV and mass photometry (MP) analyses, the tetramer population was observed in every *MtSerB2* batch we produced in *E. coli*. The ratio between populations was quite reproducible from production to production, with the tetramer population being always smaller than the dimer population. We suspect that the existence of this species was not reported in earlier work either because of its loss during the purification process, different from ours, or because it was overlooked as a non-relevant aggregate. Our work provides the first quantitative evidence for the existence of *MtSerB2* tetramer, whose stoichiometry was confirmed by SEC-MALS and SEC-SAXS.

As it proved stable enough to be separated from the dimer by preparative SEC, we were able to find that **the tetramer is an inactive species** through kinetics analyses and to undertake crystallogenesis trials. No crystallographic data could be obtained, but SAXS showed that **the tetramer is a rather globular species**, distinct from the dimer. Insight into its architecture was obtained by combining SAXS to protein-protein docking. We generated several tetramer models based on symmetric constraints and domain-swapping principles, and their scattering curves calculated by CRY SOL were then compared to the experimental SAXS profile. The single model retained is the best fit to the experimental SAXS data and exhibits a **global D2 symmetry and is formed by four monomers folded on themselves** with ACT1 and ACT2 domains interacting intramolecularly (Figure 6.3).

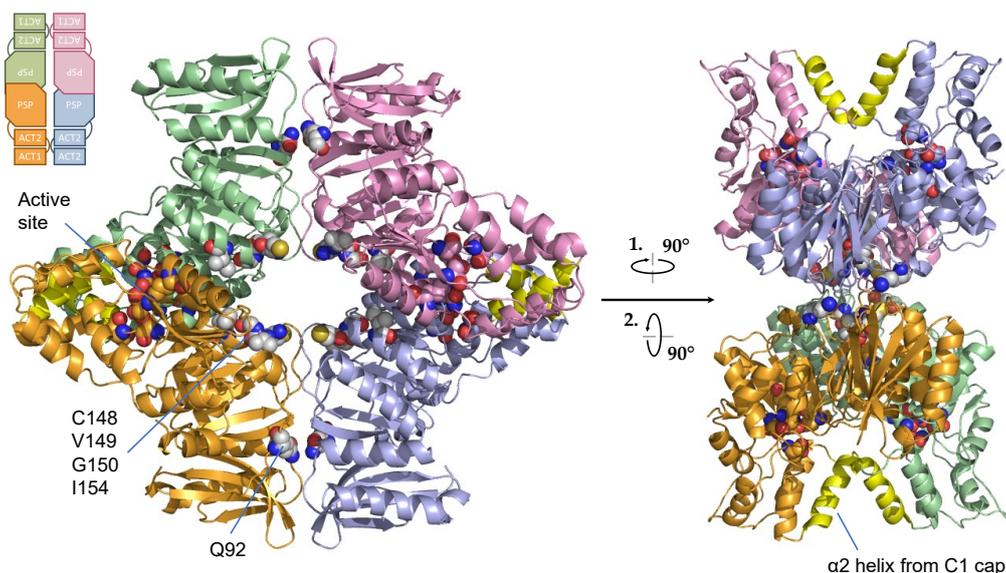


FIGURE 6.3: *MtSerB2* tetramer model obtained from D2 symmetric protein-protein docking and selected based on the best fit to the experimental SAXS data.

We do not pretend to propose here the exact architecture of *MtSerB2* tetramer, as the constraints set in the modeling process may not be representative of reality. First, the tetramer could be globally asymmetric. This phenomenon is rarely observed in nature, but is for example seen in the first enzyme of the serine pathway, *MtSerA1*. The protomers of the latter are not related by a symmetry operation. Then, the conformation adopted by the four protomers in *MtSerB2* may not be that of a monomer folded on itself as assumed according to domain-swapping theory. The hinge-loop that connects the two ACT domains confers great flexibility to the enzyme and the ACT1 domains may therefore adopt a completely different position than the one proposed in the model. This possibility has been illustrated by models generated via a rigid-body modeling algorithm, CORAL. Although the latter does not take any energy considerations into account in the modeling process, as opposed to docking, various models corresponding to the SAXS profile could be generated.

Our hypothetical tetramer must therefore be challenged further. Nevertheless, its structural features are well in line with experimental observations made throughout this work. To begin with, this model explains why the tetramer is inactive: the tetrameric interface involves the dynamic α_2 -helix of the C1 cap module known to modulate substrate access to the active site. Next, small neutral hydrophobic residues required for *MtSerB2* tetramerization (C148, V149, G150, I154) are located at the interfaces between the monomers. The patch was identified based on the sequence difference between *MtSerB2* and its close orthologs *MaSerB* and *MmSerB2*, exclusively dimeric in the same studied conditions. We observed that mutating these residues to the polar and charged residues observed at the same location in *MmSerB2* prevented *MtSerB2* from tetramerizing.

But the most important feature of our tetramer model is also the defining basis of the hypothesized morpheein model: it is formed of four monomers interacting with each other and not of two butterfly-like domain-swapped dimers. When screening tetramer models, we had obtained dimer-of-dimers that fitted well the experimental SAXS profile. However, we chose to discard them for two reasons: 1. they did not exhibit global symmetry with respect to the positioning of the monomers and therefore looked more like non-specific aggregates of two dimers 2. the formation of such aggregates could likely be forced at high protein concentration, however we have never been able to instantly reform tetramer by concentrating the dimer, even very strongly (25 mg mL^{-1}). The dimer-tetramer equilibrium appeared to be independent of the

total protein concentration. Considering this, we hypothesize that the dimer-tetramer interconversion appeals for a more complex mechanism involving the dissociation of the two oligomeric species into monomer.

The latter hypothesis is well in line with a complementary observation first reported by Yadav and Shree, then confirmed in this work: **when exposed to L-serine, the dimeric form of *MtSerB2* undergoes an oligomeric transition to a higher-order oligomer, that we quantitatively identified as a trimer by SEC-SAXS.** The stoichiometry underlying the dimer-trimer transition testifies to **the possible (transient) existence of a dissociated monomeric state in solution.**

SEC-UV analyses of *MtSerB2*, *MaSerB* and *MmSerB2* all showed the elution of what could be a monomeric population, indicated by the presence on the chromatograms of a small absorbance peak ("M?" on Figure 6.4, about 5%AUC) following that associated with the dimeric population. For this peak, the corresponding molar mass measured by MALS was comprised between those of the dimer (about 86 kDa) and the monomer (about 43 kDa), which could indicate a rapid re-equilibration from monomer to dimer after their separation. The identity of the species related to this small peak thus remains to be proven, but this observation is a first indication that SerBs domain-swapped dimers might be in equilibrium with their monomer.

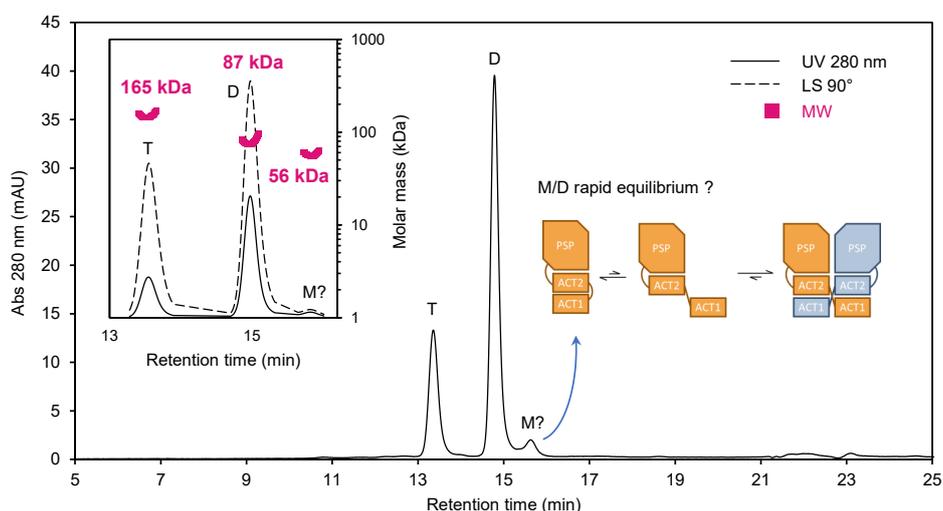


FIGURE 6.4: SEC-UV-MALS analysis of *MtSerB2*. Exclusively dimeric *MaSerB* and *MmSerB2* show similar chromatograms but without the tetramer peak. The molar mass of the "M?" peak is estimated between those of the dimer and the monomer, which could indicate a rapid monomer/dimer equilibrium.

The 12-residues hinge-loop connecting the two ACT domains confers a flexible nature to the protein and it is therefore very likely that the monomer shows a highly dynamic behavior in solution. According to the principles of domain-swapping, the most stable form would probably be a conformation where the monomer is folded on itself, with the ACT1 domain positioned below the ACT2 domain to form an intramolecular 8-stranded β -sheet. As per the mechanism proposed by Eisenberg and coworkers for proteins swapping entire domains [6], the monomer-dimer interconversion would take place through a high-energy intermediate¹ in which the monomer is in open form.

In the case of *MaSerB* and *MmSerB2*, exclusively dimeric in the studied conditions, only this monomer-dimer equilibrium would exist. For *MtSerB2*, the monomeric population would then also be in equilibrium with a tetrameric population, formed from the closed monomer form, which bears the small enabling C148 V149 G150 I154 hydrophobic patch not present in *MaSerB* and *MmSerB2*. In favor of this hypothesis is the molar mass associated with the tetramer peak, slightly underestimated by MALS compared to the theoretical mass (172 kDa). This thesis reports the result of only one indicative measurement, but we observed this trend in every measurement performed. The dimer peak was the only one whose measured mass accurately matched the theoretical mass each time.

This model of dissociative dimer-tetramer equilibrium in *MtSerB2* is in line with our experimental observations. Analyzing enriched dimer and tetramer samples, we found that the interconversion between the two populations is slow, on a scale of days to months, and that the transformation is much faster from tetramer to dimer than from dimer to tetramer. This finding is explained by our hypothetical model: from the thermodynamics and kinetics point of view, it is probably easier to break the interfacial interactions between four closed monomers, whose conformation gives them a certain stability in solution, and to overcome the kinetic barrier required for their opening than to readily disrupt a highly stable intertwined dimer into unstable open monomers (Figure 6.5). On this basis, starting from the hypothetical closed monomer formed after RNA translation, we could consider that *MtSerB2* tetramerization is favored by kinetic factors (lower activation energy) whereas dimerization is mainly driven by thermodynamic factors (more stable species).

¹Eisenberg and coworkers actually use the term "transition state" but we prefer to consider the form as a high-energy intermediate because it can potentially be isolated.

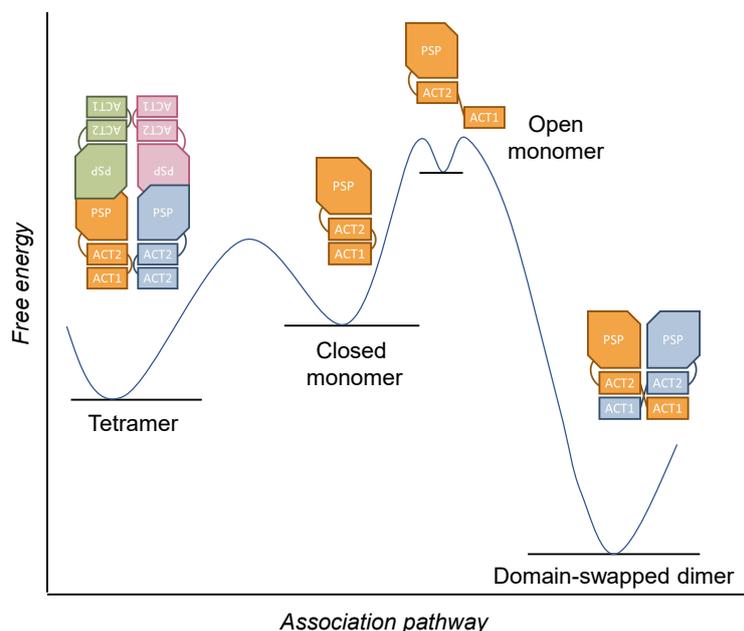


FIGURE 6.5: Hypothetical free energy diagram of *MtSerB2* tetramer/dimer dissociative equilibrium.

The dissociative equilibrium model could also explain why we observed that *MtSerB2* tetramerization propensity was impaired by the generation of a point mutation in the hinge-loop (Q92E). Although Q92 could only be involved in the stabilization of the tetramer through interactions at the protein-protein interface, its substitution by a glutamate residue could also favor dimerization by facilitating the opening of the monomer. In this context, it is possible that mutation Q92E lowers the kinetic barrier between the closed and open states of the monomer, for example by promoting an extended hinge-loop conformation, thus stabilizing the open conformation of the monomer, or by destabilizing the closed monomer form through unfavorable electrostatic contact.

In this work, we also investigated the mechanism of L-serine induced trimerization. Through analytical SEC experiments, we demonstrated that **the trimer was formed from *MtSerB2* dimeric population and that the phenomenon was specific to this enzyme.** None of the orthologs studied in parallel undergo this transition, even *MaSerB* and *MmSerB2* which have a very similar butterfly-like ACT1 domain-swapped architecture as shown by SAXS. Upon *in silico* analysis of potential L-Ser binding sites on the three mycobacterial dimeric SerBs, we discovered a site detected only in *MtSerB2* and not in *MaSerB* and *MmSerB2* (Figure 6.6). This site is located at

the ACT1-ACT2 interface, opposite the known binding pocket identified in *MaSerB* co-crystals (PDB: 5JLP and 5JLR), and is homologous to an L-Ser binding site observed in *E. coli* PGDH (PDB: 1PSD). Docking studies in this additional pocket allowed to identify residues likely responsible for the interaction between *MtSerB2* dimer and L-Ser (E33 R103 T136) and we confirmed by the generation and SEC analysis of a triple alanine mutant that they were indeed required for trimerization.

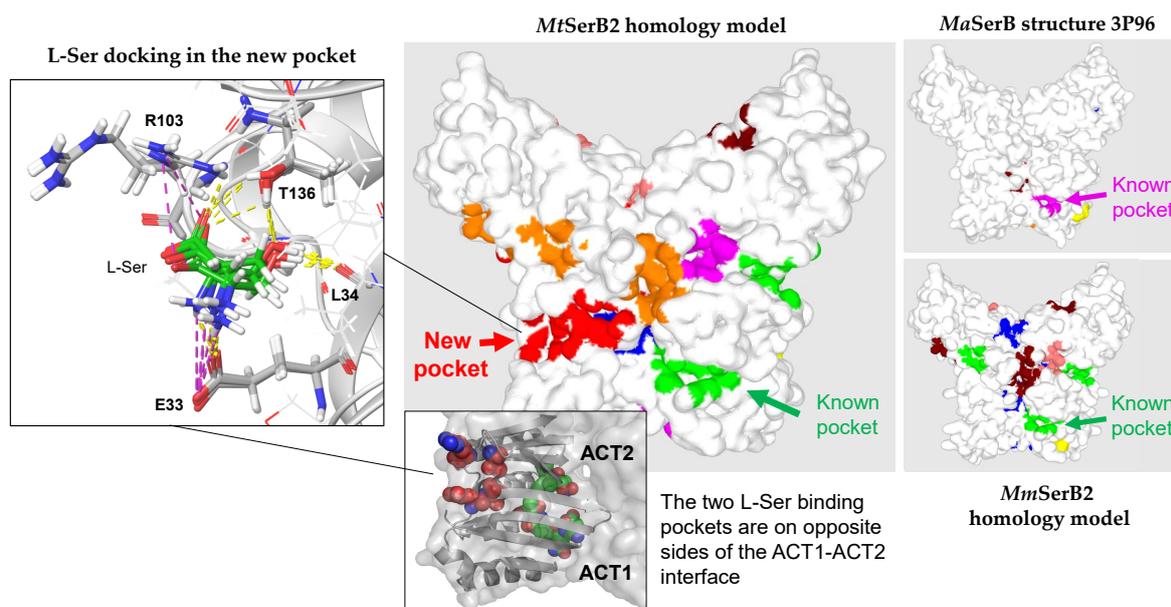


FIGURE 6.6: Identification of a new L-Ser binding pocket specific to *MtSerB2* dimer at the ACT1-ACT2 domain interface. The residues forming this pocket are necessary for *MtSerB2* to trimerize.

As in the case of dimer and tetramer, no diffracting crystal could be obtained for the trimer. However, we were able to access enlightening information about the shape of this species in solution thanks to SAXS. Based on the Guinier and Kratky plots and the $P(r)$ function, we noted that **the trimer is elongated and probably more flexible than the dimer and the tetramer**. Protein-protein docking was undertaken to get an idea of a plausible architecture and models that best fit the experimental scattering profile all exhibited a flattened shape. The great majority of these models were interacting through their ACT2-PSP part, leaving the ACT1 domain on the outside of the complex. Rigid body modeling was also undertaken using CORAL and the best-fitting generated models also exhibited these characteristics, along with the ACT1 domain separated from the rest of the structure by the hinge-loop in an extended conformation (Figure 6.7). This important feature resulted in models whose radius of

gyration was much closer to the experimental value and whose extended configuration could explain the apparent conformational flexibility of the species. Moreover, this hypothetical configuration is in line with the fact that the ACT1 domain does not appear to be strictly necessary for *MtSerB2* trimerization. We indeed showed through SEC-MALS experiments that a monomeric mutant lacking the ACT1 domain (*MtSerB2* Δ ACT1) could spontaneously form a small trimer population.

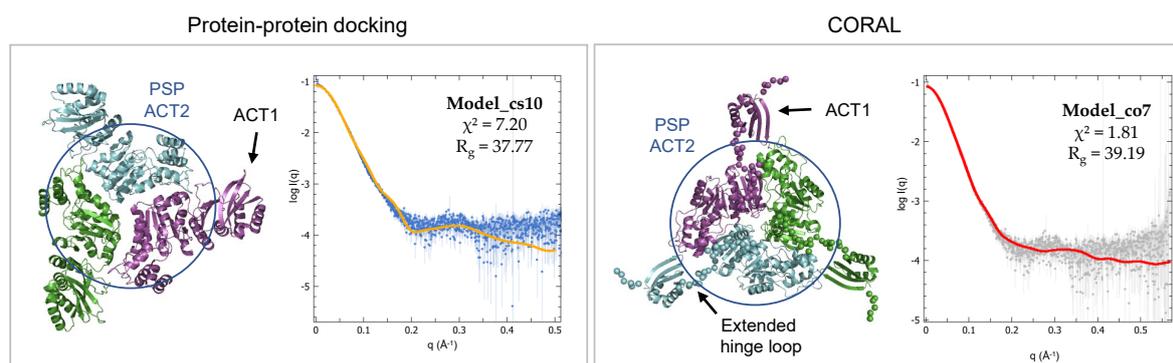


FIGURE 6.7: Best-fitting C3 trimer models obtained by protein-protein docking and rigid body modeling (CORAL), and superimposition of their calculated scattering profile (CRY SOL) to the experimental SAXS data.

Finally, we characterized the inhibition mode of mycobacterial SerBs dimers by L-Ser in enzyme kinetics experiments. The results obtained well support the existence of an L-Ser interacting pocket specific to *MtSerB2*: while *MaSerB* and *MmSerB2* are completely inhibited by the consecutive binding of two L-Ser molecules, the first one probably interacting with the known binding site and leading to the appearance of a second inhibitory site by conformational change, *MtSerB2* is partially inhibited by the interaction with a single molecule of L-Ser. To the best of our knowledge, the inhibitions kinetics of *MaSerB* and *MmSerB2* by L-Ser have never been documented before, but the results obtained regarding *MtSerB2* are consistent with the allosteric partial competitive inhibition mechanism reported by Grant [5].

Based on the structural and kinetic considerations reminded above, we propose an outline for the structural mechanism of *MtSerB2* regulation by L-Ser (Figure 6.8). According to our results, L-Ser would interact with a pocket containing residues E33, R103 and T136 located at the intermolecular ACT1-ACT2 domain interface of the domain-swapped dimer. This interaction would be destabilizing for the dimer, which would eventually be disrupted into monomers. The interaction with L-Ser would prevent these monomers from shifting to a closed conformation by re-establishing

intramolecular ACT1-ACT2 interactions. Open monomers being likely quite unstable in solution, the equilibrium would be shifted towards the formation of a trimeric population of very low catalytic activity. The latter could be explained by oligomeric interfaces mostly formed at the PSP domain, making it more difficult for the substrate to access the active site.

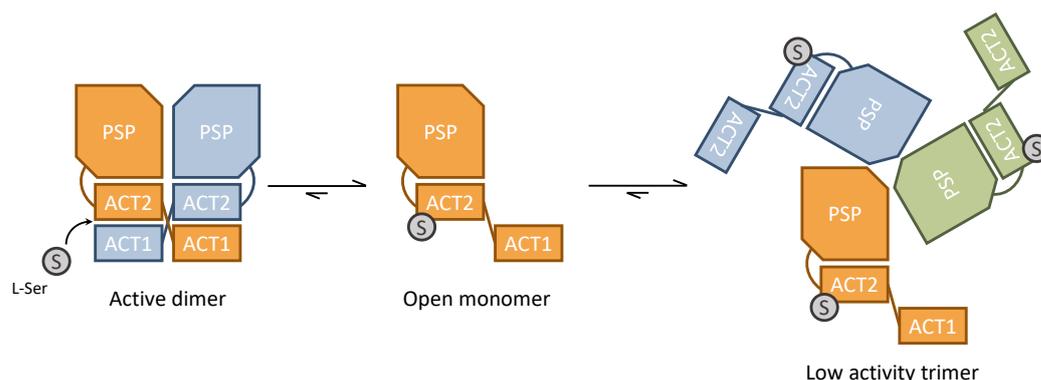


FIGURE 6.8: Hypothetical L-Ser induced trimerization mechanism in *MtSerB2*.

We observed that L-Ser acted on the dimeric population acknowledging a displacement and a deformation of the corresponding peak in SEC analyses. As for the peak related to the tetrameric population, only the area under the curve was slightly decreasing with increasing L-Ser concentration. Based on the equilibrium model proposed above and depicted in Figure 6.2, we believe that the decrease observed in the tetrameric population is due to the shift of the equilibrium towards the trimer through a stabilization of the open monomer. L-Ser would therefore not directly interact with either the tetramer or the closed monomer, or with greater difficulty. This assumption is supported by the fact that the E33 R103 T136 binding pocket is not detected in our closed monomer model that was subjected to Molecular Dynamics. This also supports again that *MtSerB2* tetramer is likely formed of four separate monomers and not of two domain-swapped dimers. If this was the case, L-Ser would probably disrupt the intermolecular ACT1-ACT2 interface of the two dimers of the dimer-of-dimers, which could lead to the total disappearance of the tetrameric population. In all cases, this *MtSerB2* specific L-Ser binding pocket does not exist in the tetramer or is not accessible.

As developed above, the experimental observations made throughout this work are consistent with and complement the upstream biophysical and biochemical work

of Yadav and Shree [2] and Grant [5], and can be explained by the morpheein system we propose for *MtSerB2*. But the latter is also well in line with the moonlighting protein behavior highlighted by Shree *et al* [3]. Along with the existence of alternate quaternary structure stoichiometries, non-trivial kinetic data or inconsistent subunit interactions in different protein crystal structures, moonlighting functions can indeed be a sign that a homo-oligomeric protein is a morpheein [7]. Porphobilinogen synthase (PBGs), the prototype morpheein, is an essential enzyme in the biosynthesis of tetrapyrrole and presents itself two moonlighting functions as a proteasome inhibitor and a cochaperone [8–10].

In the bacterial cytosol, *MtSerB2* acts as a phosphoserine phosphatase. It catalyzes the third and last, irreversible step of L-serine biosynthesis: the hydrolysis of *O*-phospho-L-serine to L-serine. This activity is crucial to the survival of *M. tuberculosis* because L-Ser is a central precursor for the synthesis of numerous metabolites and the pathogen cannot acquire it from host macrophages [11]. Based on the state of the art and our work, it is the dimeric ACT1 domain-swapped form of *MtSerB2* that performs this metabolic function. The L-serine phosphorylated pathway directly consumes an intermediate of glycolysis, 3-phosphoglycerate, so the pathway must be regulated depending on the energy needs of the bacteria. Like *MtSerA1*, the first enzyme in the pathway, *MtSerB2* is allosterically feedback inhibited by L-Ser. Our results suggest a regulation mechanism based on a change of oligomerization state towards a very low activity trimeric form, induced by the disruption of intermolecular ACT1-ACT2 interfaces stabilizing the domain-swapped dimer.

But *MtSerB2* is also found in the cytosol of host macrophages, as shown by Shree *et al.* [3]. There, it can dephosphorylate host-phosphoproteins, physically interact with anti-apoptotic proteins and modulate the expression of various genes, leading to the modification of the host's immune response. The first of these moonlighting functions is probably related to the dimeric form of *MtSerB2* as it involves dephosphorylation and this oligomeric population is the only one that shows significant activity. This assumption is supported in a recent thesis of our group which provides a protein-protein docking structural basis for the dephosphorylation of phosphocofilin by *MtSerB2* domain-swapped dimer [12]. A priori, the other moonlighting functions would not require the phosphatase activity and could be related to the other oligomeric forms identified in this work. The alternative forms adopted by the tetramer and trimer could lead to the exposure of new interfaces for the interaction

with partners from the host. The idea that *MtSerB2* may play a role in modulating the transcription of certain genes does not seem unreasonable either thanks to the presence of ACT domains. Some transcriptional regulators indeed possess ACT domains, named RAM domains (for Regulator of Amino acid Metabolism) [13]. Lastly, how *MtSerB2* is translocated from the bacterial cell to the host cell has not been determined yet, but assuming that this is due to a secretion system, it could be that one or the other morphoein form is preferentially used as a substrate.

6.2 Conclusions

In the popular imagination, tuberculosis (TB) is a forgotten disease far from current concerns. The reality is quite different: recently dethroned by SARS-CoV-2, TB is the second leading cause of death worldwide due to a single infectious agent. Although treatable, the disease is no exception to the growing global public health threat of antibiotic resistance. Through the Global Tuberculosis Programme, a global strategy of prevention, care and innovation aiming at a world free of TB, the World Health Organization calls for the development of new anti-tuberculosis agents.

In this context, numerous research initiatives have been undertaken to identify novel drug targets. The sequencing of *M. tuberculosis* genome allowed the development of methods to determine gene essentiality in the pathogen, and it is on that basis that the phosphoserine phosphatase *MtSerB2* was first proposed as a potential drug target in 2003. Further characterization of the enzyme was subsequently published in 2014, 2016 and 2017. The studies suggested the involvement of *MtSerB2* in host colonization and demonstrated its druggability through the inhibition of its phosphatase activity by small molecules. The premise of a feedback regulation mechanism by L-serine involving a transition to a higher order oligomeric state of undefined stoichiometry was also proposed.

It is in the latter observation that the present thesis takes root. We further investigated *MtSerB2* self-assembly and its involvement in the regulation of phosphatase activity through structural biology and enzymological approaches. In addition to confirming the inferred domain swapped-architecture of the dimer, we brought the first proof for the existence of a tetramer in solution and identified a trimeric stoichiometry as well as the interaction site for the L-Ser induced oligomer. At the end of

this work, we propose that *MtSerB2* is a morpheein, existing in solution as an equilibrium of dimer, trimer and tetramer populations of distinct activity levels that are interconverted through a conformationally flexible monomeric state. Our morpheein model, consistent with and complementing the literature, is yet to be fully established but already lays the foundation for allosteric drug design. It sheds light on a new reservoir of potential targets: compounds stabilizing the (almost) inactive morpheein forms of *MtSerB2* could be designed. Moreover, such an equilibrium does not exist in the human ortholog *HsPSP*, and the latter has a dimeric structure very different from that of domain-swapped *MtSerB2*, which makes a therapeutic strategy based on quaternary structure dynamics very selective.

MtSerB2 appears to have gained rather unusual properties for an essential metabolic enzyme. According to the literature, it would also be used by *M. tuberculosis* to interfere with host cell signaling processes and thus create a favorable environment for its proliferation. Although much remains to be demonstrated, the morpheein dynamics of *MtSerB2* could account for these observed moonlighting functions.

In addition to contributing to the ongoing effort to tackle TB, this work exemplifies more fundamental concepts. First, the morpheein model itself challenges the protein-folding paradigm telling us that to one primary structure corresponds a well-defined quaternary structure *in fine*. In the morpheein model, a single polypeptide chain can adopt different tertiary structures through subtle rearrangements of secondary structures, leading to functionally distinct quaternary structures.

Furthermore, during this project, we also studied several orthologous phosphoserine phosphatases all showing more than 30% sequence similarity with *MtSerB2*. Among these were the SerBs of *M. avium* and *M. marinum*, both sharing about 90% sequence similarity with *MtSerB2*. As demonstrated in this thesis, the dimeric architectures formed by these enzymes and *MtSerB2*, as well as their catalytic properties, are very similar. However, when it comes to feedback regulation by L-serine, they exhibit distinct inhibition modes where *MtSerB2* is partially inhibited by the binding of a single molecule while *MaSerB* and *MmSerB2* are totally inhibited by the interaction with two molecules. This difference is most likely due to a L-Ser interacting pocket present in *MtSerB2* but not in the two orthologs. These results not only point out that enzymes of very similar structures catalyzing the same reaction can be regulated differently in distinct species, but also reflect significant differences in the physiology

and metabolism between organisms.

On that note, this thesis also provides a new crystallographic structure: that of the essential phosphoserine phosphatase SerB from *Brucella melitensis*. The latter could be of great help to investigate molecular mechanisms at work in the bacteria and a basis for the development of novel small-molecule inhibitors against brucellosis.

Another contribution is that the characterization of the quaternary structures of all these homologous enzymes illustrates well the different oligomerization mechanisms by which an ancestral monomer can evolve in distinct species (Figure 6.9).

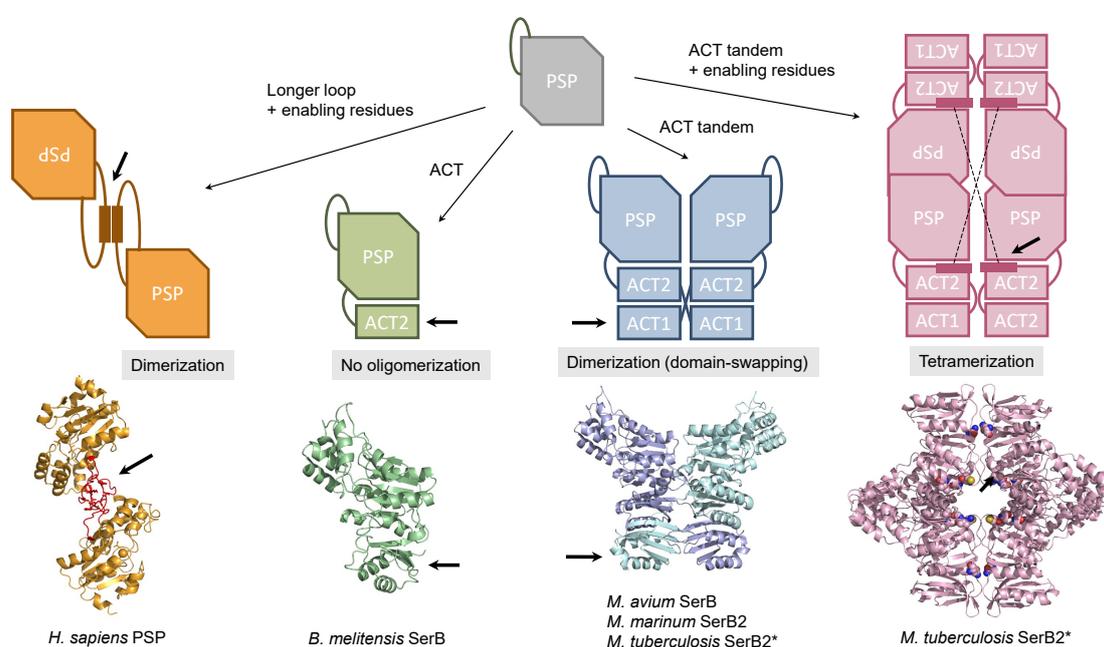


FIGURE 6.9: Features that a hypothetical monomeric common ancestor could have acquired during evolution explaining the difference in quaternary structure observed in the PSPs studied in this work. This scheme is purely illustrative and does not reflect the real evolutionary links between species. * For *MtSerB2* the depicted structures are *in silico* models.

For example, *BmSerB* is monomeric in solution while *HsPSP* is dimeric even though they both have the same Rossmann-like fold. However, when comparing the structures and sequences of the two enzymes, it can be seen that the β -hairpin loop through which *HsPSP* dimerizes is four residues longer and has Phe and Tyr residues not conserved in *BmSerB*. The difference in oligomerization between these two enzymes could therefore be explained by a combination of key insertions/deletions and substitutions. Mycobacterial SerBs, for their part, have an additional N-terminal ACT

domain grafted by a 12 residue arm to the rest of the enzyme. Not only are ACT domains known to self-assemble, but the amino acid arm acts as a hinge-loop that allows a reciprocal exchange of the N-terminal ACT domain. The presence of motifs promoting oligomerization and domain-swapping thus allows *MtSerB2*, *MaSerB* and *MmSerB2* to dimerize. But that is not all: *MtSerB2* is able to tetramerize whereas its close counterparts are not. It is the substitution of some charged and bulky residues to smaller and hydrophobic residues in an α -helix that gives it this ability.

The unexpected oligomeric behavior of *MtSerB2* compared to its very close mycobacterial counterparts emphasizes the importance to experimentally determine the quaternary structure of enzymes. Although being the object of a growing interest, the computational prediction of oligomeric states and protein-protein interactions in general remains a real challenge to this day. Homology modeling is a tool of choice to predict the structure of enzymes that have never been crystallized. However, the prediction of quaternary structures is based on what is observed in related enzymes and this thesis demonstrates that the strategy is not flawless. Although we found the architecture of *MtSerB2* dimer homology model to be coherent with its solution structure, in no case the prediction method has given the slightest clue to the existence of a tetramer. The monomeric *BmSerB* was also predicted to be dimeric by the same approach. Similarly, cutting edge AI technologies like AlphaFold, even if a major breakthrough in the protein folding problem, struggles to accurately predict homo-oligomers and complexes. Like all AI algorithms, AlphaFold is based on what has been observed before and therefore manages to find the butterfly-like domain-swapped structure of *MtSerB2* dimer. However, the predictions it makes for the tetramer is closer to plates of spirelli and spaghetti than to a well-behaved biological macromolecule (Figure 6.10). With this example among many others [14], it is obvious that experimental structural biology still has a bright future ahead.

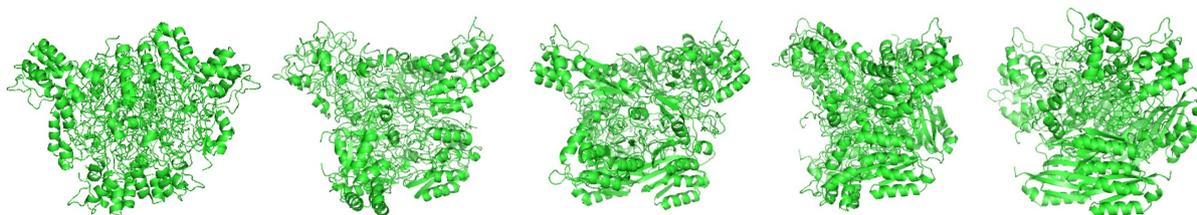


FIGURE 6.10: Five attempts of AlphaFold Multimer at modeling *MtSerB2* tetramer.

6.3 Outlooks

MtSerB2 morpheein model proposed in this thesis remains hypothetical and must therefore be challenged. The most direct way to get more information about the equilibrium would be to obtain the experimental structures of the newly discovered trimer and tetramer species. The conformation of the protomers within them could indeed be very indicative of the interconversion mechanism between the oligomers in solution. Although not diffracting, crystals have been obtained during our trials, which is encouraging as to the possibility of obtaining crystallographic structures of *MtSerB2* trimer and tetramer. It would be interesting to optimize the conditions that produced hits but a larger-scale robotic screening strategy may be considered to speed up the process. *MtSerB2* homomers are also large enough to allow the determination of their structure by Cryo-Electron Microscopy (CryoEM). However, their study by this more recent technique would need an entire project on its own, knowing that sample preparation often requires a great deal of optimization in order to collect high-resolution data sets and that current access to CryoEM facilities is limited and in high demand.

In the shorter term, the validation of *MtSerB2* homo-oligomer models proposed in this thesis is an interesting perspective. Site-directed mutagenesis would be a method of choice to probe inferred oligomeric interfaces. For example, residues located at the protein-protein interface of our D2 tetramer model, notably in the dynamic alpha helix of the C1 cap module, could be mutated to alanine. A reduction in the tetrameric population of the alanine mutant would then indicate that this interface probably exists in the solution tetramer and further validate our model. It would also be interesting to introduce the tetramerization patch identified in *MtSerB2* in its fully dimeric counterparts *MaSerB* and *MmSerB2* to see if the latter would be able to tetramerize in the studied conditions with only this minimal set of residues. The mutagenesis approach could also be used for the discrimination of trimeric models but would require a more substantial effort as several arrangements have been proposed. To start with, an alanine-scan based on the trimeric interfaces of the best fitting model could provide a basis for further sorting.

The mechanism of *MtSerB2* trimer formation from the dimer could also be investigated further. We have identified three residues (E33 R103 T136) potentially involved in this phenomenon and hypothesized that the interaction of L-Ser with these would

disrupt the intermolecular ACT1-ACT2 domain interaction, leading to dimer dissociation. Another round of site-directed mutagenesis and SEC experiments could be used to determine the minimum set of residues required for the transition. Molecular dynamics simulations could help shedding light on the disruption mechanism at the atomic-level, although the phenomenon could only be observed at the sub-millisecond timescale, probably limiting the simulation to the beginning of the conformational change. We could also use stopped-flow fluorescence resonance energy transfer (FRET) to test our hypothesis of ACT1-ACT2 domains separation while monitoring the change in conformation and oligomerization state over time upon L-Ser addition. The idea would be to graft a donor fluorophore on the ACT1 domain and an acceptor fluorophore on the ACT2 domain (or vice versa) and analyze the variation of the FRET signal over time. The major challenge of this experiment would be the site-specific double labeling of *MtSerB2* with fluorophores. However, we believe that this is feasible: there are only three cysteine residues in the enzyme (Figure 6.11A) and, according to our dimer homology model, only Cys148 of ACT2 is exposed to the solvent (Figure 6.11B). The latter could thus be specifically functionalized with an acceptor such as Alexa₅₆₈ maleimide and a FlAsH₅₂₈ donor could be specifically placed at the N-term on the ACT1 domain thanks to the introduction of a CCPGCC recognition sequence by mutagenesis like described in reference [15] (Figure 6.11C).

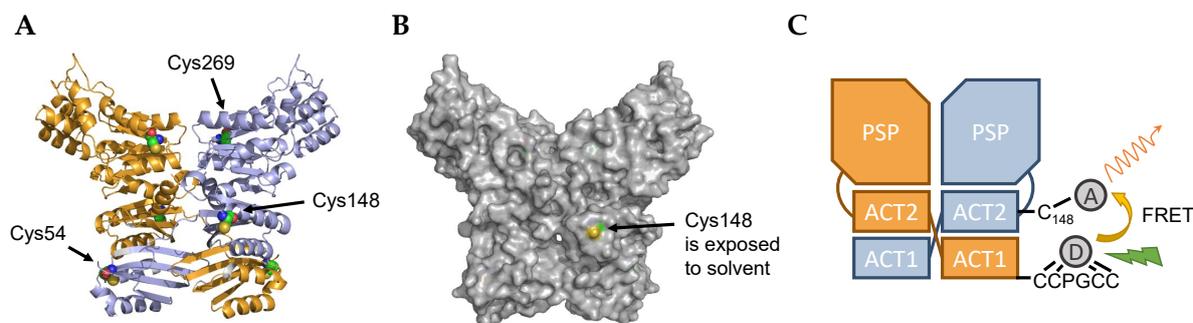


FIGURE 6.11: (A) Position of the three cysteine residues of *MtSerB2*. (B) Surface representation of *MtSerB2* showing that only Cys148 is exposed to solvent. (C) Functionalization strategy with FRET donor and acceptor to monitor conformational changes (ACT1 and ACT2 domains movement) induced by L-serine. The functionalization of only one monomer is depicted for clarity.

The proposed morpheein equilibrium for *MtSerB2* also depends on the dissociation of dimer and tetramer to a monomeric state. The presence of the monomer

in solution has proven difficult to demonstrate directly: in SEC, a small peak appears after the elution of the dimer peak, but MALS gives an intermediate molar mass between dimer and monomer, and mass photometry did not allow to observe any monomeric population because of the high background effect of Tris buffer (Appendix F). However, it would be possible to prove the dissociative equilibrium by showing that the tetramer and dimer can exchange subunits. One approach is to produce a heavier or lighter *MtSerB2* mutant by mutating one or several residues not involved in the oligomeric equilibrium (preliminary verification). Equimolar amounts of this mutant's tetramer and native *MtSerB2* tetramer would then be mixed and the mixture would be left a few days at 4 °C so that dimer could be reformed. The sample would then be analyzed by native mass spectrometry: an intermediate mass between that of a native dimer and a mutant dimer would indicate that a hetero-dimer was formed and that tetramer-dimer interconversion does indeed involve dissociation into a monomer.

With the aim of using this system for the design of new antituberculous drugs, the existence of these novel targets should also be proven *in cellulo* in *M. tuberculosis* as well as in macrophages. This could be achieved via non-denaturing electrophoresis experiments coupled to western blot. Another solution would use immunofluorescence, although more challenging to implement because it requires the production of antibodies specific for the tetramer and trimer.

It would also be important to identify the role of the different morphein forms. As a first step, one could determine whether the tetramer and trimer are important to pathogenicity by studying *M. tuberculosis* mutants unable to form these higher-order oligomers or assessing the effect of the exogenous addition of *MtSerB2* α (impaired tetramerization) and *MtSerB2* 3A (impaired trimerization) mutants on THP-1 cells as in the work of Shree *et al.* [3]. Pull-down assays combined with native electrophoresis could also help determine if a form other than the dimer physically interacts with host proteins. Another perspective would be to study the oligomeric behavior of the phosphoserine phosphatase SerB653 from *Porphyromonas gingivalis* as it has moonlighting features strongly resembling those of *MtSerB2* [16, 17].

Finally, the (quasi) inactive *MtSerB2* trimer and tetramer offer new perspectives for the design of drugs preventing deleterious phosphatase activity, whether in the pathogen metabolism or in the disruption of host cell signaling. One possibility would

be to design inhibitors that bind to *MtSerB2* tetramer. Such compounds could stabilize this inactive form and prevent its conversion to active dimer, but first determining the atomic structure would be necessary for a rational and efficient design. Another option would be to directly disrupt the active dimer by preventing the highly stabilizing intermolecular interaction between the ACT1 and ACT2 domains, using small peptides for example. The design of compounds targeting *MtSerB2* trimerization pocket (E33 R103 T136) is also a strategy that could be implemented. One could imagine molecules resembling L-Ser but bearing an electrophilic head that could react with T136 to form a covalent bond. Such compounds would maintain *MtSerB2* in a low activity trimeric state.

This thesis work allowed to unveil some of *MtSerB2*'s secrets but, as proposed above, the system still offers room for further investigation in structural, molecular and cellular biology as well as in organic synthesis and pharmacochimistry.

Bibliography

- [1] G. Arora, P. Tiwari, R.S. Mandal, A. Gupta, S. Sharma, D. and Saha, and R. Singh. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *Journal of Biological Chemistry*, 289(36):25149–25165, 2014.
- [2] G.P. Yadav, S. Shree, R. Maurya, N. Rai, D. K. Singh, K.K. Srivastava, and R. Ramachandran. Characterization of *M. tuberculosis* SerB2, an essential HAD-family phosphatase, reveals novel properties. *PLoS One*, 9(12):e115409, 2014.
- [3] S. Shree, A.K. Singh, R. Saxena, H. Kumar, A. Agarwal, V.K. Sharma, K. Srivastava, K.K. Srivastava, S. Sanyal, and R. Ramachandran. The *M. tuberculosis* HAD phosphatase (Rv3042c) interacts with host proteins and is inhibited by Clofazimine. *Cellular and Molecular Life Sciences*, 73(17):3401–3417, 2016.
- [4] E. Pierson, M. Haufroid, T.P. Gosain, P. Chopra, R. Singh, and J. Wouters. Identification and repurposing of trisubstituted harmine derivatives as novel inhibitors of *Mycobacterium tuberculosis* phosphoserine phosphatase. *Molecules*, 25(2):415, 2020.

- [5] G.A. Grant. Regulatory mechanism of *Mycobacterium tuberculosis* phosphoserine phosphatase SerB2. *Biochemistry*, 56(49):6481–6490, 2017.
- [6] M.P. Schlunegger, M.J. Bennett, and D. Eisenberg. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Advances in Protein Chemistry*, 50:61–122, 1997.
- [7] E.K. Jaffe and S.H. Lawrence. The morpheein model of allostery: evaluating proteins as potential morpheeins. In *Allostery*, pages 217–231. Springer, 2012.
- [8] E.K. Jaffe. The remarkable character of porphobilinogen synthase. *Accounts of Chemical Research*, 49(11):2509–2517, 2016.
- [9] G.G. Guo, M. Gu, and J.D. Etlinger. 240-kDa proteasome inhibitor (CF-2) is identical to delta-aminolevulinic acid dehydratase. *Journal of Biological Chemistry*, 269(17):12399–12402, 1994.
- [10] M. Gross, S. Hessefort, and A. Olin. Purification of a 38-kDa protein from rabbit reticulocyte lysate which promotes protein renaturation by heat shock protein 70 and its identification as δ -aminolevulinic acid dehydratase and as a putative DnaJ protein. *Journal of Biological Chemistry*, 274(5):3125–3134, 1999.
- [11] J. Rengarajan, B.R. Bloom, and E.J. Rubin. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences*, 102(23):8327–8332, 2005.
- [12] M. Haufroid. *Structural and inhibition study of phosphoserine phosphatase (SerB2) from Mycobacterium tuberculosis*. PhD thesis, Université de Namur, 2022.
- [13] T.J.G. Ettema, A.B. Brinkman, T.H. Tani, J.B. Rafferty, and J. Van Der Oost. A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. *Journal of Biological Chemistry*, 277(40):37464–37468, 2002.
- [14] A Perrakis and T.K. Sixma. AI revolutions in biology: The joys and perils of AlphaFold. *EMBO reports*, 22(11):e54046, 2021.
- [15] M.J. Patel, G. Yilmaz, L. Bhatia, E.E. Biswas-Fiss, and S.B. Biswas. Site-specific fluorescence double-labeling of proteins and analysis of structural changes in solution by Fluorescence Resonance Energy Transfer (FRET). *MethodsX*, 5:419–430, 2018.

- [16] G.D. Tribble, S. Mao, C.E. James, and R.J. Lamont. A *Porphyromonas gingivalis* haloacid dehalogenase family phosphatase interacts with human phosphoproteins and is important for invasion. *Proceedings of the National Academy of Sciences*, 103(29):11027–11032, 2006.
- [17] Y. Hasegawa, G.D. Tribble, H.V. Baker, J.J. Mans, M. Handfield, and R.J. Lamont. Role of *Porphyromonas gingivalis* SerB in gingival epithelial cell cytoskeletal remodeling and cytokine production. *Infection and Immunity*, 76(6):2420–2427, 2008.

Part V

Appendices

Appendix A

Brief overview of the techniques

The table on the next page (Figure A.1) lists the techniques used in this thesis to study the oligomeric states of enzymes, along with the type of data to be interpreted they provide and the parameters and information to which they gave access in this work.

In this appendix, special attention was given to Multi-Angle Light Scattering (MALS) and Small-Angle X-ray Scattering (SAXS). These characterization techniques are indeed less routinely used and further explanation about how they work and what information they provide is essential to the understanding of the analyses performed in this work. The material presented below is a synthesis from references [1–3] for MALS and [4–6] for SAXS.

A word about mass photometry (MP) - MP is a recent biophysical technique that enables the quantification and measure of the molecular mass of individual biomolecules in solution. Like MALS and SAXS, MP exploits the ability of molecules to scatter light (*vide infra*). It is essentially based on the measurement of light scattering signals (events) as biomolecules land on an illuminated measurement surface. The events are counted and the scattering signal is proportional to the molar mass. Therefore, the technique allows to produce mass histograms quantifying the species present in solution and identifying their stoichiometry. We were given the opportunity to test MP during a demonstration of a Refeyn Two MP instrument at the Laboratory for the Analysis of Medicines (LAM, ULiège). The principles will not be developed further here but detailed explanation can be found in reference [7].

	Purpose	Technique	Data to interpret	Obtained parameters	Information
Structure	Identification and quantification of the species	Native-PAGE	Colored bands on a gel	Number of bands, intensity, electrophoretic mobility	Number of species, quantity
		SEC	UV chromatogram	Number and position of peaks, area under the curve	Number of species, quantity, (stoichiometry*)
		MALS	Intensity of scattered light	I_s	Molar mass
Structure determination	Structure determination	MP	Counts (landing events on illuminated surface) and intensity of scattered light	Mass histogram*	Number of species, quantity, molar mass*
		SAXS	Scattering curve	$I(q) \rightarrow I(0), R_{gr}, P(r),$ replots (Guinier, Kratky), V_p	Size, shape, molar mass, structure verification
Activity	Activity of the species	X-ray crystallography	Diffraction pattern	Electron density map	Atomic structure
		Phosphatase assay	Absorbance values	Quantity of free phosphate as a function of substrate concentration $\rightarrow K_{M}, K_{cat}, K_i$	Enzyme kinetics, inhibition mechanisms

FIGURE A.1: Biophysical and biochemical characterization techniques used in this work to study the oligomeric state of *MtSerB2*, *MmSerB2*, *MaSerB*, *BmSerB* and *HsPSP*. PAGE: PolyAcrylamide Gel Electrophoresis, SEC: Size Exclusion Chromatography, MALS: Multi-Angle Light Scattering, MP: Mass Photometry, SAXS: Small-Angle X-ray Scattering. *achieved through calibration with proteins of known molar mass.

Multi-Angle Light Scattering (MALS)

Multi-angle light scattering (MALS) is an analytical technique based on the optical phenomenon of *light scattering*, i.e. the ability of a particle to scatter light in every direction when interacting with a light source. It allows the absolute determination of the molar mass (M_W) of a macromolecule in solution and in some cases also its radius of gyration (R_g).

In this work, we used SEC-MALS. The sample to be analyzed passes first through a SEC column, then through a UV detector before reaching the MALS cell. This ensures that the analyzed macromolecular population is homogeneous in terms of size.

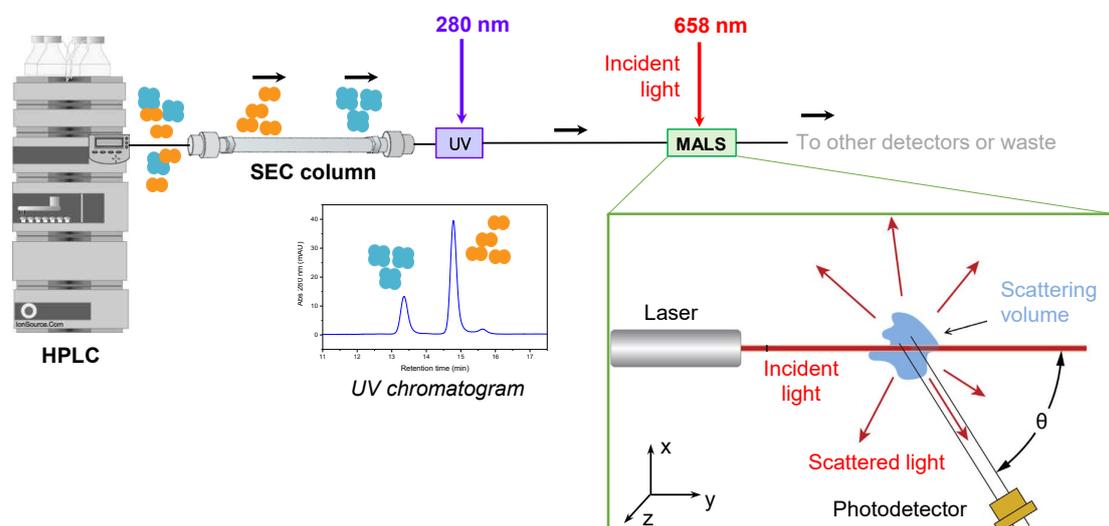


FIGURE A.2: SEC-MALS setup and data collection.

Figure A.2 illustrates the experimental setup and the phenomenon of light scattering happening in the MALS cell, leading to M_W determination. The sample is illuminated with a polarized monochromatic light from a laser ($\lambda = 658$ nm). This incident light is an electromagnetic wave composed of an electric field \vec{E} oscillating in the z direction and a magnetic field \vec{B} oscillating perpendicularly in the x direction. When the incident light encounters the electronic cloud of a particle in solution, the latter is displaced by the oscillating electric field \vec{E} . This induces an oscillating dipole \vec{p} that re-radiate light in every direction. This radiated light (*scattered light*) is collected in the x - y plane at given angles (θ) with respect to the direction of incident

light (usually 15° , 90° , but some detectors can go up to 18 angles evenly distributed from 0 to 180°).

Following Zimm's formalism and according to the Rayleigh-Gans-Debye model, the light scattering of dilute polymer solutions can be expressed as Equation A.1 [8].

$$\frac{Kc}{R(\theta)} = \frac{1}{M_W P(\theta)} + 2A_2 c \quad (\text{A.1})$$

In Equation A.1:

- $R(\theta)$ is the excess *Rayleigh ratio* of the solution. It is directly proportional to the intensity of scattered light I_s detected at the angle θ in excess of the light scattered by the solvent.
- c is the solute concentration (mg mL^{-1})
- M_W is the solute molar mass
- A_2 is a second virial coefficient taking inter-particle interactions into account
- K is a constant of value $\frac{4\pi^2(dn/dc)^2 n_0^2}{N_A \lambda_0^4}$, where n_0 is the solvent refractive index, dn/dc is the specific refractive index increment of the solution¹, N_A is Avogadro's number and λ_0^4 is the vacuum wavelength of the laser
- $P(\theta)$ is an infinite degree polynomial that describes the angular dependence of the scattered light

Knowing that $R(\theta)$ is proportional to the intensity of scattered light I_s , substituting K with $\frac{4\pi^2(dn/dc)^2 n_0^2}{N_A \lambda_0^4}$ and introducing K_{LS} as a calibration constant for the LS detector taking n_0 , λ_0^4 and scattering from the solvent into account, Equation A.2 describes the intensity of scattered light in the absence of inter-particle interactions (the virial coefficient term is negligible at the working concentrations of SEC).

$$I_s(\theta) = P(\theta) K_{LS} c (dn/dc)^2 M_W \quad (\text{A.2})$$

¹This parameter quantifies the change of the solution refractive index n with increasing molecular concentration.

The expansion of $P(\theta)$ to first order gives Equation A.3. It depends on the mean square radius $\langle r_g^2 \rangle$ (or radius of gyration R_g which is the square root of $\langle r_g^2 \rangle$) developed in Equation A.4, a parameter reflecting the size of the macromolecule.

$$P(\theta) \approx 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2} \langle r_g^2 \rangle \sin^2 \frac{\theta}{2} + O\left(\sin^4 \frac{\theta}{2}\right) - \dots \quad (\text{A.3})$$

$$\langle r_g^2 \rangle = \frac{1}{M} \sum_i r_i^2 m_i \quad (\text{A.4})$$

where r_i is the distance of element m_i from the center of mass of the molecule of total mass M .

Whether or not the scattered light shows angular dependence will depend on the difference between the size of the particle and the wavelength of the incident light.

In the case of large macromolecules ($R_g > 15$ nm), destructive and constructive interference appear at the detector because light scattered from one part of the molecule travels a different distance from another part of the molecule and therefore reach the detector with different phases. The scattering is said to be anisotropic and I_s thus varies with the angle θ . For such molecules, R_g can be determined from the Zimm plot: plotting $Kc/R(\theta)$ vs $\sin^2(\theta/2)$ yields a curve whose intercept gives M_W and whose slope at low angles gives $\langle r_g^2 \rangle$.

On the other hand, molecules whose R_g is smaller than 15 nm are much smaller than the wavelength of the incident light and can be treated as point scatterers, scattering light isotropically. In this case $P(\theta) \approx 1$ since $\langle r_g^2 \rangle$ dependent terms of the expansion are canceled out. The angular dependency of I_s therefore vanishes and only the M_W of the scatterer can be determined. Equation A.2 becomes Equation A.5.

$$I_s = K_{LS} c (dn/dc)^2 M_W \quad (\text{A.5})$$

Equation A.5 shows how the molar mass of a macromolecule in solution is obtained from the measurement of the intensity of the light it scatters. K_{LS} includes all constants and calibration parameters, while the concentration c of the analyte is

measured by the UV detector based on its extinction coefficient and the specific refractive index increment dn/dc can be measured by a differential refractometer or approximated by the mean value of 0.185 mL g^{-1} for proteins.

Small-Angle X-Ray Scattering (SAXS)

Data collection - Small-Angle X-Ray Scattering (SAXS) is a low resolution structural characterization technique that can be applied to macromolecules in solution. Like MALS, SAXS exploits the light scattering properties of the molecules and provides the molar mass M_W and radius of gyration R_g of the scatterers. However, SAXS accesses a level of structural information beyond the simple size of the molecule because X-rays are used as the light source. As the wavelength used for the measurement is close to interatomic distances ($\lambda = 1.033 \text{ \AA}$), interference phenomena between the scattered rays occur and the intensity scattered by the molecule shows angular dependency. The variation of scattered intensity with respect to the angle of measurement carries information about the shape of the scattering object in solution.

In this work, we performed SEC-SAXS experiments so that homogeneous populations could be analyzed. In our case, we know that *MtSerB2* exists mainly as dimers and tetramers in solution. It is therefore essential to separate these two populations before the SAXS analysis so that the observed scattered intensity comes from a single species. In addition to separating analytes from other populations or protein aggregates, SEC-SAXS also allows accurate subtraction of solvent scattering, which is important for the same reasons as previously mentioned.

A typical SEC-SAXS setup is represented in Figure A.3. The sample is first separated in the SEC column, then passes through a UV detector monitoring elution and finally enters the SAXS flow-cell on which a X-ray beam is focused. The analyte scatters the incident X-rays and the scattered intensity is measured as a function of the angle between the incident and scattered beam (2θ) on a two-dimensional detector placed at a fixed distance from the SAXS cell. Since particles are randomly oriented in solution, the observed scattering is radially isotropic. To obtain a 1D SAXS profile from the scattering image on the detector, the scattered intensities are radially integrated and plotted against the magnitude of the momentum transfer vector $q = 4\pi \sin \theta / \lambda$.

As pictured in Figure A.4, such $I(q)$ vs q scattering profiles are acquired at regular intervals throughout the elution (each second for example). In this context, they are

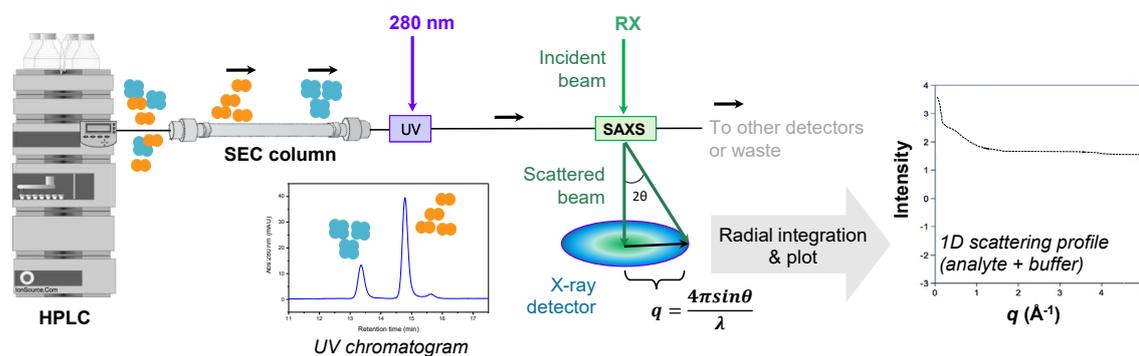


FIGURE A.3: SEC-SAXS setup and data collection.

called frames. Therefore, there are frames that correspond to the scattering of solvent and frames that correspond to the scattering of the analyte in solution. To produce the final SAXS curve corresponding to the scattering of the monodisperse analyte alone, the frames corresponding to the solvent are averaged, and the average frame is then subtracted from the average of the frames corresponding to the analyte.² It is from this final scattering profile that all the evaluation and analysis of the data can be carried out.

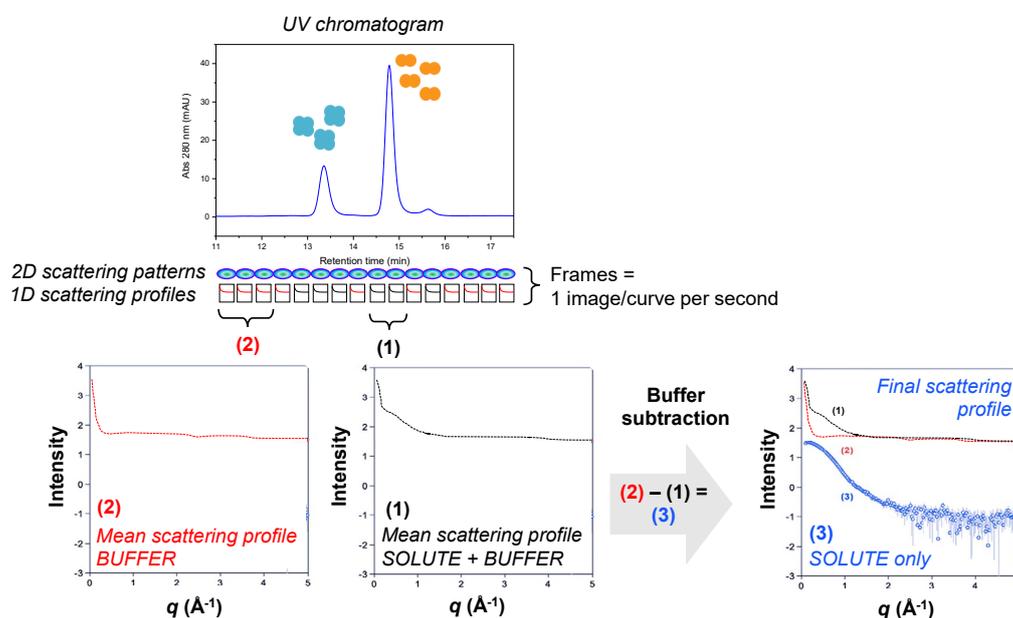


FIGURE A.4: Buffer subtraction principle.

²To each frame corresponds a value of R_g (*vide infra*) which allows to recognize the frames belonging to the same species. Frames for which the standard deviation of the R_g value is lower than a user-defined value are averaged.

Data analysis - We describe here the strategy that was followed to evaluate and process the SAXS profiles obtained for the proteins studied in this thesis. All the steps were performed using the SAXS analysis software *BioXTAS RAW* including the *ATSAS* suite, specifically designed for biological SAXS data.

The first step in the analysis of a SAXS profile is the **Guinier analysis**. It allows to **determine the parameters R_g (overall size) and $I(0)$ (the intensity at zero scattering angle, depending on M_W times the concentration)** for the analyzed macromolecule but is also of great help in the **evaluation of data quality**.

According to Guinier's approximation, at low q values, a scattering profile can be approximated by Equation A.6.

$$I(q) \approx I(0) \exp^{-q^2 R_g^2/3} \quad (\text{A.6})$$

A plot of $\ln(I)$ vs q^2 called the *Guinier plot* (Figure A.5, left) can be used to determine R_g (from the slope) and $I(0)$ (from the y-intercept). This plot will be linear only up to a q value such that $qR_g \sim 1.3$ for globular scatterers and $qR_g \sim 1.0$ for extended scatterers. The upper limit of the fit therefore already gives an indication of the overall shape of the sample.

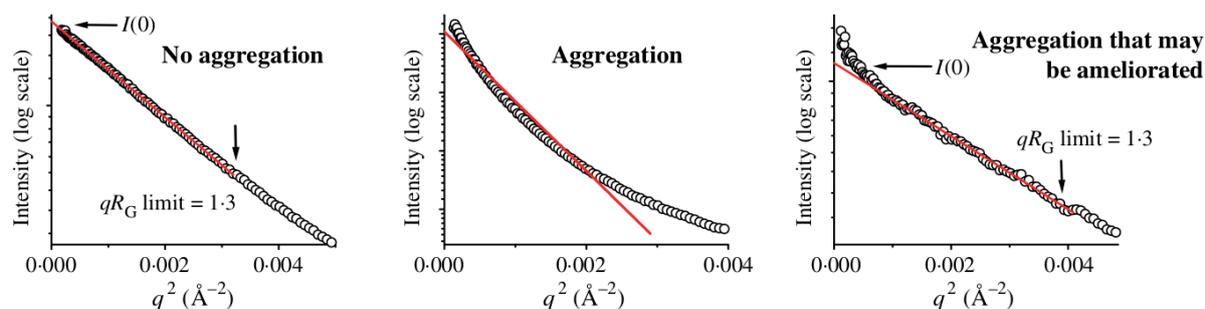


FIGURE A.5: Guinier fits of samples showing no (left), large (middle) and small (right) amounts of aggregates. Figure is from [4].

Being able to perform a Guinier fit meeting these constraints is a good sign that the data is from a monodisperse sample with no interparticle interactions. On the other hand, a non-linear Guinier plot is indicative of poor sample quality. Samples containing non-specific aggregates or changing oligomerization states will yield Guinier plots showing a characteristic upturn in intensity at low q values (Figure A.5, middle and right), while inter-particle repulsion will show as a downturn. A faulty buffer

subtraction will also cause such effects. Trying to fit such curves lead to distorted estimations of R_g and I_0 and data showing these deviations should not be treated further.

The best thing to do in this case is to collect the data again after improving the sample conditions. As inter-particle interactions are often linked to the sample concentration, scattering profiles can be collected at several concentrations to allow the identification of a sample concentration where interactions are negligible. Based on the variation of the contribution of these interactions to the scattering intensity as a function of concentration, it is also possible to extrapolate the data to an *infinite* dilution to obtain an ideal curve for further analysis.

The problem with data obtained from diluted samples is that the curve is very noisy at high q values, which is not ideal for accurate further characterization of the solution structure. To obtain a high-quality scattering curve for the entire q -range, the scattering profile at low q from a diluted sample showing no inter-particle interactions can be merged with the more accurate scattering profile at high q obtained at the highest measured concentration.

When R_g and $I(0)$ have been defined on the basis of good quality data, the **molecular weight** M_W can then be determined. In this work, we used two concentration independent determination methods: (1) from Porod volume and (2) from Bayesian inference.

The **Porod volume** V_p (\AA^3), or the hydrated particle volume of the macromolecule in solution, can be directly calculated from the scattering profile by calculating the Porod invariant Q_p (Equation A.7) and injecting it in Equation A.8. The M_W is then calculated by multiplying V_p by the macromolecule's density ($\text{kDa}/\text{\AA}^3$). BioXTAS RAW uses an average protein density of $0.00083 \text{ kDa}/\text{\AA}^3$ but a known value for the macromolecule can be inputted.

$$Q_p = \int_0^\infty q^2 I(q) dq \quad (\text{A.7})$$

$$V_p = \frac{2\pi^2 I(0)}{Q_p} \quad (\text{A.8})$$

This method is relatively accurate yielding M_W estimates showing about 10% uncertainty but may need that the protein density is adjusted in some cases. It gives less accurate estimates for flexible and elongated proteins.

The **Bayesian inference approach** is a probabilistic method for calculating the molecular weight that was developed by Svergun and coworkers [9]. Essentially, the authors estimated the M_W associated to 165 982 scattering profiles calculated from high-resolution protein structures using four different concentration-independent M_W determination methods. From this large test dataset and for each determination method, they created a probability distribution describing the probability of obtaining a particular molecular weight given the true molecular weight. The probabilities were then combined across all methods in a final Bayes formula that enables the estimation of the most likely M_W for the analyzed macromolecule.

The shape and folding of the macromolecule can be assessed by a **Kratky analysis**. As shown in Figure A.6, a visual inspection of the Kratky plot $q^2 I(q)$ vs q can help distinguish whether the protein is rather compact and globular (bell-shaped Gaussian peak), unfolded/highly flexible (plateau at high q) or partially unfolded/flexible (combination). In this thesis we use the dimensionless Kratky plot $(qR_g)^2 I(q)/I(0)$ vs qR_g , where the intensity $I(q)$ is normalized to the forward scattering intensity $I(0)$ and q is normalized to R_g . This makes the analysis independent of the protein's size and M_W and allows for the direct comparison of Kratky plots from different proteins. Moreover, whatever the size of the protein, the dimensionless Kratky plot exhibits a maximum value of 1.104 at $qR_g = \sqrt{3}$ for a compact globular protein or a curve rising with increasing angle for random chains. For partially disordered proteins, such as flexible multi-domain protein for example, the profile will be comprised between the two extremes.

Another method to know more about the shape of the analyte is to calculate **the pair distance distribution function** $P(r)$, a real space representation of the scattering data. As the SAXS profile is measured in reciprocal space, a Fourier transform is applied to the data to yield the $P(r)$ function (Equation A.9).

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin(qr)}{qr} dq \quad (\text{A.9})$$

The $P(r)$ is essentially a r^2 weighted histogram of all possible pairs of electrons in

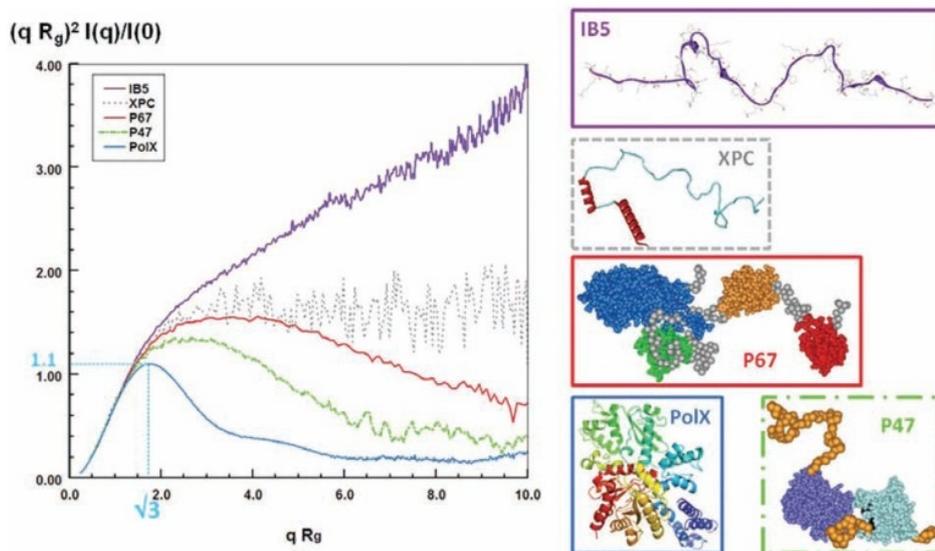


FIGURE A.6: Dimensionless Kratky plots of proteins showing various folding degrees. Figure is from reference [10].

the sample, graphically displaying the traits of the particle shape (globular, rod-like, multi-domains,...) as illustrated in Figure A.7.

However, given the finite extent of the measurement and experimental errors, the direct Fourier transform of Equation A.9 cannot be used to calculate the $P(r)$ function from the scattering data. The problem is then solved the other way round by trying to find the $P(r)$ function that will best fit the experimental scattering data. The GNOM program, incorporated in BioXTASRaw, determines this by calculating the scattering intensity for trial $P(r)$ functions with Equation A.10.

$$I(q)_{calc} = 4\pi \int_0^{D_{max}} P(r) \frac{\sin(qr)}{qr} dr \quad (\text{A.10})$$

The fitting criteria include the goodness of fit between the experimental data and that calculated by the direct transform of the $P(r)$ function estimated by a χ^2 value (Equation A.11) and regularization parameters such as the smoothness and the positivity of the $P(r)$ function and whether the solution changes significantly when changing the weighting (by a parameter α) of these parameters.

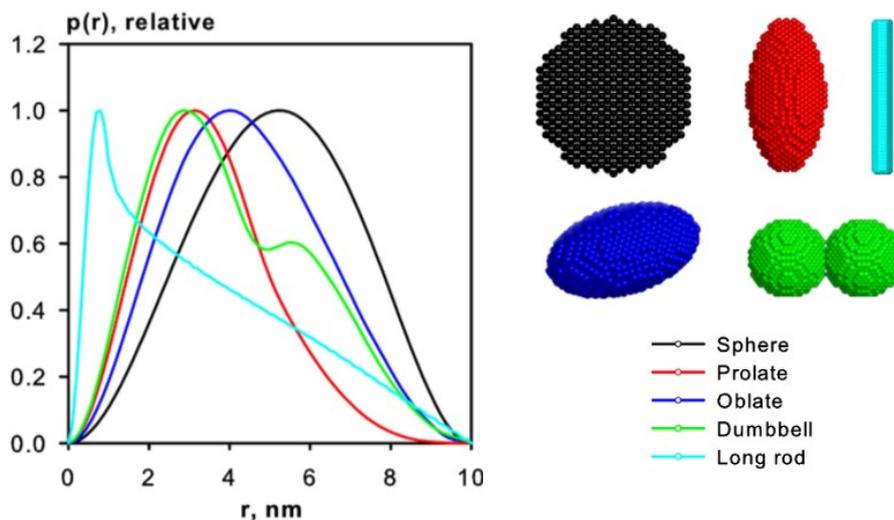


FIGURE A.7: $P(r)$ functions calculated for typical geometric shapes. Figure is adapted from reference [5].

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I(q_j)_{exp} - I(q_j)_{calc}}{\sigma(q_j)} \right]^2 \quad (\text{A.11})$$

where N is the number of data points and σ is the standard deviation.

To determine $P(r)$ in this way thus requires determining the **maximum dimension of the sample** D_{max} (upper bound of the integral in Equation A.10.) This is an iterative problem: the optimal D_{max} value will be the one yielding the best $P(r)$ function i.e. gradually falling to zero at D_{max} , fitting the measured scattering profile and being zero at $r = 0$ and $r \geq D_{max}$. BioXTAS and GNOM fortunately allow the user to easily identify a reliable solution through a convenient graphical user interface that enables to see the effect of the modification of the D_{max} and the α parameter on the shape of the $P(r)$ function. D_{max} is usually determined with a deviation of about 5 to 10% from the real value.

R_g and $I(0)$ can be calculated from the determined $P(r)$ function and provide a complementary approach to the Guinier fit. If the values do not agree well between both methods, it may suggest either a faulty Guinier fit or $P(r)$ function, except in the case of flexible and disordered systems for which larger (and more reliable) values are obtained from the $P(r)$ function.

Characterization of the solution structure - In addition to the shape parameters described above, it is also possible to obtain more information about the three-dimensional structure of the target macromolecule from its scattering profile.

As in this work, the system under study sometimes allows the creation of plausible atomic models from known high-resolution structures. In such cases, the most direct characterization method is to **compute the predicted scattering curves from these atomic models and compare them to the experimental data**. This can be achieved using the program CRY SOL, which approximates the scattering profile of the molecule and its hydration shell using spherical harmonics and fits it to the experimental data by minimizing the discrepancy (χ^2 , see Equation A.11) between them. The fit is done by varying the average displaced solvent volume per atomic group, the contrast of the hydration shell and relative background. Readers interested in more detailed explanations can find the mathematical formalism in the [11]. A model that provides a good fit to the data is considered a possible valid description of the solution structure among others.

Models, like high-resolution crystallographic structures, may be close to but not accurately represent the conformation that the macromolecule adopts in solution. This results in slight discrepancies between the calculated and experimental SAXS profiles from q values around 0.15 \AA^{-1} . For such cases, programs like DADIMODO and SREFLEX can be used to refine inaccurate atomic models against experimental SAXS data and provide a model that better describes the protein in solution. Both algorithms explore the conformational space of the atomic model, SREFLEX using normal mode analysis and DADIMODO through a stochastic selection process of conformational mutations introduced in user-defined flexible parts, and refine the structure guided by the agreement with the experimental SAXS data.

Rigid body modeling is another method that can be used when the structure of subunits or domains is known but not their relative positions and orientations in space. Essentially, rigid body modeling algorithms also employ CRY SOL to calculate predicted scattering profiles and are based on an iterative approach: starting from an arbitrary positioning of subunits, the algorithm (e.g. SASREF) performs a series of random rigid body movements and rotations, using a simulated annealing protocol to search for the configuration whose computed scattering profile gives the best fit to the experimental data. Constraints are taken into account to make sure that the models

are interconnected and show no main-chain clashes. In some cases, the structures of linkers or even entire domains of the macromolecule to model are unknown. To remedy this, some algorithms like BUNCH or CORAL (used in this work) combine rigid body modeling with *ab initio* methods (*vide infra*), where missing regions are modeled using dummy residues. The optimization process then takes into account the determination of the optimal local conformation of these dummy residues.

When no structural information is available, *ab initio*³ modeling can provide a low resolution shape reconstruction of the sample. There are different algorithms (DAMMIN/F, GASBOR, SASMODEL,...), based on different formalisms and shape representations, but all of them apply the following general approach: (1) proposition of shapes (continuous objects) (2) calculation of the scattering curves or $P(r)$ functions (3) optimization of the agreement to the experimental data. The limitation of *ab initio* modeling is that the reconstruction of a 3D shape from a 1D scattering pattern can yield multiple solutions. To one scattering profile can indeed correspond several possible shapes and this means that different models are obtained when performing independent runs. The common practice to reduce ambiguity is therefore to generate several models, cluster them according to their structural similarity, identify the most representative cluster and average its members into a smoothed model. The model generated in this way represents a low-resolution shape (about 20 to 50 Å) having the same scattering profile as the macromolecule in solution. It is best treated as an hypothesis to be tested by additional experimentation. In this work, since we had information on the folds of the domains and information provided by *ab initio* modeling is rather ambiguous, we preferred the two approaches mentioned above.

Bibliography

- [1] Wyatt Technology Corporation. Understanding multi-angle static light scattering. <https://www.wyatt.com/library/theory/multi-angle-light-scattering-theory.html>, Accessed online July 6, 2022.
- [2] Agilent Technologies. *Light Scattering Operation Guidelines Manual, Rev 5.0*. 2016.
- [3] E. Foltá-Stogniew. Light scattering theory. <https://medicine.yale.edu>, Accessed online July 6, 2022.

³Here, *ab initio* refers to the lack of a pre-defined input structure.

- [4] C.D. Putnam, M. Hammel, G.L. Hura, and J.A. Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, 40(3):191–285, 2007.
- [5] H.D.T. Mertens and D.I. Svergun. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1):128–141, 2010.
- [6] J.B. Hopkins. BioXTAS RAW documentation. <https://bioxtas-raw.readthedocs.io/>, Accessed online July 8, 2022.
- [7] G. Young, N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E.G. Marklund, M.P. Collier, et al. Quantitative mass imaging of single biological macromolecules. *Science*, 360(6387):423–427, 2018.
- [8] P.J. Wyatt. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta*, 272(1):1–40, 1993.
- [9] N.R. Hajizadeh, D. Franke, C.M. Jeffries, and D.I. Svergun. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Scientific Reports*, 8(1):1–13, 2018.
- [10] V. Receveur-Bréchet and D. Durand. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current Protein and Peptide Science*, 13(1):55–75, 2012.
- [11] D. Svergun, C. Barberato, and M.H.J. Koch. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, 28(6):768–773, 1995.

Appendix B

Experimental procedures

B.1 Protein production

B.1.1 Plasmids

Wild-type enzymes

Plasmids encoding bacterial enzymes (*MtSerB2*, *MaSerB*, *MmSerB2* and *BmSerB*) are based on the AVA0421 vector (derived from pET14b and described in ref. [1]). They carry ampicillin resistance and a gene coding for proteins bearing a N-terminal hexahistidine tag (His₆) cleavable by human rhinovirus 3C protease (HRV 3C) that is under control of the T7 RNA polymerase promoter. They were kindly provided by the Seattle Structural Genomics Center for Infectious Disease (www.SSGCID.org) which is supported Federal Contract No. HHSN272201700059C from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services.

His₆-HsPSP is encoded by a pET28a plasmid that was kindly provided by collaborators from UCLouvain for a previous project.

Mutant enzymes

Primers are detailed in Table B.1. It was ensured that each mutant had exactly the same plasmid construct as the wild-type enzymes via a specific primers design, i.e.

the open reading frame is located at exactly the same position in each plasmid.

MtSerB2*Δ*ACT1 - Plasmid pAVA0421 encoding His₆-*MtSerB2*Δ*ACT1* was produced following the one-step site directed mutagenesis strategy of Liu *et al.* [2]. The polymerase chain reaction (PCR) of 50 μL contained 10 μL 5X Q5 reaction buffer (*NEB*), 22 μL milli-Q water, 10 μL Q5 High GC enhancer (*NEB*), 50 ng template pAVA0421-*MtSerB2*, 0.5 μM forward primer, 0.5 μM reverse primer, 400 μM dNTPs and 0.5 μL Q5 High-fidelity DNA polymerase (*NEB*). The PCR cycles were initiated at 98 °C for 5 minutes to denature template DNA, followed by 35 amplification cycles. Each cycle consisted of 98 °C for 30 seconds, 65 °C (T_m) for 30 seconds and 72 °C for 4 minutes. The PCR cycles were finished with an extension step at 72 °C for 15 minutes. A volume of 3.5 μL of PCR reaction was run on a 1% agarose gel stained with ethidium bromide to verify that the reaction worked. The PCR product was subsequently treated with 1 μL *DpnI* and 10 μL CutSmart buffer (*NEB*) at 37 °C for 1.5 hours. A volume of 25 μL of PCR product was transformed into 50 μL *E. coli* TOP10 competent cells by heat shock at 42 °C for 1 minute. The transformed cells were grown in 1 mL liquid lysogeny broth (LB) at 37 °C for 45 minutes, spread at 1X and 10X dilutions on LB agar plates containing 100 mg L⁻¹ ampicillin and incubated at 37 °C overnight. To check for deletion, four colonies from each plate were picked with a toothpick, saved on another LB plate and mixed in a total 20 μL volume containing 6 μL 5X GoTaq Green buffer (*Promega*), 19.4 μL 0.5 μM forward check primer (AM-F), 0.5 μM reverse check primer (AV-R), 400 μM dNTPs and 0.2 μL GoTaq DNA polymerase (*Promega*). The PCR cycles were initiated at 94 °C for 10 minutes to denature DNA, followed by 30 amplification cycles. Each cycle consisted of 94 °C for 30 seconds, 64 °C (T_m) for 30 seconds and 72 °C for 2 minutes. The PCR cycles were finished with an extension step at 72 °C for 10 minutes. Volumes of 3.5 μL of the PCR reactions were run on a 1% agarose gel stained with ethidium bromide to check for deletion. Colonies actually showing the deletion were grown overnight in liquid LB at 37 °C and the plasmid DNA was isolated using GeneJET Plasmid Miniprep kit (*ThermoFischer*) according the manufacturer's instructions. For the final check, 1.2 μg plasmid DNA was mixed with 20 nmol of sequencing primer (Q29R-AM-F and T7-term, in separated tubes) in a total volume of 17 μL and the DNA sequencing was carried out using the sequencing service Mix2Seq of *Eurofin Genomics*.

MtSerB2 Q92E - Plasmid pAVA0421 encoding His₆-MtSerB2SerB2 Q92E was also produced following the one-step site directed mutagenesis strategy of Liu *et al.* [2]. The production protocol is identical to that described above for pAVA0421-MtSerB2ΔACT1. No band was observed on the agarose gel when checking the reaction, but the entire PCR reaction volume (50 μL) was still transformed into *E. coli* TOP10 competent cells. Only one colony was obtained but sequencing confirmed that the mutation had been successful.

TABLE B.1: Primers used for mutagenesis, check and sequencing. The primers were synthesized by Integrated DNA Technologies IDT.

MtSerB2ΔACT1 forward	5' TCC TGG TTC GTC CAC CCA CAC CAT TTT CG 3'
MtSerB2ΔACT1 reverse	5' TGT GGG TGG ACG AAC CAG GAC CCT GGG T 3'
MtSerB2 Q92E forward	5' ATT CGG GAA CCG TCC ACC CAC ACC ATT TTC GTG CTG 3'
MtSerB2 Q92E reverse	5' GGA CGG TTC CCG AAT GAT TGG CAG ATC GTC GCT GC 3'
T7-term	5' CTA GTT ATT GCT CAG CGG T 3'
AM-F	5' AAA GAT CTC GAT CCC GCG AAA TTA ATA CG 3'
AV-R	5' TTG GAT CCT CGA GAA GCT TGG CT 3'
Q29R-AM-F	5' ATC GGT GAT GTC GGC GAT ATA G 3'

MtSerB2 Q92E C148T V149Y G150R I154T (MtSerB2 α) - Plasmid pAVA0421 encoding His₆-MtSerB2SerB2 α was produced by GenScript.

MtSerB2 E33A R103A T136A (MtSerB2 3A) - Plasmid pAVA0421 encoding His₆-MtSerB2SerB2 3A was produced by GenScript.

B.1.2 Protein overexpression

Bacterial enzymes and mutants were all produced using the following protocol. The pAVA0421 plasmid was transformed into 50 μL *E. coli* BL21 DE3 pLysS competent cells by heat shock at 42 °C for 1 minute. The transformed cells were grown in 1 mL liquid lysogeny broth (LB) at 37 °C for 45 minutes, spread at 1X and 10X dilutions on LB agar plates containing 100 mg L⁻¹ ampicillin and 34 mg L⁻¹ cloramphenicol and incubated at 37 °C overnight. Colonies were grown overnight in 10 mL liquid LB at 37 °C. The transformed *E. coli* BL21 DE3 pLysS cells were aliquoted in 20% glycerol for storage at -80 °C.

A volume of 50 μL of transformed cells (glycerol stock) was grown overnight at 37 °C in 10 mL liquid LB containing 100 mg L^{-1} ampicillin and 34 mg L^{-1} cloramphenicol. A larger volume of liquid LB containing both antibiotics was then inoculated with the starter culture (1:35 volumic ratio) and the cells were further grown at 37 °C until reaching an optical density of 0.6-0.8 at 600 nm. The cells were put to rest at 4 °C until induction. Protein expression was induced at 20 °C by the addition of 0.2 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). The cells were grown for 18 hours and harvested by centrifugation (30 min, 4 °C, 4000 rpm in a JA-14 *Beckman Coulter* fixed-angle rotor). The pellet was either used directly or stored at -20 °C.

B.1.3 Protein purification

The pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mg mL^{-1} lysozyme, supplemented with cOmplete EDTA-free protease-inhibitor cocktail from *Roche*) and disrupted by sonication over ice (6 cycles of 30 seconds at 20 W with 30 seconds of rest on ice in between). The soluble protein fraction was recovered by centrifugation (1 hour, 4 °C, 12500 rpm in a JA-20 *Beckman Coulter* fixed-angle rotor) and loaded onto a 5 mL HisTrap FF crude column (*Cytiva*) connected to an AKTA Purifier 10 FPLC system. Unbound protein was washed away with buffer A (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 30 mM imidazole). Bound protein was then eluted in buffer B (50 mM Tris-HCl pH 7.4, 300 mM NaCl, 200 mM imidazole). Fractions containing the eluted protein were pooled and exchanged for buffer A2 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP)). To remove the hexahistidine tag, His₆-HRV 3C protease was added to the protein solution (1 μg per 200 μg protein) and the mix was incubated overnight at 4 °C. The free hexahistidine tag was then separated from the cleaved protein by another IMAC using a 1 mL HisTrap FF crude column (*Cytiva*) and buffer A2. The cleaved protein was recovered in the flow through and concentrated up to about 1 mg mL^{-1} (or to higher concentrations depending on the downstream use). The concentration was determined by a measure of absorbance at 280 nm using the extinction coefficient of the reduced form of the protein as calculated on ProtParam (<https://web.expasy.org/protparam/>) based on the protein primary sequence preceded by the four residues GPGS remaining after cleavage. The purity of the protein was assessed by SDS-PAGE using a 12% polyacrylamide gel and Coomassie Brilliant Blue R staining. The protein solution was finally aliquoted and flash-frozen in liquid nitrogen for storage at -80 °C.

B.2 Observation and identification of oligomeric states

B.2.1 Native polyacrylamide gel electrophoresis (Native-PAGE)

Native-PAGE analyses were conducted on 10% polyacrylamide gels prepared by mixing 4.170 mL distilled water, 2.5 mL of 1.5 M Tris-HCl pH 8.8 buffer, 3.330 mL acrylamide mix 30%, and 50 μ L 10% ammonium persulfate (APS) and 10 μ L N,N,N',N'-tetramethylethylenediamine (TEMED) to catalyze polymerization. Usually, the samples to load consisted of 5 μ g of protein mixed with 5 μ L sample buffer (0.12 M Tris-HCl pH 6.8, 0.008% bromophenol blue, 30% glycerol) completed with distilled water to a final volume of 20 μ L. Samples were migrated at 110 V and room temperature for 80 minutes. The gels were stained with Coomassie Brilliant Blue R (0.25% Coomassie Brilliant Blue R, 45% vol MeOH, 10% acetic acid).

B.2.2 Mass photometry (MP)

The landing of the protein species present in *MtSerB2* sample was recorded using a Refeyn Two MP instrument (*Refeyn Ltd*) by adding 1 μ L of a diluted protein stock solution (230 nM) directly into a 19 μ L drop of filtered buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl). Movie acquisition was performed during 60 s with the DiscoverMP software (version 2022 R1, *Refeyn Ltd*) and data were analyzed using the default settings. Prior to the experiments was performed a contrast-to-mass calibration using a solution of protein standards with molecular weights of 66, 146, 480 and 1048 kDa.

B.2.3 Size exclusion chromatography (SEC)

Semi-preparative SEC

Semi-preparative SEC was conducted at the CBS laboratory using a Superdex 200 10/300 GL (*GE Healthcare*) column, or a more recent Superdex 200 Increase 10/300 GL (*Cytiva*) column for the analysis of *MtSerB2* mutants α and 3A. The columns were connected to an AKTA Purifier 10 FPLC system. The data were recorded and processed using Unicorn 5.11 software (*GE Healthcare*).

Column calibration - The Superdex 200 10/300 GL column was calibrated using the Gel Filtration Markers Kit for Protein Molecular Weights 29000-700000 Da (ref. MWGF1000, *Merck*). Standards were sampled as indicated in Table B.2, each diluted

in 60 μL of buffer A2 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP) and combined in a final volume of 500 μL to inject onto the column. A volume of 250 μL of the standard mix (full loop injection) was analyzed in buffer A2 at a flow rate of 0.2 mL min^{-1} . Column dead volume (7.54 mL) was checked by injecting 250 μL of a 2 mg mL^{-1} blue dextran solution. The elution volume at peak maximum was plotted against the logarithm of the molar mass to construct the calibration curve.

TABLE B.2: Sampling of the standards used for Superdex 200 10/300 GL column calibration.

Standard	Molar mass (kDa)	Weighing (mg)	Sampling (μL)
Albumin	66	6.0	-
Apoferritin	443	-	200
Thyroglobulin	669	5.3	-
Alcohol dehydrogenase	150	3.0	-
β -amylase	200	2.1	-
Carbonic anhydrase	29	3.0	-

Analyses - Depending on the experiment, volumes of 100 or 250 μL of protein sample (0.75-3.75 mg, pre-incubated with the desired L-Ser concentration where applicable) were injected onto the column and analyzed in buffer A2 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP) or buffer A2Ser (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP, 10 mM L-Ser) at a flow rate of 0.2 mL min^{-1} (Superdex 200 10/300 GL) or 0.5 mL min^{-1} (Superdex 200 Increase 10/300 GL).

Preparative SEC

Preparative SEC for crystallization assays was conducted at the CBS laboratory using a HiLoad 16/600 Superdex 200 pg (*Cytiva*) column connected to an AKTA Purifier 10 FPLC system. A volume of 2.5 mL of MtSerB2 sample (25 mg) was injected onto the column and separation of MtSerB2 dimer and tetramer was conducted in buffer A2 at a flow rate of 0.5 mL min^{-1} . Fractions containing the same species were pooled and concentrated by centrifugation up to about 25 mg mL^{-1} .

Analytical SEC

See SEC-MALS in section B.2.4.

B.2.4 SEC coupled to multi-angle light scattering (SEC-MALS)

SEC-MALS experiments were conducted at the LAM laboratory (CIRM, ULiège) using a BioResolveSEC mAb 200 Å 2.5 µm 7.8 x 300 mm column (*Waters*) preceded by a BioResolve Sec mAb 200 Å 2.5 µm 4.6 x 30 mm precolumn (*Waters*) mounted on a LC 1260 Infinity II Bio-Inert (*Agilent*) system connected in-line to a 1260 Infinity II Bio-SEC Multi-Detector system equipped with a MDS LS dual angle (15° / 90°) light scattering detector and a MDS DLS dynamic light scattering detector (90°). The detectors were normalized to 5 mg mL⁻¹ BSA as recommended per the manufacturer ($M_W=66463 \text{ g mol}^{-1}$, $\epsilon=0.670 \text{ mL mg}^{-1} \text{ cm}^{-1}$, $dn/dc=0.670 \text{ mL g}^{-1}$). The data were recorded and processed using Bio-SEC software (*Agilent*).

Column calibration - The system (column + pre-column) was calibrated using BEH200 SEC Protein Standard Mix (ref. 186006518, *Waters*) containing thyroglobulin, IgG, BSA, myoglobin and uracil. Upon reception, the BEH200 vial was solubilized in SEC buffer (50 mM Tris-HCl pH 7.4 150 mM NaCl). The mix to inject consisted of 25 µL of the vial contents diluted in 25 µL of SEC buffer. A volume of 10 µL of this diluted mix were injected onto the column and analyzed at a flow rate of 0.5 mL min⁻¹. The retention time at peak maximum was plotted against the logarithm of the molar mass to construct the calibration curve.

Analyses - Depending on the experiment, volumes of 10 or 30 µL of protein sample (10-50 µg, pre-incubated with the desired L-Ser concentration where applicable) were injected onto the column and analyzed at a flow rate of 0.5 mL min⁻¹ in SEC buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl) or SEC buffer containing L-Ser. The desired concentration of L-Ser in the mobile phase was obtained through the HPLC's mixing system by combining SEC buffers with (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP, 2.5 or 25 mM L-Ser) and without L-Ser. The weight averaged molar mass was determined for each protein species eluting as a monodisperse peak at 280 nm, using a dn/dc ratio of 0.183 mL g⁻¹ for *MtSerB2*¹ or a default value of 0.185 mL g⁻¹ for the other enzymes.

¹As determined by Wyatt Technology Corporation in February 2021.

B.3 Structural characterization

B.3.1 *In silico* modeling of oligomeric species

Homology modeling of *MtSerB2* and *MmSerB2* dimers

MtSerB2 and *MmSerB2* dimers were modeled using SWISS-MODEL (<https://swissmodel.expasy.org/interactive>) [3]. On the basis of only the primary sequences, the algorithm proposed a homo-dimer model based on *MaSerB* 3P96 structure for *MtSerB2* and a homo-dimer model based on *MaSerB* 5JJB structure for *MmSerB2*. These two models were used for representation of the enzymes and the following modeling steps.

In silico modeling of *MtSerB2* theoretical closed monomer using Molecular Dynamics (MD)

MtSerB2 theoretical closed monomer was modeled starting from *MtSerB2* dimer homology model described in the previous paragraph. The corresponding PDB file was modified using PyMOL to keep only the coordinates of the ACT2-PSP part (HIS96-ASP398) of chain A and the ACT1-hinge part (ALA3-THR95) of chain B. The hinge-loop of chain B was then reconnected to chain A using the 3D builder function of Maestro 11.9.011 software (*Schrodinger*) and minimization through MacroModel with OPLS3e as the force field and constraining the distance between THR95 and HIS96. This manipulation allowed the generation of a new PDB file with all residues belonging to the same chain that was used as the starting point for MD simulation.

The MD simulation was run using GROMACS 2020 [4] with CHARMM27 force field [5] and CMAP corrections for the protein. The protocol was based on that of Mirgaux *et al.* [6]. Hydrogen atoms were added using GROMACS and solvation was accounted for using all-atom TIP3P and coarse-grained SIRAH water particles [7–9]. A cubic box was built around the protein with at least 2.0 nm between the box edges and the protein atoms. TIP3P water molecules were placed in a 1.0 nm thick shell around the molecular system. Coarse-grained SIRAH water particles were then placed between this shell and the edges of the box. Sodium ions were randomly placed in the bulk of the SIRAH water particles to neutralize the total charge of the system. The optimization and MD trajectories were generated under the particle mesh Ewald periodic boundary conditions. A cutoff value of 1.2 nm was applied for Coulomb and

van der Waals interactions. Temperature and pressure were respectively fixed using the Parrinello–Rahman [10] and V-Rescale algorithms [11]. Covalent bonds involving H atoms were constrained using the LINCS algorithm [12]. The resulting system was optimized using the steepest-descent algorithm for a maximal number of 2500 steps with an initial step size of 0.05 nm. During the equilibration stage of the system, the temperature was progressively increased from 50 to 310 K using short MD runs. The first run consisted of a 10 ps simulation at 50 K on the system obtained after optimization. Afterwards, the system was relaxed for two runs of 20 ps at 150 and 310 K. Finally, a run of 50 ps at 310 K and 1 bar was performed to finalize the relaxation of the system. The equilibration was extended for 60 ns with a time step of 2 fs at 310 K and 1 bar. The production step was run for 200 ns (100×10^6 steps) with a time step of 2 fs. The evolution of the system during the equilibration and production stages was followed using energy and r.m.s.d. profiles. The structure after 200 ns of simulation was extracted and used for the next modeling steps.

***In silico* modeling of *MtSerB2* tetramer and trimer using symmetrical protein-protein docking**

MtSerB2 trimer and tetramer models were generated by protein-protein under symmetry constraints using M-ZDOCK (<https://zdock.umassmed.edu/m-zdock/>) [13], ClusPro (<https://cluspro.bu.edu/>) [14–17] and GalaxyTongDock (<https://galaxy.seoklab.org/>) [18]. Depending on the oligomer to be modeled, a symmetry of 2,3 or 4 was selected on M-ZDOCK, GalaxyTongDock-C with C2, C3 and C4 symmetries or GalaxyTongDock-D with D2 symmetry was used or ClusPro was run with 2 or 3 subunits in the multimer docking mode available in the advanced options. The input PDB files consisted of *MtSerB2* dimer homology model, *MtSerB2* theoretical closed monomer model described above, or the half of *MtSerB2* dimer without hinge loop (coordinates of chain A ACT2-PSP part (HIS96-ASP398) and chain B ACT1 part (ALA3-ARG83) of *MtSerB2* dimer homology model).

B.3.2 *In silico* study of L-Ser binding sites

Ligand binding site prediction

Prior to ligand binding site prediction, *MtSerB2* and *MmSerB2* dimer homology models as well as *MaSerB* structure 3P96 were prepared in the same way using the Protein Preparation Wizard of Maestro 12.9.137 software (*Schrodinger*). The simulation pH

was set to 7.4 and missing side chains were filled in. No restrained minimization was performed. Hydrogen atoms were subsequently deleted. The prepared structures were submitted to PrankWeb server (<https://prankweb.cz/>) [19, 20] and the prediction was run without the use of conservation.

Induced-fit docking of L-Ser in *MtSerB2* dimer model

Induced-fit docking of L-Ser was performed in *MtSerB2* dimer homology model described in section B.3.1. The model was first prepared and minimized using the Protein Preparation Wizard of Maestro 11.9.011 software (*Schrodinger*). The protonation state of the residues was adjusted using Epik at pH 7.4 and the global structure was refined with the OPLS3e force field. L-Ser structure was prepared at pH 7.4 (Epik) by minimization with OPLS3e force field using LigPrep. The prepared structures were then entered into the Induced Fit Docking protocol. The receptor box center was defined as the centroid of residues R103, P104, D132, T136 of molecule 2 (chain B) and A3, E33, L34, L35, S53, I89 of molecule 1 (chain A) based on PrankWeb prediction. The box size was set to dock ligands similar in size to L-Ser. Residues within 5.0 Å of ligand poses were refined with an optimization of the side chains.

B.3.3 SEC coupled to small-angle X-ray scattering (SEC-SAXS) and SAXS-based model evaluation/modelisation

SEC-SAXS data collection

SEC-SAXS experiments were carried out on the SWING beamline at SOLEIL Synchrotron (Saint-Aubin, France). The X-ray wavelength (λ) was set to 1.033 Å and the sample to detector (17x17 cm² low-noise Avix charge-coupled detector) distance was set to 2000 nm. Those parameters corresponded to a scattering wave-vector range of $0.0036 \text{ \AA}^{-1} < q < 0.5 \text{ \AA}^{-1}$, where $q = 4\pi \sin \theta / \lambda$ and 2θ is the scattering angle. The sample solutions were circulated in a thermostated quartz capillary with a diameter of 1.5 mm and a wall thickness of 10 µm inserted in a vacuum chamber.

All protein samples (9.3-14.5 mg mL⁻¹) were thawed at room temperature and centrifuged for 5 minutes at 6,000g before the SEC-SAXS experiment. For each analysis, 50 µL of sample (0.46-0.73 mg, pre-incubated with the desired L-Ser concentration where applicable) were injected onto a BioResolveSEC mAb 200 Å 2.5 µm 7.8 x 300 mm column (*Waters*) preceded by a BioResolve Sec mAb 200 Å 2.5 µm 4.6 x 30 mm

precolumn (*Waters*) pre-equilibrated with buffer A2 ou A2Ser (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP, 0 or 10 mM L-Ser). The column was mounted on an Agilent high-performance liquid chromatography (HPLC) system allowing for the elution of the samples at a controlled flow-rate of 0.5 mL min^{-1} and a temperature of 10°C . The elution was monitored at 280 nm by a UV-diode array detector installed just downstream of the column, before the SAXS flow cell where the sample was exposed to X-rays.

A total of 1140 scattering patterns were collected during the elution of the sample, with a frame duration of 1 second. The scattering signal of the buffer was collected in 180 frames before the void volume. To generate individual 1D curves, the frames were radially averaged, divided by the transmitted intensity and normalized to absolute units with water scattering as a reference using the image analysis software Foxtrot (courtesy of SWING beamline). The software was also used to generate plots corresponding to $I(0)$ and R_g as a function of frames. Curves from consecutive images corresponding to the analysis of a single protein species and showing similar R_g (± 0.2 or 0.3 \AA) were averaged and the same operation was performed for the buffer. The averaged buffer scattering curve was subtracted from the averaged sample scattering curve to generate the final SAXS curve to be analyzed.

SAXS data treatment

The final SAXS curves were treated and analyzed using BioXTAS RAW 2.1.1 software [21]. Guinier analysis was first performed for each final SAXS curve to detect signs of interparticle interaction. In the absence of such signs, $I(0)$ and R_g parameters were calculated from the Guinier plot and the curve was further analyzed by determining M_W , performing Kratky analysis and calculating the $P(r)$ function as described in reference [22].

Evaluation of the models generated by symmetric protein-protein docking and comparison of crystallographic structures with solution structures

Atomic models and crystallographic structures were evaluated by determining the discrepancy (χ^2) between their calculated SAXS curve and the experimental SAXS data using CRY SOL [23] in primus/qt ATSAS 3.0.4 software [24] with default settings (51 points, 15 spherical harmonics, order of Fibonacci grid: 17, solvent density: 0.33 e/\AA^3).

Refining of models and crystallographic structures based on SAXS data

DADIMODO - *MaSerB* crystallographic structure (PDB: 3P96) was refined against the experimental SAXS data using DADIMODO software [25]. The input was prepared by modifying the file "ConfFile.txt" found on the website <https://dadimodo.synchrotron-soleil.fr/submission> for rigid body definition as follows

body1 = A: 352-400

body2 = B: 352-400

and allowing for CRY SOL to be run instead of Pepsi-SAXS for the evaluation of the discrepancy between the theoretical and experimental SAXS curves.

SREFLEX - *BmSerB* crystallographic structure and *MtSerB2* tetramer model T4-D2-7 were refined against the experimental SAXS data using the ATSAS online (<https://www.embl-hamburg.de/biosaxs/atsas-online/>) version of SREFLEX software [26] with default settings.

Rigid body modeling

MtSerB2 trimer and tetramer were modeled by rigid body modeling using the ATSAS online version of CORAL software [27]. Trimer was run with an overall P3 symmetry and tetramer was run with P4 and P222 symmetries. The number of domains was defined as 2, with domain 1 being the ACT1 domain (ALA7 to GLU86) and domain 2 being the ACT2-PSP part (HIS100 to ASP402). Domain 1 was preceded by a 6-residues long N-terminal chain and connected to domain 2 by a 13-residues linker. Domain 2 was followed by a 11-residues long C-terminal chain. Both domains were defined as free.

B.3.4 Crystallization assays

Screening in 96-well plates

After purification, protein solutions were exchanged for buffer χ (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM TCEP, 2.5 mM MgCl₂) and concentrated by centrifugation to about 25 mg mL⁻¹. A volume of 0.4 μ L protein solution mixed with 0.4 μ L of well solution were set up at 20 °C in sitting-drop vapor diffusion experiments against the

conditions of the crystallization screens Crystal Screen 1 and 2 (*Hampton Research*), JCSG Plus, PACT Premier and BCS (*Molecular Dimensions*).

***BmSerB* crystallization**

After purification, *BmSerB* was exchanged for buffer χ and concentrated by centrifugation to 25.7 mg mL⁻¹. Crystals of *BmSerB* were grown at 20 °C by the hanging-drop vapor diffusion method by mixing 1 μ L protein solution with 1 μ L well solution consisting of 0.2 M sodium formate, 0.1 M sodium phosphate pH 6.2-6.8, 14-24% w/v PEG smear medium (12.5% w/v PEG 3350, 12.5% w/v PEG 4000, 12.5% w/v PEG 2000, 12.5% w/v PEG 5000 MME). The crystals had a dodecahedral shape with dimensions around 0.3 mm. Prior to data collection, the crystals were soaked for 2 minutes in well solution supplemented with glycerol to a final concentration of 21.25% for cryoprotection. The crystals were subsequently harvested with cryoloops and flash-frozen in liquid nitrogen.

B.3.5 X-ray data collection and analysis

X-ray diffraction data were collected at 100 K on the PROXIMA-1 beamline at SOLEIL Synchrotron (Saint-Aubin, France) using a Dectris EIGER-X-16M detector and a single wavelength of 0.979 Å. The data were processed using XDS [28] through autoPROC 1.0.5 (*Global Phasing*) [29]. Initial phases were calculated by molecular replacement using PHASER [30], with two parts (res 4-79 and res 80-290) of a *BmSerB* model predicted by Alphafold2 [31] as search models. Model building and structure refinement were performed using PHENIX [32] and COOT [33], with the exception of the first three refinement cycles performed with BUSTER (*Global Phasing*) [34]. The ligands were added to the structure using eLBOW [35].

B.4 Enzyme activity

B.4.1 Steady-state kinetics with and without L-Ser

Malachite green-based phosphatase assay

Enzyme activity was assayed by free orthophosphate (Pi) determination using a malachite green-based phosphatase assay based on Itaya's colorimetric method [36, 37].

The enzyme (about 1 pmol) was incubated at 37 °C in a total volume of 180 µL containing 25 mM Tris-HCl pH 7.4, 5 mM MgCl₂, 1 mM DTT and the desired L-Ser concentration. The reaction was initiated by adding 20 µL of a *O*-phospho-L-serine (PS) solution at 10x the final well concentration (0-200 mM). After incubation for 10 min at 37 °C, the reaction was stopped by mixing 150 µL of the reaction volume with 50 µL of dye composed of 1.7% ammonium heptamolybdate and 0.22% malachite green in 2 M HCl. The absorbance of the solution was measured at 660 nm. Absorbance due to PS was quantified by replacing the enzyme by the same volume of buffer for each assayed PS concentration and the obtained value subtracted from the total absorbance. The activity (released Pi) was calculated from a calibration curve constructed using dilutions of a phosphate standard solution. All the measurements were made in triplicate.

Evaluation of the kinetic parameters

Analysis of kinetic data and curve fitting was performed using GraphPad Prism 5 (GraphPad Software, La Jolla California USA). Values of $K_{M,(app)}$, $V_{max,(app)}$ and $k_{cat,(app)}$ were determined by fitting either the equation for uncompetitive substrate inhibition (B.1) for kinetics in the absence of L-Ser or the compact form of the general modifier equation (B.2) to the initial velocity curves v vs $[S]$.

$$v = \frac{V_{max}[S]}{K_M + [S](1 + \frac{[S]}{K_{is}})} \quad (B.1)$$

$$\frac{v}{[E]_t} = \frac{k_{cat,app}[S]}{K_{M,app} + [S]} = \frac{k_{cat} \left(1 + \beta \frac{[I]}{\alpha K_i}\right) [S]}{K_M \left(1 + \frac{[I]}{K_i}\right) + [S] \left(1 + \frac{[I]}{\alpha K_i}\right)} \quad (B.2)$$

The nature of the inhibition mechanisms were determined following Baici's methodology (<https://www.enzyme-modifier.ch>) [38]. The dependence of $k_{cat,app}$, $1/k_{cat,app}$, $K_{M,app}$, $k_{cat,app}/K_{M,app}$, and $K_{M,app}/k_{cat,app}$ on L-Ser concentration was assessed from the slope and intercept replots of the double-reciprocal form of Equation B.2 (Equation B.3).

$$\frac{[E]_t}{v} = \frac{K_{M,app}}{k_{cat,app}} \frac{1}{[S]} + \frac{1}{k_{cat,app}} = slope \frac{1}{[S]} + intercept \quad (B.3)$$

Bibliography

- [1] R. Choi, A. Kelley, D. Leibly, S. Nakazawa Hewitt, A. Napuli, and W. Van Voorhis. Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 67(9):998–1005, 2011.
- [2] H. Liu and J.H. Naismith. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, 8(1):1–10, 2008.
- [3] M. Waterhouse, A. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1):W296–W303, 2018.
- [4] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.
- [5] A.D. MacKerell Jr, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers: Original Research on Biomolecules*, 56(4):257–265, 2000.
- [6] M. Mirgaux, L. Leherte, and J. Wouters. Influence of the presence of the heme cofactor on the JK-loop structure in indoleamine 2, 3-dioxygenase 1. *Acta Crystallographica Section D: Structural Biology*, 76(12):1211–1221, 2020.
- [7] L. Darré, M.R. Machado, A.F. Brandner, H.C. González, S. Ferreira, and S. Pantano. SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *Journal of Chemical Theory and Computation*, 11(2):723–739, 2015.
- [8] L. Darré, A. Tek, M. Baaden, and S. Pantano. Mixing atomistic and coarse grain solvation models for MD simulations: let WT4 handle the bulk. *Journal of Chemical Theory and Computation*, 8(10):3880–3894, 2012.

- [9] H.C. Gonzalez, L. Darré, and S. Pantano. Transferable mixing of atomistic and coarse-grained water models. *The Journal of Physical Chemistry B*, 117(46):14438–14448, 2013.
- [10] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.
- [11] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.
- [12] B. Hess, H.J.C. Bekker, H. Berendsen, and J.G.E.M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [13] B. Pierce, W. Tong, and Z. Weng. M-ZDOCK: a grid-based approach for C_n symmetric multimer docking. *Bioinformatics*, 21(8):1472–1478, 2005.
- [14] I.T. Desta, K.A. Porter, B. Xia, D. Kozakov, and S. Vajda. Performance and its limits in rigid body protein-protein docking. *Structure*, 28(9):1071–1081, 2020.
- [15] C. Vajda, S. Yueh, D. Beglov, T. Bohnuud, S. E Mottarella, B. Xia, D.R. Hall, and D. Kozakov. New additions to the Clus Pro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*, 85(3):435–444, 2017.
- [16] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, D. Beglov, and S. Vajda. The ClusPro web server for protein-protein docking. *Nature Protocols*, 12(2):255–278, 2017.
- [17] D. Kozakov, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, and S. Vajda. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*, 81(12):2159–2166, 2013.
- [18] T. Park, M. Baek, H. Lee, and C. Seok. GalaxyTongDock: Symmetric and asymmetric ab initio protein-protein docking web server with improved energy parameters. *Journal of Computational Chemistry*, 40(27):2413–2417, 2019.
- [19] L. Jendele, R. Krivak, P. Skoda, M. Novotny, and D. Hoksza. FrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Research*, 47(W1):W345–W349, 2019.

- [20] R. Krivák and D. Hoksza. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):1–12, 2018.
- [21] R.E. Hopkins, J.B. and Gillilan and S. Skou. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, 50(5):1545–1553, 2017.
- [22] J.B. Hopkins. Bioxtas raw documentation. <https://bioxtas-raw.readthedocs.io/>, Accessed online July 8, 2022.
- [23] D. Svergun, C. Barberato, and M.H.J. Koch. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, 28(6):768–773, 1995.
- [24] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J Koch, and D.I. Svergun. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, 2003.
- [25] O. Rudenko, A. Thureau, and J. Perez. Evolutionary refinement of the 3D structure of multi-domain protein complexes from small angle X-ray scattering data. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 401–402, 2019.
- [26] A. Panjkovich and D.I. Svergun. Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Physical Chemistry Chemical Physics*, 18(8):5707–5719, 2016.
- [27] M.V. Petoukhov, D. Franke, A.V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P.V. Konarev, and D.I. Svergun. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(2):342–350, 2012.
- [28] W. Kabsch. XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):125–132, 2010.
- [29] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, and G. Bricogne. Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):293–302, 2011.

- [30] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, and R.J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4):658–674, 2007.
- [31] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, pages 1–4, 2022.
- [32] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L-W Hung, G.J. Kapral, R.W. Grosse-Kunstleve, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221, 2010.
- [33] P. Emsley, B. Lohkamp, W.G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501, 2010.
- [34] G. Bricogne, E. Blanc, M. Brandl, C. Flensburg, P. Keller, W. Paciorek, P. Roversi, A. Sharff, O.S. Smart, C. Vonrhein, and T.O. Womack. BUSTER. 2017, Cambridge, United Kingdom: Global Phasing Ltd.
- [35] N.W. Moriarty, R.W. Grosse-Kunstleve, and P.D. Adams. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallographica Section D: Biological Crystallography*, 65(10):1074–1080, 2009.
- [36] K. Itaya and M. Ui. A new micromethod for the colorimetric determination of inorganic phosphate. *Clinica Chimica Acta*, 14(3):361–366, 1966.
- [37] AA Baykov, OA Evtushenko, and SM Avaeva. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Analytical Biochemistry*, 171(2):266–270, 1988.
- [38] A. Baici. *Kinetics of enzyme-modifier interactions*, volume 1. Springer, 2015.

Appendix C

Superimposition of *MaSerB* crystallographic structures

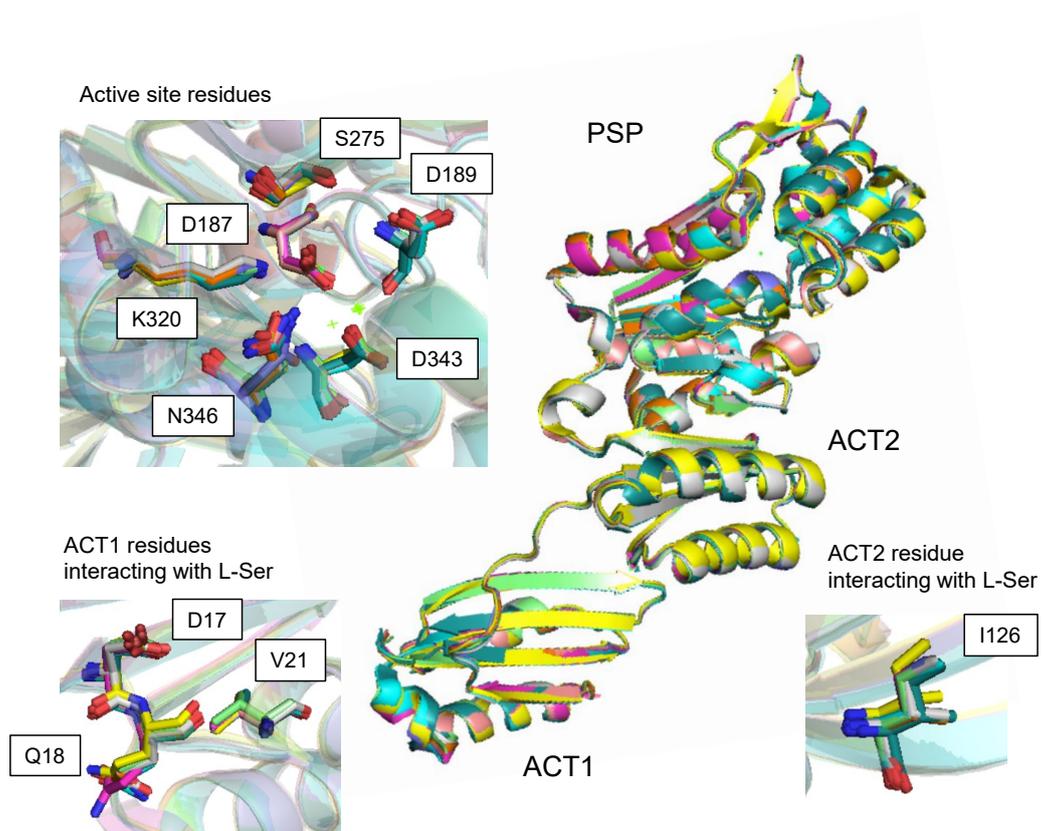


FIGURE C.1: Alignment of *MaSerB* structures 3P96, 5JJB, 5T41, 5IT0, 5JMA, 5IT4, 5IS2, 5JLR and 5JLP.

Appendix D

Biochemical characterization of phosphoserine phosphatase SerB2 from *Mycobacterium marinum*



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Biochemical characterization of phosphoserine phosphatase SerB2 from *Mycobacterium marinum*



Elise Pierson*, Johan Wouters

Laboratoire de Chimie Biologique Structurale (CBS), Namur Medicine and Drug Innovation Center (NAME DIC), Namur Research Institute for Life Sciences (NARILIS), University of Namur (UNamur), B-5000, Namur, Belgium

ARTICLE INFO

Article history:

Received 6 July 2020

Accepted 7 July 2020

Available online 8 August 2020

Keywords:

Mycobacterium marinum

SerB2

Phosphoserine phosphatase

HAD phosphatases

Clofazimine

Harmine derivatives

ABSTRACT

SerB2 is an essential phosphoserine phosphatase (PSP) that has been shown to be involved in *Mycobacterium tuberculosis* (*Mtb*) immune evasion mechanisms, and a drug target for the development of new antitubercular agents. A highly similar (91.0%) orthologous enzyme exists in the surrogate organism *Mycobacterium marinum* (*Mma*) and could have acquired similar properties. By homology modeling, we show that the two PSPs are expected to exhibit almost identical architectures. *Mma*SerB2 folds into a homodimer formed by two intertwined subunits including two ACT regulatory domains followed by a catalytic core typical of HAD (haloacid dehalogenase) phosphatases. Their *in vitro* catalytic properties are closely related as *Mma*SerB2 also depends on Mg^{2+} for the dephosphorylation of its substrate, O-phospho-L-serine (PS), and is most active at neutral pH and temperatures around 40 °C. Moreover, an enzyme kinetics study revealed that the enzyme is inhibited by PS as well, but at lower concentrations than *Mtb*SerB2. Substrate inhibition could occur through the binding of PS in the second active site and/or at the ACT domains interface. Finally, previously described beta-carboline *Mtb*SerB2 inhibitors also decrease the phosphatase activity of *Mma*SerB2. Altogether, these results provide useful information when *M. marinum* is used as a model to study immune evasion in tuberculosis.

© 2020 Elsevier Inc. All rights reserved.

The pre-print peer-reviewed version of the full paper can be accessed online on the institutional repository of UNamur at the following link:

<https://researchportal.unamur.be/en/publications/>

[biochemical-characterization-of-phosphoserine-phosphatase-serb2-f](https://researchportal.unamur.be/en/publications/biochemical-characterization-of-phosphoserine-phosphatase-serb2-f).

Appendix E

Data collection and refinement statistics for *BmSerB* structure

TABLE E.1: Data collection and refinement statistics for *BmSerB* structure on-hold PDB entry 7QPL. Statistics for the highest-resolution shell are shown in parentheses.

Wavelength (Å)	0.978565
Resolution range (Å)	38.27-1.77 (1.833-1.77)
Space group	I2 ₁ 3
<i>a</i> , <i>b</i> , <i>c</i> (Å)	143.21, 143.21, 143.21
α , β , γ (°)	90, 90, 90
Total reflections	1966903 (193653)
Unique reflections	47522 (2558)
Multiplicity	41.4 (41.3)
Completeness (%)	95.42 (54.56)
Mean I/ σ (I)	22.77 (1.43)
Wilson B factor (Å ²)	39.61
R_{merge}	0.09586 (2.942)
R_{meas}	0.09705 (2.978)
$R_{p.i.m}$	0.01509 (0.4624)
CC _{1/2}	0.999 (0.616)
CC*	1.000 (0.873)
Reflections used in refinement	45353 (2558)
Reflections used for R_{free}	2288 (123)
R_{work}	0.1578 (0.2895)
R_{free}	0.1833 (0.3178)
CC _{work}	0.973 (0.859)
CC _{free}	0.956 (0.794)
No. of non-H atoms	
Total	2573
Macromolecules	2221
Ligands	72
Solvent	280
Protein residues	295
R.m.s.d., bonds (Å)	0.010
R.m.s.d., angles (°)	1.20
Ramachandran plot	
Favored (%)	97.61
Allowed (%)	2.39
Outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.92
Average B factors (Å ²)	
Overall	54.00
Macromolecules	52.68
Ligands	68.34
Solvent	60.77
Number of TLS groups	3

Appendix F

Mass photometry measurement on unseparated *MtSerB2* sample

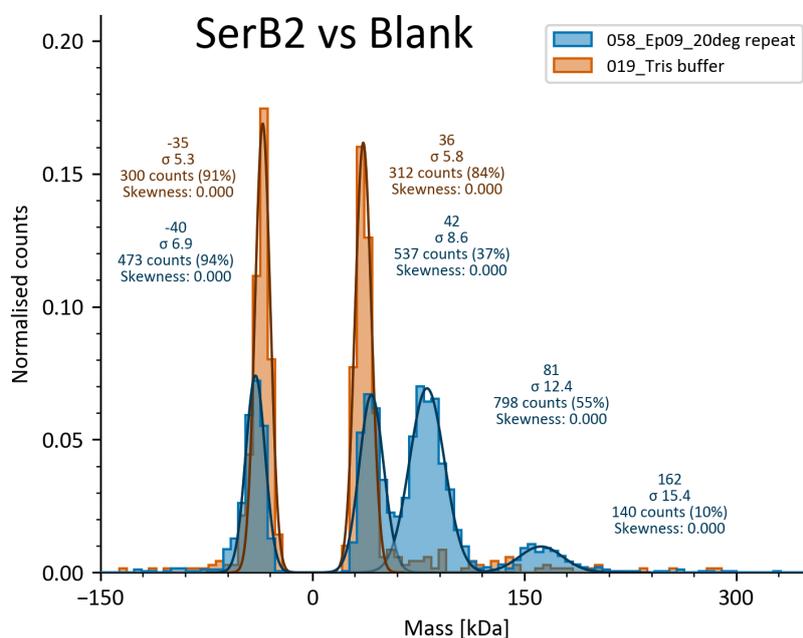
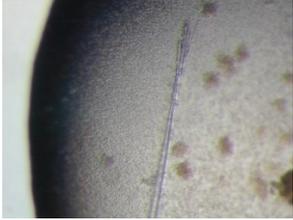
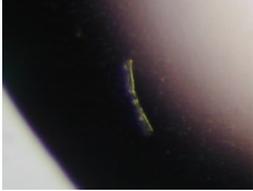
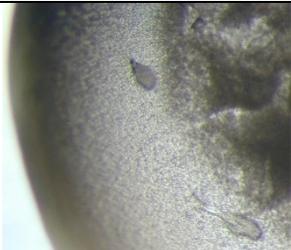
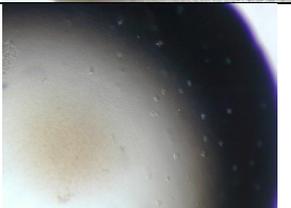


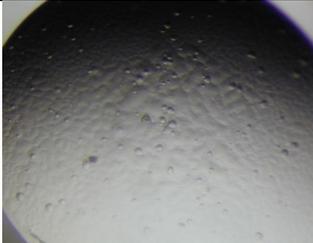
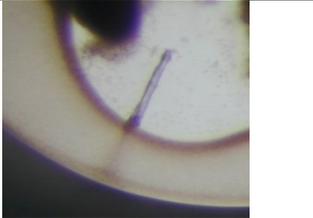
FIGURE F.1: Superimposition of mass photometry analyses of an unseparated *MtSerB2* sample and Tris buffer (blank).

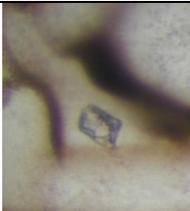
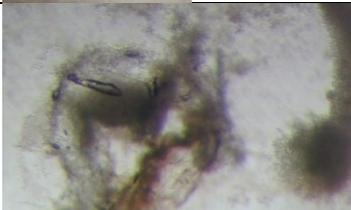
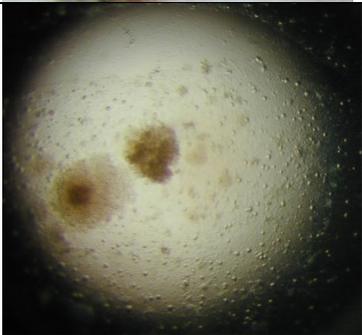
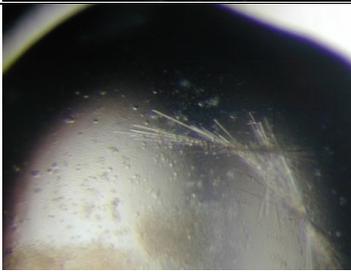
Appendix G

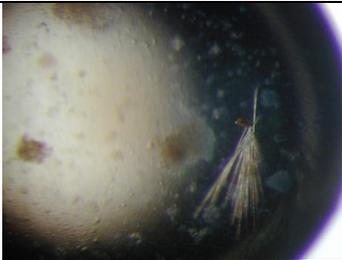
Crystals obtained from *MtSerB2* crystallization assays

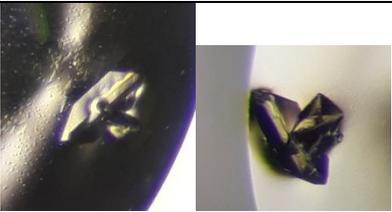
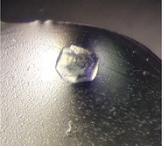
Screen	Set up	Exp	Well	Lig.	Conditions	Cryo	Crystals	Diffraction
Pic1_PACT_H10 24 wells around PACT 2-46	28/04/21	05/06/21	D1 Tetramer 24.3 mg mL ⁻¹	-	0.02M Sodium/potassium phosphate 16% PEG 3350 0.1M Bis Tris Propane pH 8.7 24.3 mg/mL	20% GOL		No
Pic1_PACT_H10 24 wells around PACT 2-46	28/04/21	05/06/21	C4 Tetramer 24.3 mg mL ⁻¹	-	0.02M Sodium/potassium phosphate 22% PEG 3350 0.1M Bis Tris Propane pH 8.5 24.3 mg/mL	20% GOL		Salt
Pic1_PACT_H10 24 wells around PACT 2-46	28/04/21	05/06/21	D5 Tetramer 24.3 mg mL ⁻¹	-	0.02M Sodium/potassium phosphate 24% PEG 3350 0.1M Bis Tris Propane pH 8.7 24.3 mg/mL	20% GOL		Not harvested
Pic2_JCSG_E12_StSeed 24 wells with streak seeding around JCSG 2- 45	12/05/2021	05/06/21	C3 Dimer 12-22 mg mL ⁻¹	-	0.1M Lithium sulfate 0.1M Bis-Tris pH 5.5 15% PEG 3350 28 mg/mL	based on C2 10% GOL		Lost
Pic2_JCSG_E12_StSeed 24 wells with streak seeding around JCSG 2- 45	12/05/2021	05/06/21	C4 Dimer 12-22 mg mL ⁻¹	-	0.1M Lithium sulfate 0.1M Bis-Tris pH 5.9 17% PEG 3350 28 mg/mL	based on C2 10% GOL		Not harvested

Pic2_JCSG_E12 24 wells around JCSG 2-45	28/04/2021	05/06/21	B3 Dimer 28.2 mg mL ⁻¹	-	0.1M Lithium sulfate 0.1M Bis Tris pH 5.5 23% PEG 3350 28 mg/mL	based on B2 12.5% GOL		Salt Salt
Pic2_BCS_A11 24 wells around BCS 1-11	07/05/2021	05/06/21	C2 Dimer 28.2 mg mL ⁻¹	-	0.1M Phosphate/Citrate pH 5.7 18% PEG Smear Broad 24.2 mg/mL	-		Not harvested
Pic2_BCS_A11 24 wells around BCS 1-11	07/05/2021	05/06/21	D3 Dimer 28.2 mg mL ⁻¹	-	0.1M Phosphate/Citrate pH 5.9 20% PEG Smear Broad 24.2 mg/mL	based on D2 12.5% GOL		Salt
Pic1_JCSG_H7 24 wells around JCSG 2-8	07/05/2021	05/06/21	B3 Tetramer 24.3 mg mL ⁻¹	-	0.8 M ammonium phosphate dibasic 0.1 M sodium acetate pH 4.5 24.2 mg/mL	based on B2 20% GOL		Salt
Pic1_JCSG_H7 24 wells around JCSG 2-8	07/05/2021	05/06/21	C3 Tetramer 24.3 mg mL ⁻¹	-	0.8 M ammonium phosphate dibasic 0.1 M sodium acetate pH 4.7 24.2 mg/mL	based on B2 20% GOL		Lost

Screen JCSG – Pic 2	30/03/2021	05/06/21	B2 1-10 Dimer 25.5 mg mL ⁻¹	-	0.2 M Potassium formate 20% PEG 3350	20% GOL		Not harvested
Screen JCSG – Pic 1	30/03/2021	05/06/21	C2 1-11 Tetramer 24.5 mg mL ⁻¹	-	0.2 M ammonium phosphate monobasic 0.1 M Tris pH 8.5 50% w/v MPD	Cryo		Salt
Screen BCS – Pic1	30/03/2021	05/06/21	A6 1-6 Tetramer 24.5 mg mL ⁻¹	-	0.1M MES pH 6.5 20% PEG smear high	Cryo		Not tested
Screen BCS – Pic1	30/03/2021	05/06/21	D6 1-42 Tetramer 24.5 mg mL ⁻¹	-	0.1M potassium thiocyanate 0.1M sodium bromide 0.1M MES pH 6.5 12% PEG smear high	10% GOL		Not harvested
Screen BCS - Pic1	30/03/2021	05/06/21	E7 2-7 Tetramer 24.5 mg mL ⁻¹	-	0.1 M magnesium formate dehydrate 0.1M rubidium chloride 0.1M PIPES pH 7.0 25% PEG smear high	10% GOL		Salt

Screen BCS – Pic 1	30/03/2021	05/06/21	F8 2-20 Tetramer 24.5 mg mL ⁻¹	-	0.15 M ammonium acetate 0.01M calcium chloride dihydrate 0.1M Tris pH 8.5 28% PEG smear broad	5% GOL			Not tested
Screen PACT – Pic 2	31/03/2021	05/06/21	C11 1-35 Dimer 25.5 mg mL ⁻¹	-	0.2M calcium chloride hexahydrate 0.1M HEPES pH 7.0 20% PEG 6000	10% GOL			Salt
Pic1_PACT_H10_2 24 wells around PACT 2-46	20/05/2021	05/06/21	B5 Tetramer 24.2 mg mL ⁻¹	-	0.02M Sodium/Potassium 0.1M BisTris Propane pH 8.7 16% PEG 3350	20% GOL			Salt
Pic1_PACT_H10_2 24 wells around PACT 2-46	20/05/2021	05/06/21	C5 Tetramer 24.2 mg mL ⁻¹	-	0.02M Sodium/Potassium 0.1M BisTris Propane pH 8.9 16% PEG 3350	20% GOL			Not harvested

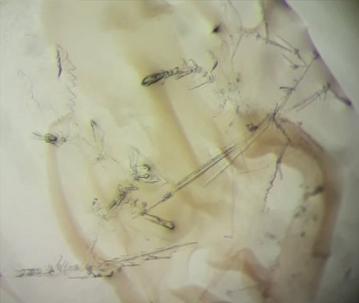
Pic1_PACT_H10_2	20/05/2021	05/06/21	C3 Tetramer 24.2 mg mL ⁻¹	-	0.02M Sodium/Potassium 0.1M BisTris Propane pH 8.9 14% PEG 3350	20% GOL		Salt
BCS Screen	30/03/21	25/09/21	C9 1-33 Tetramer 24.5 mg mL ⁻¹	-	0.15M Magnesium acetate tetrahydrate 0.1M Sodium citrate pH 5.6 20% w/v PEG Smear Broad (1-33)	20% GOL		Salt Salt
BCS Screen	30/03/21	25/09/21	E11 2-11 Tetramer 24.5 mg mL ⁻¹	-	0.05M Magnesium chloride hexahydrate 0.05M Sodium citrate tribasic dehydrate 0.1M Bis-Tris Propane pH 7.8 22.5% w/v PEG Smear High (2-11)	15% GOL		Salt
BCS Screen	30/03/21	25/09/21	E11 2-11 Dimer 25.5 mg mL ⁻¹	-	0.05M Magnesium chloride hexahydrate 0.05M Sodium citrate tribasic dehydrate 0.1M Bis-Tris Propane pH 7.8 22.5% w/v PEG Smear High (2-11)	15% GOL		Fiber diffraction pattern

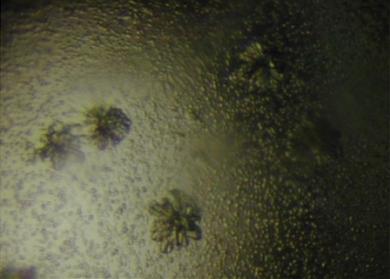
BCS Screen	30/03/21	25/09/21	G8 2-32 Dimer 25.5 mg mL ⁻¹	-	0.075M Magnesium chloride hexahydrate 0.075M Sodium citrate tribasic dehydrate 0.1M Bis-Tris pH 6.0 18% w/v PEG Smear Broad (2-32)	20% GOL		Salt
BCS Screen	30/03/21	25/09/21	H12 2-48 Tetramer 24.5 mg mL ⁻¹	-	0.075M Sodium bromide 0.05M Sodium Fluoride 0.1M HEPES pH 7.8 22.5% w/v PEG Smear Broad 0.075M Sodium iodide (2-48)	15% GOL		Salt
BCS Screen	30/03/21	25/09/21	H12 2-48 Dimer 25.5 mg mL ⁻¹	-	0.075M Sodium bromide 0.05M Sodium Fluoride 0.1M HEPES pH 7.8 22.5% w/v PEG Smear Broad 0.075M Sodium iodide (2-48)	15% GOL		Not harvested
PACT Screen	31/03/21	25/09/21	H1 2-37 Tetramer 24.5 mg mL ⁻¹	-	0.2M Sodium fluoride 0.1M Bis- Tris Propane pH 8.5 20% w/v PEG 3350 (2-37)	15% GOL		No

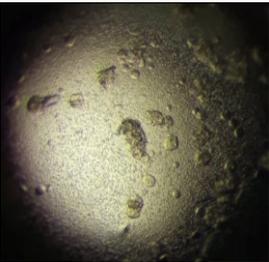
PACT Screen	31/03/21	25/09/21	H1 2-37 Dimer 25.5 mg mL ⁻¹	-	0.2M Sodium fluoride 0.1M Bis- Tris Propane pH 8.5 20% w/v PEG 3350 (2-37)	15% GOL		Not tested, too much ice
PACT Screen	31/03/21	25/09/21	H10 2-46 Dimer 25.5 mg mL ⁻¹	-	0.02M Sodium potassium phosphate 0.1M Bis-Tris propane pH 8.5 20% w/v PEG 3350 (2-46)	-		Salt
Screen PACT LSer	15/10/21	12/02/22	G2 2-26 27.6 mg mL ⁻¹	L-Ser	0.2M sodium bromide 0.1M BisTris propane pH 7.5 20% w/v PEG 3350	20% GOL		Not tested, loading problem
Screen PACT LSer	15/10/21	12/02/22	C4 1-28 27.6 mg mL ⁻¹	L-Ser	0.1M PCTP* pH 7.0 25% w/v PEG 1500 * Sodium propionate, Sodium cacodylate trihydrate, Bis-Tris propane	20% GOL		No
Screen PACT LSer	15/10/21	12/02/22	A4 1-4 27.6 mg mL ⁻¹	L-Ser	0.1M SPG* pH 7.0 25% w/v PEG1500 *Succinic Acid, Sodium Dihydrogen Phosphate and Glycine	20% GOL		Salt

Screen PACT L-Ser	15/10/21	12/02/22	A5 1-5 27.6 mg mL ⁻¹	L-Ser	0.1M SPG pH 8.0 25% w/v PEG1500	20% GOL			Salt
Screen PACT L-Ser	15/10/21	12/02/22	G5 2-29 27.6 mg mL ⁻¹	L-Ser	0.2M sodium nitrate 0.1M BisTris propane pH 7.5 20% w/v PEG 3350	20% GOL			Not harvested
MtSerB2 Screen 5mM L-Ser	13/10/20	12/02/22	F11 2-24 27.6 mg mL ⁻¹	L-Ser	0.02M magnesium sulfate heptahydrate 0.2M potassium chloride 0.1M BICINE pH 8.8 22.5% PEG smear broad	20% GOL			Salt

MytuSerB2 TPic1 BCS Screen	24/11/2020	12/02/22	H12 2-48 Tetramer 18.8 mg mL ⁻¹	-	0.075M sodium bromide 0.05M sodium fluoride 0.1M HEPES pH 7.8 22.5% w/v PEG smear broad 0.075M sodium iodide	20% GOL		Salt
JCSG+ Screen LSer	30/03/2021	12/02/22	A5 1-33 27.6 mg mL ⁻¹	L-Ser	0.1 M sodium potassium phosphate pH 6.2 25% v/v 1.2-propanediol	20% GOL		Fiber diffraction pattern
JCSG+ Screen LSer	30/03/2021	12/02/22	G5 1-39 27.6 mg mL ⁻¹	L-Ser	0.2M NaCl 0.1M sodium/potassium phosphate pH 6.2 50% w/v PEG 200	20% GOL		Fiber diffraction pattern
JCSG+ Screen LSer	30/03/2021	12/02/22	A8 2-9 27.6 mg mL ⁻¹	L-Ser	1.6M Magnesium sulfate heptahydrate 0.1M MES pH 6.5	20% GOL		No

JCSG+ Pic1/Pic2	30/03/2021	12/02/22	A1 Tetramer 1-1 24.5 mg mL ⁻¹	-	0.2M Lithium sulfate 0.1M sodium acetate pH 4.5 50% w/v PEG 400	20% GOL		Not shot, technical issue
JCSG+ Pic1/Pic2	30/03/2021	12/02/22	A1 Dimer 1-1 25.5 mg mL ⁻¹	-	0.2M Lithium sulfate 0.1M sodium acetate pH 4.5 50% w/v PEG 400			Not harvested but probably salt
JCSG+ Pic1/Pic2	30/03/2021	12/02/22	G5 Dimer 1-39 25.5 mg mL ⁻¹	-	0.2M NaCl 0.1M sodium/potassium phosphate pH 6.2 50% w/v PEG 200			Not harvested
JCSG+ Pic1/Pic2	30/03/2021	12/02/22	C6 Tetramer 1-43 24.5 mg mL ⁻¹	-	0.2M Lithium sulfate 0.1M Tris pH 8.5 40% w/v PEG 400			Salt

JCSG+ Pic1/Pic2	30/03/2021	12/02/22	H7 Tetramer 2-8 24.5 mg mL ⁻¹	-	1M ammonium phosphate dibasic 0.1M sodium acetate pH 4.5			Salt
BCS Pic1/Pic2	30/03/2021	12/02/22	G2 Tetramer 2-26 24.5 mg mL ⁻¹	-	0.15M lithium sulfate 0.05M magnesium chloride hexahydrate 0.1M BisTris pH 6.8 25% w/v PEG smear low			Not harvested but probably salt
BCS Pic1/Pic2	30/03/2021	12/02/22	D6 Tetramer 1-42 24.5 mg mL ⁻¹	-	0.1M Potassium thiocyanate 0.1M sodium bromide 0.1M MES pH 6.5 12% w/v PEG smear high			Not harvested
Opti PACT LSer F11 4°C 24 wells around PACT screen with LSer condition 2-23	19/11/2021	12/02/22	A4 14.1 - 28.2 mg mL ⁻¹	-	0.2M sodium citrate tribasic dihydrate 0.1M BisTris propane pH 6.3 24% w/v PEG 3350			Not harvested

Opti PACT LSer F11 4°C 24 wells around PACT screen with LSer condition 2-23	19/11/2021	12/02/22	D4 (+seed) 14.1 - 28.2 mg mL ⁻¹	-	0.2M sodium citrate tribasic dihydrate 0.1M BisTris propane pH 6.9 20% w/v PEG 3350			Not harvested
Opti PACT F11 (2) 24 wells around PACT screen with LSer condition 2-23 with seeding at lower conc.	04/01/2022	12/02/22	B4 (seed from A6 seeded from opti PACT F11 TA) 7.1 – 14.2 mg mL ⁻¹	-	0.2M sodium citrate tribasic dihydrate 0.1M BisTris propane pH6.5 24% w/v PEG 3350			Not harvested
Opti PACT LSer F11 4°C 24 wells around PACT screen with LSer condition 2-23	19/11/21	12/02/22	B4 (seed) 14.1 - 28.2 mg mL ⁻¹	L-Ser	0.2M sodium citrate tribasic dihydrate 0.1M BisTris propane pH 6.5 20% w/v PEG 3350	20% GOL but crystal was dissolving		No
Opti PACT F11 (2) 24 wells around PACT screen with LSer condition 2-23 with seeding at lower conc.	04/01/22	12/02/22	A3 (seed) 7.1 – 14.2 mg mL ⁻¹	L-Ser	0.2M sodium citrate tribasic dihydrate 0.1M BisTris propane pH 6.3 21% w/v PEG 3350	10% GOL		No

Appendix H

Calculation of the assumed amount of active subunit in a solution of *MtSerB2* dimer and tetramer

Consider a solution of *MtSerB2* composed of a mixture of the dimeric and tetrameric forms of the enzyme. The dimer is made up of two subunits and the tetramer is made up of four subunits. If t is the amount of tetramer (moles), d is the amount of dimer, the Equation H.1 describes the total amount of subunits, m , in the mixture:

$$4t + 2d = m \quad (\text{H.1})$$

The ratio, a , between the amount of dimer and tetramer is defined by Equation H.2:

$$t = ad \quad (\text{H.2})$$

Combining and rearranging the two equations, we obtain:

$$4ad + 2d = m \Leftrightarrow d(4a + 2) = m \Leftrightarrow d = \frac{m}{4a + 2} \quad (\text{H.3})$$

In our case, the ratio a of the sample used for enzyme kinetics was measured by mass photometry and is equal to $845/716 = 1.1802$. The total amount of subunit m in the same sample was quantified by a measure of absorbance at 280 nm. The total amount of subunit present in the well is estimated to be 0.836 pmol.

Knowing this, the amount of dimer in a well (pmol) can be calculated:

$$d = \frac{0.836}{(4 * 1.1802) + 2} = 0.124 \quad (\text{H.4})$$

We consider that if the dimer is active and the tetramer is inactive, then the dimer is made up of two active subunits and the tetramer is made up of four inactive subunits. The quantity of active subunits is then twice the amount of dimer:

$$0.124 * 2 = 0.249 \quad (\text{H.5})$$

The corrected amount of active species to be counted per well for the estimation of the enzymatic activity is therefore 0.249 pmol.

