Understanding ribosome binding interactions and conformational changes of the EngA bacterial GTPase, a potential target for new antibiotics

Catarina da Silveira Tomé

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Physique pour les sciences du vivant**

Arrêté ministériel : 25 mai 2016

Présentée par

Catarina DA SILVEIRA TOMÉ

Thèse dirigée par **Dominique HOUSSET** et
codirigée par **Jean-Michel JAULT**

préparée au sein de l'**Institut de Biologie Structurale**
dans l'**École Doctorale de Physique**

Intéractions avec le ribosome et changements conformationnels de la GTPase bactérienne EngA, une cible potentielle pour de nouveaux antibiotiques

Thèse soutenue publiquement le **05/12/2016**,
devant le jury composé de :

M Franz BRUCKERT
Professeur des universités, Institut Polytechnique de Grenoble, Président
Mme Jacqueline CHERFILS
Directeur de recherche, École Normale Supérieure Paris-Saclay, Rapporteur
Mme Mirjam CZJZEK
Directeur de recherche, Station Biologique de Roscoff, Rapporteur
M Jean-François GUICHOU
Professeur des universités, Centre de Biochimie Structurale, Examinateur
M Dominique HOUSSET
Chercheur CEA, Institut de Biologie Structurale, Directeur de thèse
M Jean-Michel JAULT
Directeur de recherche, Institut de Biologie et Chimie des Protéines, Co-directeur de thèse

Understanding ribosome binding interactions and conformational changes of the EngA bacterial GTPase, a potential target for new antibiotics

Acknowledgements

I would like to thank the members of my committee, Jacqueline Cherfils, Mirjam Czjzek, Franz Bruckert and Jean-François Guichou, for accepting to evaluate and discuss this work.

I thank my thesis advisors, Dominique Housset and Jean-Michel Jault, for welcoming me to this project. I'm grateful for their guidance and for the support and trust they gave me.

A special thanks to Anne-Emmanuelle Foucher, for her coaching, advice, discussions and companionship.

I want to thank the people who contributed to our project with their expertise: our collaborators from Institut de Biologie Structurale and Université Grenoble-Alpes, the scientists from the HTX Laboratory and ESRF beamlines, and all the members of the IRPAS group for always being available to help me throughout this project.

Finally, I want to express my gratitude to my family and friends with whom I shared these three years. In particular, a special thank you to my parents and Sara for their encouragement and motivation, and Nuno for his constant presence.

Le résumé en français

Le traitement des infections bactériennes est devenu possible au XX^{e} siècle grâce au développement des premiers antibiotiques. La découverte de nombreuses nouvelles familles d'antibiotiques entre les années 1940 et 1970 s'est accompagnée de leur usage généralisé ce qui a conduit à l'émergence de phénomènes de résistance à ces médicaments. Aggravé par le ralentissement du développement de nouveaux antibiotiques après les années 1980, la résistance aux antibiotiques constitue aujourd'hui, selon l'OMS, «l'une des plus graves menaces pesant sur la santé mondiale». Des nouveaux outils envisageant la prévention et le traitement des infections bactériennes sont actuellement en développement. La prophylaxie (dont les vaccins), les traitements complémentaires aux antibiotiques (dont les probiotiques) et les traitements alternatives aux antibiotiques (y compris les bactériophages et les anticorps monoclonaux) sont parmi les voies thérapeutiques en développement. La reprise de la recherche sur des nouveaux antibiotiques pour lesquels les bactéries n'ont pas encore acquis de mécanismes de résistance est aussi nécessaire. Le point de départ de cette recherche passe par l'identification des nouvelles cibles moléculaires. Une cible pharmacologique idéale doit comporter trois caractéristiques principales : être essentielle à la survie cellulaire, être conservée au sein des bactéries, et être absente chez les eucaryotes. Parmi les machineries cellulaires ciblées par les antibiotiques aujourd'hui disponibles, le ribosome bactérien représente environ 40% des cas. Alors que tous ces antibiotiques agissent directement sur le fonctionnement du ribosome (par interaction avec le ribosome et inhibition de la synthèse protéique), la voie de biogenèse du ribosome reste toujours inexplorée.

La biogenèse du ribosome est un processus cellulaire essentiel et complexe, où plus de cinquante protéines et trois ARN doivent être synthétisés, modifiés, repliés et assemblés pour former une particule 70S fonctionnelle. En plus des protéines et ARN ribosomaux, des facteurs d'assemblage non ribosomiques participent à ce processus pour le faciliter et l'accélérer. Parmi eux on trouve des endonucléases responsables par la maturation des ARN, des enzymes responsables des modifications chimiques des ARN et protéines, des hélicases et chaperonnes qui jouent un rôle dans le repliement, et des GTPases dont le rôle exact reste encore

méconnu mais qui, chez les bactéries, représentent une grande partie des facteurs impliqués dans la biogenèse du ribosome.

Les GTPases sont des enzymes qui lient et hydrolysent le GTP. Elles possèdent un domaine catalytique conservé – le domaine G – composé d'un feuillet β central entouré par les hélices α et des boucles. Cinq boucles font partie du site actif et présentent des motifs conservés, nommés G1 à G5, qui jouent un rôle dans la fixation et hydrolyse du nucléotide. Grâce à leur flexibilité, ces boucles adoptent plusieurs conformations en fonction du nucléotide présent dans le site catalytique qui caractérisent différents états de la GTPase. Notamment, les boucles nommées Switch I et Switch II, contenant respectivement les motifs G2 et G3, sont en grande partie responsables de ces changements structuraux. Le cycle catalytique s'initie avec la fixation du GTP, qui engendre un changement conformationnel qui permet l'activation de l'enzyme et l'interaction avec des effecteurs. Les principaux contacts protéine-ligand présents dans les GTPases se produisent entre la nucléobase et les motifs G4 et G5, les phosphates α et β du nucléotide et le motif G1, et le phosphate γ du GTP et les motifs G2 et G3. L'hydrolyse du GTP en GDP entraîne un changement de conformation de la GTPase, qui passe ainsi de l'état actif à l'état inactif, avec une faible affinité pour les effecteurs. En particulier, les éléments responsables par la coordination du phosphate γ du GTP sont perdus, ce qui engendre une déstabilisation et repositionnement des motifs G2 et G3. Le GDP est finalement échangé contre du GTP pour que l'enzyme puisse entrer dans un nouveau cycle. Malgré une structure et un mécanisme catalytique conservés, les GTPases présentent une grande diversité fonctionnelle. Dans les bactéries, elles participent dans la réplication du DNA, la division cellulaire, la réponse au stress, les mécanismes de pathogenèse, la synthèse protéique et l'assemblage du ribosome.

Depuis quelques années, plusieurs GTPases bactériennes essentielles ont été identifiées comme étant impliquées dans la biogenèse du ribosome. C'est le cas d'EngA, identifiée en 2000 et associée à la division cellulaire. Des études ultérieures ont montré qu'elle interagit de façon nucléotide-dépendante avec le ribosome et que sa déplétion entraîne l'accumulation de précurseurs ribosomaux, suggérant un rôle dans la biogenèse du ribosome. Le caractère essentiel d'EngA chez les bactéries et son absence chez les humains font de cette enzyme une bonne cible pharmacologique pour le développement de nouveaux antibiotiques. Cependant, les

particularités structurales d'EngA la rendent plus difficile à étudier. En effet, malgré sa conservation et la présence des motifs caractéristiques de la famille des GTPases, EngA est la seule GTPase connue à posséder deux domaines G. Dans sa structure, EngA présente trois domaines : un domaine G N-terminal (nommé GD1) lié par une boucle à un deuxième domaine G (GD2) et un domaine C-terminal non catalytique (KH). Ainsi, la présence de deux domaines catalytiques chez EngA apporte une diversité conformationnelle plus grande. Plusieurs structures d'EngA ont été résolues par cristallographie et cryo-microscopie électronique et ont suggéré qu'outre les changements structuraux locaux au niveau des motifs G, typiques des GTPases, des changements conformationnels au niveau de la structure tertiaire de la protéine peuvent aussi avoir lieu. Trois conformations différentes d'EngA ont été observées et associées à des différents états du cycle catalytique. Ces trois structures présentent une position conservée des domaines GD2 et KH, pendant que le domaine GD1 se repositionne en fonction de l'étape du cycle. Cependant, les informations structurales disponibles présentent aussi des aspects peu clairs, notamment l'absence d'une structure avec du GTP fixé simultanément aux deux domaines G d'EngA ou la faible résolution des structures de cryo-microscopie électronique. Malgré les données disponibles actuellement, les événements qui régulent les changements conformationnels d'EngA, la fixation/hydrolyse du nucléotide et les interactions avec le ribosome restent encore mal compris.

Notre objectif a donc été de mieux comprendre les mécanismes structuraux, biochimiques et fonctionnels d'EngA, et de mettre en place un protocole de criblage de composés inhibiteurs. Pour cela, nous avons combiné des données structurales en solution et dans l'état cristallin et des données d'interaction, et développé un test pour identifier des inhibiteurs d'interactions EngA-ribosome.

Des expériences de SAXS (small-angle X-ray scattering), protéolyse limitée, séquençage N-terminal et spectrométrie de masse ont été réalisées pour étudier l'effet du nucléotide dans les changements conformationnels d'EngA. Les données de SAXS ont montré que la protéine dans son état apo ou lié au GDP adopte la conformation observée par des structures cristallographiques EngA-GDP. Par contre, un changement structural a été observé en présence d'analogues du GTP, suggérant que les nucléotides triphosphate sont capables d'induire une nouvelle conformation. Cependant, cette conformation ne correspond à aucune structure

connue : la conformation active (liée au GTP) d'EngA reste toujours indéterminée. En outre, ce changement est observé uniquement en présence de hautes concentrations de nucléotide (10 mM), indiquant une faible affinité de l'enzyme pour le GTP. Des expériences de protéolyse limitée corroborent un changement de conformation nucléotide-dépendant : le profil de fragmentation d'EngA par la trypsine en absence et présence de GDP est identique, alors qu'en présence d'analogues du GTP la cinétique de protéolyse est modifiée. Ceci suggère que le nucléotide triphosphate engendre des changements structuraux qui rendent la protéine plus susceptible à l'action de la protéase. Les fragments issus de la protéolyse ont été analysés par séquençage N-terminal et spectrométrie de masse. Cinq régions ont été identifiées, dont une semble contribuer significativement à la différente cinétique de protéolyse : l'étiquette histidine N-terminale est plus activement clivée en présence d'analogues du GTP, suggérant un changement de la structure tertiaire d'EngA, possiblement par un mouvement du GD1, vers une conformation où l'extrémité N-terminale est plus accessible. Un test ELISA a été développé pour étudier les interactions entre EngA et le ribosome bactérien. Ces études ont montré que les changements conformationnels nucléotide-dépendants chez EngA modifient son affinité pour le ribosome : alors que le GDP déstabilise l'interaction EngA-ribosome, les analogues du GTP (à 10 mM) la renforcent. La faible affinité qu'EngA semble avoir pour les nucléotides, de l'ordre du millimolaire, est proche des concentrations basales intracellulaires de GTP (1–2 mM). Le rôle d'EngA pourrait alors être la détection des niveaux intracellulaires de nucléotides, de façon à moduler ses interactions avec le ribosome – et donc son assemblage – selon l'état énergétique de la cellule. Des expériences de cryo-microscopie électronique ont débuté afin de visualiser le complexe EngA-50S. Les données enregistrées sont encore en train d'être analysées.

Des études cristallographiques ont été menées avec l'objectif de déterminer la structure de l'état actif d'EngA liée au GTP. Plusieurs cristaux ont diffracté à une résolution suffisante pour pouvoir résoudre la structure d'EngA. Toutefois, malgré les nombreuses conditions de cristallisation essayées et les différents groupes d'espace obtenus, EngA est toujours observée dans sa conformation inactive. Nonobstant, nous avons pu rassembler quelques nouvelles informations sur EngA. Une analyse comparative de ces structures suggère une affinité distincte pour les deux domaines

catalytiques : alors que le domaine GD2 présente toujours un nucléotide fixé, des ions phosphate et sulfate présents dans la solution de cristallisation occupent facilement le site catalytique du GD1. Une de ces structures, obtenue à partir d'un cristal préparé dans l'absence de nucléotide, a révélé la présence d'un GDP lié au GD2, suggérant une très forte fixation du GDP. D'après cette analyse, le GD2 présenterait une forte affinité pour les nucléotides en opposition au GD1 (qui dans ce cas contribuerait pour la faible affinité d'EngA observé par SAXS). Ces observations s'opposent aux données biochimiques présentes dans la littérature qui montrent une affinité identique pour le GD1 et le GD2 tronqués. Ces résultats contradictoires pourraient être expliqués par une coopérativité entre les deux domaines dans la protéine *full-length* qui, quand les deux domaines sont étudiés séparément, n'est plus présente. Une de nos structures a, pour la première fois, montré un analogue du GTP fixé dans les deux domaines G. L'observation détaillée du site catalytique du GD1 révèle un résidu Gln en amont du le motif G3 qui stabilise le phosphate γ du nucléotide. Ainsi, cette glutamine pourrait avoir un rôle catalytique identique aux GTPases Ras-like, ce qui impliquerait un mécanisme d'hydrolyse différent pour le GD1 et le GD2 – pour lequel un tel résidu catalytique n'a pas été identifié. Malgré la présence d'un analogue du GTP dans les deux domaines, la conformation d'EngA reste inchangée. Cette conformation inactive d'EngA semble être très stable : l'analyse de l'empilement cristallin des plusieurs cristaux a montré une interface conservée entre les molécules symétriques qui est possible uniquement avec cette conformation. Ainsi, la conformation inactive serait sélectionnée pendant la nucléation au détriment d'autres conformations possibles. Cette prédisposition pourrait être modifiée par des mutations qui perturbent ces contacts cristallins.

Le test ELISA a été adapté au criblage de composés afin d'identifier des molécules capables d'interférer avec les interactions EngA-ribosome. Parallèlement, des essais de cristallisation ont été réalisés afin de déterminer la structure des complexes EngA-ligands. Le caractère hydrophobe des composés testés a rendu difficile soit les essais biochimiques, soit les efforts de cristallisation. Cependant, des données préliminaires suggèrent que certains composés seraient capables d'interférer avec les interactions d'EngA. Deux scénarios peuvent être considérés : soit les inhibiteurs empêchent les interactions EngA-ribosome, soit ils empêchent la dissociation des deux partenaires. En tous les cas, la fonction d'EngA, et donc l'assemblage du

ribosome, serait compromise. EngA semble donc être une bonne cible thérapeutique pour le développement de nouveaux antibiotiques. Des efforts pour dévoiler ses mécanismes sont encore nécessaires. La compréhension de tels mécanismes serait un pas important pour les travaux futurs visant la lutte contre la résistance aux antibiotiques.

Contents

List of tables

List of figures

Abbreviations

A. thaliana *Arabidopsis thaliana*
aa amino acid
A-site aminoacyl-site
AlFx aluminium fluoride
AMR antimicrobial resistance
AU absorbance unit
B. subtilis *Bacillus subtilis*
BLI bio-layer interferometry
BSA bovine serum albumin
C. crescentus *Caulobacter crescentus*
CDC Centers for Disease Control and Prevention
CDDEP Center for Disease Dynamics, Economics & Policy
CEB-II-5x chemical library of compounds (CEB-II-54, CEB-II-55, CEB-II-56)
CRE carbapenem-resistant Enterobacteriaceae
D_{max} maximum dimension
DMSO dimethyl sulfoxide
DNA deoxyribonucleic acid
dNTP deoxynucleotide triphosphate
DOC deoxycholate
DTT dithiothreitol
E. coli *Escherichia coli*
E-site exit-site
ECDC European Centre for Disease Prevention and Control
EDTA ethylenediaminetetraacetic acid
ELISA enzyme-linked immunosorbent assay
EM electron microscopy
FKB ForteBio kinetics buffer
GAP GTPase-activating protein
GDI Guanine nucleotide dissociation inhibitor
GEF Guanine nucleotide exchange factor
GD G-domain
GD1 G-domain 1
GD2 G-domain 2
GDP guanosine diphosphate
Gln^{cat} catalytic glutamine
GMPPCP/GMPPNP GTP analogues
GTP guanosine triphosphate
GST glutathione S-transferase
GV EngA variant G139C/V388C
H. influenza *Haemophilus influenza*
HAS-GTPase hydrophobic amino acid substituted GTPase
HBS HEPES-buffered saline
HPLC high-performance liquid chromatography
I_0 forward scattering intensity
IA interface area
IEC ion exchange chromatography
IMAC immobiliseed metal-affinity chromatography
IPTG isopropyl β-D-1-thiogalactopyranoside
ITC isothermal titration calorimetry
K. pneumoniae *Klebsiella pneumoniae*

k_{cat} catalytic constant
k_a association rate
K_D dissociation constant
KH K-homology
K_M Michaelis constant
LB Luria broth cell culture medium
LC-ESI-TOF liquid-chromatography electrospray-ionisation time-of-flight
LDH lactate dehydrogenase
LLG log likelihood gain
M. smegmatis *Mycobacterium smegmatis*
Mc main chain
mAb monoclonal antibody
mRNA messenger RNA
MS mass spectrometry
MRSA methicillin-resistant Staphylococcus aureus
MW molecular weight
N. benthamiana *Nicotiana benthamiana*
N. gonorrhoeae *Neisseria gonorrhoeae*
NAD^+/NADH oxidised/reduced nicotinamide adenine dinucleotide
OD optical density
(p)ppGpp guanosine (penta)tetraphosphate
P-site peptidyl-site
PCR polymerase chain reaction
PDB Protein Data Bank
PEG polyethylene glycol
PEP phosphoenolpyruvate
PK pyruvate kinase
PMSF phenylmethylsulfonyl fluoride
PTC peptidyltransferase center
PVDF polyvinylidene fluoride
r-protein ribosomal protein
RA-GTPase ribosome assembly GTPase
R_g Radius of gyration
rmsd root mean square deviation
RNA ribonucleic acid
rRNA ribosomal RNA
RU response units
tRNA transfer RNA
S. aureus *Staphylococcus aureus*
S. pneumoniae *Streptococcus pneumoniae*
S. typhimurium *Salmonella typhimurium*
SA specific enzymatic activity
SAXS small-angle X-ray scattering
SDM site-directed mutagenesis
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC size exclusion chromatography
SIMIBI signal-recognition particle, MinD, BioD related
SPR surface plasmon resonance
SwI / SwII switch I / switch II regions
t_{50} time at which full-length protein is reduced by half upon proteolysis
T. maritima *Thermotoga maritima*
TBS tris-buffered saline
TBST tris-buffered saline tween-20

TCA trichloroacetic acid
TFA trifluoroacetic acid
Tm melting temperature
TMB 3,3',5,5'-tetramethylbenzidine
TRAFAC translation factor-related
TSA thermal shift assay
$\mathbf{V_{max}}$ maximal velocity
$\mathbf{V_p}$ Porod volume
WHO world health organisation
WT wild-type
YT EngA variant Y51C/T169C

Context

The development of new therapeutic options against bacterial infections has aroused great interest over the last years in the context of drug resistance. In the drug discovery and development pipeline, a long path stands before a molecule can enter clinical trials and lastly be approved. The starting-point in the pursuit of new antibiotics for which bacterial resistance mechanisms do not pre-exist is the identification of novel cellular targets. An ideal drug target should have three main characteristics: play a critical role in an essential metabolic pathway in the bacterial cell; be conserved among pathogens (in the context of the search for broad-spectrum antibiotics); and be absent from the human genome to avoid adverse drug reactions in the human host. Academic research is a key component in the early stages of drug discovery: genetics has been of central importance in the identification of bacterial genes whose products represent potential drug targets, and the structural and biochemical characterisation of such enzymes is a major contributor for the implementation and validation of hit discovery approaches.

Genetics studies in the early 2000s have identified *engA* as a conserved bacterial gene that is not present in the human genome and whose product is an enzyme essential for cell survival. These initial studies revealed that EngA exhibits characteristics of an ideal drug target and several studies have since contributed to its biochemical and structural characterisation.

EngA is now known to be a GTPase that acts as an assembly factor for the bacterial ribosome. Although the bacterial ribosome is one of the major bacterial targets for the action of antibiotics, the pathway of ribosome assembly has remained unexploited.

The aim of this work was to study EngA as a target for inhibitors of the assembly process. A brief bibliographic summary of what is known on bacterial ribosome biogenesis and EngA is presented. The results from our work and future perspectives are discussed next and will hopefully give some useful insights into the characterisation of EngA as a potential bacterial drug target.

INTRODUCTION

1. Antibiotic resistance

The advent of the first antibiotics, approximately 70 years ago, introduced a period where it became possible to treat bacterial infectious diseases.

The antibiotic era began in the first half of the twentieth century, with the introduction by Paul Ehrlich of the idea that chemical compounds could selectively target disease-causing microorganisms. The discovery of penicillin by Alexander Fleming and the first sulfonamide by Gerhard Domagk and co-workers were the breakthrough for the “golden” antibiotic era, between the 1940s and 1970s, when many new classes of antibiotics were discovered. Antibiotic discovery was one of the major medical advances, contributing significantly to the public (control of endemic and epidemic diseases) and individual health (reduction of morbidity, mortality and increase of life expectancy). However, the discovery rate of novel classes has decreased since the 1980s, with very few new antibiotics being discovered or approved for clinical use since then.

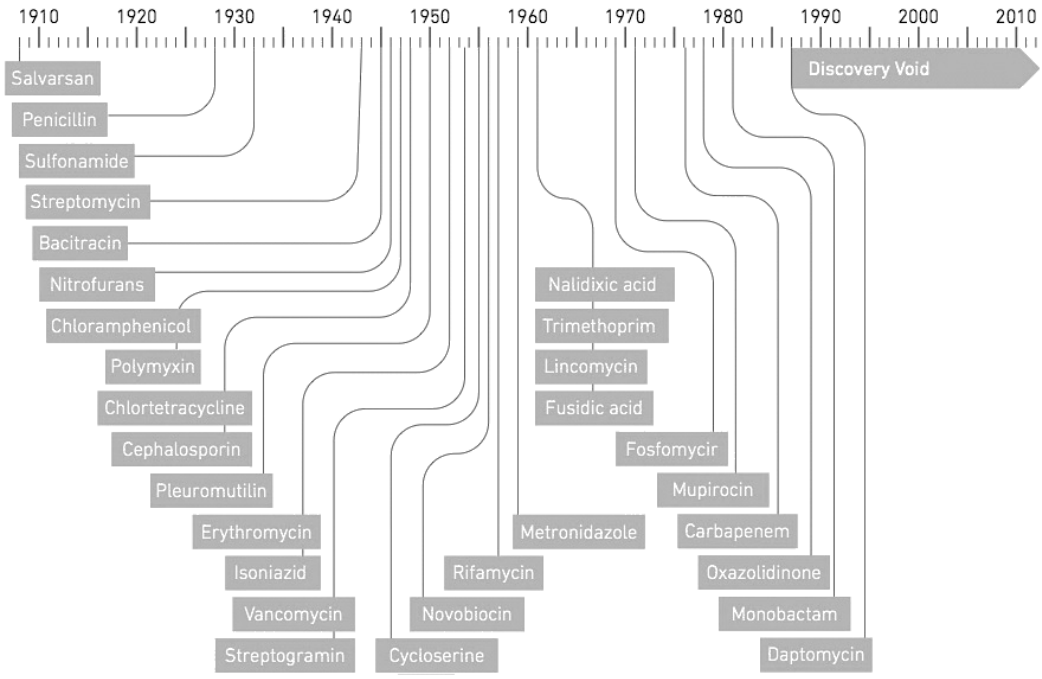

Fig. 1 Discovery dates of classes of antibacterial drugs. While between 1940 and 1980 about 20 new classes of antibiotics were identified from natural substances, no major advances have been made since then. Since the 1980s, most of the approved antibiotics were members of already known classes. Adapted from WHO, 2014.

Along with the introduction of antibiotics came the emergence of resistance in pathogenic bacteria. Antimicrobial resistance (AMR) is the ability of a microorganism to resist to drugs that would normally kill it or limit its growth. Early after penicillin discovery, Fleming warned for the possibility of microbes becoming resistant in case of penicillin underdosage (Nobel Prize Lecture: *Penicillin*, 1945). Indeed, resistance started to emerge years after the large-

scale production of penicillin. Since then it has been transmitted to many pathogenic bacteria and it concerns all antibiotic classes.

Evolution of resistant strains occurs spontaneously in nature, either by spontaneous mutations or by acquisition of genes (i.e. horizontal gene transfer) from environmental microbiota, and it precedes the introduction of antibiotics (Martínez, 2012). However, since their introduction, the use and misuse of antibiotics has accelerated this process by elimination of susceptible organisms and selection of the resistant ones. A study by Alonso and collaborators reported an increase in *P. aeruginosa*'s mutation rate by 10^5 times after a 4-days incubation with tetracyclines (Alonso et al., 1999). In *E. coli*, a linear relationship is observed between the mutation rate and the antibiotic concentration used (Long et al., 2016).

As a direct consequence of antibacterial resistance, standard treatments become ineffective and infections persist and spread to others, with a huge impact in public health. *E. coli* is the most common cause of urinary tract infections. *E. coli* resistance to third generation cephalosporins and fluoroquinolones has attained more than 50% in some countries (WHO, 2014). For such strains, carbapenems remain the only available treatment. Public health implications for such high resistance proportions imply less specificity of treatment and increased adverse-side effects (by the use of broad-range therapy), longer hospital stays, high medical costs and expansion of resistance to last-resort antibiotics (such as carbapenems). Reports of carbapenem-resistant Enterobacteriaceae (CRE) confirm this reality, with a rising number of deaths from untreatable infections caused by CRE *E. coli* and *K. pneumonia* all around the world (CDDEP, 2015; WHO, 2014). *S. aureus* is another common cause of healthcare-associated infections. Over the last decade, methicillin-resistant *S. aureus* (MRSA) strains, which were initially limited to hospital-acquired infections, now account for more than 80% of community-acquired infections in some countries (CDDEP, 2015).

Facing this situation, international nongovernmental health agencies together with governments and private entities around the world have launched initiatives to stop the spread of antimicrobial resistance. In the occasion of the 2011 World Health Day, dedicated to the combat against drug resistance, the WHO introduced a policy package with critical actions to struggle against the spread of antimicrobial resistance, calling on «all key stakeholders (…) to act and take responsibility for combating antimicrobial resistance». In November 2015, several worldwide activities such as the European Antibiotic Awareness Day, the United States Get Smart About Antibiotics Week, the Canadian Antibiotic Awareness Week and the Australian Antibiotic Awareness Week were simultaneously held in a global effort to warn for antibiotic resistance. In September 2016, antimicrobial resistance

was addressed for the first time in the United Nations General Assembly to define a coordinated approach across multiple sectors to be adopted and reported in a two-years time.

Despite the global efforts to tackle antimicrobial resistance, this is still a mounting problem that is far from being solved. Upon the *2014 Surveillance Report on Antimicrobial Resistance* (WHO, 2014), the WHO stated that «without urgent, coordinated action, the world is heading towards a post-antibiotic era, in which common infections and minor injuries (…) can once again kill».

The development of new initiatives for the treatment of bacterial infections is crucial to stop antibiotic resistance. Alternative therapies to complement antimicrobials are being developed. Therapies that target the host are based on the fact that signs and symptoms upon infection result from a response driven by the host's immune system. By targeting the host rather than the microbe, immunomodulation can avoid selective pressure for the development of resistance (Hancock et al., 2012). The use of anti-inflammatory monoclonal antibodies (mAbs) is a potential strategy to reduce pathogen-mediated tissue damage and help recovery (Spellberg et al., 2013). Conversely, immune stimulation has been suggested as a way of enhancing antibiotics' action by promoting an appropriate immune response and is currently in clinical development (Czaplewski et al., 2016). Vaccines have also been in development, although their action is restricted to prophylaxis. mAbs can have a direct antimicrobial mechanism, by binding to microbes, virulence factors or toxins. This strategy is currently in clinical trials (Czaplewski et al., 2016). Their high specificity for their target and the exploitation of mechanisms different from those of antibiotics, make mAbs unlikely to select for mAbs resistance among non-targeted microorganisms or for cross-resistance (Bebbington and Yarranton, 2008; Saylor et al., 2009). Modulation of the microbe growth environment can potentially be done by metal-chelation sequestration of nutrients, so that microbes do not proliferate. Another approach that has reached phase III clinical development is the use of probiotics that compete with microbial growth (Czaplewski et al., 2016; Spellberg et al., 2013). Phage lysins that destroy bacterial cell wall and reduce formation of biofilms can be used in combination with antibiotics. Bacteriophages could eventually replace antibiotics as not only they directly attack bacteria, but also have the ability to replicate and evolve to adapt to the pathogenic strain (Czaplewski et al., 2016).

The return to the investment on new antibiotics is also critical to overcome resistant microbes. Different strategies can be employed in the discovery of new antibiotics. Examples include the exploration of new sources of natural antibiotics, such as marine or extremophilic bacteria, or the search for novel cellular targets (Ventola, 2015). Genetics studies have revealed genes that are critical to the survival of bacteria or that are responsible for virulence

mechanisms. Such genes represent potential new targets for antibiotic design. Identification of bacterial essential enzymes that do not present susceptibility to resistance mechanisms has become an important approach for antibiotic development. Such is the case of the *engA* gene, now known to encode for an essential bacterial GTPase, which was initially discovered in a study aiming at identifying genes responsible for pilin variation and adaptation in bacteria (Mehr et al., 2000).

This work focuses on the EngA bacterial GTPase as a potential target for antibiotic design. EngA is conserved in bacterial species and essential for cell viability. Since its identification in 2000 by Mehr and collaborators, studies on EngA have revealed a role in bacterial ribosome assembly (Bharat and Brown, 2014; Bharat et al., 2006; Foucher et al., 2012; Hwang and Inouye, 2006; Jeon et al., 2014; Lamb et al., 2007; Schaefer et al., 2006; Tan et al., 2002; Tomar et al., 2009; Zhang et al., 2014). While the ribosome itself is a drug target for many of the available antibiotics (Fig. 2), its biogenesis process has remained unexploited from a pharmacological perspective.

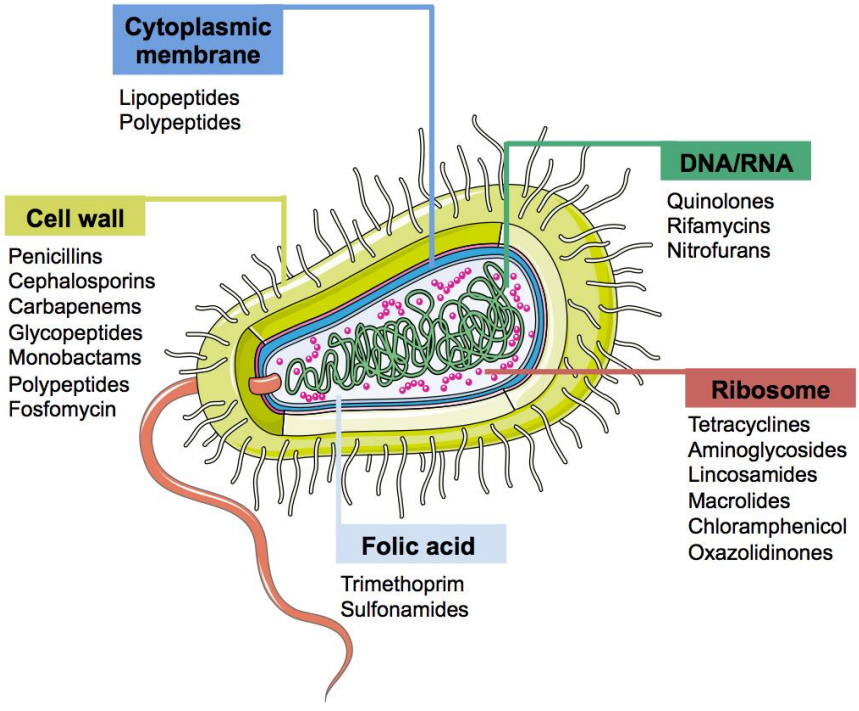

Fig. 2 Major targets for the action of antibacterial drugs. Antibiotics can interfere with the cell wall, mainly by inhibiting the synthesis of peptidoglycan. Others injure the cytoplasmic membrane, leading to the depolarisation and leakage of cell content. Folic acid is essential for the synthesis of nucleic acids in cells and can be inhibited by the action of antibacterials that interfere with its synthesis pathway. Inhibition of nucleic acid synthesis is achieved by antibiotics that interfere with enzymes such as DNA gyrases, topoisomerases or RNA polymerases. Finally, the protein biosynthesis pathway in bacteria is inhibited by drugs that directly bind the ribosome subunits and alter their translational capability. Modified from Servier Image Bank.

The absence of exploitation of the ribosome assembly pathway together with the ubiquitous and essential nature of EngA and the lack of a human ortholog make EngA a promising therapeutic target for the development of new antibacterial agents.

2. Bacterial Ribosome

Ribosomes are ribonucleoprotein cellular machineries responsible for the translation of the genetic code into proteins. These catalytic machines are present in all domains of life, sharing a common structural core that harbours the functional centres, highlighting their importance in the fundamental process of protein biosynthesis (Yusupov, 2014).

2.1. Structure

Since the 1970s, when the shape of the bacterial ribosome and its subunits was for the first time reconstituted by electron microscopy images (Lake, 1976), a great progress has been made on the determination of the ribosome structure. During the last decades, studies on X-ray crystallography and cryo-electron microscopy, along with advances in synchrotron light sources and crystallographic hardware and software, have lead to the elucidation of the ribosome atomic structure (Ban et al., 2000; Schuwirth et al., 2005; Wimberly et al., 2000). With this information, the understanding of protein biosynthesis pathways and antibiotics' mechanisms of action was made possible.

All ribosomes – prokaryotic, eukaryotic or organellar – consist of two unequal subunits responsible for protein biosynthesis in cells: a large subunit, that contains the catalytic site for peptide-bond formation, and a small subunit, responsible for mRNA binding and decoding. With the exception of the mammalian mitochondria ribosome, that shows a reversed ratio, ribosomes are composed of one-third protein and two-thirds RNA (Ramakrishnan, 2009). The next sections will present a description of the bacterial ribosome, which can be explored as a target for antibacterial drugs. The following references contain information on the structure of chloroplast, mitochondrial and eukaryotic ribosomes, respectively: Agrawal et al., 2011; Greber and Ban, 2016; Yusupova and Yusupov, 2014.

The basic architecture of the 70S ribosome and its subunits is represented in Fig. 3. The 70S ribosome is an asymmetrical assembly of more than 50 proteins and 3 RNA chains, with a ratio by weight of 2:1 RNA:protein. It is built of a small subunit 30S and a large subunit 50S, that associate to form a functional ribosome with a molecular weight of 2.5 MDa, a sedimentation coefficient of 70S and a diameter of ~20 nm (Kaczanowska and Rydén-Aulin, 2007; Nelson and Cox, 2005).

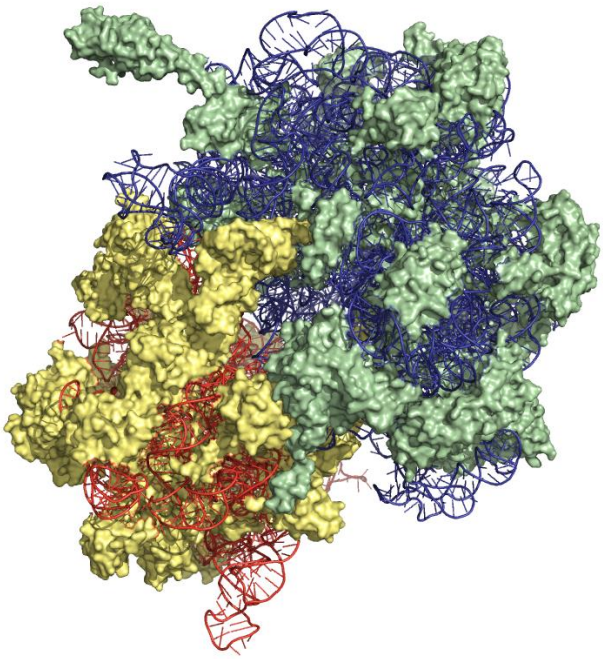

Fig. 3 Structure of the *E. coli* 70S ribosome. The 50S subunit is represented in green (proteins) and blue (rRNA) and the 30S subunit in yellow (proteins) and orange (rRNA). Image edited in PyMol from PDB file 4V4Q.

The 50S subunit (1.45 MDa) consists of two rRNA chains, 23S (~2900 nucleotides) and 5S (~120 nucleotides), and 34 r-proteins, L1–L34, that range from 5–30 kDa (Fig. 4A). It is responsible for the catalysis of peptide bonds linking, providing the peptidyltransferase-centre (PTC). Its crown-like shape encompasses a central protuberance, formed by 5S rRNA and associated r-proteins, and two flanking arms: the L1-stalk, formed by r-protein L1 and rRNA, and the L7/L12-stalk, formed by r-proteins L7 and L12 associated with L10, L11 and rRNA (Yusupov, 2014). The PTC is located in the centre of the interface side of the 50S subunit, where peptide bond formation occurs. Just below the PTC, there is the polypeptide exit tunnel of about 100 Å long and 25 Å in diameter that accommodates the nascent peptide (up to 40 amino acids) (Kaczanowska and Rydén-Aulin, 2007).

The 30S subunit (0.85 MDa) contains a single RNA chain 16S (~1500 nucleotides) and 21 r-proteins, S1–S21, that range from 8–60 kDa (Fig. 4B). The small subunit provides the decoding-centre, responsible for decoding mRNA and controlling translation fidelity. The shape of the 30S is segmented into 3 main regions, defined by the secondary structure of 16S rRNA: the body (domain I), formed by the 5’ end; the platform (domain II), formed by the central part of the rRNA; and the head (domain III), formed by the 3’ end. A narrow region between the head and the platform, called neck, forms a cleft where the decoding centre is located (Kaczanowska and Rydén-Aulin, 2007).

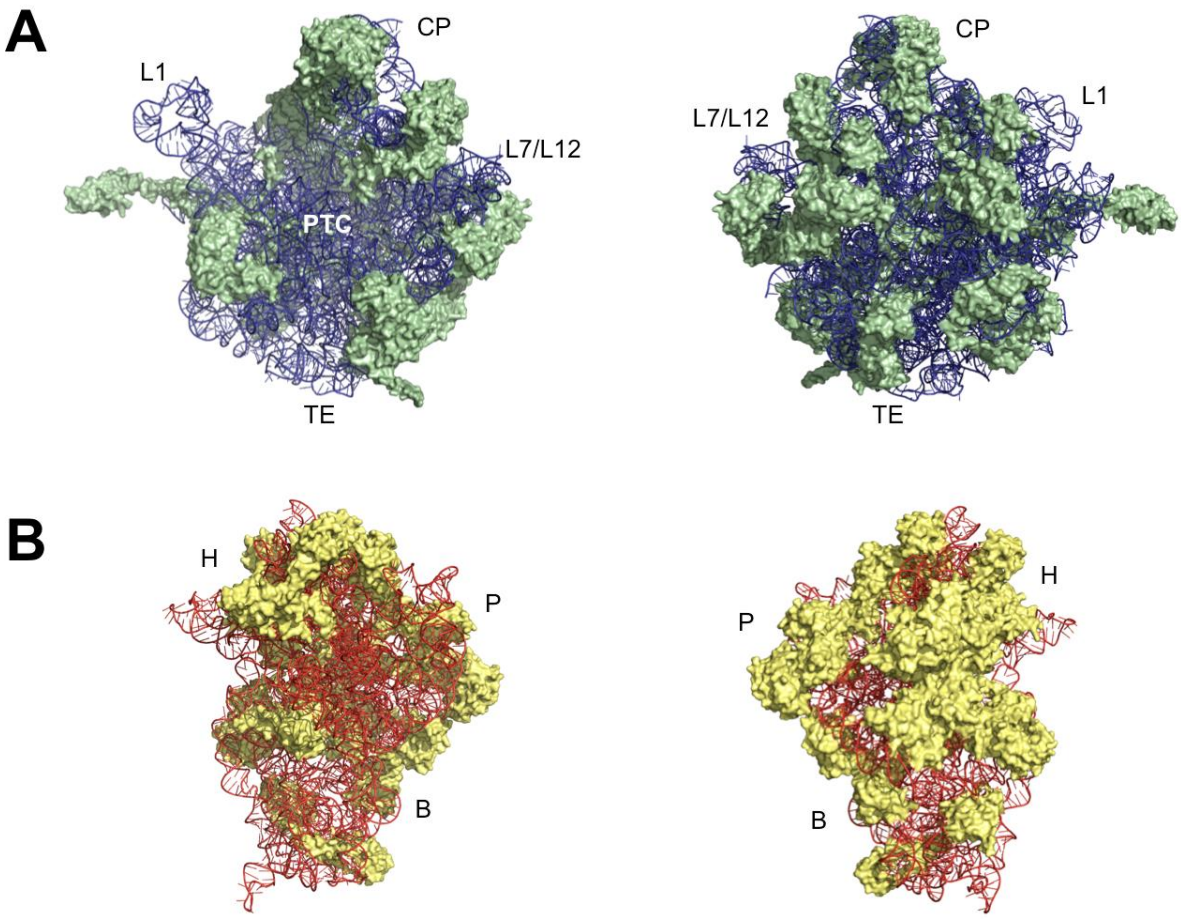

Fig. 4 Structure of the *E. coli* **(A)** 50S and **(B)** 30S ribosomal subunits. The left images represent the internal view of the subunits and the images on the right the exterior view. The 50S subunit is represented in green (proteins) and blue (rRNA) and the 30S subunit in yellow (proteins) and orange (rRNA). L1: L1-stalk; L7/L12: L7/L12-stalk; CP: central protuberance; PTC: peptidyltransferase centre; TE: tunnel exit; H: head; P: platform; B: body. Image edited in PyMol from PDB file 4V4Q.

It is at the interface between the two subunits that the functional regions of the ribosome are located. This interface consists mainly of RNA, with no protein within 18 Å of the active site, suggesting that the ribosome is a ribozyme (Yusupov, 2014). While the core of the subunits is made of rRNA chains, r-proteins are located at the surface. Most r-proteins contain a globular domain, which stands at the surface of the subunit, and an extended domain that protrude into the rRNA to stabilise its tertiary structure (Yusupov, 2014).

Association of the 30S and 50S forms a tunnel in their interface through which mRNA passes during translation and where the codons of mRNA interact with the anticodons of the tRNAs. Three tRNA-binding sites are present: aminoacyl (A)-site, peptidyl (P)-site and exit (E)-site. The A-site binds the aminoacyl-tRNA to be presented to the mRNA. The P-site carries the nascent peptide chain attached as a peptidyl-tRNA complex. The E-site harbors deacylated-tRNAs ready to leave the ribosome (Lafontaine and Tollervey, 2001; Schmeing and Ramakrishnan, 2009).

2.2. Protein synthesis

Bacterial translation can be divided into three main stages: initiation, elongation and termination (Fig. 5) (Nelson and Cox, 2005; Schmeing and Ramakrishnan, 2009).

The initiation of polypeptide synthesis in bacteria requires the 30S and 50S subunits, the mRNA coding for the peptide, the fMet-tRNA^{fMet}, three initiation factors (IF-1, IF-2 and IF-3), Mg^{2+} and GTP. Initiation starts with binding of IF-1 and IF-3 to the 30S subunit, which stimulate the release of mRNA and deacylated tRNA from the previous cycle and prevent association of the 50S subunit. The mRNA binds to the small subunit and the initiation codon (5')AUG is guided to its correct position at the P-site. This alignment occurs through the interaction between the Shine-Dalgarno sequence in the mRNA, a sequence of 4–9 purine bases (8–18bp) upstream the start codon, and a complementary pyrimidine-rich sequence in the 16S rRNA. The GTP-bound initiation factor IF-2 and the fMet-tRNA^{fMet} bind to the complex, allowing the pairing of the tRNA anticodon and the mRNA start codon. GTP hydrolysis by IF-2 induces the release of the initiation factors and the binding of the 50S subunit, forming a functional 70S ribosome – the initiation complex.

Elongation requires the initiation complex, aminoacyl-tRNAs, three elongation factors (EF-Tu, EF-Ts, and EF-G) and GTP. In the first elongation step, the incoming aminoacyl-tRNA binds to the elongation factor EF-Tu GTP-bound. The resulting aminoacyl-tRNA–EF-Tu–GTP ternary complex binds the A-site of the initiation complex. EF-Tu then hydrolyses GTP and dissociates from the complex, while the aminoacyl end of A-site tRNA moves into the PTC. Decoding is followed by peptide bond formation between the fMet and the new amino acid bound by their tRNAs to the P- and A-sites, respectively. Upon peptide bond formation, the nascent peptide on the P-site is transferred to the aminoacyl-tRNA on the A-site, leaving the P-site deacylated. Hydrolysis of GTP by EF-G catalyses translocation, where the tRNAs and the mRNA move relative to the ribosome, shifting positions of both tRNAs: the deacylated tRNA on the P-site is transferred to the E-site and the dipeptidyl-tRNA on the A-site is transferred to the P-site. The ribosome is now ready for the next elongation cycle and elongation will continue until the ribosome adds the last amino acid coded by the mRNA.

Termination is signalled by the presence of a termination codon (UAA, UAG, UGA). Once a termination codon occupies the A-site, release factors RF-1 or RF-2 will bind, leading to the cleavage and release of the nascent polypeptide from the tRNA in the P-site. RF-3 will bind to the ribosome–RF-1/2 complex and accelerate the dissociation of the release factors. At this stage, the mRNA and the deacylated tRNA on the P-site are released from the ribosome by the recycling factor RRF and EF-G. Finally, the 70S ribosome dissociates into its 30S and 50S subunits to start a new cycle.

An interactive overview of protein translation in bacteria can be found at Ramakrishnan's group web page (http://www.mrc-lmb.cam.ac.uk/ribo/homepage/mov_and_overview.html), from the MRC Laboratory of Molecular Biology.

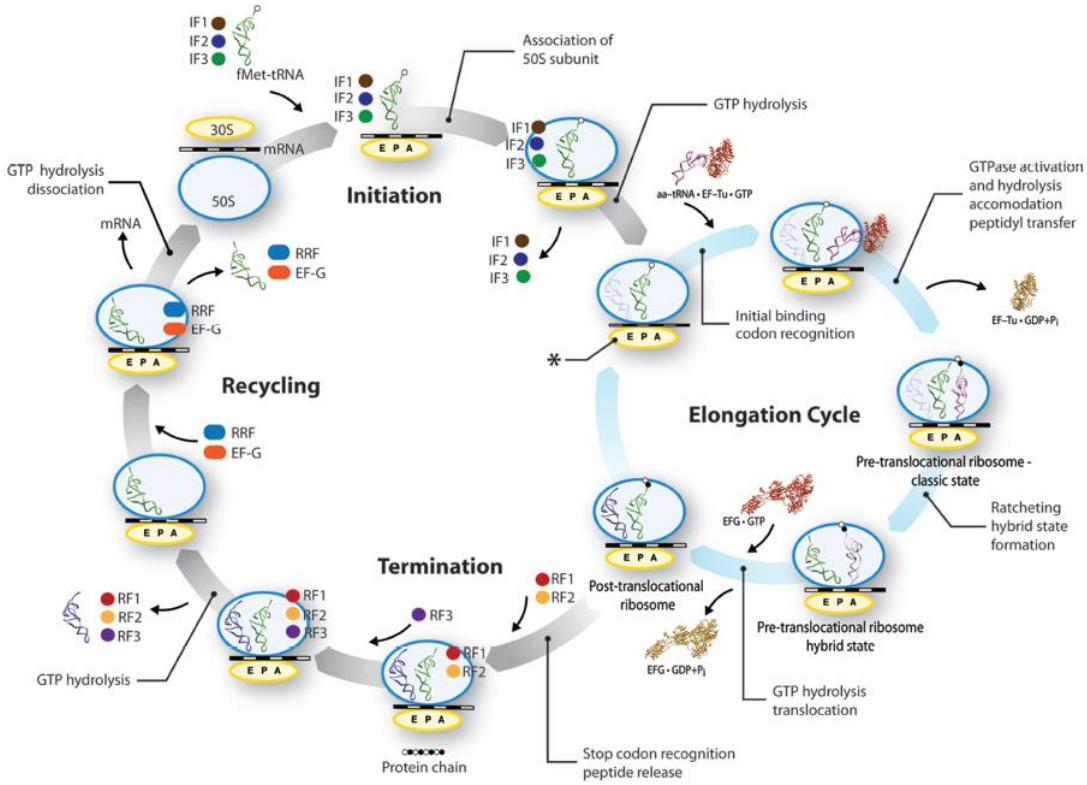

Fig. 5 Overview of bacterial translation. Adapted from Agirrezabala and Frank, 2010.

2.3. Inhibitors of bacterial translation

Protein synthesis is a major cellular target for antibiotic action, with approximately half of the available antibiotics targeting the translation machinery. Antibiotics that inhibit translation act by directly binding to the ribosomal subunits, affecting translational capability. With some exceptions, the functional centres of the ribosome – PTC, polypeptide exit tunnel and tRNA/mRNA interaction sites – constitute the binding sites for antibiotics (Wilson, 2009). A few examples of bacterial translation inhibitors are described below.

2.3.1. Inhibitors of translation initiation

The initiation stage of translation is inhibited by drugs that interfere with the formation of the initiation complex by preventing the binding of one interaction partner. Few antibiotics that inhibit the translation initiation step are available for clinical use.

GE81112 is a synthetic tetrapeptide antibiotic, representing a structurally unique class, still in research, that binds the 30S subunit and overlaps the P-site. GE81112 induces conformational changes in the 16S rRNA that lead to an incorrect binding of fMet-tRNA^{fMet} and mRNA to the 30S subunit, preventing subsequent binding of 50S and progression of the initiation stage (Fabbretti et al., 2016).

G1(MW297) is a synthetic nitrovinylfuran broad-spectrum antibiotic that binds the 30S subunit, overlapping the P-site and preventing binding of the fMet-tRNA^{fMet} (Fabbretti et al., 2012).

GSK1322322 is a synthetic hydrazide and represents a new class of antibiotics, having recently completed phase II clinical trial. It targets peptide deformylase, a bacterial enzyme required for modification of the N-terminus of the nascent polypeptide chain in prokaryotes (Corey et al., 2014).

2.3.2. Inhibitors of translation elongation

Many of the currently available antibiotics target the elongation step of translation.

Tetracyclines are broad-spectrum antibiotics that bind reversibly to the head of the 30S subunit, blocking the binding of aa-tRNAs and preventing addition of amino acids to the nascent peptide.

By irreversibly binding the 30S, aminoglycosides are bactericidal agents that interfere with a loop comprising the decoding site. Aminoglycosides interfere with regions of the 16S rRNA that are responsible for ensuring the binding of a cognate aa-tRNA to the mRNA during decoding. The presence of an aminoglycoside results in misreading and misincorporation of amino acids. Other members of this antibiotic class bind the mRNA in the P-site of the PTC, inhibiting translocation.

Chloramphenicol is a bacteriostatic drug that reversibly binds to the A-site of the PTC on the 50S subunit, overlapping the aminoacyl moiety of aa-tRNAs and inhibiting the peptidyltransferase step. Oxazolidinones, such as linezolid, have a similar mechanism.

Macrolides are bacteriostatic and bactericidal drugs that act by binding to the exit tunnel in the 50S subunit, adjacent to the PTC. This blocks aminoacyl translocation and the progression of the nascent polypeptide chain and also prevents formation of initiation complexes. Ketolides are semi-synthetic macrolides derivatives developed to escape

resistance mechanisms. Structural modification of the sugar moiety renders ketolides poor substrates for efflux pumps and increases ribosome binding affinity (Katzung, 2006).

Lincosamides bind to the PTC in the 50S subunit and overlap with both binding sites of chloramphenicol and macrolides by binding to A-site of the PTC and extending into the exit tunnel. They inhibit peptide growth, translocation of aminoacyl and formation of initiation complexes.

Streptogramines are bactericidal drugs that also bind adjacent sites on the PTC, interfering with binding of tRNA to both A- and P-sites, and on the exit tunnel.

2.3.3. Inhibitors of translation termination and recycling

Fusidic acid is a steroidal antibiotic that binds to EF-G in complex with the ribosome. Binding of fusidic acid allows GTP hydrolysis by EF-G but prevents conformational changes of the GTPase and hence its release from the ribosome.

2.4. Ribosome biogenesis

While numerous antibiotics inhibit the translating ribosome, none is directed towards the pathway of ribosome assembly. The assembly of such complex cellular machineries has been subject of intensive study for the last decades and important steps and factors have been unravelled.

2.4.1. *In vitro* assembly mechanisms

In the late 1960s, work from the laboratories of Masayasu Nomura (Traub and Nomura, 1968) and Knud Nierhaus (Nierhaus and Dohme, 1974) demonstrated that both ribosomal subunits could be broken down into rRNA and r-proteins and reconstituted to functionally active particles *in vitro*, in an orderly and self-assembly process. The observations from studies on ribosome biogenesis *in vitro* resulted in assembly maps for the 30S and 50S subunits (Fig. 6).

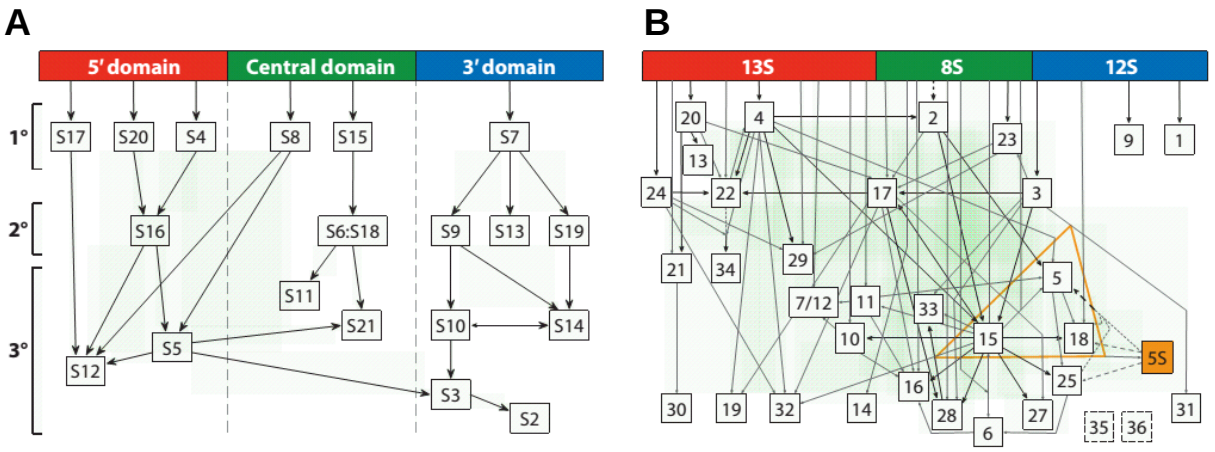

Fig. 6 Assembly maps of ribosomal subunits. **(A)** Nomura assembly map represents the hierarchy of protein binding in the 30S subunit. 16S rRNA is represented in three domains (domain I in red; domain II in green; domain III in blue) and r-proteins are classified as primary (1º), secondary (2º) or tertiary (3º). **(B)** Nierhaus assembly map represents the hierarchy of protein binding in the 50S subunit. 23S is divided in three sections analogous to the 16S domains. Three r-proteins are responsible for binding of 5S rRNA (orange). Adapted from Shajani et al., 2011.

During 30S *in vitro* assembly, two intermediates are formed (Fig. 7A). Low temperature (0ºC) results in a first intermediate that sediments at 21S–22S. An increase in temperature to 40ºC induces a rearrangement of the particle, forming a second intermediate that sediments at 25S–26S. The final 30S subunit is formed by addition of the remaining proteins after a decrease in temperature (Shajani et al., 2011). The binding of ribosomal proteins to rRNA is an interdependent and hierarchical process, where early binding events dictate binding of late stage proteins. Ribosomal proteins are classified as primary proteins, which bind directly to rRNA, secondary proteins, which depend on primary binders, and tertiary binding proteins, which depend on secondary binders (Kaczanowska and Rydén-Aulin, 2007).

With almost twice the number of proteins and RNA molecules, the assembly of the 50S subunit is more complex and involves the formation of three intermediates, also depending on temperature and ionic conditions (Fig. 7B). The first intermediate (33S) is formed at low temperature and undergoes conformational changes that result in the second intermediate (41S–43S) at high temperature. The transition to the third intermediate (48S) occurs by addition of the remaining proteins. A functional 50S subunit is obtained by further increasing temperature and magnesium concentrations (50ºC, 20 mM $MgCl_2$).

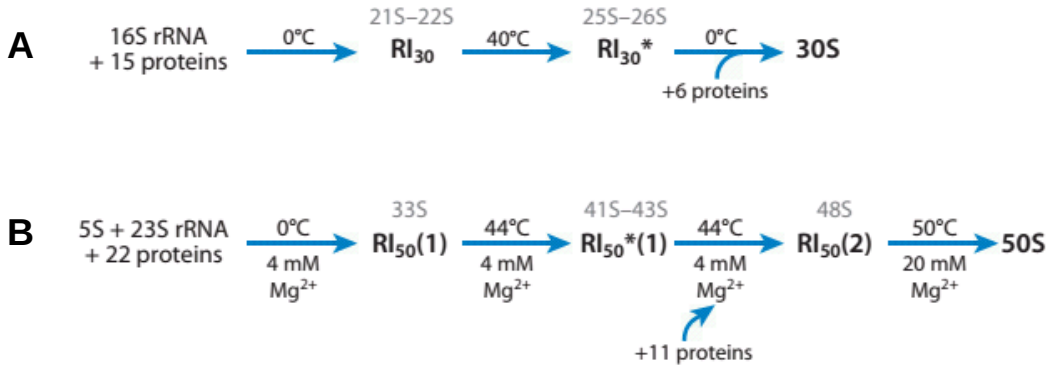

Fig. 7 Mechanism of the *in vitro* assembly of the **(A)** 30S subunit and **(B)** 50S subunit. RI stands for reconstitution intermediate. Adapted from Shajani et al., 2011.

2.4.2. *In vivo* assembly mechanisms

Although the observations obtained from *in vitro* systems yielded important information on ribosome assembly mechanisms, *in vivo* studies made clear that ribosome biogenesis in cells is a far more complex process.

In vivo ribosome biogenesis is a multi-step process regulated and assisted by protein factors, such as transcription factors, ribonucleases, rRNA helicases, chaperones, rRNA and r-protein modification enzymes and assembly factors. Biogenesis steps of the bacterial ribosome include (i) transcription, processing, modification and folding of rRNA, (ii) translation, modification and folding of r-proteins, (iii) binding of r-proteins to rRNA and (iv) binding/release of assembly factors (Shajani et al., 2011).

Transcription, processing and modification of rRNAs

In bacteria, the rRNA operon contains all three rRNAs that are transcribed as a single transcript with non-coding intervening sequences. Chemical modifications of rRNA – such as addition of carbonyl, methyl, amino or thiol groups, as well as isomerisation of uridine or pseudouridine – enable processing of the primary transcript by RNAses, generating the mature rRNA species 16S, 5S and 23S (Fig. 8) (Kaczanowska and Rydén-Aulin, 2007).

Post-translational modification of r-proteins

Ribosomal proteins are subject to post-translation chemical modifications, such as methylation or acetylation. These modifications are thought to regulate efficiency and specificity of protein binding to rRNA or to optimise binding sites of ligands (Kaczanowska and Rydén-Aulin, 2007).

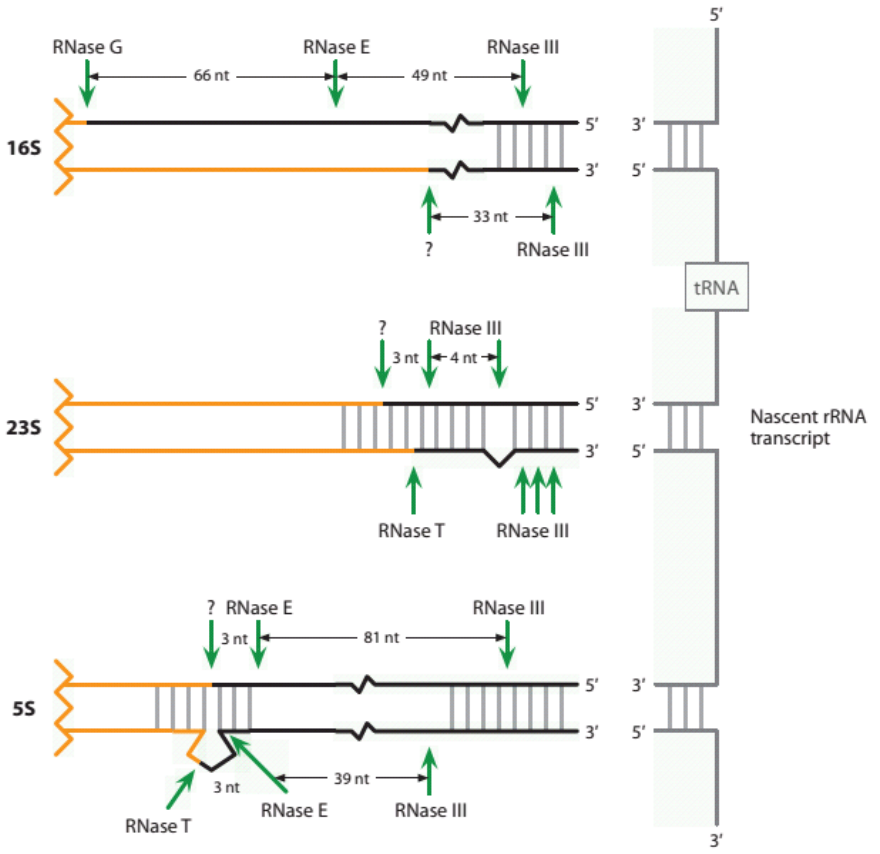

Fig. 8 Maturation of ribosomal RNA. The primary transcript is cleft by a series of at least five nucleases to obtain the three mature species. Initially, RNAse III cleaves the nascent rRNA transcript to separate precursor rRNA and tRNA, yielding the precursors of 16S (17S), 5S (9S) and 23S. The 17S precursor contains 115 additional nucleotides at the 5' end, cleft by RNAses E and G, and 33 nucleotides at the 3'end, cleft by a currently unknown RNAse. The 23S precursor contains 3–7 additional nucleotides at the 5' end that are cleft by RNAse III and an unknown RNAse, while the additional 7–9 nucleotides at the 3' end are cleft by RNAse T. The 9S precursor is further cleft at the 5' end by RNAses III, E and an unknown enzyme to remove 84 nucleotides, and at the 3' end by RNAses III, E and T to remove 42 nucleotides. Cleavage sites are represented by green arrows and mature rRNA are represented in orange. Question marks indicate unknown RNAses. Adapted from Shajani et al., 2011.

Assembly intermediates

Although in general the assembly mechanisms observed *in vivo* agree with those found *in vitro*, some discrepancies can be found in the assembly intermediates. These, despite the similar sedimentation coefficient, contain different rRNA and r-protein compositions.

The presence of rRNA precursors in cells, instead of mature rRNA used in *in vitro* systems, lowers the activation energy for assembly, suggesting that *in vivo* assembly is coupled with the transcription and maturation of rRNAs (Shajani et al., 2011). Assembly of the 30S subunit, for instance, is now known to occur from 5' to 3', when the formation of secondary structures of 16S rRNA generates binding sites for r-proteins before transcription is completed (Kaczanowska and Rydén-Aulin, 2007). Similarly, the first *in vivo* intermediate of the 50S is formed shortly after the onset of rRNA transcription (Wilson and Nierhaus, 2007).

It is still not clear whether there are one or multiple pathways for ribosome assembly *in vivo*. Recent observations support the presence of parallel assembly pathways in the ribosome. Despite the protein-binding hierarchy (depicted in the Nomura and Nierhaus assembly maps), this thermodynamic dependence between r-protein/rRNA interactions appears to be circumvented *in vivo* in strains deleted for primary binders (Shajani et al., 2011). In another study (Adilakshmi et al., 2008), 30S assembly was shown to take place in a first early stage – where it nucleates concurrently from different points along the 16S, forming different early assembly intermediates – and in a late stage – where assembly is driven by protein binding-induced fit.

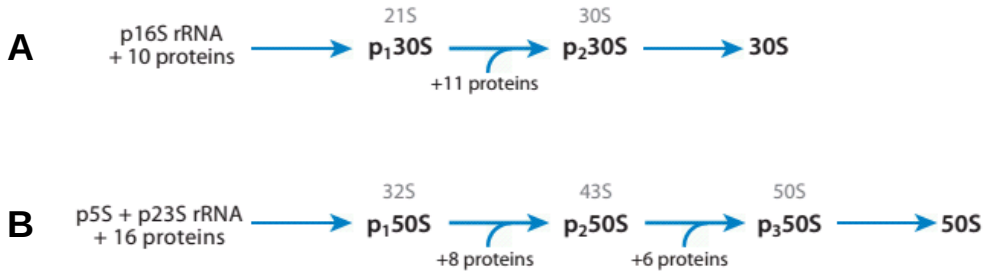

Fig. 9 Mechanism of the *in vivo* assembly of the **(A)** 30S subunit and the **(B)** 50S subunit. Adapted from Shajani et al., 2011**.**

2.5. Non-ribosomal proteins involved in ribosome subunit assembly

In vitro reconstitution of functional ribosomal subunits from purified components suggested that the information required for biogenesis was intrinsic to the rRNA and r-proteins. However, the non-physiological conditions required suggest the presence of additional exogenous factors that facilitate this process *in vivo* by preventing kinetic traps (Kaczanowska and Rydén-Aulin, 2007; Shajani et al., 2011; Wilson and Nierhaus, 2007).

Endoribonucleases and RNA/protein modification enzymes (methylases, acetylases, pseudouridinylases) have been described in the previous paragraphs as regulators of ribosome biogenesis in bacteria. In addition to these enzymes, other protein-factors participate directly in the subunit assembly process. It is the case of RNA helicases, chaperones and GTPases.

DEAD-box proteins are a family of conserved RNA helicases characterised by a specific motif Asp-Glu-Ala-Asp. They participate in the regulation of ribosome biogenesis, either by unwinding RNA secondary structures – helicase activity – or by assisting RNA folding and rearranging protein-RNA interactions – chaperone activity (Kaczanowska and Rydén-Aulin, 2007; Shajani et al., 2011; Wilson and Nierhaus, 2007).

The heat-shock proteins DnaK/DnaJ/GrpE and GroEL/GroES are chaperones that participate in later stages of 30S and 50S assembly, respectively. Chaperones act by getting RNA out of kinetic traps in the folding pathway, and by preventing intramolecular structures (Kaczanowska and Rydén-Aulin, 2007; Shajani et al., 2011; Wilson and Nierhaus, 2007).

GTPases constitute the largest class of essential ribosome assembly factors in bacteria (Karbstein, 2007). Although their precise function is still unclear, they are thought to act as checkpoints proteins by coupling the assembly of the ribosome with intracellular GTP concentrations (Britton, 2009). EngA, the subject of this work, belongs to this class of ribosome-assembly GTPases. The structural and biochemical features of GTPases and their possible roles in assembly are described in the next section.

3. GTPases

GTPases constitute a superclass within the P-loop NTPases fold that selectively binds and hydrolyses guanosine-5'-triphosphate (GTP) to guanosine-5'-diphosphate (GDP) and inorganic phosphate. Although there are fewer members of this superfamily in prokaryotes than in eukaryotes, they form a group of highly conserved and often essential enzymes in bacteria. Cellular functions for bacterial GTPases range from DNA replication, cell division, stress response, pathogenesis, protein synthesis and ribosome assembly (Verstraeten et al., 2011).

3.1. Structural and sequence features

Despite this functional diversity, GTPases share a conserved core structure and mechanism of action. A conserved ~20 kDa catalytic domain, called the G-domain, adopts an α/β fold, with a central β-sheet of at least 6 β-strands (β1–β6), forming the hydrophobic core, flanked by 5 α-helices (α1–α5) and 5 hydrophilic polypeptide loops (Fig. 10) (Paduch et al., 2001). Strong sequence identity between different GTPases subfamilies is observed in these loops, which define the GTPase superfamily. GTPases can consist solely of the G domain or have additional domains on the N- and C-termini (Caldon et al., 2001).

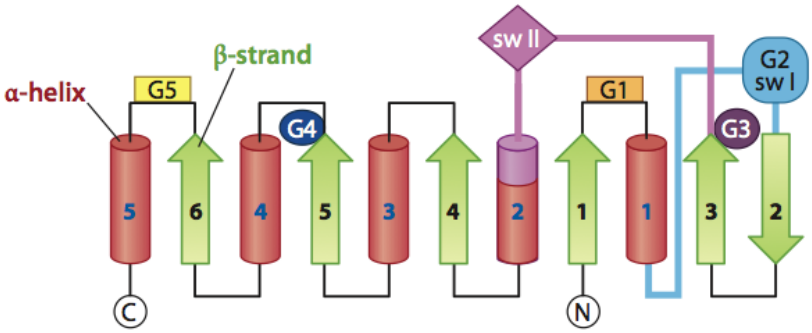

Fig. 10 Diagram of the G-domain. α-helices are represented in red (α1 to α5). β-strands are represented in green (β1 to β6). G motifs or loops are represented as G1 to G5. Adapted from Wittinghofer and Vetter, 2011.

The conserved sequences in GTPases, called G motifs (G1–G5), contain residues involved in nucleotide and metal binding and are critical for nucleotide exchange and hydrolysis and for conformational changes in the protein structure (Bourne et al., 1991; Wittinghofer and Vetter, 2011).

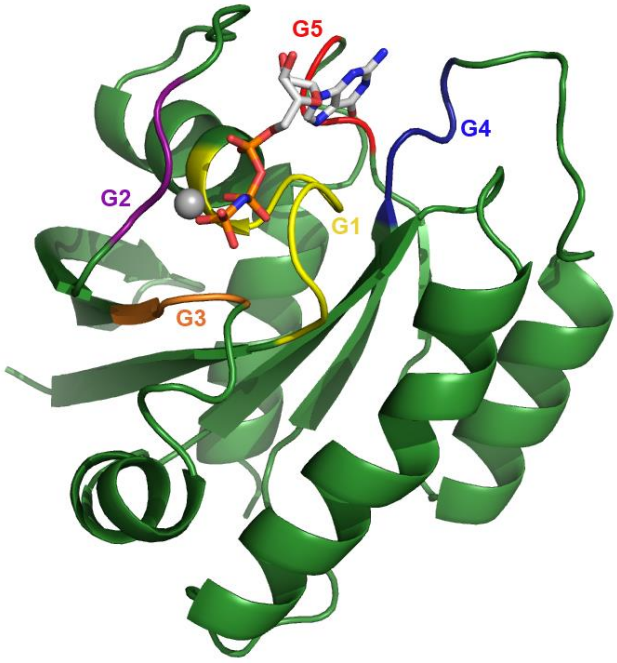

Fig. 11 3D structure of the G-domain from human Ras GTPase. The structure consists of a central β-sheet composed of six β-strands and five α-helices surrounding the core (Pai et al., 1990). The five G-motifs are highlighted in yellow (G1/Walker A/P-loop), violet (G2/Switch I), orange (G3/Walker B/Switch II), blue (G4) and red (G5). In this structure, Ras is bound to a GTP analogue (GMPPNP, represented in sticks) and to a magnesium ion (represented as a grey sphere). Image edited using PyMol from PDB 5P21.

The G1 motif corresponds to the Walker A motif or P-loop (sequence GXXXXGKS/T) common to P-loop NTPases and is located between the β1-strand and the α1-helix. The main role of the G1 motif is to stabilize the phosphate moiety thanks to main chain NH groups and the positively charged lysine side chain.

The G2 motif connects the α1-helix and the β2-strand and is involved in interactions with effector molecules. Its sequence differs among different GTPase subfamilies but it is highly conserved within the same subfamily. A conserved threonine residue changes its orientation upon nucleotide binding, allowing for conformational changes of the loop. This threonine is involved in the coordination of a Mg^{2+} ion through its side chain hydroxyl group. Due to these altered conformations, this loop is also called the switch I region.

The G3 motif, located at the N-terminus of the α2-helix, corresponds to a variation of the Walker B motif and has a characteristic hhhhDXXG sequence (h being a hydrophobic residue). Its aspartate is essential to the catalytic activity as it indirectly stabilises a Mg^{2+} ion through hydrogen bonds with the G2 threonine and a water molecule. Like G2, the loop harbouring G3 motif undergoes large conformational changes following binding to GTP or GDP and is therefore called the switch II region. This is the most flexible element of the catalytic domain (Paduch et al., 2001).

The G4 motif is located between the β5-strand and the α4-helix and contains a conserved sequence N/TKXD. It stabilises the guanine moiety of the nucleotide and is thus responsible for the specificity of GTPases for GDP/GTP over other nucleotides.

The G5 motif is located in a loop between the β6-strand and the α5-helix and contains the sequence SAK/L. Like G4, it has a role in the stabilisation of the guanine base through hydrogen bonds.

3.2. Guanine nucleotide binding

Binding of GTP/GDP occurs via interaction with residues from the five G motifs and with a Mg^{2+} ion (Fig. 12).

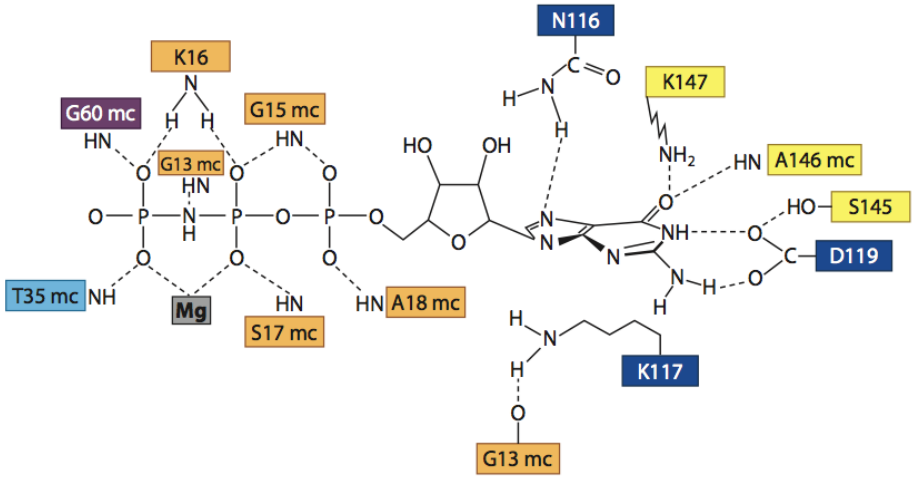

Fig. 12 Guanine nucleotide binding site. Represented as G1 residues: orange; G2 residues: light blue; G3 residues: purple; G4 residues: dark blue; G5 residues: yellow; Mg: magnesium ion; mc: main-chain. Adapted from Wittinghofer and Vetter, 2011.

The α-/β- or β-/γ-phosphate groups of the di- or tri-phosphate nucleotide, respectively, are stabilised by an electrostatic interaction with the side-chain of the conserved Lys of the P-loop (GXXXGKS/T) as well as the P-loop main-chain amide groups.

The γ-phosphate of GTP is further stabilised by the formation of a hydrogen bond with the Ser/Thr of the P-loop (GXXXGKS/T) and the conserved Gly of the switch II region (hhhhDXXG) and conserved Thr of the switch I region (xTx) upon hydrolysis.

The magnesium ion is hexacoordinated and further stabilises the β- and γ-phosphates of the nucleotide, contributing to the rigidity of the switch regions when GTP is bound. Magnesium coordinates a water molecule, the oxygen atoms of β- and γ-phosphates, the Ser/Thr of the

P-loop (GXXXGKS/T), the Thr of the switch I (xTx) and the Asp of switch II (hhhhDXXG) (through a water molecule).

The elements responsible for coordination of the γ-phosphate and the magnesium ion are lost in the GDP-bound state. These include the coordination of magnesium with the γ-phosphate, as well as the hydrogen bonds between this phosphate and Ser/Thr (P-loop), Thr (switch I) and Gly (switch II). Although the P-loop conformation is not affected after GTP hydrolysis, the switch I and II regions become destabilised and undergo large conformational changes. Due to the loss of coordination of the magnesium ion, the affinity for magnesium decreases in the GDP-bound state.

The guanine ring is stabilised by interactions with the G4 and G5 motifs, which confer specificity for guanine nucleotide (Fig. 13). The carboxyoxygens of the conserved Asp and the amino group of Asn of G4 (N/TKXD) form hydrogen bonds with the guanine base. The interaction of the conserved Ala of G5 (SAK/L) with the oxygen of the base ring further contributes to guanine specificity. The lack of affinity for adenine nucleotides is due to loss of the interactions between G4 Asp and the base and between G5 Ala and the O6. Further destabilisation comes from a steric hindrance between the amino group of G4 Asn and the exocyclic NH_2 of the adenine base, making impossible the binding of an ATP molecule (Bourne et al., 1991; Paduch et al., 2001; Sprang, 1997).

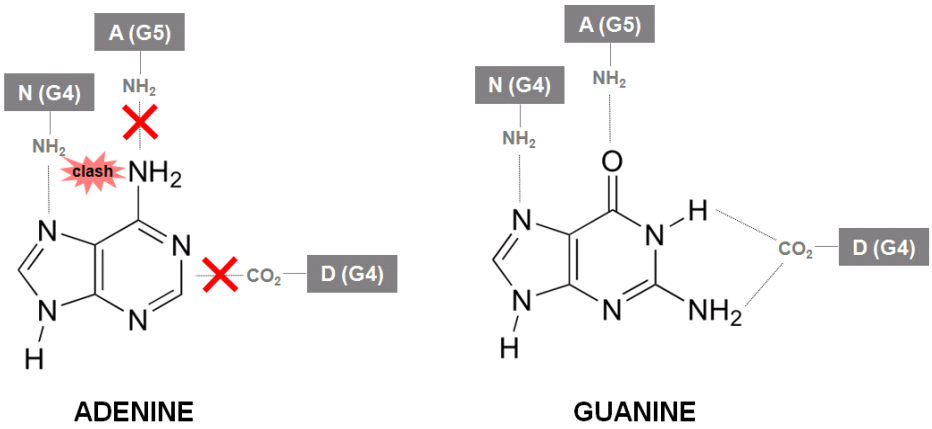

Fig. 13 Nucleoside interactions with GTP-binding proteins. In GTPases, specificity for guanine nucleotides is conferred by interactions between conserved residues from G4 and G5 motifs and the guanine base (right). The same interactions are not possible when an adenine base is present (left), hence lacking affinity for adenine nucleotides.

3.3. Catalytic cycle

GTP binding and hydrolysis induce conformational changes in GTPases such that different states are present during the catalytic cycle. Depending on the nucleotide that is bound,

GTPases may switch between a GTP-bound “active” conformation and a GDP-bound “inactive” conformation (Fig. 14).

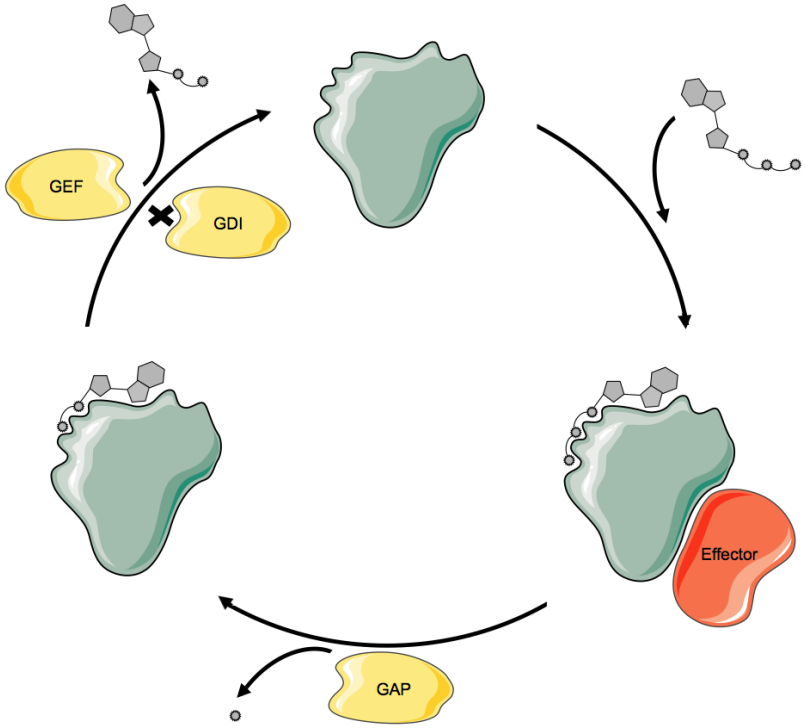

Fig. 14 GTPase catalytic cycle. When GTP binds the GTPase, the enzyme becomes activated (the “on”-state) and has affinity for an effector molecule. GTP hydrolysis, in some cases mediated by a GAP (GTPase-Activating Protein), induces conformational changes in the GTPase that reduce its affinity for the effector, which dissociates. Hydrolysis of GTP generates an “off”-state GDP-bound GTPase. Activation of the GTPase, which can be mediated by a GEF (Guanine nucleotide Exchange Factor), occurs by exchange of GDP by a GTP molecule and allows the GTPase to start a new cycle. GDIs (Guanine nucleotide Dissociation Inhibitor) can also act by inhibiting nucleotide exchange and blocking the GTPase in the inactive state. Modified from Servier Image Bank.

A catalytic cycle starts with the activation of the enzyme via binding of GTP. The presence of GTP allows the formation of hydrogen bonds between the γ-phosphate and the conserved Thr and Gly of the switch I and switch II regions, respectively. These contacts contribute to a larger rigidity of the switch regions, which in turn triggers a change in conformation in which the protein binds an effector target (the “on”-state). Upon GTP hydrolysis, the energy stabilising the switch regions is dissipated and these loops become destabilised, triggering a rearrangement of the G-domain (Fig. 16). This conformation change is described as a “loaded-spring mechanism” (Vetter and Wittinghofer, 2001), where the switch regions become relaxed following release of the γ-phosphate (Fig. 15). In this GDP-bound conformation, the GTPase has no longer affinity for the effector (the “off”-state), which dissociates. GDP is subsequently released from the active site, setting the protein in its apo-state for another cycle to take place (Sprang, 1997).

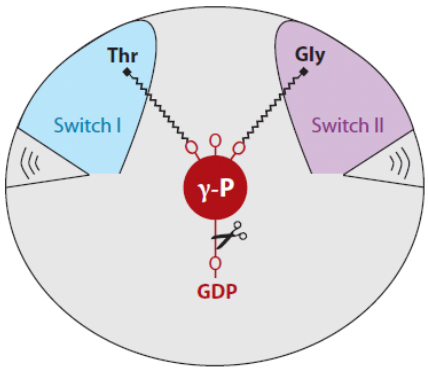

Fig. 15 The universal switch mechanism of P-loop GTPases, described as the loaded-spring mechanism. When a GTP molecule is bound to the GTPase, the switch I and II regions are stabilised by interactions between the conserved Thr and Gly residues from these motifs, respectively, and the γ-phosphate. Upon GTP hydrolysis, these regions become destabilised by the departure of the γ-phosphate and a change of conformation in the protein is induced. Adapted from Wittinghofer and Vetter, 2011.

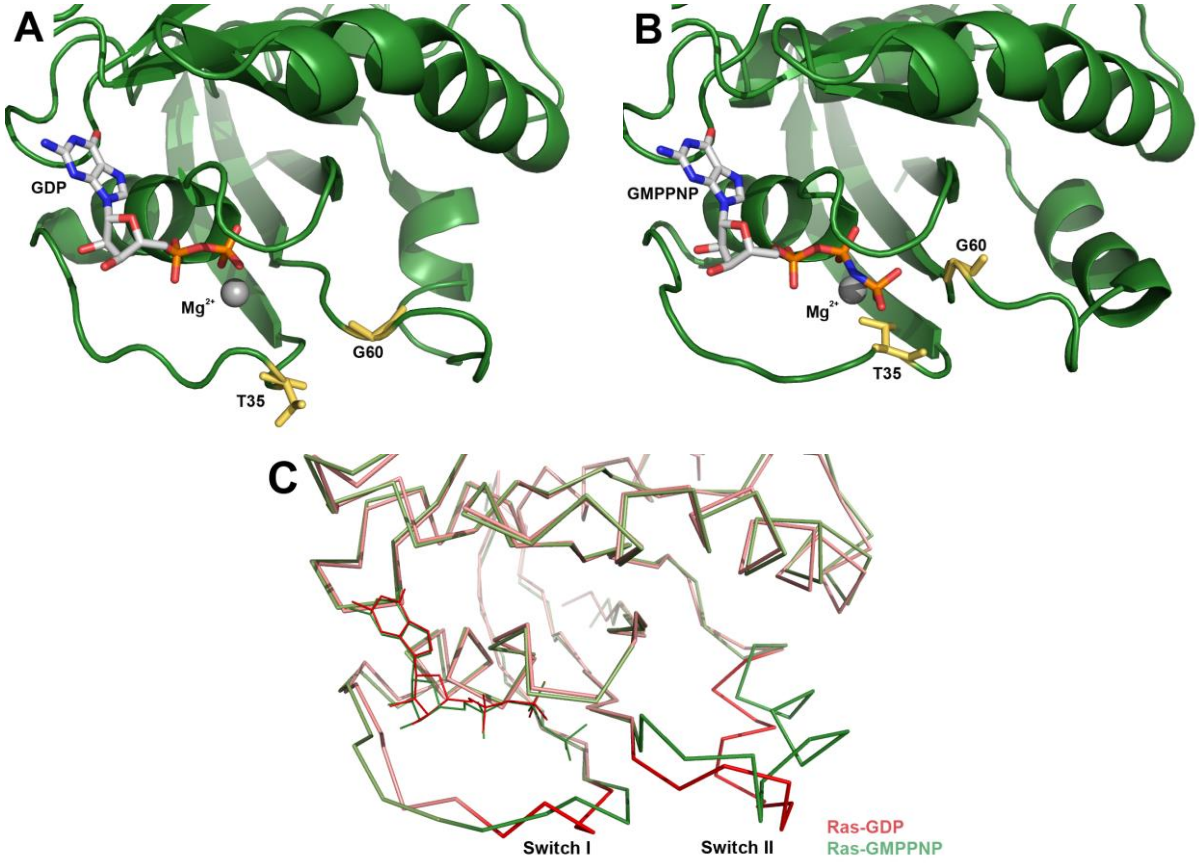

Fig. 16 3D structure of the human GTP-binding protein Ras in the **(A)** GDP-bound form and **(B)** GTP-bound form. The change in the conformation of Ras is due to different positioning of Threonine 35 and Glycine 60 (represented in yellow) from the switch I and switch II regions, respectively. When a triphosphate nucleotide is bound – in this structure, the non-hydrolysable analogue of GTP, GMPPNP – these two residues become orientated towards and stabilise the γ-phosphate. In image **(C)** the Cα of both structures are superposed showing the different orientation of switch I and II regions for the GDP-bound (red) and GTP-bound (green) forms of Ras (Milburn et al., 1990). Images edited in PyMol from PBD files 4Q21 and 5P21.

The duration of each state – on, off and apo – during the catalytic cycle is regulated by intrinsic properties of the GTPase. Both the cleavage of the β-/γ-phosphate bond and the release of GDP are limiting steps in the cycle. However, unlike classical catalysts, the majority of GTPases have low activity rates (κ_{cat} ~ 10^{-3} – 10^{-6} s^{-1}) and slow dissociation constants for nucleotides (K_D in the range of nanomolar) (Mishra and Lambright, 2016).

Additional interacting proteins assist GTPases during their cycle. These include GTPase-activating proteins (GAPs), guanine nucleotide exchange factors (GEFs), and guanine nucleotide dissociation inhibitors (GDIs).

3.3.1. GAPs, GEFs and GDIs

For some GTPases, the low intrinsic activity is bypassed by the action of GAP proteins. GAPs increase the GTP hydrolysis rate by several orders of magnitude by participating in the hydrolysis reaction as a GTPase-GAP complex.

GAPs are specific for each GTPase subfamily. The Ras-specific GAP is one of the most extensively studied GAP protein. The complex Ras-GAP^{Ras} is formed by docking switch I and switch II regions in a groove between two helices of GAP. GAP^{Ras} acts by providing an Arg residue, called the “arginine finger” that stabilises negative charges that develop during the transition state of the hydrolysis reaction (Fig. 17). Further stabilisation of the switch II region is attained when GAP^{Ras} is present, by providing a conserved Gln residue that coordinates an attacking water (Sprang, 1997). Rho-specific GAPs, although sharing little sequence homology to GAP^{Ras}, share a similar mechanism of GAP-assisted hydrolysis.

Despite being present for many GAPs, the “arginine finger mechanism” does not apply to all GTPases. Other GAP mechanisms have been identified that do not contain a catalytic arginine or glutamine. Ran and Rap are examples of such GTPases (Cherfils and Zeghouf, 2013; Mishra and Lambright, 2016). Ran-specific GAP does not provide an arginine finger. The mechanism of GAP^{Ran} involves the stabilisation of switch II conformation, where a conserved Gln residue is present similarly to Ras and Rho. A conserved tyrosine in the switch I of Ran induces a closed conformation of this region, preventing the access of a potential arginine finger. Rap GTPase lacks the conserved Gln residue in the switch II and its specific GAP does not contain the arginine finger. In the case of Rap-GAP^{Rap} complex, GAP^{Rap} supplies an “asparagine thumb” that acts as the Gln residue in Ras, while a Tyr residue in Rap GTPase stabilises the transition state. In addition, a similar closed conformation of switch I to Ran GTPase is observed for Rap.

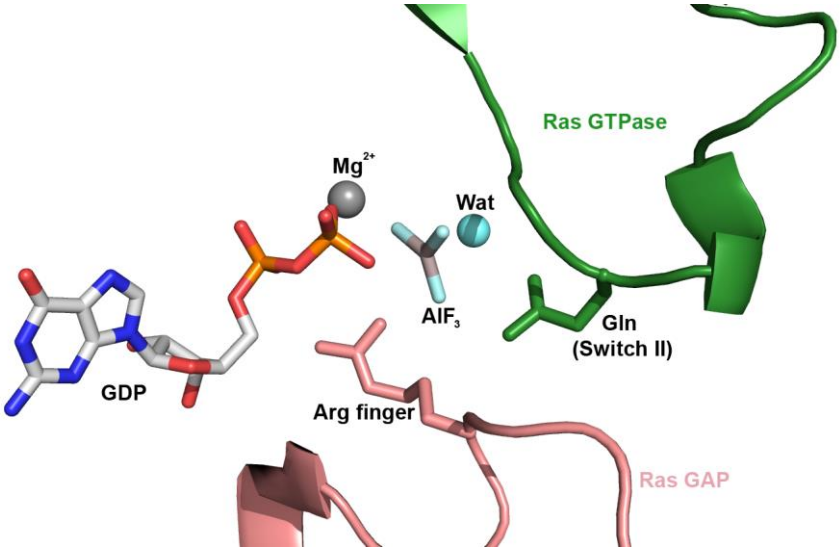

Fig. 17 Conventional mechanism of GAP-mediated GTP hydrolysis. The GAP^{Ras} provides the arginine finger that contacts the β-phosphate and the AlFx that can mimic the transition state. The conserved glutamine from switch II coordinates the water molecule for nucleophilic attack. Image edited with PyMol from PDB 1WQ1.

The slow dissociation constants of GTPases can be increased by the action of GEFs that accelerate nucleotide release. The general mechanism of nucleotide exchange mediated by GEFs is based on the distortion of the nucleotide-binding site, resulting in a reduced affinity.

GEFs form a ternary low affinity complex with the GTPase-GDP. Interactions with the switch I and II regions of GTPases induce conformational changes in the P-loop and in the Mg^{2+} binding area. The adopted conformations clash with the binding of the nucleotide and the magnesium, and promote release of the nucleotide. GEF is displaced through binding of a triphosphate nucleotide. Exchange for a GTP molecule occurs due to higher intracellular concentrations of GTP (Cherfils and Zeghouf, 2013; Vetter and Wittinghofer, 2001). As for GAPs, each GTPase subfamily has its specific GEF and differences in their mechanism may be observed. While some GEFs sterically block the metal/nucleotide binding sites, others contribute with hydrophobic residues that generate hydrophobic repulsion of the nucleotide and/or acidic residues that generate electrostatic repulsion of the Mg^{2+} ion (Cherfils and Zeghouf, 2013).

GDIs inhibit the release of GDP. As these regulatory proteins are only present for certain GTPases, namely for the Rho and Rab families, that alternate between the cytosol and the cell membrane (Cherfils and Zeghouf, 2013), their mechanism will not be discussed here.

Although GAPs and GEFs are frequently present in eukaryotes, their presence in prokaryotes is less common (Verstraeten et al., 2011). The absence of known bacterial GEFs may arise from the low nucleotide affinities that most bacterial GTPases present, therefore likely sparing the need for these regulators. Similarly, GAPs have also not been

identified for the majority of bacterial GTPases. Interestingly, other mechanisms have been identified that accelerate the rate of GTP hydrolysis, such as G-domain dimerisation, coordination with ions, or interaction with ribosomal elements. These observations may indicate that a different regulation mechanism may be present for bacterial GTPases.

3.4. Classification of GTPases

Koonin and colleagues (Leipe et al., 2002) developed a phylogenetic classification of GTPases and divided this superfamily into two large classes, based on their structural and sequence features.

The first class is designated TRAFAC class, for *translation factor-related*. It includes translation factors, signal transducing proteins, heterodimeric G proteins, some ATPases and the Ras family. Members of this class are mainly involved in mRNA translation, signal transduction, cell motility and intracellular transport. In terms of their structure, TRAFAC GTPases are characterised by an adjacent β-strand antiparallel to the β-strand of the Walker B region. As for their sequence, TRAFAC GTPases contain a conserved threonine residue in the loop N-terminal to β2-strand (G2 motif) and a conserved serine residue in the last β-strand (G5 motif).

The second class is the SIMIBI class, for *signal-recognition particle, MinD, BioD*, and includes signal recognition particle-associated GTPases, the G3E, XAB1 and CLP1 families, the MPR/MinD-related superfamily and several metabolic enzymes such as BioD-related enzymes. Structurally, opposing to TRAFAC class, SIMIBI GTPases are characterised by a parallel β-strand adjacent to the Walker B motif. At the sequence level, SIMIBI members have an additional glycine residue in the Walker A motif (GXXGXGKS/T), a conserved aspartate residue in the switch I region (instead of the conserved threonine in TRAFAC), a glutamate in the Walker B motif (instead of the aspartate) (hhhhEXXG), and an aspartate residue in the N-terminus of the Walker B-containing strand. Furthermore, the G4 motif that provides guanine specificity shows large sequence variation.

4. Bacterial ribosome assembly GTPases

The evolutionary genomic study of the GTPase superclass by (Leipe et al., 2002) and subsequent studies done on P-loop GTPases allowed the identification of a group of universally conserved GTPases in bacteria. The conservation and the phenotypic dependence observed for these enzymes suggest they play an essential role in bacterial physiology and survival. Indeed, many of the universally conserved bacterial GTPases have been implicated in ribosome biogenesis. Bacterial ribosome assembly GTPases (RA-GTPases) form a cluster in the GTPase superclass that is conserved throughout evolution and shares common unique properties.

4.1. Properties of RA-GTPases

All the bacterial GTPases involved in ribosome assembly are TRAFAC GTPases. Additionally, all of them belong to the family of *hydrophobic amino acid substituted for catalytic glutamine GTPases* (or HAS-GTPases) (Goto et al., 2013). While in classical GTPases a Gln is conserved in the switch II and is proposed to act as a catalytic residue during hydrolysis, in HAS-GTPases a hydrophobic residue is present instead.

A circularly permuted G-domain is also present for some of the RA-GTPases, such as RsgA, YqeH and YlqF (Goto et al., 2013). These G-domains are characterised by a G4-G5-G1-G2-G3 motif pattern in their sequence instead of the canonical order G1-G2-G3-G4-G5.

In addition to the G-domain, the presence of extra domains is also a common feature of RA-GTPases (Karbstein, 2007). These non-catalytic domains are thought to be associated with rRNA or r-protein binding.

RA-GTPases present low enzymatic activity and low nucleotide affinity. In contrast to the nanomolar range nucleotide affinity of most GTPases, ribosome assembly GTPases have affinities in the order of the micromolar range (Britton, 2009). This low nucleotide affinity suggests ribosome assembly GTPases may not need the action of GEFs and that their catalytic cycle may be regulated by a different mechanism.

A potassium-dependency is also observed for some RA-GTPases, for which enzymatic activity is enhanced by coordination to a potassium ion (Goto et al., 2013). Potassium is thought to stabilize the transition state during catalysis, replacing the arginine finger supplied by some GAPs. Coordination takes place in a region of the switch I, called the K-loop, and is further determined by a conserved Asn residue in the P-loop (Fig. 18) (Rafay et al., 2012).

Potassium dependent GTPases form a GTPase family called TEES (for TrmE-Era-EngA-YihA-Septin).

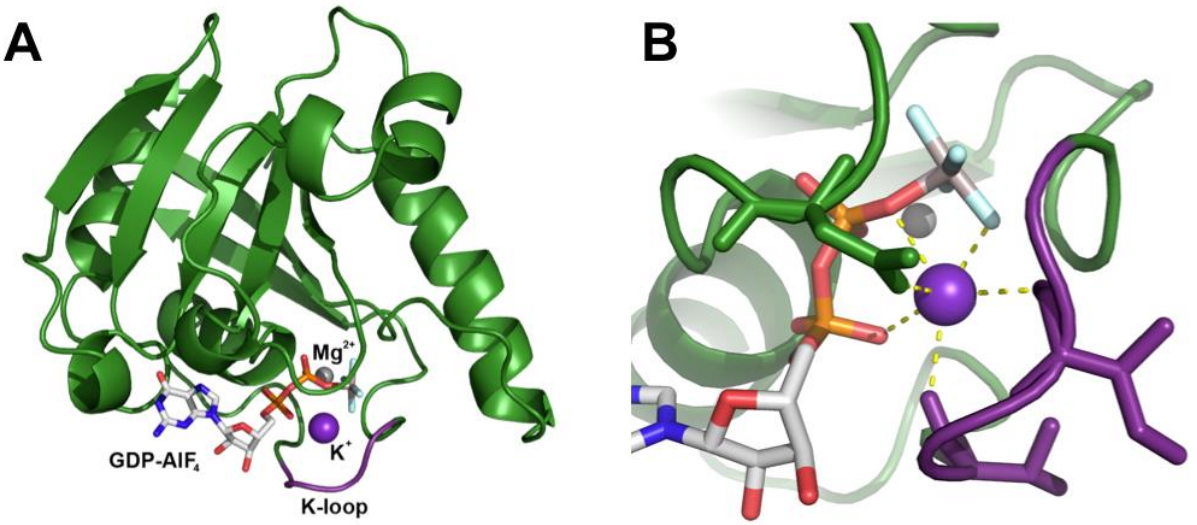

Fig. 18 (A) 3D structure of the G-domain of *E. coli* GTP-binding protein MnmE. In this structure, MnmE is bound to a GDP molecule and an AlF_4^- ion (both represented in sticks) that mimics the triphosphate nucleotide. Mg^{2+} and K^+ ions are represented by spheres in grey and violet, respectively. The K-loop, located upstream of switch I, is coloured in violet. **(B)** Close view of the catalytic centre. K^+ ion is hexacoordinated by the K-loop, an Asn residue from the P-loop, the α- and β-phosphate of GDP and the AlF_4^- ion. Images edited in PyMol from PBD file 2GJ8.

Studies on the essential character of RA-GTPases for cell growth and survival have revealed a similar phenotype on mutants of these enzymes. This includes decreased rate of cell growth or total absence of growth for lethal phenotypes and reduced cellular levels of 70S (Goto et al., 2013).

Finally, all RA-GTPases interact with the ribosome (either 70S or individual subunits). For most GTPases, this interaction occurs in the GTP-bound form and for some it has been shown to enhance enzymatic activity. It is the case of RsgA where an increase in GTPase activity of more than 100-fold was observed upon binding to the ribosome (Goto et al., 2013).

4.2. Possible roles of GTPases in ribosome assembly

Despite the amount of structural and biochemical information available on RA-GTPases, how they act on ribosome assembly as maturation factors is still not completely understood. Possible mechanisms for the action of RA-GTPases were reviewed by K. Karbstein and R. Britton (Britton, 2009; Karbstein, 2007) and are presented here.

One possible role of RA-GTPases in ribosome assembly can be the direct recruitment of r-proteins to nascent ribosomal subunits. The GTPase could bind both the r-protein and the

ribosome in the GTP-bound state, delivering the r-protein and abandoning the complex upon hydrolysis. Alternatively, it could induce conformational changes required for the binding of an r-protein, stabilising the complex and promoting association of the r-protein to the ribosome. In both cases, the energy generated from GTP hydrolysis can act as an energy input for the association of ribosomal elements.

Similarly, RA-GTPases could use the energy from GTP hydrolysis to regulate the delivery or removal of other assembly factors. Release can be achieved by inducing conformational changes in the ribosome that promote dissociation of the element, or by its mechanical displacement.

Besides participating in the recruitment of proteins, RA-GTPases could also regulate rRNA structure by directly promoting conformational changes in the rRNA or by controlling the action of RNA helicases.

Another possible role for RA-GTPases is assembly quality control. The complexity of *in vivo* ribosome biogenesis requires quality control pathways to ensure ribosomal quality. RA-GTPases could act as monitors at specific checkpoints during the assembly process in order to assess proper achievement of a certain step, allowing or avoiding progression of the assembly. One mechanism for this could be binding to an r-protein binding site and blocking it until favourable conditions for assembly progression are met.

RA-GTPases could also act as sensors of the cellular energy status, thereby regulating the assembly process. Under starvation, bacteria initiate a stringent response by producing the alarmone (p)ppGpp. In such conditions, GTPase activity can be targeted by two mechanisms. One involves the direct binding of (p)ppGpp to GTPases resulting in competitive inhibition with GTP and GDP, as for many GTPases binding affinities of the three ligands are similar (Kanjee et al., 2012). A second mechanism involves the decrease of intracellular GTP concentrations due to its consumption for (p)ppGpp synthesis and to inhibition of GTP biosynthesis enzymes by (p)ppGpp. Both mechanisms result in a decrease of the GTPase:GTP population. As most RA-GTPases bind the ribosome in the GTP-bound state, this association would be affected in such conditions. Hence, RA-GTPase function could be sensitive to the intracellular GTP/GDP ratio, controlling ribosome assembly. The fact that RA-GTPases do not need the action of GAPs and GEFs suggests that a different mechanism of cycle regulation may be present, such as regulation by GTP/GDP ratio.

4.3. GTPases involved in ribosome biogenesis

Era (Bex, Sgp, RbaA, Pra) was the first bacterial GTPase to be identified (Brown, 2005). Multiple studies have shown its essential nature for some bacterial species, while in others depletion results in growth defects. Era binds 16S rRNA (Sharma et al., 2005) and is associated with maturation of the 30S subunit, which accumulates as free-30S along with accumulation of a 16S rRNA precursor in depleted cells (Brown, 2005; Goto et al., 2013). Era belongs to the TEES family that requires K^+ for optimal activity. Binding to RNA has also shown to stimulate Era's GTPase activity up to 12-fold (Tu et al., 2009).

RsgA (YjeQ, YloQ, CpgA) is another bacterial GTPase involved in the maturation of the 30S subunit. Despite not being essential, depletion of this enzyme in *E. coli* (Daigle and Brown, 2004) and *B. subtilis* (Campbell et al., 2005) resulted in slow growth, filamentous phenotypes, decreased levels of 70S, accumulation of both ribosomal subunits and accumulation of 16S rRNA precursors. Binding to the ribosome occurs in the presence of GTP and increases its low intrinsic activity over 160-fold (Daigle and Brown, 2004).

YqeH is a bacterial GTPase involved in the initial stages of the biogenesis of the 30S subunit (Anand et al., 2009). Although it is not universally conserved in bacteria, being found only in some gram-positive bacteria, it is present in higher organisms. In *B. subtilis*, YeqH is essential for cell growth (Morimoto et al., 2002) and its depletion results in decreased levels of 16S rRNA, 30S subunit, 70S ribosome and accumulation of 50S subunits (Uicker et al., 2007). Binding to the 30S occurs in the GTP-bound state, but in contrast to RsgA its intrinsic activity is not affected by the interaction (Loh et al., 2007).

YlqF (RbgA) is an essential GTPase distributed through the three domains of life. YlqF is thought to play a role in the late stage of 50S maturation. Depletion results in reduction of 70S ribosomes and accumulation of 50S subunit precursors (Uicker et al., 2006). In the presence of GTP, binding to the 50S subunit results in an enhanced GTPase activity (Achila et al., 2012). YlqF is a potassium-dependent GTPase, with potassium increasing its intrinsic activity by approximately 100-fold (Achila et al., 2012).

Obg (ObgE, YhbZ, CgtA) is an essential GTPase conserved in prokaryotes and eukaryotes. It binds to the 50S subunit in *C. crescentus* (Lin et al., 2004) and to the 50S and 30S subunits in *E. coli* with higher affinity in the GTP-bound form (Jiang et al., 2006). In *E. coli*, mutations in its gene result in decreased levels of 70S and accumulation of both 16S and 23S rRNAs precursors.

YsxC (YihA, EngB) is a GTPase essential in gram-negative and gram-positive bacteria (Goto et al., 2013). In *B. subtilis*, YsxC depletion results in decreased levels of 70S ribosome and accumulation of a 50S precursor (Schaefer et al., 2006). YsxC participates in the maturation of the functional core of the large subunit (Ni et al., 2016), which stimulates its GTP-hydrolysing activity by a 4- to 9-fold factor (Wicker-Planquart and Jault, 2015).

EngA (Der, YphC, YfgK) is a ribosome assembly GTPase conserved in bacteria. The present work focus on EngA, with a detailed description of this RA-GTPase following in the next chapter.

5. The Bacterial GTPase EngA

The EngA family (also known as Der, YphC or YfgK) is a ubiquitous GTPase family in eubacteria. EngA (named for Essential neisseria GTP-binding protein A) was firstly identified by Mehr and collaborators (Mehr et al., 2000) as a product of a gene essential for *N. gonorrhoeae* growth, where transformants with a disruption in this gene could not be obtained. Since this finding, subsequent studies on EngA orthologs have shown that it is conserved and essential as well for other gram-positive and gram-negative bacteria and have suggested a role in ribosome assembly. With the exception of some higher plants, such as *A. thaliana*, where it has an organellar localisation (Suwastika et al., 2014), EngA is not present in archaeal and eukaryotic organisms.

5.1. Tandem GTP-binding domains

Genomic analysis of EngA orthologs showed that EngA is highly conserved in eubacterial species (Hwang and Inouye, 2001). Comparison of EngA and Era showed sequence homology between G-domains of both enzymes, placing EngA in the GTPase superfamily. High sequence similarity is present between homologues, in particular in the consensus G motifs (Fig. 19).

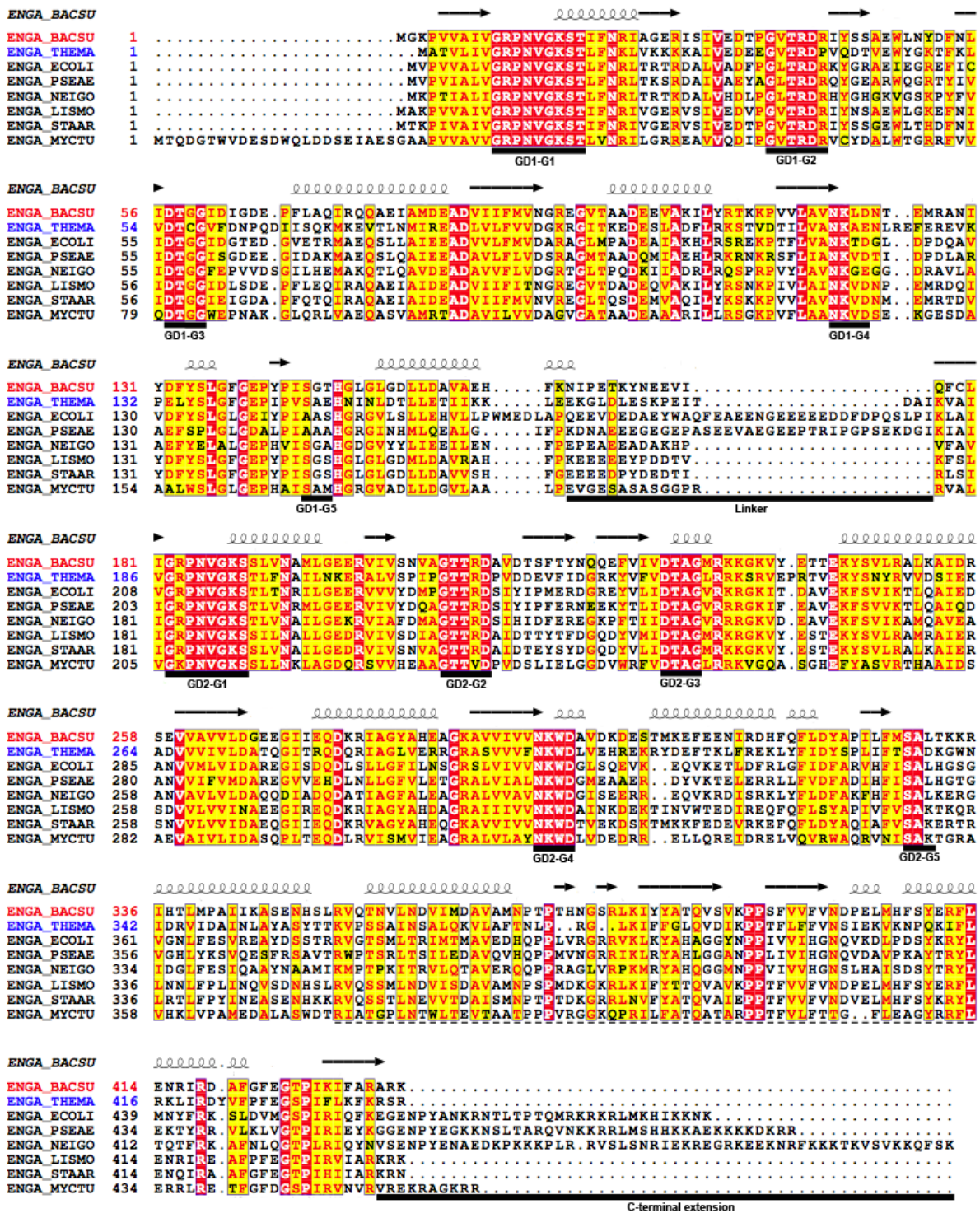

Fig. 19 Amino acid alignment of EngA orthologs from *Bacillus subtilis* (BACSU), *Thermotoga maritima* (THEMA), *Escherichia coli* (ECOLI), *Pseudomonas aeruginosa* (PSEAE), *Neisseria gonorrhoeae* (NEIGO), *Listeria monocytogenes* (LISMO), *Staphylococcus aureus* (STAAR) and *Mycobacterium tuberculosis* (MYCTU). Secondary structures are represented above the aligned sequences. G motifs, linker and C-terminal extension are represented and labelled in black.

Despite displaying the conserved G-domain common to GTPases, EngA is unique among all the other prokaryotic and eukaryotic members of the GTP-binding superfamily: it has two homologous and adjacent N-terminal G-domains connected by an acidic linker (Fig. 20).

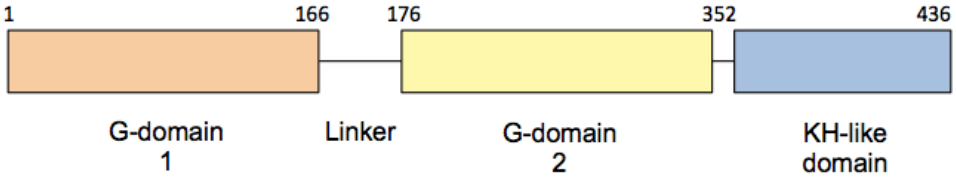

Fig. 20 Representation of the domain organisation of EngA. The numbers delimiting domains correspond to the *B. subtilis* ortholog. The two homologous G-domains of 160-170 amino acids are present at the N-terminus: G-domain 1 (GD1) is represented in orange and G-domain 2 (GD2) in yellow. A basic C-terminal KH-like domain of 80-90 amino acids is represented in blue. The G-domains are connected by an acidic linker that can range between 9-65 amino acids.

Each G-domain has 160 – 170 amino acids and both share ~30% sequence identity (Hwang and Inouye, 2001; Muench et al., 2006). With the exception of G5 motif in GD1 (which is structurally conserved but presents sequence variation), G1–G4 motifs in G-domain 1 (GD1) and G1–G5 motifs in G-domain 2 (GD2) are well conserved (Fig. 19). A linker of variable extension (up to 65 amino acids) connects GD1 and GD2. This linker is highly acidic, comprising mainly Asp and Glu residues and lacking basic residues (Tomar et al., 2012).

A non-catalytic domain of 80–90 amino acids is present at the C-terminus in EngA orthologs. This C-terminal domain displays an αββαβ topology that resembles a K-homology domain (KH-domain). KH-domains are RNA-binding domains that contain a conserved sequence VIGXXGXXI responsible for interaction with nucleic acids (Nicastro et al., 2015). In Era, at the C-terminus, a KH-domain follows the G-domain and is known to bind 16S, being implicated in rRNA processing and maturation (Tu et al., 2009). In EngA, however, despite the KH-like topology, the C-terminal domain has no sequence similarity with classical KH-domains and the consensus sequence for RNA-recognition is absent (Fig. 21). Thus, the KH-like domain of EngA may be implicated in a different function other than RNA-binding.

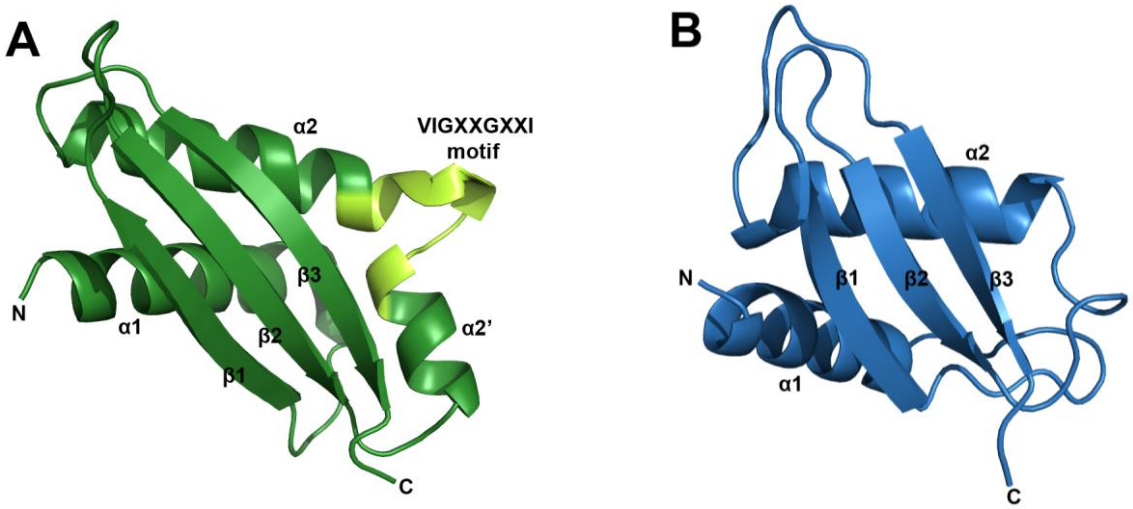

Fig. 21 (A) 3D structure of the KH-domain of *E. coli* Era. KH domains have a αββαβ topology. The α2-α2' turn contains the RNA recognition sequence VIGXXGXXI, represented in light green. **(B)** 3D structure of the KH-domain of *T. maritima* EngA. Although the αββαβ fold is present, the domain lacks the RNA recognition sequence. Image edited in PyMol from PDB files 1EGA and 1MKY.

Sequence analysis of EngA orthologs has identified variations in the linker and in the KH-domain that appear to be related to ribosome-binding regulation (Tomar et al., 2012). Phylogenetic analysis has suggested that these variations are species-specific and have clustered EngA orthologs into two groups. One group comprises mostly gram-negative bacteria, which possess a longer linker (15–65 residues) connecting G-domains and a highly basic extension in C-terminal (20–70 residues) following the KH-domain. The second group comprises mainly gram-positive bacteria, that possess a short linker (<15 residues) and lack the C-terminal extension. Regardless of the species, the basic character of the C-terminal extension, when present, and the acidic nature of the linker seem to be conserved. Experiments on ribosome binding have suggested a structural regulation by the C-terminal extension that seems to stabilise the long linker and support certain conformations of EngA during interaction with an effector (Tomar et al., 2012).

5.2. Functional studies

5.2.1. Essential nature

After the identification of EngA in *N. gonorrhoeae*, genomic analysis suggested that the *engA* gene would make part of the “minimal bacterial genome” (Mittenhuber, 2001). Studies with mutants on different bacteria confirmed the essential nature of EngA in *E. coli* (Hwang and Inouye, 2001), *B. subtilis* (Morimoto et al., 2002), *S. pneumonia*, *S. aureus* and *H. influenza* (Zalacain et al., 2003). Attempts to obtain deletion mutants resulted in absence of cell growth, while conditional mutants generated defects in cellular morphology (such as cell filamentation and abnormal cell curvature) and in chromosome segregation. These observations suggested that EngA would play a role in cell cycle and division.

5.2.2. Role in ribosome biogenesis

The first evidence that linked EngA to ribosome biogenesis appeared in a study by Tan and collaborators (Tan et al., 2002) where EngA rescued a null mutation of an rRNA methyltransferase. RrmJ is a heat-shock protein that methylates a residue in 23S rRNA that contacts the A-site tRNA. Deletion of *rrmJ* results in impaired cell growth and in increased levels of unassembled ribosome subunits and decreased levels of the 70S ribosome. Tan et al. observed suppression of the *rrmJ* deletion phenotype upon overexpression of EngA, suggesting functional interactions with the 50S subunit. Since this observation, efforts have been made to further understand the cellular function of EngA in ribosome biogenesis.

Several studies in *E. coli* and *B. subtilis* have revealed abnormal polysome profiles in EngA-depleted cells, with decreased levels of 70S and accumulation of free ribosomal subunits 30S and 50S (Bharat et al., 2006; Hwang and Inouye, 2006; Schaefer et al., 2006). Precursors and aberrant 50S subunit were also present in such strains, indicating that EngA might play a role in the maturation and stabilisation of the large ribosomal subunit.

Sucrose density gradient sedimentation experiments in *E. coli*, *B. subtilis* and *M. smegmatis* showed cofractionation of EngA with the 50S subunit (Agarwal et al., 2012; Bharat et al., 2006; Hwang and Inouye, 2006; Tomar et al., 2009) and with the 70S ribosome (Schaefer et al., 2006; Tomar et al., 2009) in the presence of GTP analogues. This interaction was strongly inhibited by GDP, suggesting that EngA is a GTP-dependent 50S-associating protein.

A role for EngA on the maturation of the A-site was proposed after the analysis of ribosomes produced by EngA-depleted cells that revealed sensitisation to aminoglycosides (Bharat and Brown, 2014). This class of antibiotic interferes with decoding by impairing binding of an aa-tRNA to the mRNA on the A-site. This observation is supported by the results obtained previously on the suppression of RrmJ loss phenotype by overexpression of EngA.

Recently, the structure of *E. coli* EngA bound to the 50S ribosomal subunit was determined by cryo-EM (Zhang et al., 2014). In this structure, EngA is bound to the PTC and the 50S lacks certain elements, resembling an intermediate subunit. A mechanism was hence suggested, where EngA acts as an rRNA chaperone by binding to the PTC and facilitating its assembly. A more recent cryo-EM analysis of the large subunit from EngA-depleted *B. subtilis* (Ni et al., 2016) revealed the presence of a late-assembly 45S intermediate presenting immature conformations of the A-, P- and E-sites, central protuberance and L7/12 stalk. Analysis of these intermediates suggested that EngA must be critical for the remodelling of RNA helices of the functional PTC core.

In the early functional studies on EngA, analysis of the rRNA processing in EngA-depleted cells detected accumulation of precursors of 23S and 16S rRNA (Hwang and Inouye, 2006) and EngA was reported to coelute with both 23S and 16S (Tomar et al., 2009). This suggested that EngA interacts not only with the 50S subunit, but also with the 30S subunit. A study by Lamb and co-workers revealed an interaction of *S. Typhimurium* EngA with the 30S ribosomal protein S7 by pull-down and ITC (Lamb et al., 2007). S7 is a 20 kDa r-protein from the small subunit, located at the subunit interface close to the decoding centre (Fig. 22). It is a primary r-protein that directly binds to 16S rRNA, nucleating the initiation of the 30S assembly (Nowotny and Nierhaus, 1988). During ITC experiments, however, binding of EngA

and S7 was accompanied by protein aggregation and the thermodynamic parameters for this interaction could not be obtained.

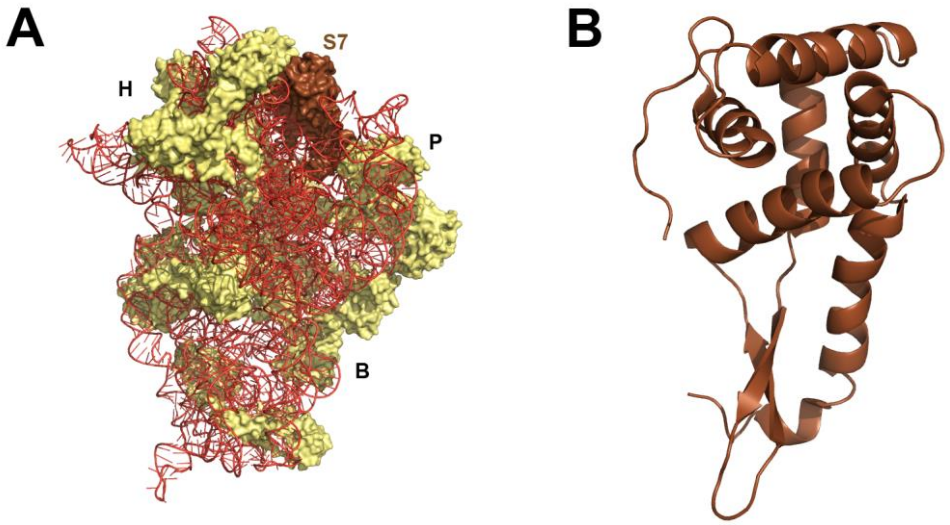

Fig. 22 (A) 3D structure of the *E. coli* 30S subunit. The interface of the subunit is towards the reader. H, head; P, platform; B, body. R-proteins are represented as surface in yellow and rRNA as cartoon in orange. Ribosomal protein S7 is represented as surface in brown. **(B)** 3D structure of the *T. maritima* ribosomal protein S7. Images were edited in PyMol from PDB files 4V4Q and 3GTY.

Binding of *E. coli* EngA to the 30S subunit was demonstrated by Tomar and co-workers (Tomar et al., 2009). The interaction was proposed to occur in a nucleotide-specific manner, where GD1 plays a regulatory role for the interaction with different ribosomal partners. According to this mechanism, interaction with the 50S subunit occurs when EngA has a GD1:APO / GD2:GTP occupancy and is further stabilised when GD1 binds GTP. When GDP is bound to GD1, however, inter-domain interactions between GD1 and KH domains shift the affinity of EngA towards the 30S subunit. Conversely, a study done on *M. smegmatis* EngA indicates that binding of EngA to the 30S occurs regardless of the presence of nucleotides (Agarwal et al., 2012).

A recent study by Lee and co-workers reported binding of *E. coli* EngA to cellular membranes (Lee et al., 2011). EngA from cell lysates cosedimented in sucrose gradients with the membrane protein OmpC. The complex EngA-OmpC was reconstituted using purified components, suggesting a specific and high affinity interaction, and is suggested to involve the C-terminal KH-domain. Based on these observations, a mechanism of regulation of the GTPase cycle was proposed, where EngA cycles between a GTP-bound state that interacts with the ribosome and a GDP-bound state that binds cell membrane.

Chloroplasts ribosomes are very similar to those of bacteria and many of the assembly factors are conserved between both these organelles and bacteria. A recent study on *N.*

benthamiana reported a cellular localisation of EngA in chloroplasts (Jeon et al., 2014). Similarly to the observations from bacteria, EngA-depleted plants presented a lethal phenotype. Chloroplast development was impaired in these strains and showed deficient levels of chloroplast proteins, accumulation of precursors of 23S, 5S and 16S rRNAs and decreased levels of polysomes. Also, EngA from chloroplasts was able to bind the 50S subunit and the 23S and 16S rRNAs. These observations indicate that EngA may be involved in rRNA processing in chloroplasts, suggesting a conserved function in plants and bacteria.

5.3. Biochemistry

Biochemical characterisation of EngA shows it binds guanine nucleotides and has an intrinsic GTPase activity.

Studies on different orthologs of EngA indicate that it possesses a low intrinsic activity. Low intrinsic activities are common to bacterial RA-GTPases in the absence of an effector protein or molecule. The kinetic parameters of GTP hydrolysis for *B. subtilis* EngA in the absence of potassium were determined as V_{max} = 0.15 µM/min, K_M = 111 µM and κ_{cat} = 0.3 min^{-1} (Foucher et al., 2012). A similar maximum rate of hydrolysis was determined for *M. smegmatis* (0.32 µM/min), with lower values for K_M and κ_{cat} (39 µM and 0.005 min^{-1} respectively) (Agarwal et al., 2012).

The intrinsic activity of EngA is strongly stimulated by potassium. Addition of 300 mM KCl increases the hydrolysis rate of *B. subtilis* EngA about 20-fold (V_{max} = 3.6 µM/min and κ_{cat} = 7.4 min^{-1}, compared to 0.15 µM/min and 0.3 min^{-1} in the absence of potassium) (Foucher et al., 2012). The values of K_M, reported to be 111 µM (no KCl) and 153 µM (300 mM KCl) in *B. subtilis* EngA, do not seem to be influenced by potassium, which suggests that potassium has an impact on the catalytic efficiency rather than on the K_M for GTP (Foucher et al., 2012). In *T. maritima*, the GTPase activity was stimulated about 4-fold by the addition of 400 mM KCl, with V_{max} = 3.5 µM/min, K_M = 110 µM and κ_{cat} = 0.87 min^{-1} (Hwang and Inouye, 2001). In the same potassium concentrations, *E. coli* EngA presents K_M = 143 µM and κ_{cat} = 1.17 min^{-1} (Bharat et al., 2006).

Potassium-dependency has been reported for several GTPases. YlqF and YqeH, two RA-GTPases, have also been reported to be activated by potassium (Achila et al., 2012; Anand et al., 2010). MnmE, a conserved bacterial GTPase involved in tRNA modification, has an intrinsic activity in the same range as *B. subtilis* EngA (κ_{cat} = 0.33 min^{-1}) that is stimulated by

potassium (κ_{cat} = 7.8-26 min^{-1}) (Meyer et al., 2009a). In the particular case of MnmE, the K^+ ion induces nucleotide-dependent dimerisation of G-domains which is accompanied by large movements that allow transition between an open and closed states necessary for its function (Meyer et al., 2009b).

Table 1 Kinetics parameters for EngA orthologs. Values of k_{cat} for Ras with and without GAP^{Ras} are presented as a reference. Values taken from references: Agarwal et al., 2012; Bharat et al., 2006; Foucher et al., 2012; Gideon et al., 1992; Hwang and Inouye, 2001.

	V_{max} (µM min^{-1})		K_M (µM)		κ_{cat} (min^{-1})	
	No K^+	K^+	No K^+	K^+	No K^+	K^+
B. subtilis	0.15	3.6	111	153	0.3	7.4
E. coli	-	-	-	143	-	1.17
T. maritima	-	3.5	-	110	-	0.87
M. smegmatis	0.32	-	39	-	0.005	-
Ras					0.007	
Ras-GAP					1140	

Kinetics studies were also performed on individual G-domains of EngA and revealed that GD1 and GD2 present distinct enzymatic activities. Mutations inserted in the G4 motif of GD1 (N118D) of *T. maritima* EngA led to a reduction in the GTPase activity to 10% of that of the wild-type whereas the same mutation in the GD2 (N300D) did not change the overall activity of the protein (Robinson et al., 2002). Conversely, the activities determined for truncated GD1 and GD2 revealed that both domains are active, with GD1 approximating the activity of the full-length and GD2 being 40% of it. This inconsistency is proposed to be due to different behaviours of the G-domains in the context of the full-length protein, suggesting a regulation mechanism within EngA.

Similarly, mutations on the G motifs of *B. subtilis* EngA resulted in a similar behaviour from *T. maritima*, where GD1 mutants present a markedly decreased activity and GD2 mutants remain unaffected (Foucher et al., 2012). Truncated GD1 has an activity similar to that of the full-length, while that from GD2 is lower (κ_{cat}^{GD1} = 0.6 min^{-1} / κ_{cat}^{GD2} = 0.004 min^{-1} without K^+; κ_{cat}^{GD1} = 3.4 min^{-1} / κ_{cat}^{GD2} = 0.45 min^{-1} with 1M KCl).

These observations suggest that GD1 contributes to the overall activity of EngA, while GD2 has a marginal GTPase activity. Regardless of the catalytic efficiency, both domains are stimulated by potassium (Foucher et al., 2012).

For many GTPases, such as Ras, the low activity is strongly stimulated by GAPs. YihI was recently identified as a GAP for EngA (Hwang and Inouye, 2010). YihI is a 169 aa protein, conserved in eubacteria, that showed a 2- to 3-fold stimulation of EngA GTPase activity (Hwang and Inouye, 2010; Naganathan and Moore, 2013). Interestingly, the stimulatory effect of YihI is inversely proportional to potassium stimulation, suggesting that YihI acts as a GAP at low potassium levels (Naganathan and Moore, 2013).

Another study suggested that the 50S ribosomal subunit would act as a GAP, increasing by 1.4-fold the GTPase activity of EngA (Zhang et al., 2014). GTPase stimulation by ribosome binding has also been reported for other GTPases (Achila et al., 2012; Daigle and Brown, 2004), for which the ribosome is thought to act as a GAP.

In either cases, however, the increase in GTPase activity of EngA by YihI or the 50S subunit is far from the stimulatory effect known for other GAPs, such as GAP^{Ras} which increases more than 10^5-fold the activity of Ras (Wittinghofer et al., 1997).

Studies to determine the affinity of EngA for nucleotides show some inconsistencies. Analysis by isothermal titration calorimetry on the full-length EngA of *S. typhimurium* determined two binding-sites (Lamb et al., 2007). While for GDP the two binding-sites were reported to have a similar affinity range (K_D = 0.6 µM and 3.2 µM), for GTP the difference was marked, with a K_D = 12.7 µM for one binding site and a K_D > 100 µM for the second. The authors suggested that GD1 would have a strong binding of GTP while GD2 would have low affinity. Yet, to avoid any GTPase activity, these measurements were performed in the absence of magnesium. In *B. subtilis*, biochemical studies on individual G-domains show similar affinity ranges for both GDP (K_D^{GD1} = 4.5 µM; K_D^{GD2} = 2.9 µM) and GTP analogues (K_D^{GD1} = 35.7 µM; K_D^{GD2} = 6.8 µM) (Foucher et al., 2012).

5.4. Conformational changes

The first solved structure of EngA, obtained by X-ray crystallography, belongs to *T. maritima* (Robinson et al., 2002). This structure shows the three domains arranged in a cloverleaf shape, where the C-terminal KH-domain is packed between the two G-domains (Fig. 23). GD1 and GD2 are connected by a flexible 13 residues linker.

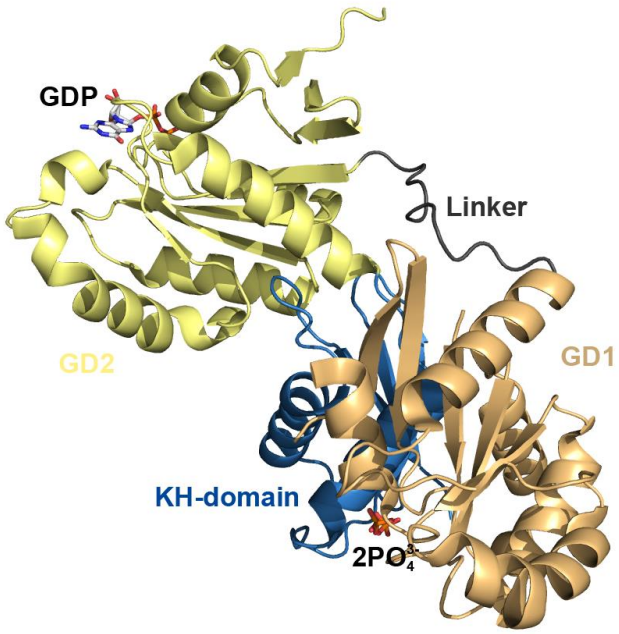

Fig. 23 3D structure of *T. maritima* EngA obtained by X-ray crystallography at 1.9 Å. The structure shows the KH-domain (blue) packed between the two G-domains (orange, GD1; yellow, GD2) that do not contact each other. Two interdomain interfaces, GD1/KH and GD2/KH, are present. An acidic linker of 13 residues (dark grey) connects both G-domains. A GDP molecule (sticks) is bound to GD2, while two phosphate ions (sticks) close to the positions of the β- and γ-phosphates of a GTP are bound to the GD1. Image edited in PyMol from PDB 1MKY.

The interdomain interfaces are formed between the KH domain and each individual G-domain, which do not interact with each other. The interface GD1/KH shows hydrogen bonds between residues of the switch II loop and the first α-helix of the KH-domain. These contacts are thought to be altered upon binding of a nucleotide, which induces a change in conformation. In the case of the interface GD2/KH, the nucleotide-binding site in GD2 is not facing the KH-domain, but is in the opposite side.

The secondary structural elements of both G-domains show it belongs to the TRAFAC class of GTPases, as the whole MnmE-FeoB-Era-EngA-YihA-Septin family, with an antiparallel β-strand adjacent to the Walker B motif.

With the exception of the switch II region of GD1, all other switch regions are mobile and are not defined in this structure, possibly due to lack of coordination of a γ-phosphate and a Mg^{2+} ion. The GD1 switch II region is stabilised by intramolecular contacts with the KH-domain and contacts with neighbouring molecules.

In this structure, a GDP molecule is seen bound to the GD2. The absence of GDP in the crystallisation conditions suggest a tight binding of GDP to GD2 and a slow exchange rate (Robinson et al., 2002). This finding is in contradiction with the biochemical studies that determined a micromolar range affinity of GDP for both GD1 and GD2 (Foucher et al., 2012).

GD1 has two phosphate ions bound in the nucleotide-binding site. The authors claimed these phosphates were located close to the positions of the β- and γ-phosphates of a GTP, and this structure has since been considered to mimic the GTP-bound conformation, thus representing the "on-state" of EngA-GD1 (Robinson et al., 2002). Also supporting this hypothesis is the presence of a water molecule in the expected position of the coordinating Mg^{2+}. However, further observations from a scrutinised analysis of this structure may challenge this assumption. While one of the phosphates ions does overlap the β-phosphate binding site, the second ion is further away from the position of the γ-phosphate (Fig. 24). Electrostatic repulsion and steric hindrance would prevent the two phosphate ions from mimicking the β-γ anhydride bond. In addition, the conserved Gly residue from switch II (DXXG) in this structure is pointing in the opposite direction of that found in GTP-bound GTPases, where it is stabilising the γ-phosphate. Hence, the "on-state" of EngA is still to be confirmed and the *T. maritima* structure is more likely to represent either a post-hydrolytic conformation where the additional phosphate might occupy a "leaving" position, or a conformation induced by the specific crystallisation condition.

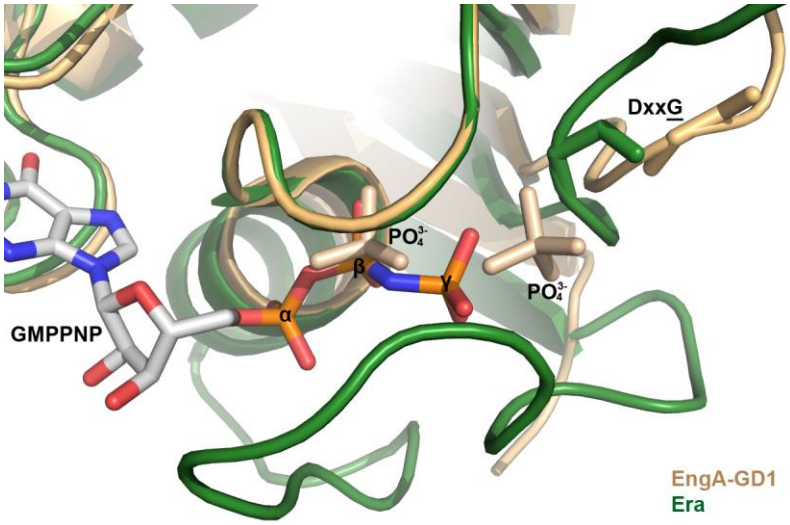

Fig. 24 Superposition of the 3D structures of the GTP-bound Era (green) and phosphate-bound *T. maritima* EngA (orange). The two phosphates bound to GD1 of EngA were suggested to mimic the β- and γ-phosphates of GTP. However, a close analysis of the positions of the phosphate ions shows that the γ-phosphate position is not occupied by any of the phosphate ions. The conserved Gly residue from the G3 motif also adopts a different orientation where it does not stabilise a potential γ-phosphate. Image edited in PyMol from PDB files 1MKY and 3IEV.

Several X-ray crystallography structures from *B. subtilis* ortholog show EngA in a different conformation (Foucher et al., 2012; Muench et al., 2006). In opposition to *Thermotoga* crystal structure, the overall fold of *Bacillus* EngA shows the two G-domains packing on either side of the C-terminal KH-domain and facing each other (Fig. 25).

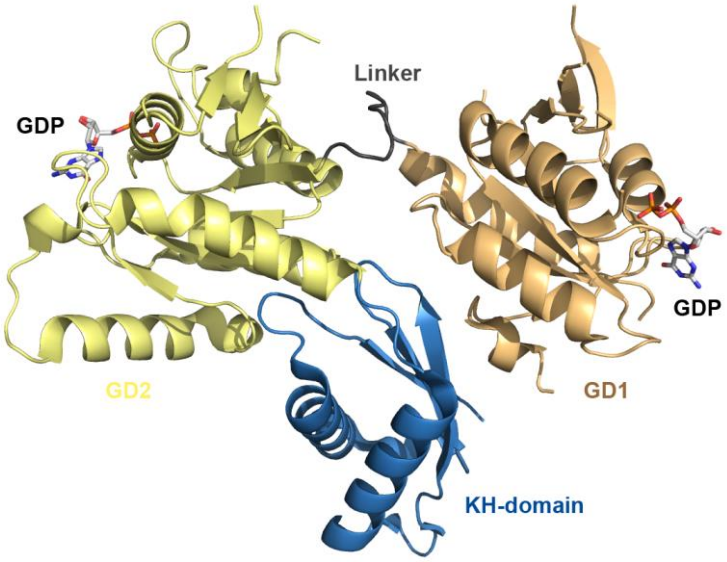

Fig. 25 3D structure of *B. subtilis* EngA obtained by X-ray crystallography at 2.0 Å. This structure reveals a rearrangement of the GD1 relative to the GD2-KH moiety as compared to *T. maritima* structure. An acidic linker of 9 residues (dark grey) connects both G-domains. A GDP molecule (sticks) is bound to GD1 and GD2. Image edited in PyMol from PDB 4DCU.

Although the fold of each domain is similar between the two structures, the overall structure of *B. subtilis* EngA reveals a big rearrangement. Here, the GD1 shows a rotation and a shift of 60 Å from the position found in *T. maritima*. As both G-domains have a bound GDP molecule, this was suggested to represent the “off-state” of EngA (Muench et al., 2006).

Comparison of the GD1 from *T. maritima* and *B. subtilis* structures shows local conformational changes in the switch II that are possibly a response to the binding of a di- or triphosphate nucleotide. *Thermotoga* structure shows a GD1/KH interface formed by switch II and an α-helix of KH-domain. Superposition of *Bacillus* GD1 onto *Thermotoga* structure (Fig. 26A) reveals a steric clash between these two elements, suggesting that GTP hydrolysis might induce a rearrangement of the switch II that destabilise the interface and promote GD1 repositioning (Muench et al., 2006). On the other hand, superposition of *Thermotoga* GD1 onto *Bacillus* structure (Fig. 26B) shows a steric clash between the switch II and the acidic linker between the G-domains, suggesting that binding of GTP induces conformational changes in the switch II region that destabilise the interface GD1/linker and promote a new GD1 repositioning (Muench et al., 2006).

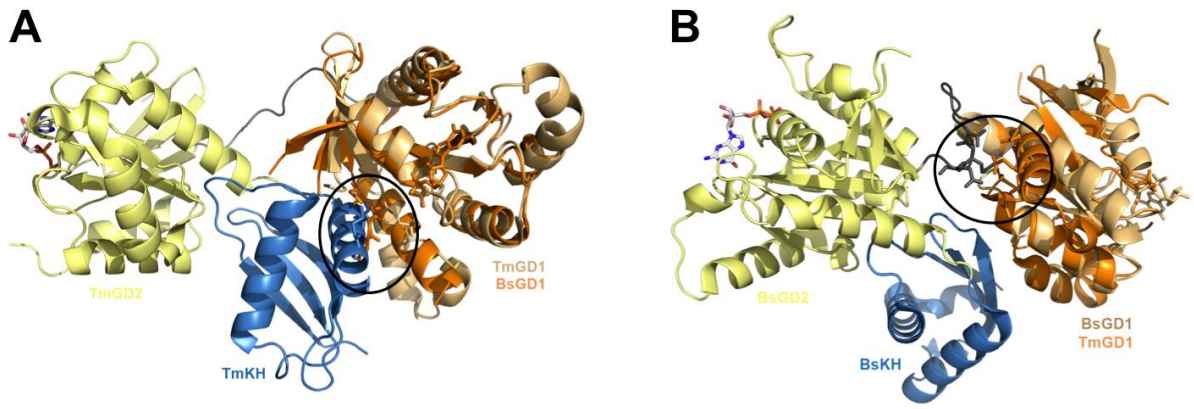

Fig. 26 (A) Superposition of the *B. subtilis* GD1 "off-state" (dark orange) and the full-length *T. maritima* EngA (with GD1 "on-state" represented in light orange). A steric clash (circled in black) can be seen between the α-helix of the KH-domain and the switch II of GD1:GDP, where residues are represented as sticks. **(B)** Superposition of the *T. maritima* GD1 "on-state" (dark orange) and the full-length *B. subtilis* EngA (with GD1 "off-state" represented in light orange). A steric clash (circled in black) can be seen between the acidic linker (grey) connecting the two G-domains and the α2-helix that bears the switch II of GD1:"GTP", where residues are represented as sticks.

Further analysis of the on- and off-conformations indicates that the rearrangement of GD1 exposes or buries a putative RNA-binding site. In the GDP-bound form, the position of GD1 is hiding a positively charged surface in the KH-domain that becomes exposed after the repositioning of GD1 in the claimed "on-state". This region is suggested to be an RNA recognition site (Muench et al., 2006).

Unexpectedly, a crystal structure of *B. subtilis* EngA with a GTP analogue bound to GD2 did not confirm the conformational changes suggested to occur in the switch II. This structure, bearing a GD1:apo and a GD2:GMPPCP reveals the same overall conformation as the GDP-bound form (Foucher et al., 2012). The different occupancy of the nucleotide-binding sites in this structure also suggests different affinities for GD1 and GD2 towards nucleotides.

Recently, the structure of *E. coli* EngA bound to the 50S ribosomal subunit was solved by cryo-EM at a resolution comprised between 5 and 6.5 Å (Zhang et al., 2014). In this structure, EngA interacts with the interface of the two ribosomal subunits and overlaps with the PTC, in the P- and E-sites for tRNA passage (Fig. 27).

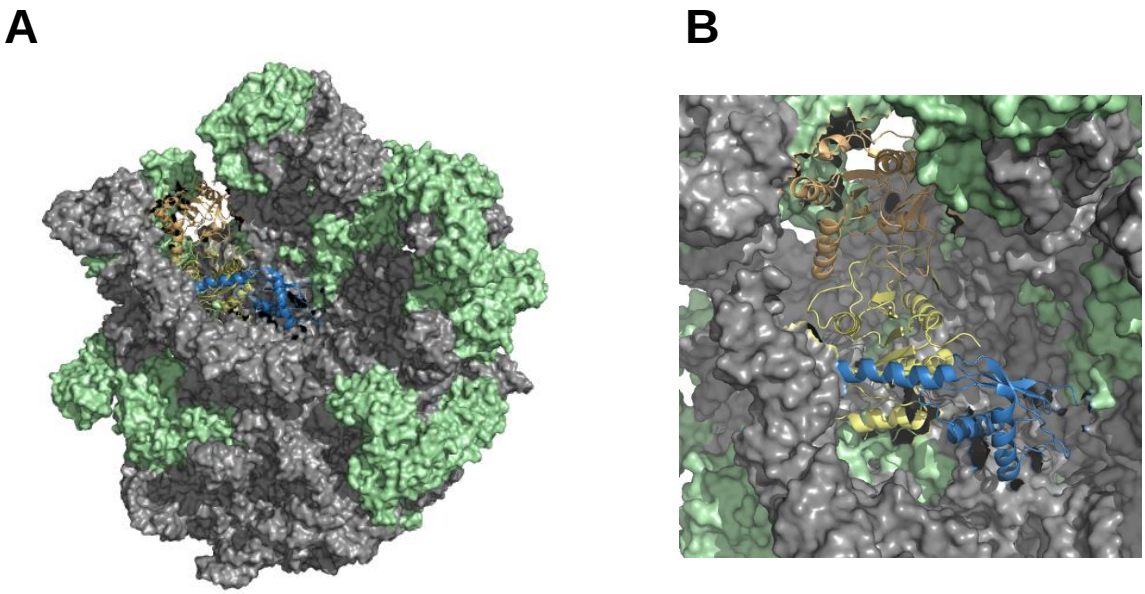

Fig. 27 (A) 3D structure of the *E. coli* 50S:EngA complex obtained by cryo-EM at a resolution of 5–6.5 Å. The ribosomal subunit is represented as a surface, with r-proteins in green and rRNA in grey. EngA, represented in cartoon, is bound to the PTC. EngA domains' colours follow the same scheme as in previous images (GD1, orange; GD2, yellow; KH-domain, blue). Image edited in PyMol from PDB 3J8G. **(B)** Close view of the EngA binding-site on 50S.

The three domains of EngA participate in direct contacts with the 23S rRNA. Both G-domains interact with rRNA helices and r-proteins that compose the PTC and that participate in interactions with tRNA and with the 30S subunit. The KH-domain shows strong interactions with 23S rRNA through an ensemble of basic residues (such as arginines and lysines) that create a positively charged surface in the PTC. The C-terminal extension of *E. coli* EngA, although not visible in the cryo-EM structure, is suggested to interact with the PTC and to be essential for ribosome interaction because deletion of this sequence resulted in reduced binding (Zhang et al., 2014).

The position of EngA on the 50S subunit reveals a steric clash upon association of the 30S subunit. Another interesting feature of this complex lies in the rearrangements observed in the 50S subunit. The displacement of rRNA helices and absence of r-proteins upon binding of EngA render the 50S subunit structure similar to that of a 45S precursor. The repositioning of rRNA helices involved in intersubunit bridges seems to also prevent association of the 30S subunit. These observations suggest that EngA may bind immature 50S subunits and facilitate the maturation of the PTC region during the *in vivo* assembly (Zhang et al., 2014).

A close analysis of EngA structure in complex with the 50S subunit reveals a third different conformation. In the cryo-EM structure, EngA adopts an extended conformation, where the two G-domains form an interface with both nucleotide-binding sites facing each other in an antiparallel orientation (Fig. 28). This extended conformation is achieved by a shift of the GD1 position in relation to the GD2 and KH-domains. The GD2-KH moiety, in turn, retains

the same conformation observed for the *Thermotoga* and *Bacillus* structures. The rearrangement of the G-domains suggests a pseudo-two-fold symmetry (Zhang et al., 2014). Dimerisation is known to activate some GTPases, such as MnmE, where the formation of a homodimer stabilises the switch regions (Scrima and Wittinghofer, 2006). However, the resolution of the cryo-EM structure does not allow comparison of local conformations on the G-domains nor the presence of a bound nucleotide.

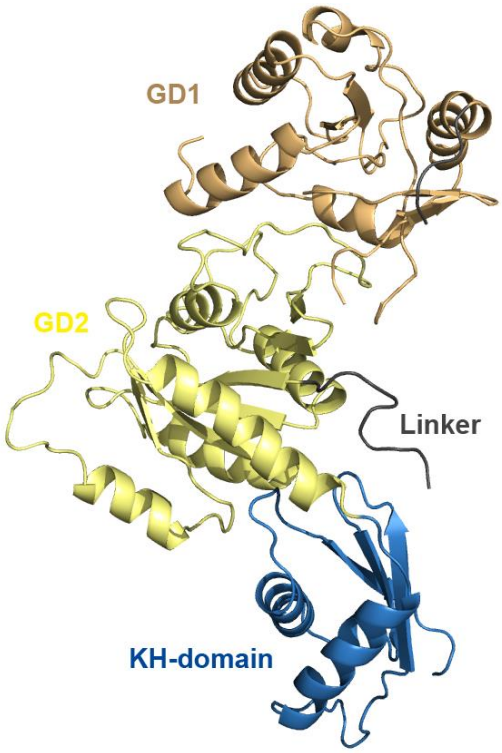

Fig. 28 3D structure of *E. coli* EngA obtained by cryo-EM at 5.0 – 6.5 Å. The original structure is that of the complex of EngA and the ribosomal 50S subunit. When bound to the large subunit, EngA adopts an extended conformation, where the nucleotide-binding sites of the GD1 and GD2 form an interdomain interface and the two G-domains are present in pseudo-two-fold symmetry. Image edited in PyMol from PDB 3J8G.

The superposition of the three structures by their GD2-KH moiety illustrates well the large rearrangements that GD1 undergoes in different bound-states (Fig. 29).

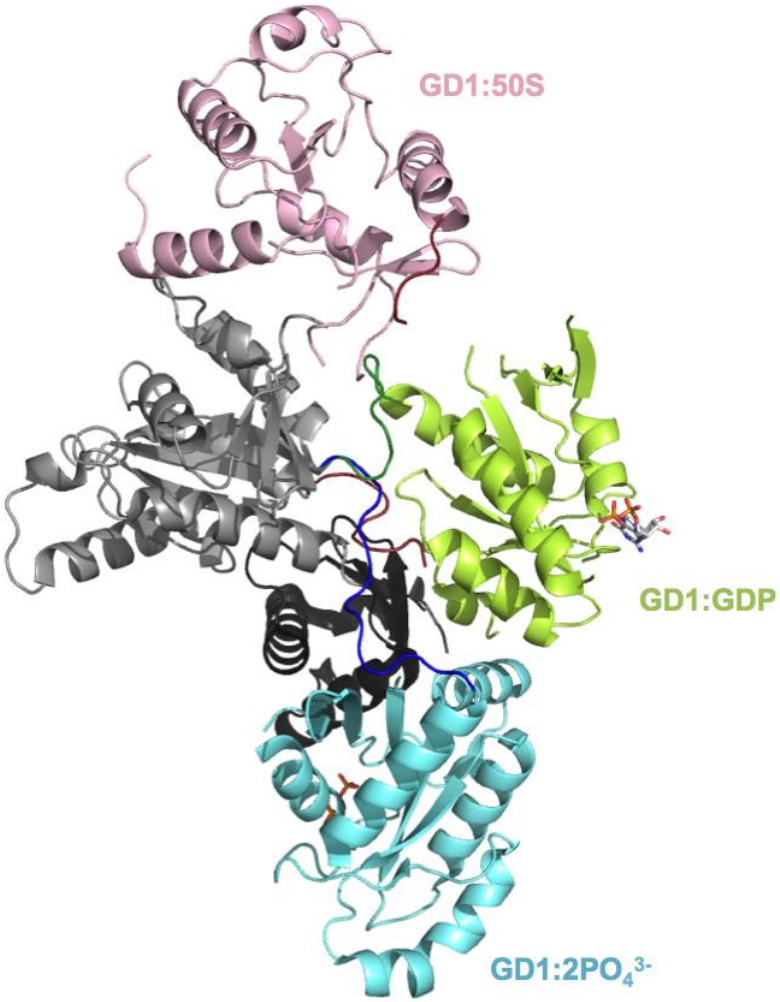

Fig. 29 Superposition of the 3D structures of EngA in different ligand-bound forms. The three structures are superposed by their GD2 and KH domains, which are represented in grey and black, respectively. The GD1s have different relative positions, showing a rearrangement upon binding of different ligands. *B. subtilis* GD1 with a GDP molecule ("off-state") is represented in green. *T. maritima* GD1 has two bound phosphates suggested to mimic the β- and γ-phosphates ("on-state" mimicry) and is represented in cyan. *E. coli* GD1 bound do the 50S ribosomal subunit is represented in pink. GD2 in this same conformation has been resolved bearing either a GDP or a GTP analogue. Image edited in PyMol from PDB files 4DCU, 1MKY and 3J8G.

During the catalytic cycle, conformational changes in GTPases are associated with nucleotide binding/hydrolysis and association/dissociation to an effector molecule. In EngA, the presence of two G-domains renders the cycle more complex and raises questions on whether and how coordination events between G-domains take place. Biochemical and structural data have suggested that GTP-bound EngA adopts a certain conformation that binds the 50S through a regulatory nucleotide-dependent mechanism between both G-domains; subsequent GTP hydrolysis induces a change in conformation to an "off-state" that dissociates from the 50S. Nevertheless, the solved structures of *B. subtilis* EngA with a GDP or GTP analogue bound to the GD2 did not reveal any change in conformation in this domain, which seems to form a rigid structure together with the C-terminal KH-like domain. Also surprisingly, the structure of EngA with a GD1 occupancy mimicking the GTP-bound state and the one of EngA bound to the 50S subunit are located in opposite sites. However, the

low resolution of the cryo-EM structure and the limited mimicry of two phosphates ions for a GTP molecule are factors to be taken into account. These inconsistencies regarding the interpretation of the three states of EngA and their role in the cell are still to be understood.

Thesis project: background and aims

The EngA project has been initiated in 2002, in the early stages of the identification of EngA, within the team of Jean-Michel Jault. This project focused on EngA as a candidate target for the development of antibacterial drugs. The main objective initially defined for my thesis was to develop an ELISA assay to screen ligands and identify inhibitors of the interaction EngA:ribosome. In parallel, crystallisation trials would be employed to determine the structural mechanisms of protein-protein interactions and the binding-mode of inhibitors for helping in structure-based optimisation.

Despite the multitude of information on EngA that has become available since its identification, its function in cells has not yet been resolved and some of its structural characteristics are elusive and intriguing. Thus, during our project, while working towards our main goals, several questions arose that lead us to explore other paths. Along with the development of the screening assay, we have assessed structural aspects of EngA and tried to understand their implications in EngA function.

The next section of this manuscript presents a description of the methods and the results obtained during the course of my thesis. These results are divided in five parts:

1. The first chapter describes the expression and purification of all proteins used in our experiments, namely EngA, S7, 70S, 50S and 30S.

2. The second chapter addresses the aspects of conformational changes of EngA. EngA has the unique feature among GTPases of comprising two G-domains. The mechanisms involved in its switch cycle and how they relate with ribosome binding are still unclear. An integrated approach was employed based on biochemical and structural biology techniques to study conformational changes of EngA in solution.

3. The results described in chapter 2 led us to search for different nucleotide-bound conformations of EngA. The third chapter describes crystallography studies performed on EngA in order to obtain the atomic structure of these conformations. In the same section, crystallisation attempts of obtaining the structure of the EngA:S7 complex and EngA bound to inhibitors are described.

4. As nucleotide-induced conformational changes on EngA have been associated with binding to the ribosome, interaction studies were performed based on our results from the structural investigations of EngA. These are discussed in chapter 4.

5. Finally, chapter 5 describes preliminary inhibitor screening assays targeting the interaction of EngA with the bacterial ribosome.

MATERIAL AND METHODS

1. Molecular biology methods

1.1. Vectors for recombinant protein expression

We have used the EngA ortholog from *B. subtilis*. The vector pET15b (Novagen) was used to express EngA with an N-terminal His-tag. This vector is under control of a T7 promotor induced by IPTG and contains an ampicillin resistance gene.

The cloning of $EngA_{WT}$ was done previously by Anne-Emmanuelle Foucher (IBS, Grenoble) (method described in (Foucher, 2010)).

The vector pET22b(+) (Novagen) was used to express S7 with a C-terminal His-tag and the vector pGEX-4T1 (GE Healthcare) was used to express S7 with an N-terminal GST-tag. Both vectors are under the control of a T7 promotor induced by IPTG and contain an ampicillin resistance gene. pGEX-4T1 contains a thrombin recognition site for cleavage of the fusion product.

The cloning of $S7_{WT}$ was done previously by Jean-Baptiste Reiser (IBS, Grenoble).

1.2. Mini-prep for preparation of plasmid DNA

Extraction and purification of plasmid DNA was done using the NucleoSpin® Plasmid method (Macherey-Nagel) that consists on three steps: clearing of bacterial lysate, adsorption of DNA onto a silica membrane and washing and elution of plasmid DNA.

A single colony of transformed *E. coli* Top10 was inoculated in 4 mL LB/ampicillin 100 µg/mL. The cell suspension was incubated at 37ºC, 180 rpm, overnight. Plasmid DNA was prepared according to the protocol provided by the kit and finally eluted with 50 µL Buffer AE (5 mM Tris pH 8.5).

Quantification of DNA was done by measuring OD_{260} on a spectrophotometer NanoDrop™ 2000 (Thermo Scientific), using 2 µL sample and 2 µL Buffer AE as blank.

1.3. Site-directed mutagenesis

Site-directed mutagenesis (SDM) is an *in vitro* method that creates specific mutations in a double stranded plasmid DNA of known sequence. This method makes use of oligonucleotide primers designed to be complementary to the template DNA while carrying the desired mutation.

Two EngA mutants were generated by SDM: $EngA_{G139C/V388C}$ and $EngA_{Y51C/T169C}$. These double mutants were obtained by first creating single mutants on which the second mutation was inserted. Single mutants $EngA_{G139C}$ and $EngA_{V388C}$ had been previously cloned by Anne-Emmanuelle Foucher (IBS, Grenoble).

The oligonucleotides used were the following:

Table 2 Oligonucleotide primers used to generate EngA cysteine mutants.

Mutation	Sequence (5' – 3')	Tm (°C)
Y51C-for	GAATGGCT**C**AATT**G**TGATTTTAATTTG	57.4
Y51C-rev	CAAATTAAAATCA**C**AATT**G**AGCCATTC	57.4
G139C-for	GGCTTT**T**GCGAGCCGTATCCAATT	62.7
G139C-rev	CGGCTCGC**A**AAAGCCTAGCGAATAAAAATCATA**T**ATATT CGCTCTC	72.1
T169C-for	CATTCCTGAA**TGT**AAATACAATGAAG	56.9
T169C-rev	CTTCATTGTATTT**ACA**TTCAGGAATG	56.9
V388C-for	GCGACTCAA**TGC**TCGGTAAAGCCGCCAAG**T**TTCGTT	77.4
V388C-rev	TACCGA**GCA**TTGAGTCGCATAGTAAATTTTC	63.6

The mutations were created using the QuikChange® II XL site-directed mutagenesis kit (Agilent). For the PCR reaction, 10 ng of template DNA (pET15b-$EngA_{WT}$ for single mutants), 5 µL buffer, 1 µL dNTPs, 1.5 µL of primers forward and reverse and 1 µL *Pfu* polymerase were used in the following thermocycling conditions:

Table 3 Thermocycling conditions used for site-directed mutagenesis of EngA cysteine mutants.

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	1 minute	1
Denaturation	95	50 seconds	18
Annealing	50	50 seconds	
Extension	68	7 minutes	
Final extension	68	7 minutes	1

After the PCR reaction, the product was incubated 1 h at 37°C with restriction enzyme *DpnI* (10 U) to digest the template plasmid. The final reaction was transformed into competent XL10-Gold cells from the kit and spread onto a LB/ampicillin plate. The resulting colonies were used to make mini-preps and the plasmid DNA was sequenced (GATC Biotech) to confirm the presence of the mutations.

1.4. Preparation of Ca^{2+}-competent E. coli strains

Bacteria are rendered chemically competent by treatment with calcium chloride, which increases the permeability of the cell membrane, enabling cells to easily attach plasmid DNA to their surface.

In our work, three *E. coli* competent strains were used: Top10 for plasmid propagation during transformation, C41(DE3) for recombinant EngA overexpression and BL21(DE3) for recombinant S7 overexpression.

A preculture of 5 mL LB was incubated at 37°C, 180 rpm, overnight. The next day, this pre-culture was diluted 100x and incubated at 37°C, 180 rpm, until it reached an OD_{600} of 0.5–0.6. The cells were harvested by centrifugation at 1800 g, 10 minutes, 4°C, resuspended in 100 mL $MgCl_2$ 100 mM and kept on ice for 30 minutes. After a second centrifugation, the cell pellet was resuspended in 10 mL $CaCl_2$ 15% glycerol. Competent cells were kept in aliquots at -80°C.

1.5. Transformation of competent cells

1 µL of plasmid was added to 100 µL of competent cells and incubated for 10 minutes on ice for the DNA to come into close contact with bacteria. Heat shock in a water bath at 42ºC for 60 seconds creates pores in the cell membrane, allowing the plasmid to enter the cell. Bacteria were then incubated on ice for 2 minutes to retain the plasmid inside cells. 200 µL of LB medium were added to the cell suspension, followed by incubation at 37ºC, 180 rpm, 1h and spreading 150 µL onto a LB-agar plate with the appropriate antibiotic (ampicillin 100 µg/mL). The plates were kept at 37ºC during the night and transferred to 4ºC the next day.

2. Expression and purification of recombinant EngA

2.1. Overexpression

One colony from an LB-agar plate with *E. coli* C41(DE3) / pET15b-EngA transformants was inoculated in 20 mL LB/ampicillin and grown at 37ºC, 180 rpm, overnight. This preculture was added to 1L LB/ampicillin and incubated at 37ºC, 180 rpm until it reached an OD_{600} of 0.6–0.8. Protein overexpression was induced by addition of IPTG 1 mM and the culture was continued for 4h. The cells were harvested by centrifugation at 1800 g, 20 minutes, 4ºC and kept at -20ºC.

2.2. Purification

Purification of His-EngA was done by two chromatographic steps: immobilised metal-affinity chromatography (IMAC) followed by size exclusion chromatography (SEC).

Cell lysis

After overexpression, the pellet of 1L culture of *E. coli* C41(DE3) / pET15b-EngA was resuspended in 20 mL Lysis Buffer (50 mM Tris pH 8.0, 250 mM NaCl, 10 mM imidazole, 1mM DTT, 1 mM PMSF, 2 µg/mL DNAse I) and cells were disrupted using a microfluidizer (3 cycles, 1000 bars). Centrifugation at 20000 g, 45 minutes, 4ºC removed cell debris and the supernatant containing soluble His-EngA was recovered.

Immobilised metal affinity chromatography

IMAC is based on the affinity of transition metal ions (such as Zn^{2+}, Ni^{2+}, Cu^{2+} and Co^{2+}) to histidine and cysteine residues in aqueous solutions.

Recombinant EngA contains an N-terminal His-tag that corresponds to a sequence of six consecutive histidine residues. This histidine tail has affinity for Ni^{2+} ions present in a stationary phase in a column, allowing retention of EngA and elimination of contaminant proteins.

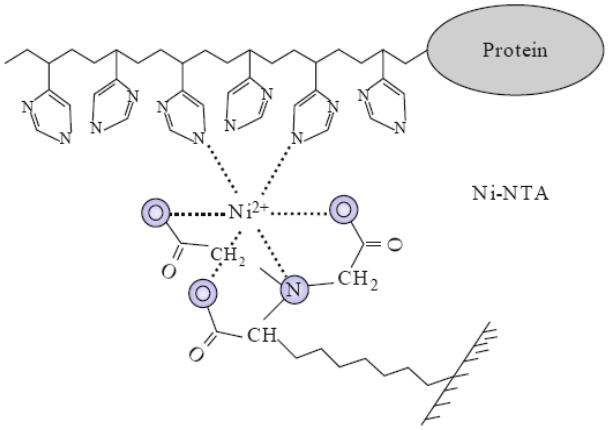

Fig. 30 Representation of the interaction between residues in the His-tag and nickel ions. The NTA ligand coordinates the Ni^{2+} with four valencies, leaving two valencies available for interaction with imidazole rings of histidine residues. Adapted from Block et al., 2009.

In a column, the soluble fraction obtained after cell lysis was added to 10 mL Ni-NTA Agarose (Qiagen) previously washed with water and equilibrated in Washing Buffer (50 mM Tris pH 8.0, 250 mM NaCl, 20 mM imidazole). The flow-through was kept for further analysis. The resin was then washed with 10 volumes of Washing Buffer. His-EngA was eluted in five fractions of 10 mL Elution Buffer (50 mM Tris pH 8.0, 250 mM NaCl, 300 mM imidazole, 10%

glycerol). For EngA double mutants, both Washing and Elution buffers were complemented with 1 mM DTT.

The flow-through, washing and elution fractions were analysed in a 12% acrylamide gel by SDS-PAGE.

Size exclusion chromatography

SEC separates components of a sample according to differences in size, as molecules pass through a porous matrix of spherical particles that form a packed bed.

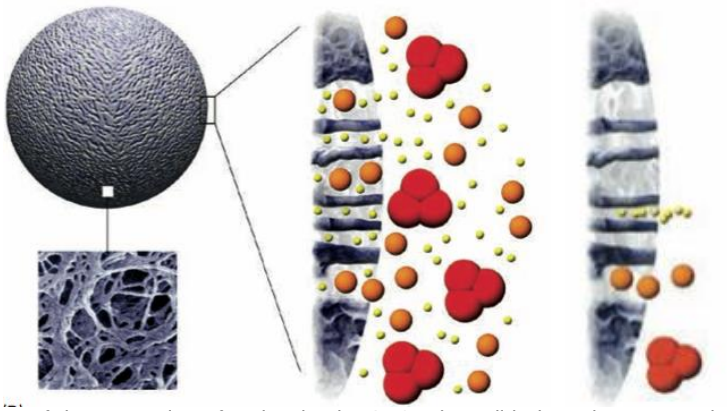

Fig. 31 Representation of the separation of molecules by SEC. The solid phase is composed of a matrix with pores. Small molecules from the sample diffuse into the pores and are more delayed than molecules with higher molecular weight that are eluted first. Image adapted from GE Healthcare Life Sciences.

The purified protein was centrifuged (30 minutes, 16000 g, 4°C) to eliminate aggregates and injected (10 mL each time) in a HiLoad® 26/60 Superdex® 75 prep grade (GE Healthcare) previously equilibrated in 50 mM Tris pH 8.0, 250 mM NaCl, 10% glycerol. Elution was done with the same buffer, used as storage buffer, at a flow 2.5 mL/min in 1 mL fractions, using the system ÄKTA™ Purifier (GE Healthcare) and the software UNICORN™ (GE Healthcare). For EngA double mutants, SEC buffer was complemented with 1 mM DTT.

After gel filtration, the fractions of interest were identified based on the chromatogram and loaded on a 12% acrylamide gel. The eluted protein was concentrated using a Vivaspin® centrifugal concentrator (Sartorius) with a cut-off of 30 kDa until a final concentration of ~4 mg/mL.

2.3. Determination of enzymatic activity

After purification of EngA, the enzymatic activity was determined to confirm the folding and function of the enzyme. GTPase activity is measured using a coupled assay were one molecule of hydrolysed GTP corresponds to the formation of one molecule of NAD^+, that is

detected at 340 nm. The amount of GTP consumed is given by the decrease in absorbance. During the assay, GTP is regenerated to maintain saturating substrate concentrations, allowing the reaction to take place at V_{max}.

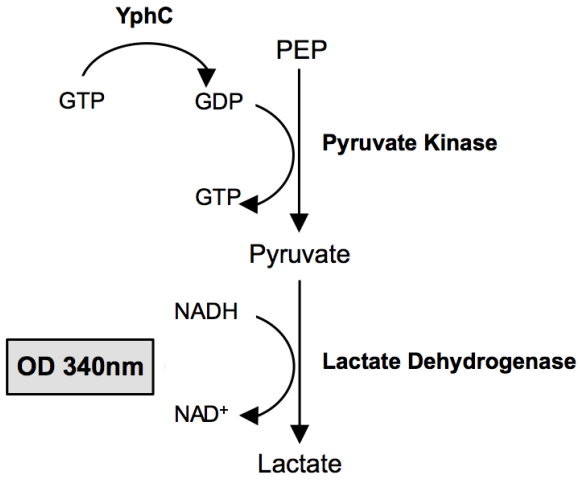

Fig. 32 Schematic representation of the PK/LDH coupled assay for determination of the GTPase activity of EngA.

A reaction mix was prepared with 4 mM phosphoenolpyruvate (PEP), 20 µg pyruvate kinase (PK), 10 µg lactate dehydrogenase (LDH), 0.8 mM NADH, 1 mM $MgCl_2$, 1 mM GTP and 300 mM KCl in buffer 50 mM Tris pH 8.0, 250 mM NaCl and incubated 5 minutes at 37°C. EngA was added to a final concentration of 0.1 mg/mL and the absorbance at 340 nm followed during 10 minutes.

The specific enzymatic activity (SA) (expressed in mol/min/mg) is calculated based on the equation:

$$SA = \frac{\Delta OD}{s} \times \frac{Vr}{m\,\varepsilon NADH} \times 60$$

where $\Delta OD/s$ is the variation of absorbance at 340 nm per second, $\varepsilon NADH$ is the molar extinction coefficient of NADH (6220 $M^{-1}cm^{-1}$), Vr is the reaction volume (expressed in L) and m is the mass of protein (expressed in mg).

3. Expression and purification of recombinant S7

3.1. Overexpression

One colony from an LB-agar plate with *E. coli* BL21(DE3) / pET22b(+)-S7 or *E. coli* BL21(DE3) / pGEX-4T1-S7 transformants (generating S7-His and GST-S7, respectively) was inoculated in 20 mL LB/ampicillin and grown at 37ºC, 180 rpm, overnight. This preculture

was added to 1L LB/ampicillin and incubated at 37ºC, 180 rpm until it reached an OD_{600} of 0.6–0.8. Protein overexpression was induced by addition of IPTG 1 mM and the culture was continued for 5h. The cells were harvested by centrifugation at 1800 g, 20 minutes, 4ºC and kept at -20ºC.

3.2. Purification

S7 was purified by two chromatographic steps: affinity chromatography followed by size exclusion chromatography.

Cell lysis

After overexpression, the pellet of 1L culture was resuspended in 20 mL Lysis Buffer (50 mM Tris pH 8.0, 500 mM NaCl, 0.5 mM PMSF, 2 µg/mL DNAse I, RNAse A) and cells were disrupted by sonication. The cell lysate was centrifuged at 20000 g, 45 minutes, 4ºC and the supernatant containing soluble proteins was recovered.

Immobilised metal affinity chromatography (for S7-His)

In a column, the soluble fraction obtained after cell lysis was added to 5 mL Ni-NTA Agarose (Qiagen) previously washed with water and equilibrated in Washing Buffer (50 mM Tris pH 8.0, 500 mM NaCl). After 1h agitation at 4°C, the flow-through was eluted and the resin washed with 50 mL of Purification buffer. A volume of 25 mL of Purification Buffer complemented with increasing concentrations of imidazole (125 mM, 250 mM, 500 mM) was passed through the column.

The flow-through, washing and elution fractions were analysed in a 15% acrylamide gel by SDS-PAGE.

Glutathione S-transferase affinity chromatography (for GST-S7)

The principle of GST affinity chromatography is the same as for IMAC, although this variant uses immobilised glutathione on the matrix to bind the GST-tag.

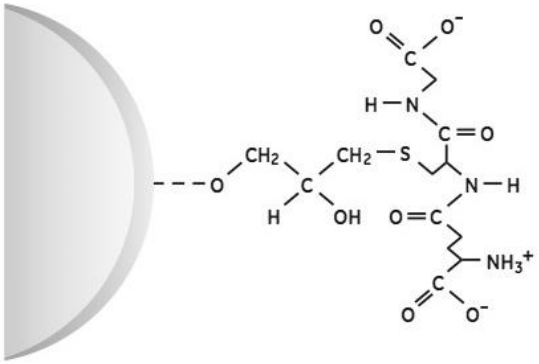

Fig. 33 Terminal structure of glutathione immobilised in a sepharose matrix. This structure is complementary to the gluthatione S-transferase (GST) binding site. Reduced glutathione competes with the resin and causes dissociation of the GST-tagged proteins. Adapted from GE Healthcare Life Sciences.

In a column, the soluble fraction obtained after cell lysis was added to 5 mL Glutathione Sepharose 4B (GE Healthcare) previously washed with water and equilibrated in Washing Buffer (50 mM Tris pH 8.0, 500 mM NaCl). After 1h agitation at 4°C, the flow-through was eluted and the resin washed with 50 mL of Purification buffer. A volume of 25 mL of Purification Buffer complemented with 10 mM reduced gluthatione was passed through the column.

The flow-through, washing and elution fractions were analysed in a 12% acrylamide gel by SDS-PAGE.

Size exclusion chromatography

The purified S7-His and GST-S7 proteins were centrifuged (30 minutes, 16000 g, 4°C) to eliminate aggregates and injected (10 mL each time) in a HiLoad® 26/60 Superdex® 75 prep grade (GE Healthcare) previously equilibrated in 50 mM Tris pH 8.0, 500 mM NaCl. Elution was done in with the same buffer, at a flow 2.5 mL/min in 1 mL fractions, using the system ÄKTA™ Purifier (GE Healthcare) and the software UNICORN™ (GE Healthcare).

After gel filtration, the fractions of interest were identified based on the chromatogram and loaded on a 12–15% acrylamide gel.

Tag cleavage and ion exchange chromatography

Recombinant S7 without a tag was obtained by cleavage of the tag of GST-S7 by thrombin. S7 was purified by ion exchange chromatography (IEC).

IEC separates proteins on the basis of differences in their net surface charge. The stationary phase is composed of charged ionic groups that interact with molecules of opposite charge.

At a certain pH, proteins that display different net charges can be separated by controlling interactions between stationary and mobile phases.

Cleavage of GST was done overnight, at room temperature, with 2 U thrombin/mg GST-S7. After the digestion, the buffer was changed with a PD10 column (GE Healthcare) to 50 mM Tris pH 8.0, 50 mM NaCl. The protein solution was centrifuged (30 minutes, 16000 g, 4°C) to eliminate aggregates and injected (5 mL each time) in a Resource S® column (GE Healthcare) previously equilibrated in 50 mM Tris pH 8.0, 50 mM NaCl. Elution was done with a salt gradient from 50 mM to 1 M NaCl, at a flow 4 mL/min in 0.5 mL fractions, using the system ÄKTA™ Purifier (GE Healthcare) and the software UNICORN™ (GE Healthcare).

After ion exchange chromatography, the fractions of interest were identified based on the chromatogram and loaded on a 15% acrylamide gel.

4. Purification of ribosomes

Ribosomes were purified from a culture of *Bacillus subtilis* grown in LB medium at 37°C, 180 rpm until a final OD_{600} of 0.6.

B. subtilis cells were harvested by centrifugation at 1800 g, 20 minutes, 4ºC. Cells were resuspended in 20 mL Buffer A (20 mM Tris pH 7.5, 20 mM $MgCl_2$, 200 mM NH_4Cl, 0.1 mM EDTA, 6 mM β-mercaptoethanol, 1mM PMSF, 1 U/mL DNAse I) and disrupted using a microfluidizer (3 cycles, 1000 bars). Centrifugation at 30000 g, 30 minutes, 4°C removed cell debris. The supernatant was then centrifuged at 150000 g, 3h30, 4°C. This step forms two pellets: a clear one, on the bottom, formed by ribosomes and a brown one, on top, formed by membrane proteins. The two pellets were resuspended in Buffer A, loaded on a 42% sucrose cushion prepared in Buffer B (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 500 mM NH_4Cl, 0.1 mM EDTA, 6 mM β-mercaptoethanol) and centrifuged at 100000 g, 16h, 4°C to obtain a crude risobomal pellet. The purified ribosomes were resuspended in Buffer C (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 50 mM NH_4Cl, 0.1 mM EDTA, 6 mM β-mercaptoethanol), fast frozen in liquid nitrogen and kept at -80°C.

To obtain the individual subunits 30S and 50S, the purified 70S ribosomes were dialysed in Buffer D (20 mM Tris pH 7.5, 3 mM $MgCl_2$, 300 mM NH_4Cl, 0.15 mM EDTA, 2 mM DTT) during the night at 4°C. The same buffer was used to prepare 5–40% sucrose gradients. Gradients were prepared using a gradient mixer, where a 5% sucrose solution is loaded in a reservoir chamber and the 40% sucrose solution is loaded on a mixing chamber, allowing

both solutions to slowly mix and form a linear gradient. A volume of 500 µL sample was loaded on 11 mL sucrose gradient and centrifuged at 25000 g, 15h, 4°C. Fractions of 300 µL were collected in 96-well plates and the OD_{260} and OD_{280} were measured in a plate reader. The fractions corresponding to each subunit were gathered and dialysed to Buffer E (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT) using an Amicon Ultra centrifugal concentrator (Millipore) with a cut-off of 100 kDa.

Quantification of the ribosomes was estimated by absorbance at 260 nm, using the following values: 1 AU_{260} corresponds to 23 pmol of 70S ribosomes, 34 pmol of 50S subunit and 69 pmol of 30S subunit (Daigle and Brown, 2004).

5. Methods for protein structural characterisation

5.1. Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) allows the determination of structural parameters of a molecule in solution. More particularly, SAXS provides information regarding size and shape of molecules. During a SAXS experiment, a beam of X-rays illuminates a protein sample in solution and is scattered. The scattered radiation is recorded on a detector, from which the intensities are plotted against the scattering angle after circular averaging of the recorded signal, providing the scattering curve. From this scattering curve we can determine few structural parameters of the molecules in solution (Jacques and Trewhella, 2010).

We employed SAXS to analyse conformational changes of EngA in solution, upon binding of different nucleotides.

Sample preparation

$EngA_{WT}$ in different nucleotide bound forms and cross-linked double-cysteine mutants were analysed in solution by SAXS. Frozen samples of EngA were thawed and centrifuged for 30 minutes at 16000 g, 4°C to remove aggregates. EngA was washed through a PD10 gel filtration column (GE Healthcare) to exchange the buffer to 50 mM Tris, 250 mM NaCl (pH 8.0). This buffer was used as a reference for SAXS experiments and for preparing sample dilutions to obtain a range of protein concentrations between 0.5 and 4 mg/mL. The protein concentration was determined by absorption spectroscopy using the molar extinction coefficient (ε = 33350 $M^{-1}.cm^{-1}$).

For the preparation of nucleotide-bound protein, the elution buffer was complemented with $MgCl_2$. Nucleotides (GDP, GMPPNP and GMPPCP) were added to the protein solution to a final concentration of 1 mM or 10 mM. The same was done for the buffer, which was then used for preparing sample dilutions and as the reference for SAXS experiments. As the nucleotides interfere with visible absorption spectroscopy at 280 nm, protein concentration was estimated based on the experimental dilution and concentration factors and further confirmed by the I_0 values obtained by the SAXS scattering curve analyses.

Data collection and processing

Small angle X-ray scattering data from solutions of different nucleotide bound forms of His-$EngA_{WT}$ were collected at the BM29 beamline (ESRF, Grenoble, France) (Pernot et al., 2013) using a Pilatus 1M detector. The detector-distance of 2.9 m covered a momentum transfer range of $0.025 < q < 5$ nm^{-1}. SAXS experiments were carried at 20ºC in the automatic sample changer with sample concentrations ranging from 0.5 to 4 mg/mL and sample volumes of 50 µL in a quartz glass capillary.

For each protein sample, data from the corresponding buffer was collected to provide a reference for the scattering background. For each measurement, ten frames were collected for which radiation damage was investigated. An initial trial using a SEC system coupled with SAXS was done to check for the presence of aggregates. For this trial, we used a Superdex® 75 10/300 GL column (GE Healthcare) equilibrated in 50 mM Tris pH 8.0, 250 mM NaCl. As scattering curves were nearly identical for both SEC and sample changer set up (see *Results, section 2.1.1*), the latter was used for all experiments.

Data processing was done using the ATSAS suite of program.

The forward scattering I_0 and the radius of gyration R_g ($R_g^2 = \frac{\int r^2 \rho(\vec{r})\, d\vec{r}}{\int \rho(\vec{r})\, d\vec{r}}$) were calculated from the Guinier approximation using the program package PRIMUS (Konarev et al., 2003) as:

$$\ln I(q) = \ln I(0) - \left(\frac{{R_g}^2}{3}\right) q^2$$

where I represents the intensity and q the scattering angle. The linear fit of ln I_q versus q^2 yields I_0 and R_g from the y-axis intersection and the slope, respectively.

The I_0 is proportional to the mass of the scattering molecule:

$$MW = N_A \frac{I_0}{c} \frac{1}{(\Delta\rho_M)^2}$$

Since $\Delta\rho_M$ is constant for proteins, a calibration either with water or a known protein (such as BSA) enables the determination of the molecular weight of the scattering molecule if the molecule concentration is known.

The distance distribution function *P(r)* and maximum dimensions D_{max} were obtained using the indirect transform package GNOM (Svergun, 1992). *P(r)* is proportional to the Fourier transform of the scattered intensity and can be calculated directly from I_q:

$$P(r) = \frac{r^2}{2\pi^2} \int_{q=0}^{\infty} q^2\, I(q)\, \frac{sin(qr)}{qr}\, dq$$

D_{max} is the maximum interatomic distance in the molecule and corresponds to the value of *r* at which *P(r)* returns to zero after reaching its maximum value. In practice, a proper value of D_{max} corresponds to the minimum value for which the intensity calculated from the Fourier transform of *P(r)*:

$$I(q) = 4\pi \int_0^{D_{max}} P(r)\, \frac{\sin qr}{qr}\, dr$$

fits perfectly the experimental scattering curve.

The distance distribution *P(r)* can be used to calculate the R_g taking the entire *q* range into account:

$$R_g^2 = \frac{\int_0^{D_{max}} r^2\, P(r)\, dr}{\int_0^{D_{max}} P(r)\, dr}$$

This estimate of the R_g should be consistent with the one derived from the Guinier's law.

The Porod invariant, only valid for folded molecules:

$$Q = \int_{q=0}^{\infty} q^2\, I(q)\, dq$$

corresponds to the surface below the Kratky plot ($q^2.I(q) = f(q)$) and is related to the volume of the scattering molecule – the Porod volume – by:

$$V = 2\pi^2\, \frac{I(0)}{Q}$$

An estimate of the molecular weight can be derived from the Porod volume:

$$MW \approx \frac{V}{1.6}$$

Fitting of the experimental data and the theoretical scattering curves computed from the PDB entries 4DCU, 1MKY and 3J8G was done using CRYSOL (Svergun et al., 1995). Statistical validation of the fitting analysis was done by evaluation of the χ^2 values between the different models:

$$\chi^2 = \sum_q \frac{\left(I_{sample}(q) - I_{buffer}(q) - \alpha.I_{model}(q)\right)^2}{\sigma\left(I_{sample}(q)\right)^2 + \sigma\left(I_{buffer}(q)\right)^2}$$

For samples containing nucleotides, we observed that the concentration of nucleotide in the sample and the buffer did not exactly match, despite dialysis or the use of the concentration flow-through as reference buffer. As the data could not be used as such, we developed a correction program to take into account a nucleotide concentration difference between sample and buffer, considering that the measured sample scattering should be considered as the sum of protein and nucleotide contribution:

$$I_{protein}(q) = I_{sample}(q) + \beta.I_{nucleotide}(q) - I_{buffer}(q)$$

Therefore, a scattering curve for the nucleotide $I_{nucleotide}(q)$ was measured by subtracting scattering curves corresponding to the same buffer either containing 10 mM nucleotide or without nucleotide. Then, a program was written to estimate the difference in nucleotide content between the buffer and the sample. The program calculates the α and β coefficients in order to minimize the χ^2 value:

$$\chi^2 = \sum_q \frac{\left(I_{sample}(q) + \beta.I_{nucleotide}(q) - I_{buffer}(q) - \alpha.I_{model}(q)\right)^2}{\sigma\left(I_{sample}(q)\right)^2 + \sigma\left(I_{nucleotide}(q)\right)^2 + \sigma\left(I_{buffer}(q)\right)^2}$$

with I_{model} being the theoretical scattering intensity of one of the available EngA structures (PDB entries 4DCU, 1MKY, 3J8G) calculates by CRYSOL. The χ^2 is calculated in the q range 0.8–2.5 nm^{-1}, where the fit between measured and calculated scattering curves is optimal. The corrected experimental scattering curve for the nucleotide-bound EngA is then $I_{sample}(q) + \beta.I_{nucleotide}(q) - I_{buffer}(q)$.

5.2. Limited Proteolysis by trypsin

Limited proteolysis is a biochemical technique that allows probing conformational features – tertiary structure, folding and dynamics – of proteins in solution. This method is based on the ability of proteases to interact and cleave polypeptide substrates in exposed regions of backbone chain flexibility (Fontana et al., 2004).

Limited proteolysis of His-$EngA_{WT}$ was performed at 20°C in 50 mM Tris pH 8.0, 250 mM NaCl and with an enzyme to substrate ratio of 1:600. The assay was done after incubation of EngA for 15 minutes at 37°C in the absence of nucleotides or in the presence of 10 mM

GDP/$MgCl_2$ or 10 mM GMPPCP/$MgCl_2$. Aliquots were collected at time intervals and the reaction stopped by addition of 2x denaturing loading dye and incubation at 95°C for 10 minutes. The digestion pattern was visualised by Tricine-SDS-PAGE on 15% acrylamide gels stained with Coomassie brilliant blue G-250 solution. The gels were processed for fragment analysis using ImageJ 1.49 and ImageLab^{TM} 4.0 softwares.

5.3. N-terminal sequencing

N-terminal sequencing was employed to determine the tryptic cleavage sites of His-EngA_{WT}. This method allows determination of ten residues of sequence from the N-terminus of each peptide present in the sample.

After digestion of His-EngA_{WT} (see method description above), the fragments in solution were analysed at the Platform Seq3A of ISBG at IBS by Jean-Pierre Andrieu.

Amino acid sequence determination was done using the Procise® Sequencing system 492 (Applied Biosystems), based on the Edman degradation reaction. During a cycle, phenylisothiocyanate reacts with the N-terminus residue of peptides which is subsequently cleft to generate a phenylthiohydantoin-amino acid derivative (Speicher et al., 2009). N-terminal residues were identified using a 140C HPLC system (Applied Biosystems). Retention times and integration values of peaks are compared to the chromatographic profile of a standard mixture of derivatised amino acids.

5.4. Mass spectrometry

Mass spectrometry (MS) analysis was employed for identification of the fragments generated by tryptic cleavage of His-EngA_{WT}.

Proteolytic cleavage of His-EngA_{WT} by trypsin was done according to the protocol described above. For MS analysis purposes, the digestion was stopped by acidification with 5% trifluoroacetic acid. Fragments in solution were analysed at the Mass Spectrometry Platform of ISBG at IBS by Luca Signor.

8 µL of digested sample at a final protein concentration of 8 µM were diluted with 95 µL of 0.03% TFA. The entire volume was injected for analysis. The sample was desalted on a reverse-phase HPLC column. MS analysis was done by electrospray ionisation (ESI) coupled to a time-of-flight (TOF) mass spectrometer, using the system 6210 LC-ESI-TOF (Agilent Technologies).

In LC-ESI-TOF, a liquid sample is injected into the apparatus via a liquid chromatography interface. The sample is ionised directly from the liquid phase, generating charged particles. A TOF mass spectrometer accelerates the released ions that are separated according to their mass-to-charge ratio (m/z) in function of time. At the end of the trajectory, the ions reach a detector that generates a spectrum of masses calculated from the “time of flight” (Jordana-Lluch et al., 2012). LC-ESI-TOF allows mass determination with high accuracy (≤ 50 ppm) (Signor and Boeri Erba, 2013).

5.5. Cross-linking assay

A cross-linking assay was done to induce formation of disulfide bonds on the two double Cys mutants generated for EngA.

Both $EngA_{G139C/V388C}$ and $EngA_{Y51C/T169C}$ were dialysed against 50 mM Tris pH 8.0, 300 mM KCl in a PD10 column (GE Healthcare) to remove DTT. Before the dialysis, the buffer was vigorously mixed. After removal of DTT, EngA was incubated with 10 mM $MgCl_2$, 10 mM GDP/$MgCl_2$ or GMPPCP/$MgCl_2$ for 15 minutes at 37°C. $CuSO_4$ 0.2 mM was added to promote cysteine oxidation and samples were again incubated for 2h at RT. During the 2h incubation, $CuSO_4$ was further added at times 45 and 90 minutes. $EngA_{WT}$ was used as a negative control.

A volume corresponding to 1 µg of protein was loaded onto a 12% gel of 37.5:1 acrylamide:bisacrylamide ratio, using a Laemli buffer with or without a reducing agent. The formation of disulfide bonds was visualised by Coomassie staining.

5.6. X-ray crystallography

Protein X-ray crystallography makes use of the electromagnetic wave nature of X-rays to determine the three-dimensional structure of proteins at atomic resolution. In X-ray diffraction, a beam of incident X-rays hits the sample and is scattered in different directions by the electrons of the atoms forming the sample. The intensity of the diffracted waves is recorded by a detector. Since the diffracted waves are proportional to the Fourier transform of the electron density of the sample, the recorded data can be processed in order to reconstruct the image of the object, i.e. its electron density. As the signal generated by the scattering from a single protein molecule is too weak to be observed, a three-dimensional ensemble of protein molecules arranged in an ordered fashion – i.e. a protein crystal – is used so that the scattering signal is highly amplified and concentrated in specific directions (Blundell T.L. and Johnson Lewis, 1994). From such a crystal, the intensity of the diffracted waves can be

recorded and the electron density of the macromolecules included in the unit cell can be calculated. Since the late 50s, when the first protein structures were determined by X-ray crystallography by Max Perutz and John Kendrew, more than 110,000 structures of macromolecules have been solved so far by this technique.

We have employed X-ray protein crystallography to study the three-dimensional structure of EngA. Several crystal structures of EngA orthologs are available. Our approach had three main goals:

(i) identify different conformations for different nucleotide-bound states, namely the GTP-bound conformation;

(ii) determine the mode of binding of inhibitors to EngA; and

(iii) obtain the structure of the complex EngA:S7.

5.6.1. Protein crystallisation

Protein purification

Frozen aliquots of purified His-EngA_{WT} and S7-His were thawed and centrifuged for 30 minutes, 4ºC, 16000 g. These aliquots were injected in a Superdex® 75 10/300 GL column (GE Healthcare) equilibrated in 50 mM Tris pH 8.0, 250 mM NaCl (for His-EngA_{WT}) or 50 mM Tris pH 8.0, 500 mM NaCl (for S7-His) prior to crystallisation assays. His-EngA_{WT} was concentrated until a final concentration of ~4 mg/mL and S7-His to ~10 mg/mL. A last centrifugation step for 30 minutes, 4ºC, 16000 g was done to remove possible aggregates or dust.

Sample preparation: co-crystallisation

Assays of co-crystallisation of EngA with S7, nucleotides or inhibitors were performed in order to obtain crystals of the complexes.

For the crystallisation of the complex EngA:S7, a mixture of both proteins to a molar ratio of 1:1 was prepared and incubated at room temperature for 30 minutes prior to the preparation of drops.

For crystallisation of EngA in the presence of nucleotides, GMPPNP or GMPPCP were tested at concentrations of 1, 10 and 20 mM. $MgCl_2$ was added at the same concentration as

nucleotides. The effect of KCl at 700 mM was also tested. The mixtures were incubated at room temperature for 30 minutes prior to the preparation of drops.

Three inhibitors were tested in crystallisation trials of EngA (Fig. 34): CEB-II-54, CEB-II-55 and CEB-II-56. The complexes were prepared by adding compounds to EngA at a final concentration of 0.5 mM and 10–20% DMSO. To avoid DMSO, which may harm the protein and affect crystallogenesis, additional samples were prepared where 10 µL of compounds solution at 50 mM were pipetted into an eppendorf tube and the solvent was evaporated using a SpeedVac system. The powder remaining in the bottom of the tube was resuspended with 50 µL EngA solution to reach a maximum concentration of 10 mM inhibitors.

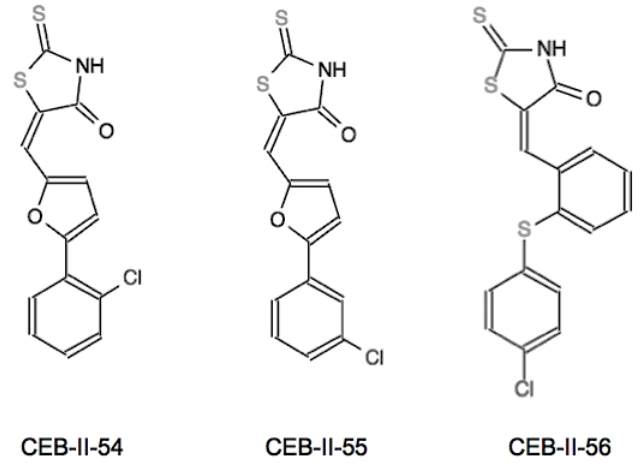

Fig. 34 Structure of three GTPase inhibitors CEB-II-54, -55 and -56. These ligands were used in crystallisation trials in order to determine their mode of binding on EngA, and in screening trials, to test their potential inhibitory effect on the interaction EngA:ribosome.

Sample preparation: soaking and dry co-crystallisation

In addition to co-crystallisation, two other strategies were used to obtain crystals of EngA-inhibitor complexes, using the three CEB-II-5x compounds. These were soaking and dry co-crystallisation.

For ligand soaking, EngA crystals were soaked in a solution of 0.2 mM CEB-II-5x compounds and 20% DMSO for 2 days.

In dry co-crystallisation (Gelin et al., 2015), a volume corresponding to 1–3 times the crystallisation drops volume of a 10 mM inhibitors solution was pipetted onto the crystallisation plates. These were placed at room temperature or at 37°C until all solvent had evaporated. Drops of EngA were afterwards prepared on top of the ligand-coated surfaces.

Screening of crystallisation conditions

To identify conditions that yield crystalline material, different crystallisation conditions were screened using the automated crystallisation robot from the High Throughput Crystallization Laboratory (HTX Lab) from the PSB.

The HTX platform prepares sitting drops using the vapour diffusion technique (Fig. 35). The drop set up consists on preparing nanoliter drops (100 nL of protein sample mixed with 100 nL of crystallisation condition) in 96-well plates.

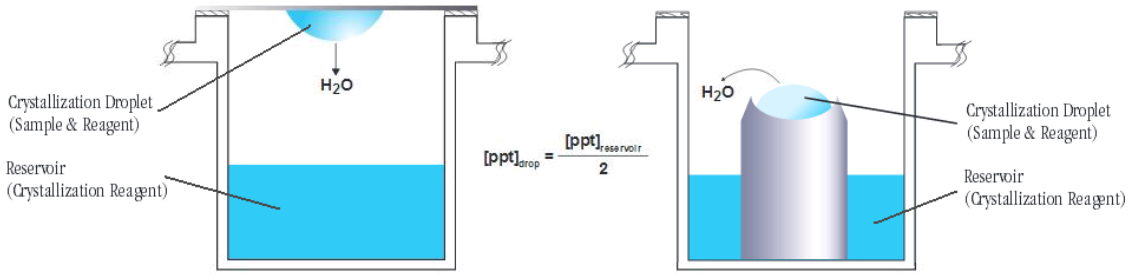

Fig. 35 Schematic representation of the vapour diffusion method by hanging drop (left) and sitting drop (right). A drop of protein and crystallisation solution is equilibrated by vapour diffusion with a reservoir containing a large volume of crystallisation solution. As the precipitant concentration ([ppt]) in the drop is lower than in the reservoir, water evaporates from the drop, decreasing its volume and increasing protein concentration until conditions of supersaturation allow nucleation and growth of crystals. Adapted from https://hamptonresearch.com/default.aspx.

Optimisation of crystallisation conditions

Promising crystallisation conditions identified from the HTX screens were optimised by preparing hand-made drops in 15-well plates. Vapour diffusion hanging drops (Fig. 35) were prepared by pipetting 1 µL protein sample and 1 µL crystallisation solution. Ranges of salt concentration, precipitant concentration and pH were tested spanning the screen hit conditions. Different temperatures (4ºC, 20ºC and 30ºC) were also tested, as well as the effect of additives (Additive Screen™ kit, Hampton Research).

Further refinement of crystallisation conditions was done at the HTX platform, using 96-well plates. These screens were prepared with the Formulator 16 robot (Formulatrix). Drops were prepared by the automated crystallisation robot as described above.

Additional trials at the HTX platform were done using commercial screens in 96-well plates that test the effect of salts and additives in a base solution. The drop set-up was done by vapour diffusion nanoliter sitting drops, as described above.

Optimisation of crystal growth: microseeding

Microseeding was employed in the optimisation of crystal growth (Fig. 36). EngA crystals were transferred into 50 µL of a stabilising mother liquor containing high precipitant concentration and smashed into crystalline particles by the use of a crystal crusher and seed beads (Hampton Research) and vortexing. A dilution series was prepared from the seeds stock. 1 µL of each dilution was directly pipetted into new 15-well plates with 1 µL protein solution to form hanging drops. To move the supersaturation state to lower levels and to avoid spontaneous nucleation, the precipitant concentration was lowered to 80–90% of the initial one.

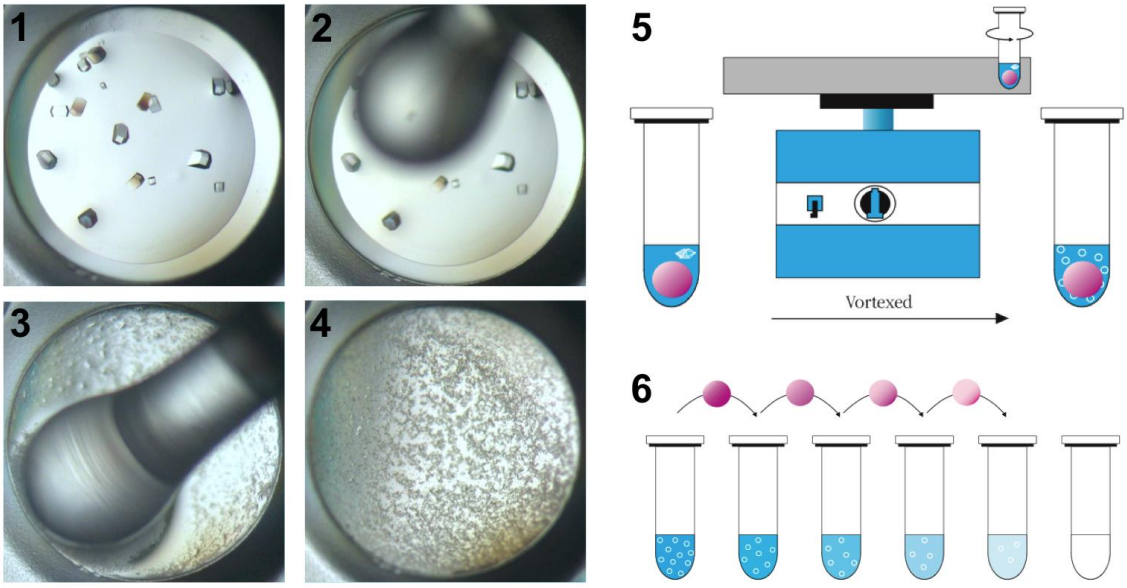

Fig. 36 Schematic representation of the microseeding technique. With the aid of a crystal crusher, crystals in a drop are smashed by an up and down motion (1–4). To make sure all crystals are reduced to small crystalline particles, the drop is transferred to a stabilising solution in a seed bead tube and vortexed (5). A single crystal can instead be directly fished from a drop and added to the seed bead tube. From this seed stock, serial dilutions were prepared (6) to optimise the number of seeds to add to the new drops. Adapted from https://hamptonresearch.com/default.aspx.

5.6.2. Data collection and processing

Freezing of protein crystals

Crystals to be analysed at the synchrotron were mounted in 35–200 µm loops, according to their size. With the exception of crystals where the crystallisation solution was already cryoprotecting, a cryoprotecting agent was added before freezing. Different cryoprotectants were used depending on the initial composition of the crystallisation solution and on the behaviour of the crystal. The cryoprotecting agents used are listed in Table 4.

Freezing of the crystals was done in liquid nitrogen or directly under the cryostream at the beamline.

Table 4 Cryoprotectants and respective concentration used when fishing EngA crystals.

Cryoprotecting agent	Concentration
PEG 3350-8000	25-30%
Glycerol	30%
Fomblin® Y Oil (Sigma)	100%
LV Cryo Oil™ (MiTeGen)	100%

Data collection

Protein crystals were analysed at the ESRF (Grenoble) beamlines ID23-1, ID23-2, ID29, BM30 and ID30B. An initial characterisation of the crystals was done by collecting four frames in order to know whether they diffract enough to collect a data set. Based on the observations from the initial characterisation (such as the characteristics of the crystal or the maximum resolution of the data), full data sets were collected by adjusting the crystal-to-detector distance, the exposure time, the transmission, the oscillation angle and the number of frames to collect.

Data processing

Data collected at the ESRF beamlines were processed with the XDS program (Kabsch, 2010). Phasing was done by molecular replacement using the program Phaser from the CCP4 suite (McCoy et al., 2007; Winn et al., 2011). The different structures of EngA and orthologs available in the PDB (4DCU, 1MKY, 3J8G) were used as models, either by using the entire structure or by searching individual domains. Cycles of refinement were done using the programs Refmac5 from the CCP4 suite (Murshudov et al., 1997; Winn et al., 2011) and Coot (for manual modifications of the model) (Emsley et al., 2010) until full interpretation of the electron density and validation of the model.

The quality of the model was analysed by computing the R-factor as:

$$Rf = \frac{\Sigma \, ||Fobs| - |Fcalc||}{\Sigma \, |Fobs|}$$

which calculates the agreement between the model and the experimental data, using 90–95% of the data (on which refinement is performed). This value should decrease during refinement. The R-free is computed similarly, using the 5–10% of the data not used in refinement for cross-validation. If the refinement is appropriate, the R-free should drop along with the R-factor of the working set.

The standard deviation from ideal values for bond lengths and bond angles are also checked in order to maintain a good stereochemistry. The program Procheck (Laskowski et al., 1993) was lastly used for final validation of the stereochemical quality of the protein structure, notably the Ramachandran Plot.

5.7. Cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) allows the three-dimensional reconstruction of macromolecules of molecular weight greater than ~100 kDa. In cryo-EM, imaging is done to vitrified samples, avoiding the need of sample staining and dehydration. It also reduces radiation damage due to the electrons (Carroni and Saibil, 2016).

We have employed cryo-EM to assess the structure of the complex 50S:EngA. Data acquisition and data processing have been done by Emmanuelle Neumann using the Electron Microscopy Platform of ISBG located at the IBS.

Negative staining preliminary analyses

The sample was firstly tested by negative stain technique, which gives high contrast images and allows for a preliminary visualisation of the sample (namely, for checking the concentration and homogeneity). Samples are stained with a heavy metal solution that forms a footprint of the structure of interest (Carroni and Saibil, 2016).

50S subunits were freshly prepared according to the protocol described in *Material and Methods, section 4*. 4 µL of sample were applied on the clean side of carbon on mica (carbon/mica interface), negatively stained with uranyl acetate 2% and air-dried. Images were collected using a Tecnai12 microscope operating at 120 kV (FEI Eindhoven, Netherlands) equipped with an Orius 1000 camera (Gatan).

Preparation of the complex

Frozen His-EngA_{WT} was thawed and centrifuged for 30 minutes at 16000 g, 4°C. GMPPNP and $MgCl_2$ were added to a final concentration of 10 mM and incubated for 15 minutes at 37°C.

Frozen 50S subunit was thawed at room temperature and centrifuged for 30 minutes at 16000 g, 4°C. Buffer E (see *Material and Methods, section 4*) was used to dilute it to a final concentration of ~200 nM. EngA previously incubated with nucleotide was added in excess to 50S to a molar ratio of 1:30. After a second incubation for 30 minutes at room temperature,

the complex was centrifuged (30 minutes, 16000 g, 4°C) to remove possible aggregates and taken to the EM platform.

Cryo-EM microscopy

Unlike negative staining, cryo-EM does not require any sample staining or fixation. Samples are applied on a holey carbon grid, blotted and rapidly frozen in liquid ethane in order to trap the molecules in vitrified ice (Carroni and Saibil, 2016).

3.5 µL of sample (50S or 50S:EngA complex) were loaded onto a quantifoil R 1.2/1.3 holey carbon film (Micro Tools GmbH, Germany). The grid was blotted to remove excess liquid and then vitrified using a Mark IV Vitrobot (FEI Eindhoven, Netherlands). Several grids were imaged using a Tecnai Polara microscope (FEI Eindhoven, Netherlands) operating at 300 kV and equipped with a K2 Summit electron counting direct detection camera (GATAN).

6. Methods for protein-protein interactions characterisation

6.1. ELISA assay

An enzyme-linked immunosorbent assay (ELISA) is a plate-based assay technique that enables detection and quantification of protein-protein interactions. During an ELISA assay, several steps are performed. Coating of a solid support with a first protein occurs by passive adsorption and immobilisation. The free spaces of the solid phase are blocked with a non-reactive protein that prevents non-specific binding. The interaction partner is subsequently added and probed with a specific antibody. This antibody is conjugated to an enzyme able to convert a substrate to a molecule with colour or luminescence properties that can be measured by a detector. During the assay, washing between different steps allows removal of non-bound molecules.

A direct ELISA assay was performed to assess the interaction of EngA with the ribosome and with the individual S7 protein from the small subunit.

Interaction 70S:EngA

A frozen aliquot of 70S ribosome was thawed at room temperature and centrifuged for 15 minutes at 16000 g to remove aggregates. The sample was diluted in 20 mM Tris pH 7.5, 60 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT to a final concentration of 24 nM. The

wells of a 96 well-plate Maxisorb Nunc-Immuno F96 (Nunc) were coated with 50 µL 70S solution for 2h at room temperature under agitation.

After immobilisation, the wells were blocked with 300 µL of TBS buffer (50 mM Tris pH 7.6, 150 mM NaCl) complemented with 3% BSA and 10 mM $MgCl_2$ at 4°C, overnight.

A frozen aliquot of His-$EngA_{WT}$ was thawed and centrifuged for 30 minutes at 16000 g, 4°C. EngA was diluted to a final concentration of 1 µM in TBST buffer (TBS buffer, 0.05% Tween-20) complemented with 10 mM $MgCl_2$. Successive dilutions were prepared to obtain a range of varying concentrations. EngA was preincubated for 15 minutes at room temperature in the absence of nucleotides or in the presence of 1 mM or 10 mM of GDP and GMPPCP/GMPPNP. 50 µL of EngA were added to the wells and the plate was incubated for 2h30 at room temperature under agitation.

Binding of EngA was probed with a monoclonal anti-polyHistidine antibody peroxidase conjugate (Sigma-Aldrich) diluted to 1:2000 in TBST complemented with 10 mM $MgCl_2$. The plate was incubated for 2h at room temperature under agitation. The antibody recognises polyhistidine-tagged fusion proteins. Several controls were made to confirm that it does not unspecifically recognise 70S ribosome or BSA.

Detection was done by addition of 100 µL of 3,3',5,5'-tetramethylbenzidine (TMB) (Sigma-Aldrich). TMB is a substrate for horseradish peroxidase, which converts TMB to a blue-coloured product. The plate was incubated in the dark until development of the blue colour. The reaction was stopped by addition of 200 µL of H_2SO_4 0.5 M, which turns the end product to yellow, and the absorbance was read at 450 nm. The absorbance is proportional to the amount of bound EngA. Between each step, the wells were washed with 300 µL of TBST.

<u>Interaction 30S/50S:EngA</u>

Initial trials were performed to assess the interaction of EngA with the individual subunits 50S and 30S. A range of concentrations between 0 – 10 nM ribosomal subunits was used to coat the plate and binding of EngA was tested in the absence of nucleotides. All other parameters of the assay are identical to the one described above.

<u>Interaction S7:EngA</u>

A similar set up was also applied for detecting interactions with S7 protein. A frozen aliquot of tagless S7 was thawed on ice and centrifuged for 15 minutes at 16000 g to remove aggregates. The sample was diluted in TBS buffer to a final concentration of 25 µM. The

wells of a 96 well-plate Maxisorb Nunc-Immuno F96 (Nunc) were coated with 50 µL S7 solution for 2h30 at room temperature under agitation.

After blocking the plate with BSA (as described above), His-EngA_{WT} was centrifuged and diluted to a final concentration of 20 µM in TBST buffer, from which serial dilutions were prepared. To test the effect of nucleotides, GDP/$MgCl_2$ and GMPPNP/$MgCl_2$ were added at 1 mM. After 2h30 incubation at room temperature under agitation, probing and detection was performed as described previously.

6.2. Sucrose density gradient ultracentrifugation

In this technique, a sucrose gradient is used to fractionate a sample based on different sedimentation rates of particles.

Frozen His-EngA_{WT} and 70S ribosome were thawed and centrifuged for 30 minutes at 16000 g to remove aggregates. A reaction mix was prepared with 55 pmol 70S and 1100 pmol EngA (molar ratio of 70S to EngA 1:20) in buffer 20 mM Tris pH 7.5, 3 mM $MgCl_2$, 300 mM NH_4Cl, 0.15 mM EDTA, 2 mM DTT. The interaction was tested in the absence of nucleotides, with 1 mM GDP and with 1 mM GMPPNP. The same buffer was used to prepare 10–40% sucrose gradients as described in *Material and methods, section 4*. For each reaction mix, the gradients were complemented with 1 mM of the corresponding nucleotide. A volume of 500 µL reaction mix was loaded on 11 mL sucrose gradient and centrifuged for 4h at 200000 g, 4°C. Fractions of 300 µL were collected in 96-well plates and the OD_{260} and OD_{280} were measured in a plate reader.

TCA/DOC protein precipitation

The fractions were afterwards analysed by SDS-PAGE and western-blotting. To be able to load whole sample volumes on the acrylamide gel, each two fractions were gathered and precipitated by using trichloroacetic acid (TCA) and deoxycholate (DOC). To one volume of sample, 1/100 V of 2% DOC was added, vortexed and incubated for 30 minutes at 4°C. 1/10 V TCA was then added and the samples were incubated overnight at 4°C. The next day, samples were centrifuged for 15 minutes at 16000 g, 4°C. The supernatant was rejected and the pellets containing precipitated proteins were resuspended in 10 µL 1 M Tris pH 8.0.

SDS-PAGE / Western-Blot

Upon addition of reducing Laemli buffer, a minimum volume of NaOH was added to samples that turned yellow due to TCA acidification. Samples were denatured for 5 minutes at 95°C and loaded onto a 12% acrylamide gel.

After migration of SDS-PAGE, proteins were transferred for 1h at 45 mA/membrane to a PVDF membrane (Immobilon-P, Millipore) previously activated in methanol, using the system Amersham™ TE 77 Semi-Dry Transfer Unit (GE Healthcare). After transfer, the membrane was blocked in TBST buffer (50 mM Tris pH 7.6, 150 mM NaCl, 0.05% Tween-20) complemented with 5% milk for 1h under agitation. A monoclonal anti-polyHistidine antibody peroxidase conjugate (Sigma-Aldrich) diluted to 1:5000 in TBST was used to probe His-EngA, for 1h under agitation. Detection was done using the SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific) and the membrane was revealed with chemiluminescent films Amersham Hyperfilm™ ECL (GE Healthcare).

6.3. Surface plasmon resonance

Surface plasmon resonance (SPR) is a technique for detection of interactions between molecules. In SPR, a ligand is immobilised on a thin metal film while the analyte is loaded in solution. A source of polarised light is directed to the back-side of the metal film and the reflected beam is recorded in a detector. If an interaction occurs, the increase in mass bound to the film will change its refractive index (Fig. 37A). This change in refracting index induces a change in the angle at which the intensity of the reflected light is minimum, due to energy loss in the evanescent wave. This change in angle can be measured in real-time (GE Healthcare Life Sciences).

We employed SPR to study the interaction between $EngA_{WT}$ and S7 ribosomal protein. SPR experiments were performed with a Biacore™ T200 instrument (GE Healthcare) using a Series S sensor Chip CM5 (Fig. 37B) (GE Healthcare). For these chips, a dextran matrix is already attached to the gold film and the ligands are then immobilised on the dextran matrix.

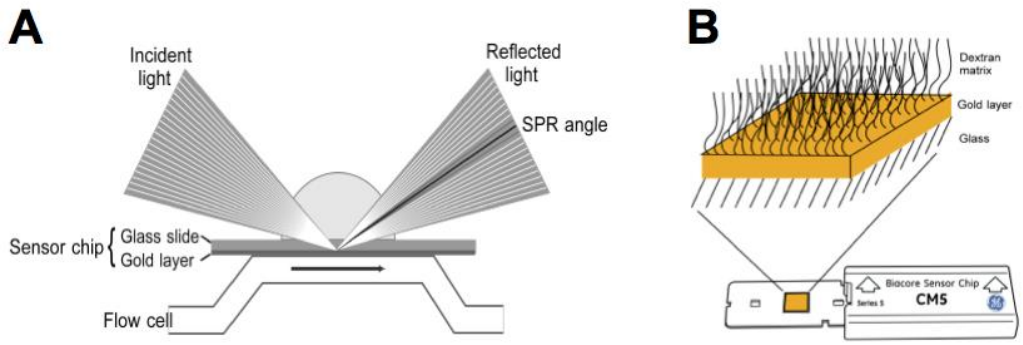

Fig. 37 (A) Principle of surface plasmon resonance. **(B)** Representation of a CM5 chip. SPR chips are coated with a dextran matrix covalently attached to a gold surface. Molecules can be covalently coupled to the sensor surface. Adapted from GE Healthcare Life Sciences.

Two immobilisation set-ups were tested, by immobilising either GST-S7 via an anti-GST antibody or S7-His by amine-coupling. His-EngA_{WT} was used as the analyte.

The system was previously equilibrated in HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4). Frozen GST-S7 and S7-His were thawed and centrifuged at 16000 g, 15 minutes, 4°C to remove aggregates. For immobilisation of GST-S7, anti-GST antibody solution (GST Capture Kit, GE Healthcare) was prepared to a final concentration of 30 µg/mL in Immobilisation Buffer (10 mM sodium acetate pH 5.0). The antibody was immobilised at 10 µL/mL in the CM5 chip using the amine-coupling chemistry, until a coupling level of 6000 RU. GST-S7 at 2 µg/mL in HBS buffer was captured on the anti-GST surface to reach an immobilisation level of 2000 RU. A flow cell with recombinant GST immobilised until 900 RU was used as negative control. S7-His at 2 µg/mL in Immobilisation Buffer was directly immobilised on the chip by amine-coupling until a coupling level of 1300 RU.

For EngA-binding analyses, the system was equilibrated in 20 mM Tris pH 8.0, 250 mM NaCl. Frozen His-EngA_{WT} was thawed and centrifuged at 16000 g, 30 minutes, 4°C. EngA was injected at 30 µL/mL in 20 mM Tris pH 8.0, 250 mM NaCl, 0.005% Tween-20. EngA concentrations of 1, 2.5 and 5 µM and the effect of 0.5 mM GDP, GMPPNP or sodium phosphate were tested. For single cycle kinetics analyses, His-EngA_{WT} was tested at 40, 20, 10, 5 and 2.5 µM in the absence or presence of 0.5 mM GDP or GMPPNP. Regeneration was done with 2 M NaCl.

The experimental data were analysed by fitting different binding models using the Biacore T200 Evaluation 1.0 software (GE Healthcare).

6.4. Bio-layer interferometry

Bio-layer interferometry (BLI) is another optical technique for detection of molecular interactions. In BLI, a ligand is immobilised in a sensor tip while the analyte is loaded in solution. Detection of interactions is done by measurement of the interference pattern of white light reflected from the two surfaces. If an interaction occurs, the interference pattern is shifted upon binding of the analyte molecules to the ligand on the sensor tip. This shift in wavelength can be measured in real-time (http://www.fortebio.com/bli-technology.html).

BLI experiments were performed to test the interaction of EngA with the 50S subunit and with the S7 protein.

Interaction 50S:EngA

The interaction 50S:EngA was tested using an Octet Red 96 (ForteBio) using 96-well plates. Two strategies were considered for this assay. In a first trial, His-$EngA_{WT}$ was diluted in ForteBio Kinetics Buffer (FKB) to a final concentration of 0.4 µM and immobilised via its His-tag on a Ni-NTA or a His1K (anti-penta-histidine antibody) biosensors soaked in the same buffer. Purified 50S subunits were added at 0.86 µM, diluted in storage buffer. The second strategy consisted on immobilising the 50S subunit and using EngA as the ligand. 50S was dialysed into PBS buffer complemented with 10 mM $MgCl_2$, biotinylated using the EZ-Link® NHS-PEO4-Biotinylation Kit (Pierce) and immobilised on in a high-precision streptavidin (SAX) biosensor at a final concentration of 28 nM. His-$EngA_{WT}$ was added in FKB buffer in a range of concentrations between 1–70 µM.

Interaction S7:EngA

The interaction S7:EngA was tested using a BLItz system (ForteBio). His-$EngA_{WT}$ was diluted in ForteBio Kinetics Buffer (FKB) to a final concentration of 0.4 µM and immobilised via its His-tag on a Ni-NTA biosensor soaked in the same buffer. Tagless S7 was passed as the ligand at a concentration of 0.7 µM in FKB.

For all assays, non-specific binding of the analyte to the biosensors was assessed by injecting the analyte onto non-coated biosensors. The signal generated was then used as the reference curve. The experimental settings for each run are described in Table 5:

Table 5 Experimental settings used in BLI experiments using the Octect Red96 and the BLItz apparatus. FKB stands for ForteBio Kinetics Buffer. Storage buffer corresponds to Buffer E from the ribosome purification protocol (20 mM Tris pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT).

	Octet Red96			**BLItz**	
	Duration (sec)	Sample (Strategy 1)	Sample (Strategy 2)	Duration (sec)	Sample
Initial baseline	60	FKB	PBS-$MgCl_2$	30	FKB
Loading	300	His-$EngA_{WT}$	50S-biotin	120	S7
Baseline	300	Storage buffer	FKB	30	FKB
Association	300	50S	His-$EngA_{WT}$	120	His-$EngA_{WT}$
Dissociation	300	Storage buffer	FKB	120	FKB

7. Methods for inhibitor screening

7.1. ELISA assay

The ELISA assay employed for detection of the interaction between EngA and the bacterial ribosome 70S was optimised in order to develop a method for screening inhibitors. The effect of three GTPase inhibitors (Fig. 34) in the binding of EngA to the ribosome were tested.

The ELISA assay was performed as described in *Material and Methods, section 6.1*. Before adding His-$EngA_{WT}$ to the ribosome-coated wells, compounds were added to EngA solution at a final concentration of 100 µM, 2% DMSO. The same amount of DMSO was added to apo-EngA.

7.2. Thermal shift assay

Thermal shift assay (TSA) allows screening of proteins stability in the presence of different ligands. When proteins become denatured, their folding is lost and the hydrophobic core residues become exposed. TSA monitors proteins thermal unfolding by using a fluorescent dye that is quenched in aqueous solution and highly fluorescent in non-polar environments. The exposure of hydrophobic residues upon thermal denaturation leads to an increase in fluorescence intensity that is measured in a real-time PCR thermocycler (Fig. 38) (Niesen et al., 2007).

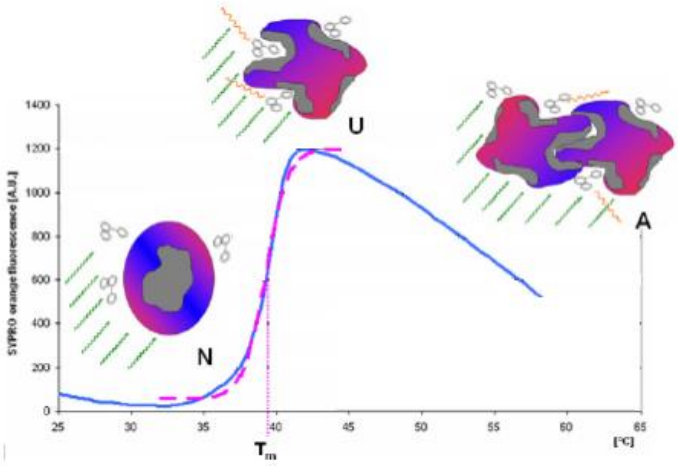

Fig. 38 Representation of a typical TSA response, where fluorescence intensity is plotted against temperature. When the protein is properly folded, the fluorescent dye is quenched. Increasing the temperature leads to protein denaturation, and the dye can bind exposed hydrophobic residues, emitting fluorescence that is recorded in real-time. A final decrease in intensity may be observed as proteins aggregate. Adapted from Niesen et al., 2007.

The stability of EngA in the presence of the three CEB-II-5x compounds used in the ELISA assay was measured by TSA, using the platform RoBioMol at IBS.

A frozen aliquot of His-EngA_{WT} was thawed and centrifuged for 30 minutes, 16000 g, 4°C to remove aggregates. EngA was diluted in storage buffer to a final concentration of 0.5 mg/mL. Ligands were added to a final concentration of 100 µM and 2% DMSO. A 96-well plate was set up for fluorescence measurements using the SYPRO Orange dye. Samples and buffers were measured in triplicates using the SYBR channel.

RESULTS

1. Protein expression and purification

1.1. Expression and purification of recombinant EngA

engA had been previously obtained from genomic DNA of *B. subtilis* and inserted into a pET15b expression vector (Foucher, 2010).

$EngA_{WT}$ pET15b construct was expressed in CD41(DE3) *E. coli* strains (Fig. 39). After lysis, the cellular supernatant from 1 L culture was purified firstly by affinity chromatography using an IMAC column, which efficiently retains his-tagged EngA and contributes to get rid of most contaminants.

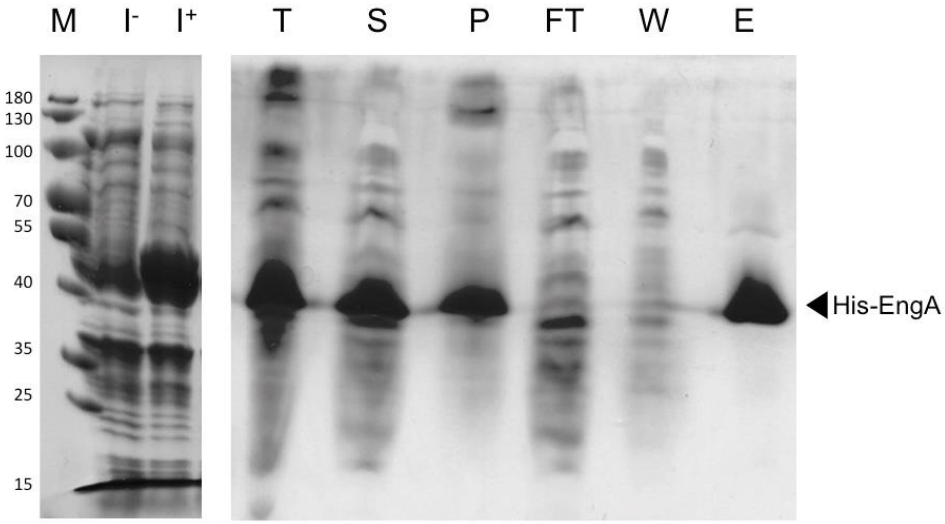

Fig. 39 SDS-PAGE analysis of the expression and IMAC of His-$EngA_{WT}$. (M) molecular weight marker; (I^-) sample before induction with IPTG; (I^+) sample after induction with IPTG; (T) total lysate; (S) supernantant; (P) cell pellet; (FT) flow-through; (W) washing step; (E) elution.

An SDS-PAGE analysis after purification with the IMAC column indicates that the purified sample migrates at the expected molecular weight of ~50 kDa. To remove remaining contaminant proteins after affinity chromatography, the sample was subjected to a size exclusion chromatography (Fig. 40). The peak corresponding to EngA was analysed by SDS-PAGE to check the purity of the sample (inset of Fig. 40).

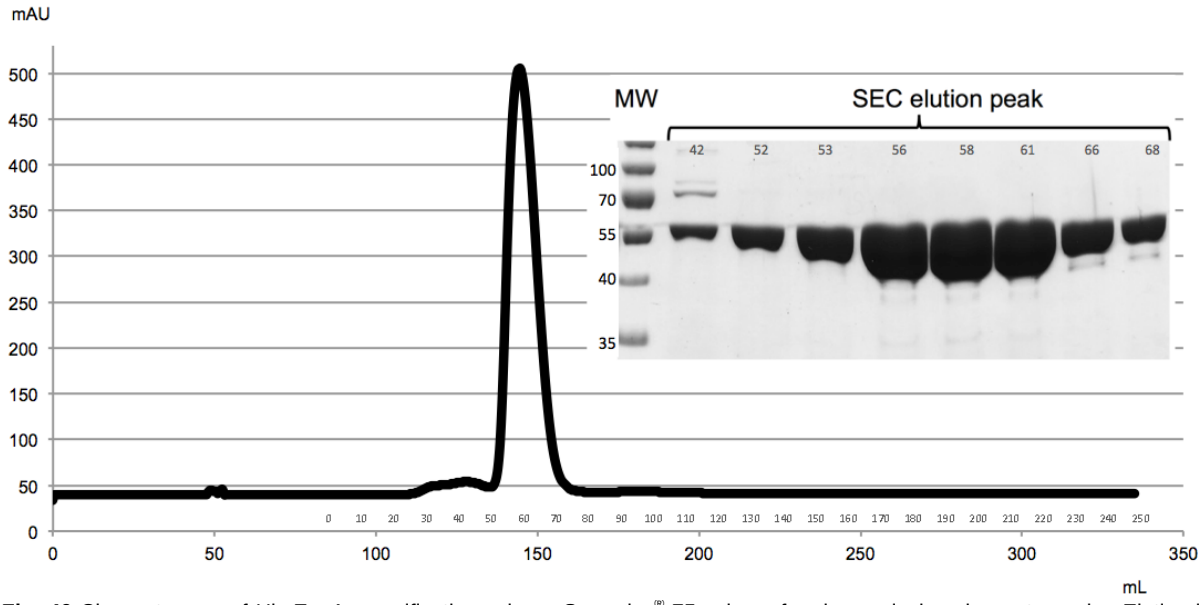

Fig. 40 Chromatogram of His-EngA_{WT} purification using a Superdex® 75 column for size exclusion chromatography. Elution is followed by measurement of UV absorbance at 280 nm collected over the elution volume. Small numbers on the *x*-axis represent fractions collected during elution. The inset represents the SDS-PAGE analysis of the elution peak with the labelled fractions analysed. (MW) molecular weight marker.

After these two steps of purification and concentration, EngA is free of significant contaminants. For 1 L culture, a mass of approximately 130 mg of protein is obtained for the His-EngA_{WT} constructs.

1.2. GTPase activity assay

After purification the GTPase activity of EngA was measured to confirm that the recombinant purified protein was functional. The assay was done using 1 mM GTP/$MgCl_2$ and a final protein concentration of 1 mg/mL. GTP hydrolysis was followed for 10 minutes by measuring the variation of OD_{340} (see *Material and Methods, section 2.3*) either in the absence or in the presence of 300 mM KCl.

The wild-type construct is active (6.9 nmol.min^{-1}.mg^{-1}) and its activity is enhanced by the presence of KCl (64.9 nmol.min^{-1}.mg^{-1}) (Fig. 41).

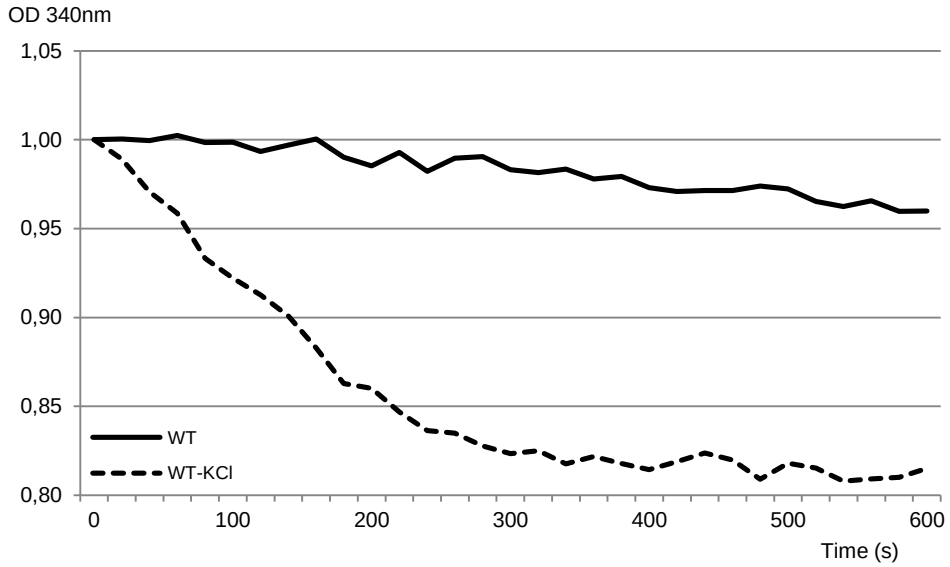

Fig. 41 GTPase activity assay for $EngA_{WT}$. The curves were corrected by a control where no GTP was used and normalised. The enzymatic activity was tested in the absence of KCl (continuous line) and in the presence of 300 mM KCl (dashed line). The specific enzymatic activities were computed from the linear region of the curves.

1.3. Expression and purification of recombinant S7 ribosomal protein

S7-His and GST-S7 pET22b(+) constructs were expressed in BL21(DE3) *E. coli* strain. After cell lysis, the cellular supernatant from 1 L culture was purified by affinity chromatography (Fig. 42A).

The SDS-PAGE analysis after affinity chromatography indicates an approximate molecular weight of 20 kDa for S7-His and 45 kDa for GST-S7, as expected (Fig. 42A). As contaminant proteins are still present in solution, a second step of purification by size exclusion chromatography was added (Fig. 42B). After size exclusion chromatography, the sample is free of significant contaminants. For 1 L culture, approximately 20 mg of S7-His and 50 mg of GST-S7 were obtained after purification.

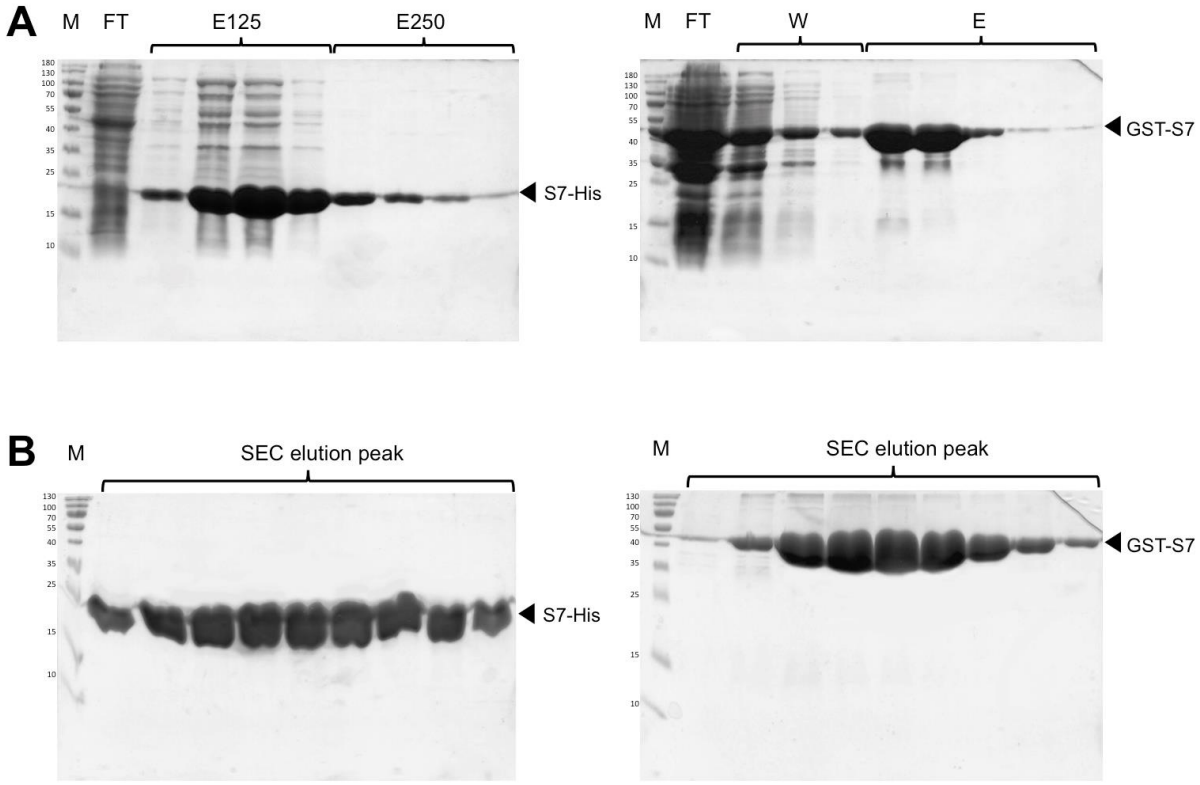

Fig. 42 (A) SDS-PAGE analysis of the affinity chromatography of S7-His (left) and GST-S7 (right). (M) molecular weight marker; (FT) flow-through; (W) washing step; (E125) elution with 125 mM imidazole; (E250) elution with 250 mM imidazole; (E) elution with 10 mM reduced glutathione. **(B)** SDS-PAGE analysis of the size exclusion chromatography of S7-His (left) and GST-S7 (right).

In the course of our experiments, a tagless S7 construct was required for interaction studies. S7 was obtained by cleavage of the GST tag from the GST-S7 construct. The presence of a thrombin recognition site allows cleavage of the fusion product.

For the purpose of obtaining tagless S7, only the affinity chromatography step was done after expression of GST-S7 for purification. After thrombin digestion, S7 is purified by ion exchange chromatography from the reaction mix containing S7, GST, thrombin and some remaining GST-S7. Since S7 is basic (pI = 9.95), a Resource S column was used. Ion exchange chromatography uses a gradient of increasing salt concentration to elute the sample from the column (in our set-up, 50 mM to 1 M NaCl). As GST-S7 had been eluted from the glutathione sepharose column in a buffer containing 500 mM NaCl, a desalting step using a PD10 column had to be added before ion exchange chromatography.

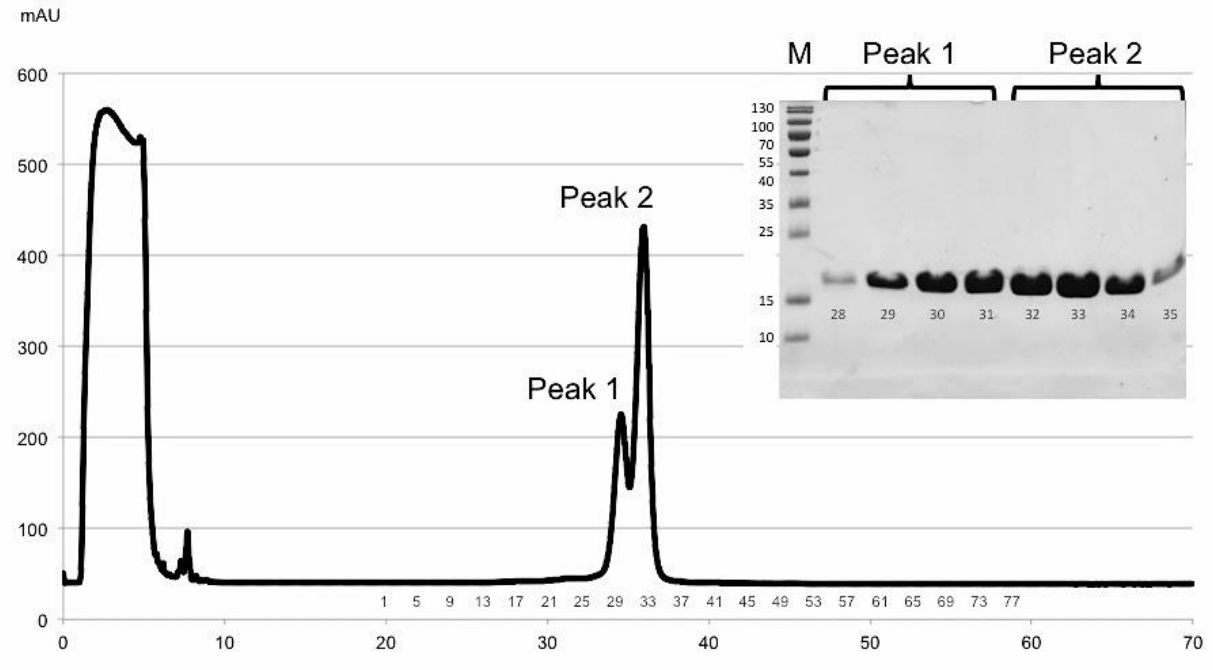

Fig. 43 Chromatogram of S7 purification using a Resource® S column for cation exchange chromatography. Elution is followed by measurement of UV absorbance at 280 nm collected over the elution volume. Small numbers on the *x*-axis represent fractions collected during elution. The inset represents the SDS-PAGE analysis of the purification with the labelled fractions analysed. (M) molecular weight marker.

After ion exchange chromatography, S7 is basically free of cloven GST (Fig. 43). During the elution, two peaks containing S7 appear at approximately 34.5 and 36 mL. SDS-PAGE analysis reveals a slight difference in size. As our aim was to use S7 for interaction studies, we considered that it would not be necessary to make further analyses to identify each species.

1.4. Purification of *B. subtilis* ribosomes

The 70S ribosomes were prepared from a culture of *B. subtilis*. Ribosomes were isolated from the cell lysate after a series of ultracentrifugations that allows pelleting the 70S particles and removal of other cellular components. From 2 L of cell culture, ~12 mg of pure ribosomes were obtained.

Lowering magnesium concentration promotes dissociation of the 50S and 30S subunits. A 5–20% sucrose gradient allows separation of both subunits based on their different size and thus different density (Fig. 44). For a mass of ~1 mg of the total pool of ribosomes, approximately 640 µg of 50S subunit and 33 µg of 30S subunit were recovered.

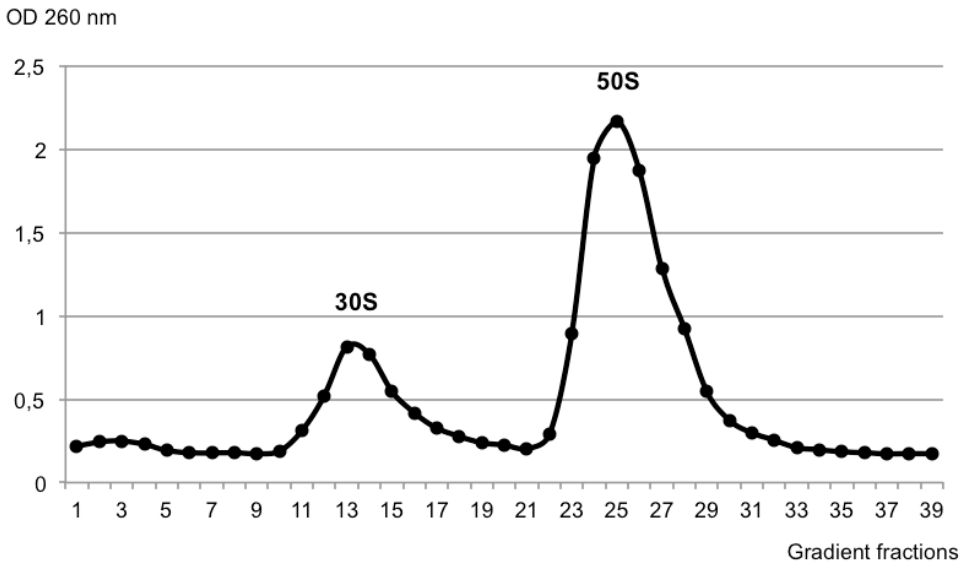

Fig. 44 5–20% sucrose gradient profile for dissociated 50S and 30S subunits. Both components are separated based on their density. Numbers in the *x*-axis represent gradient fractions from top (1) to bottom (39).

2. Analysis of conformational changes

2.1. Effect of nucleotides on EngA structure

During their catalytic cycle, GTPases undergo conformational changes which are concerted with the interaction with an effector molecule and dependent on the binding of a di- or tri-phosphate nucleotide.

Changes in protein structure upon binding or hydrolysis of GTP occur in the catalytic G-domain. Conserved flexible regions in the nucleotide-binding site can be stabilised or destabilised by contacts with the ligands and switch between different conformations.

The presence of two adjacent G-domains in EngA renders the switch mechanism unique and complex in comparison to other members of the GTPase subfamily. Crystal and cryo-EM structures have revealed so far three different conformations of EngA, where the G-domains not only undergo local conformational changes but also change the global protein structure following repositioning of one domain in relation to the other. These conformational changes have been proposed to be associated with different nucleotide-binding states and interaction with an effector molecule (i.e. the bacterial ribosome). However, since the crystal packing can always induce tertiary structural changes and since no GTP was ever observed in the GD1 domain on these structures, the exact role of the nucleotide remains to be determined.

We have thus employed structural and biochemical techniques in order to assess the nucleotides' effect on EngA structure and to obtain structural insights into the regions implicated in conformational changes.

2.1.1. Small-angle X-ray scattering (SAXS)

Conformational changes of $EngA_{WT}$ in solution were studied by small-angle X-ray scattering. Our goal was to test whether the binding of a nucleotide can trigger conformational changes on EngA and whether the nature of the bound nucleotide has an effect on these changes. To do so, the SAXS pattern of EngA was determined in the presence of different nucleotides.

When recording SAXS data, the monodispersity of the sample is an important factor so that the data is not biased by the presence of distinct species, aggregates or other contaminants (Jacques and Trewhella, 2010). For samples where the protein solution does not remain stably monodisperse during the time of sample preparation and data collection, recording SAXS data on a sample directly eluted from a column can be a solution. We have compared the behaviour of EngA in solution by performing a trial using a coupled size exclusion

chromatography system and the sample changer. EngA in the apo-form was used for this test.

The superposition of the scattering curves from apo-EngA analysed using the sample changer and the gel filtration are represented in Fig. 45, as well as the respective Guinier plots (inset of Fig. 45). Comparison of both samples shows no major changes in the scattering curves and confirms that there is no protein aggregation with either set-up. Accordingly, similar values for the radius of gyration, R_g, were computed from the Guinier region: 2.71 ± 0.07 nm for the sample changer and 2.79 ± 0.01 nm for the size exclusion chromatography analyses.

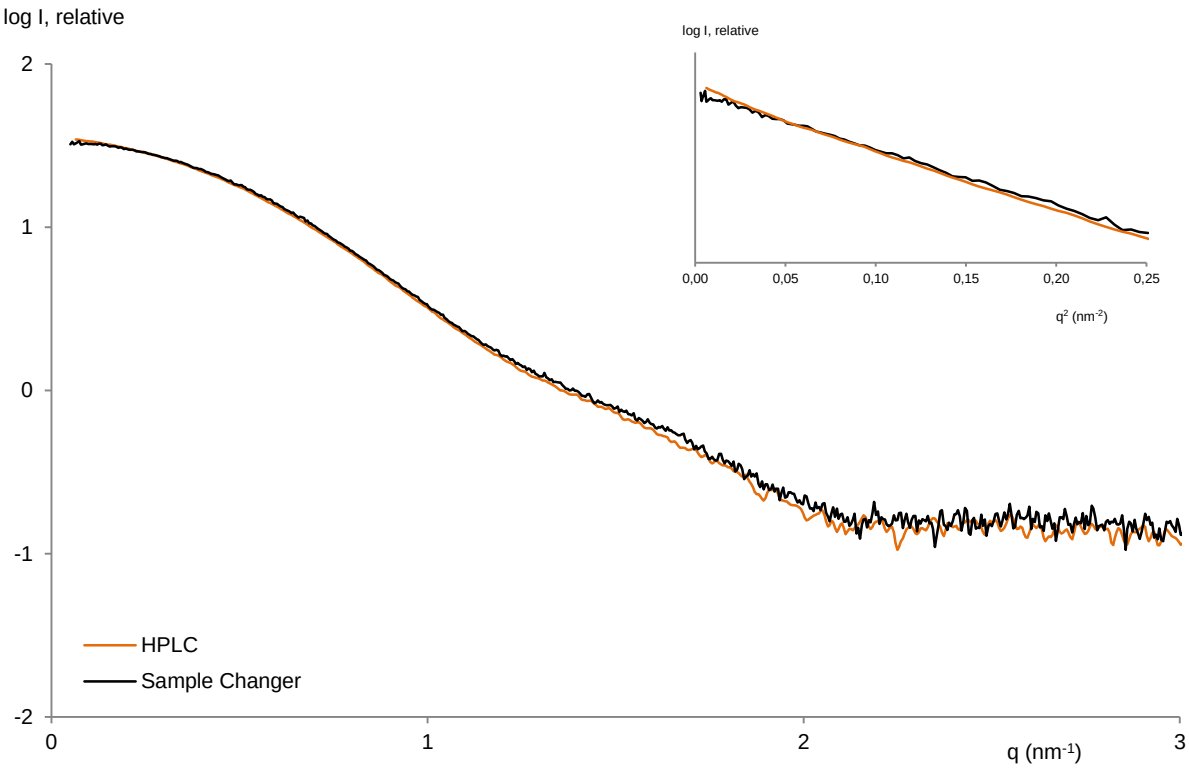

Fig. 45 Logarithmic plot of the scattering curve [log I(q) vs. q] for His-EngA_{WT} (apo) measured using the robot sample changer (black) or the eluted fraction from a size exclusion chromatography column (orange). The inset represents the Guinier plot [log I(q) vs. q^2].

As EngA seems to remain stable during the time of data collection from the sample changer, with no aggregates spoiling the data, this set-up was chosen for the subsequent experiments. All data presented in the next sections was hence collected using this system.

During SAXS data collection, radiation damage can occur due to formation of free radicals and bond breakage. To investigate the effects of radiation damage, ten frames were collected for each sample measurement. Each frame was inspected and those showing

radiation damage were excluded from the analysis. The usable frames from each exposure were combined into one dataset. This merged scattering curve was then used for the structural characterisation of EngA.

The scattering curve of apo-EngA$_{WT}$ is represented in Fig. 45 and the computed structural parameters are given in Table 6. Inspection of the Guinier plot (inset of Fig. 45), which remains linear, confirmed no protein aggregation.

Table 6 Overall parameters of apo-EngA$_{WT}$ obtained by SAXS. I_0/c is the forward scattering. R_g is the radius of gyration calculated from the Guinier approximation given in nm. V_p is the excluded volume and D_{max} the maximum particle dimensions calculated from the Porod plot.

I_0/c	R_g (nm)	V_p (nm^3)	D_{max} (nm)
37.6	2.71	85.1	9.3

From the SAXS profile, two parameters can be calculated that are related to the size and shape of the particles in solution: the forward scattered intensity (I_0) and the radius of gyration (R_g). The I_0 is the intensity of the radiation scattered at angle zero and is related to the number of particles and their volume. When normalised by the protein concentration (c) it gives a measure of the molecular weight. The R_g gives an indication of the shape of the particles, as its value will be different for particles with the same volume but different shapes.

For apo-EngA$_{WT}$, both the I_0 and R_g were computed with the Guinier approximation, as described in *Material and Methods, section 5.1*. An R_g value of 2.71 ± 0.07 nm and an I_0/c of 37.6 ± 0.05 were obtained. These values remain constant for the different concentrations, indicating that no events of concentration-dependent aggregation, repulsion or interaction were observed.

The molecular mass obtained does not completely corroborate the known MW of 50.8 kDa for His-EngA$_{WT}$. As previously stated, the forward scattering I_0 is computed based on the proportion MW = I_0/c, where c is the protein concentration. This concentration was determined based on the theoretical ε for EngA (33350 M^{-1}.cm^{-1}). However, at 280 nm, guanine nucleotides absorb significantly, with an ε of 7720 M^{-1}.cm^{-1}. Hence, the presence of a GDP molecule bound to EngA would increase its ε to a value of 41070 M^{-1}.cm^{-1}. Applying this correction factor of 0.812 ($\varepsilon_{EngA}/\varepsilon_{EngA:GDP}$) to the forward scattering I_0, a MW of 46.3kD was obtained, which approaches the expected value. These results suggest the presence of a GDP molecule bound to EngA, despite the purification steps after protein expression. A tight binding and slow rate of exchange of GDP at GD2 has been previously suggested in a crystallographic study (Robinson et al., 2002) where a GDP molecule was bound to GD2 despite its absence from both purification and crystallisation solutions.

Structures of EngA and orthologs obtained by X-ray crystallography and cryo-electron microscopy (PDB entries 4DCU, 1MKY, 3J8G) show three different conformations of the protein, suggested to be three different bound states (see *Introduction, section 5.4*). The program CRYSOL was used to compute the scattering curves from these three rigid body refinement models, using the PDB coordinates. The computed curves were fitted to the experimental data. In order to validate the fitting analysis, we computed the χ^2 values between models, which were used to assess the quality of the experimental data and of the fit. The results from the fitting analysis are presented in Table 7.

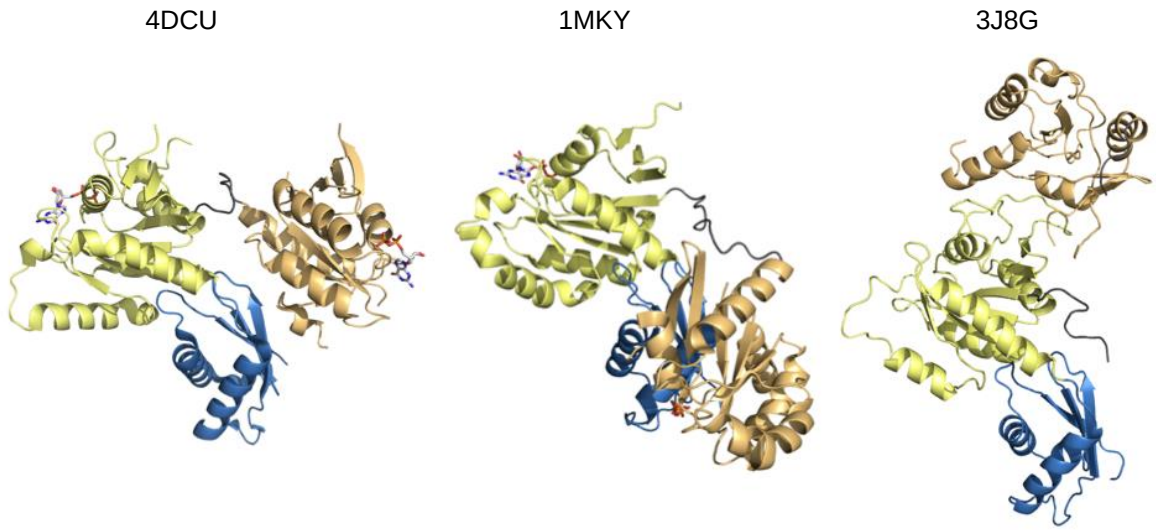

Fig. 46 3D structures of different conformations of EngA used for model validation with CRYSOL. 4DCU has a GDP molecule bound to both G-domains and corresponds to the GDP-bound (off-state) conformations. 1MKY is suggested to mimic the on-state due to the presence of two phosphates in the GD1 nucleotide-binding site close to the positions of a β- and γ-phosphates. 3J8G was obtained by cryo-EM where EngA is bound to the 50S ribosomal subunit. GD1 is represented in light orange, GD2 in yellow and KH-domain in blue. These structures were used as models for a fitting analysis of the SAXS experimental data. Images edited using PyMol from PDB entries 4DCU, 1MKY and 3J8G.

The best fit with the experimental data for apo-EngA was obtained with the 4DCU model, which shows a good agreement (χ^2 = 3.16), in opposition to the other two conformations (χ^2 = 32.14 for 1MKY; and χ^2 = 45.32 for 3J8G). The plotted fit of the scattering curves is represented in Fig. 47. These results confirm the presence in solution of the conformation obtained by crystallography.

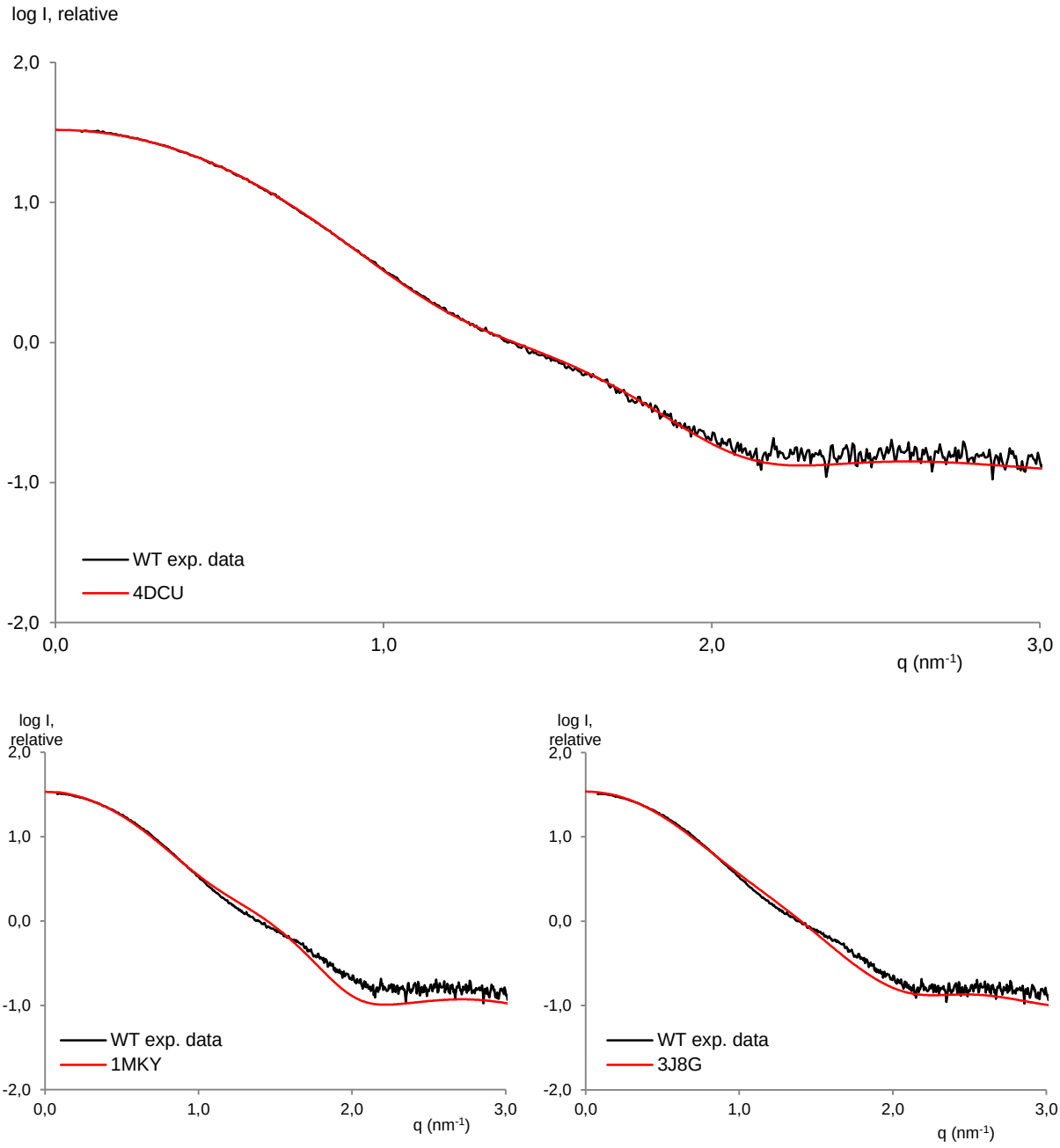

Fig. 47 (Top) Logarithmic plot of the scattering curve [log I(q) vs. q] for His-EngA_{WT} (apo). Experimental data is denoted by the black line. The red line corresponds to the scattering curve of the model 4DCU computed from the PDB coordinates. **(Bottom)** Same representation of the experimental data for His-EngA_{WT} (apo) fitted with PDB structures 1MKY (left) and 3J8G (right). The experimental data does not agree with these two models, as seen by the mismatch bertween curves.

Additional information on the conformation of proteins is given by the R_g, that provides a measure of the overall size of the particles. Given the same number of amino acid, the R_g will be smaller for proteins with a compact shape and larger for extended proteins (Kikhney and Svergun, 2015). The three structures of EngA orthologs show a different positioning of the GD1 domain compared with the GD2-KH domains, which contributes to a more compact shape in 4DCU and an elongated shape in 3J8G. This observation from the crystal/cryo-EM structures is in agreement with the R_{gT} obtained from the computed scattering curves of

these structures (4DCU 2.69 nm < 1MKY 2.81 nm < 3J8G 2.87 nm). The R_g estimated for apo-EngA (2.71 nm), close to the 4DCU value, further supports the presence of a 4DCU-like conformation in solution.

Despite the information that can be extracted from the scattering curve log*I*(*q*) vs. *q* about the shape of the particles in solution, interpretation of the data in terms of structure can not be directly obtained from it. To describe the sample in the real space we can use the pair-distance distribution function *P(r)*, which can be calculated from the scattering curve by a Fourier transform. The *P(r)* function provides the interatomic distance distribution of the molecule, describing the probable frequency of interatomic vector lengths (*r*) within a protein molecule (Jacques and Trewhella, 2010; Kikhney and Svergun, 2015).

Comparison of the *P(r)* plots from EngA obtained experimentally and from the different models can thus give additional information on how the shape and volume occupied by EngA relate to each of the conformations represented by the models. The pair-distance distribution represented in Fig. 48 for EngA in the absence of nucleotide shows that EngA clearly adopts a conformation similar to that of 4DCU, with a very good match of the two *P(r)* plots. In contrast, the two other structures, 1MKY and 3J8G, show distinct conformations, with longer interatomic distances.

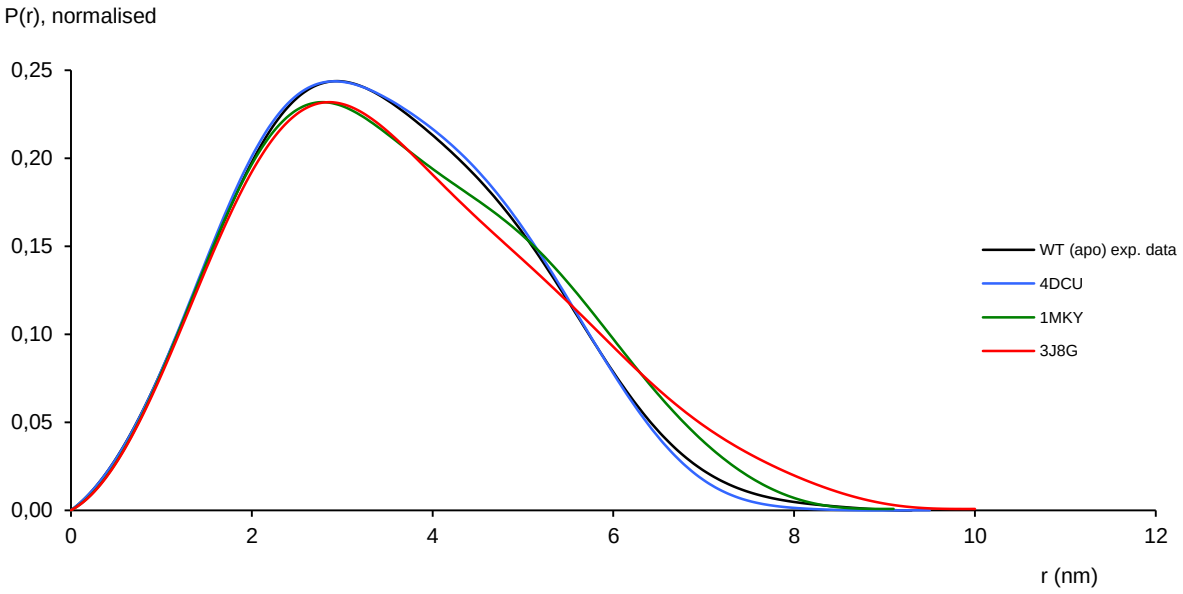

Fig. 48 Distance distribution functions of apo-EngA obtained experimentally (black line) and of the three models 4DCU (blue), 1MKY (green), 3J8G (red). EngA-apo adopts a conformation similar to the one represented by PDB 4DCU, as seen by the good match between the *P(r)* plots. The two other known EngA structures, 1MKY and 3J8G, show a distinct conformation from the experimental data, with longer interactomic distances.

GTPases go through a catalytic cycle where conformational changes occur upon nucleotide binding and hydrolysis. To assess the effect of nucleotides on EngA conformational changes,

the same analysis was done in the presence of GDP and two GTP analogues (GMPPNP and GMPPCP), tested at 1 and 10 mM.

Because the chemical composition of the buffer in samples can impact the scattering signal, the buffer alone must also be measured and its signal subtracted from the sample measurements. Hence, for this correction to be reliable, it is important that the composition of the measured buffer exactly matches the one within the protein sample. Making a gel filtration or a multi-stage dialysis are common methods to ensure we have a good reference. When preparing our samples for apo-EngA, we made a gel filtration to obtain a protein sample and a buffer sample with the same chemical composition. However, due to the unfeasible amounts of nucleotides that gel filtration or dialysis would require, a different set-up had to be developed for preparing EngA in the presence of nucleotides. Our initial approach consisted on preparing small dialysis volumes, using 100-200 μL sample against 3 mL dialysis buffer containing nucleotides. However, possibly due to a small exchange surface of this dialysis set-up, the differences in sample and buffer were too large to be able to process the SAXS data. Our second strategy consisted on making a dilution followed by a concentration step. One volume of apo-protein solution was diluted with two volumes of a buffer solution containing 3x concentrated nucleotide. This diluted sample was then concentrated in a 30 kDa cut-off centrifugal filter and the flow-through was used as a reference. When analysing the data obtained from this preparation, a higher concentration of nucleotide was observed in the sample solution than in the buffer solution. We associated this difference to the possibility of the nucleotide not passing completely freely through the filter. In order to extract reliable information from this data, a program to correct the difference in nucleotide concentration between sample and buffer had to be applied to the data (see *Material and methods, section 5.1*). A third attempt to minimise these differences consisted on preparing the protein sample and buffer individually while being careful with pipetting. Although preparing both solutions separately may introduce pipetting errors that could significantly impact the difference signal, the data showed a much smaller deviation compared to the one observed for the previous preparations. Thus, this methodology was chosen for the preparation of the samples and the correction program applied for minor corrections.

EngA could hence be prepared and tested in the presence of nucleotides. In a 1 mM GDP solution, EngA shows similar overall parameters to the apo protein, with an R_g of 2.71 nm and a D_{max} of 9.13 nm (Table 7). Accordingly, the experimental and 4DCU model scattering curves fit very well (χ^2 = 2.78; Table 7). These observations suggest that EngA adopts the same conformation in the apo and in the GDP-bound states.

The presence of GTP analogues was next tested to know whether they could trigger a change in conformation. In the presence of 1 mM GMPPNP, SAXS analysis did not show any significant conformational change in EngA, with the overall parameters remaining the same (Table 7). However, in the presence of 10 mM GMPPNP and GMPPCP, a new conformation was detected: R_g and D_{max} values increased to 2.93 nm and 12.0 nm, respectively, and the experimental scattering curve does no longer agree with the 4DCU model (Fig. 49, Table 7). These results suggest that high concentrations of GTP analogue induce a more elongated protein shape. When the same analysis is made by fitting the experimental curve and the two other PDB models, 1MKY and 3J8G, we can see that none of these models can fully explain the experimental data (Table 7), suggesting that these are neither the conformations adopted in solution by the GTP-bound *B. subtilis* EngA. A close observation of the results from the fit analysis seems to indicate that the GTP-bound conformation has structural properties that are closer to that of 1MKY, although the visual fit and the χ^2 values from Table 7 demonstrate properties intermediate to 4DCU and 1MKY (Fig. 49). The presence of protein molecules in both conformations in the sample is also a possibility that can be considered.

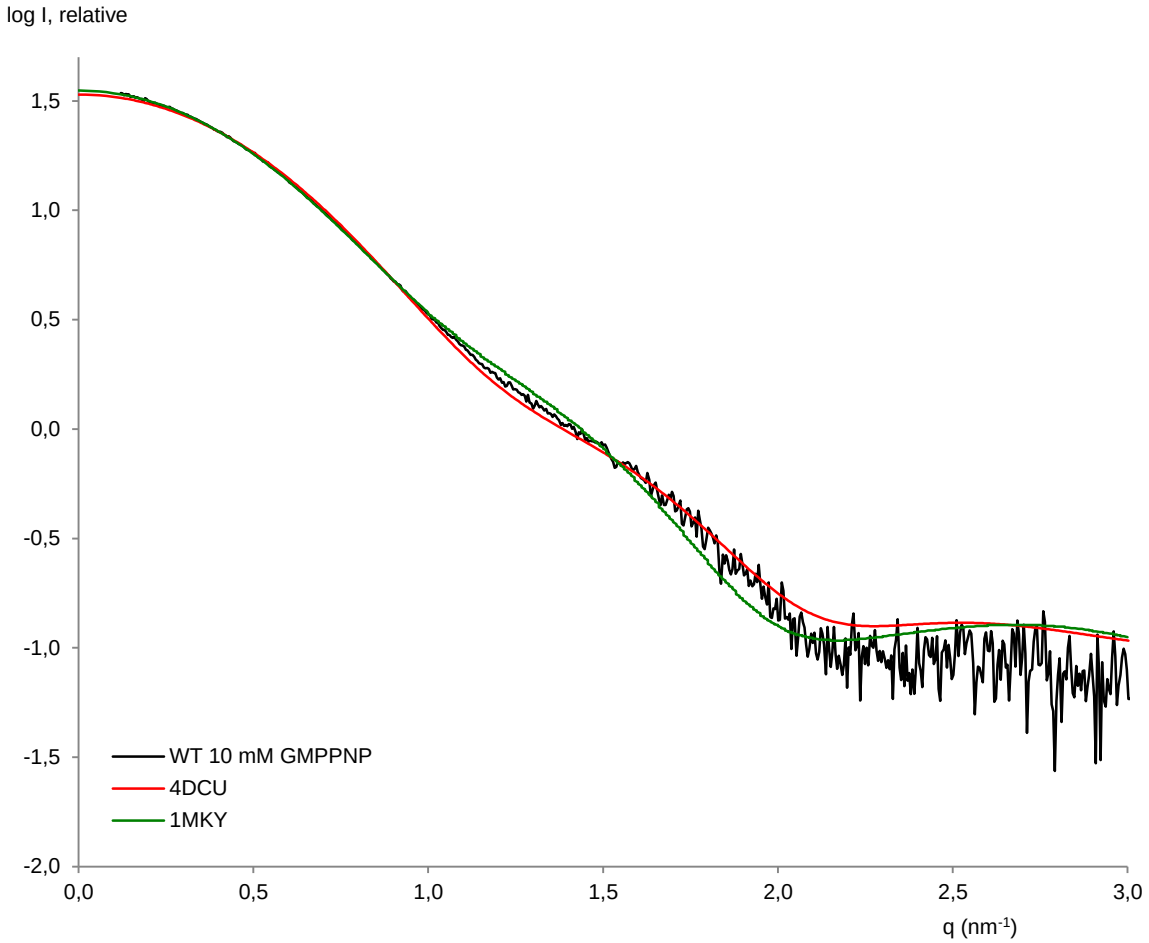

Fig. 49 Logarithmic plot of the scattering curve [log I(q) vs. q] for His-EngA_{WT} with 10 mM GMPPNP. Experimental data are denoted by the black line. The red line corresponds to the fit with the crystal structure PDB 4DCU and the green with 1MKY. The induced conformation does not match any of the structures, but its structural properties are an intermediate of the models.

Comparison of the *P(r)* functions from the experimental data of EngA in the absence and presence of 1 mM nucleotides shows the same profile, indicating a similar shape (Fig. 50). In contrast, the *P(r)* distribution for EngA in the presence of 10 mM GMPPNP and GMPPCP displays a different profile at larger distances. In the GTP-bound state, EngA exhibits a marked asymmetry with a progressive decrease to zero after the maximum peak, typical of an extended particle. This observation, supporting the high R_g and D_{max} values, further suggests the presence of a more extended conformation for EngA in the GTP-bound state.

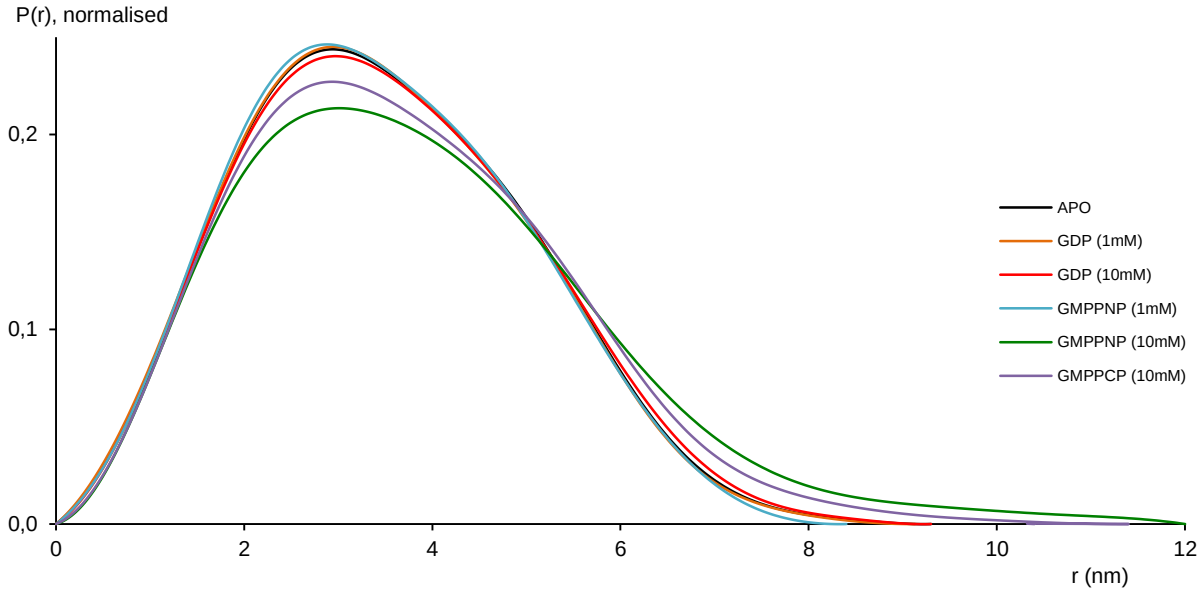

Fig. 50 Distance distribution functions of EngA in the absence of nucleotide (black) and in the presence of 1 mM GDP (orange), 10 mM GDP (red), 1 mM GMPPNP (cyan) and 10 mM GMPPNP (green) or GMPPCP (violet). Triphosphate nucleotides at 10 mM induce a conformational change on EngA towards a more extended conformation. In the presence of high concentrations of GTP analogues, the 4DCU conformation is no longer observed in solution.

From the previous analysis, we saw that fitting the experimental scattering curves of EngA (10 mM GMPPNP) and the theoretical models did not converge to any known conformation. Further analysis of the *P(r)* plot confirms this observation, as superposition of the experimental data with the three models shows no agreement (Fig. 51), suggesting a possible new unknown conformation for *B. subtilis* EngA GTP-bound form.

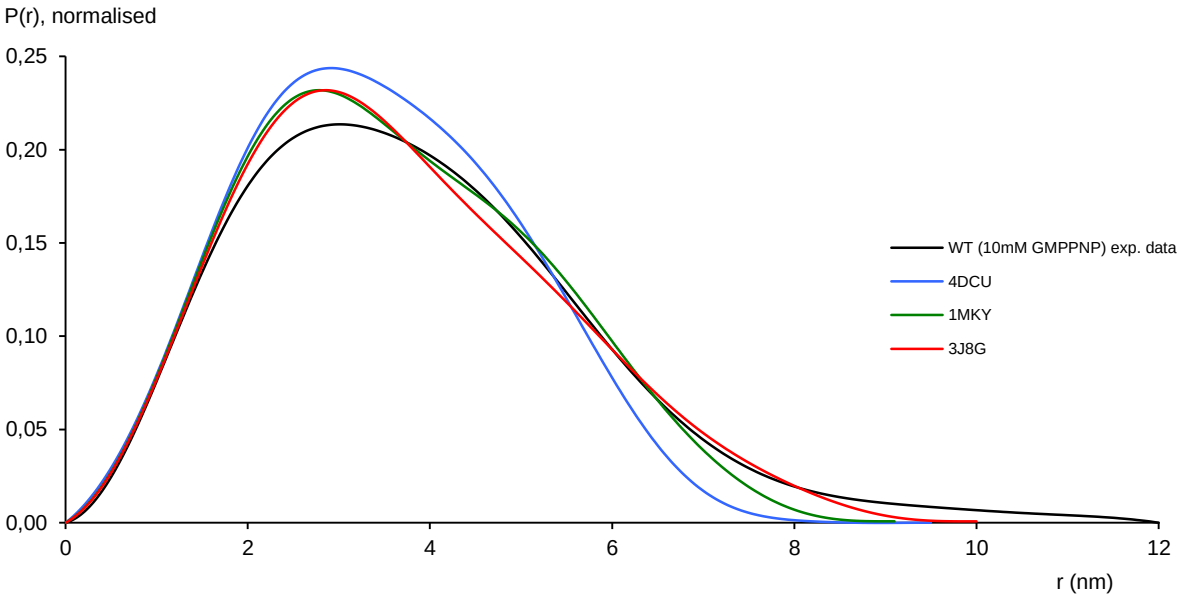

Fig. 51 Distance distribution functions of EngA (10 mM GMPPNP) obtained experimentally (black line) and of the three models 4DCU (blue), 1MKY (green), 3J8G (red). The *P(r)* plot of EngA bound to GTP analogues does not match any of the known PDB structures, suggesting the presence of a new unknown conformation in solution triggered by triphosphate nucleotides.

Analysis of the shape of the scattering curve on a Kratky plot can be used to assess the flexibility of proteins, namely conformational switching (Rambo and Tainer, 2011). Representation of a Kratky plot for fully folded, monodomain globular proteins will show a bell-shaped curve with a maxima peak at $qR_g \approx \sqrt{3}$, whereas unfolded proteins will present a rising curve with increasing angle until it reaches a plateau, as represented in Fig. 52:

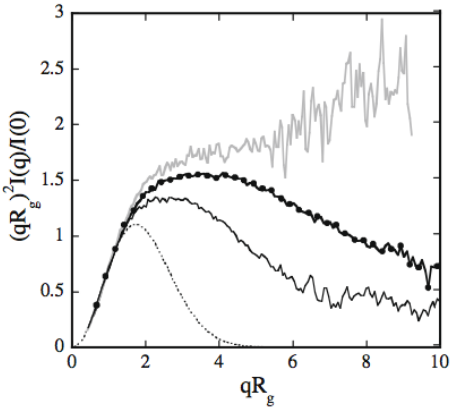

Fig. 52 Schematic representation of a Kratky plot for a globular protein (dotted line), partially folded proteins (thin and thick continuous lines) and unfolded protein (grey). Adapted from Durand et al., 2010.

EngA presents a profile that exhibits the typical behaviour of a folded compact protein, with a parabolic shape and a maximum value close to 1 at $qRg \approx 1.75$ (Fig. 53). A small tail at high qR_g suggests the presence of small unstructured regions, which possibly correspond to loops from the nucleotide-binding site and linkers connecting domains. The presence of nucleotide does not seem to have a big impact on the flexibility of EngA.

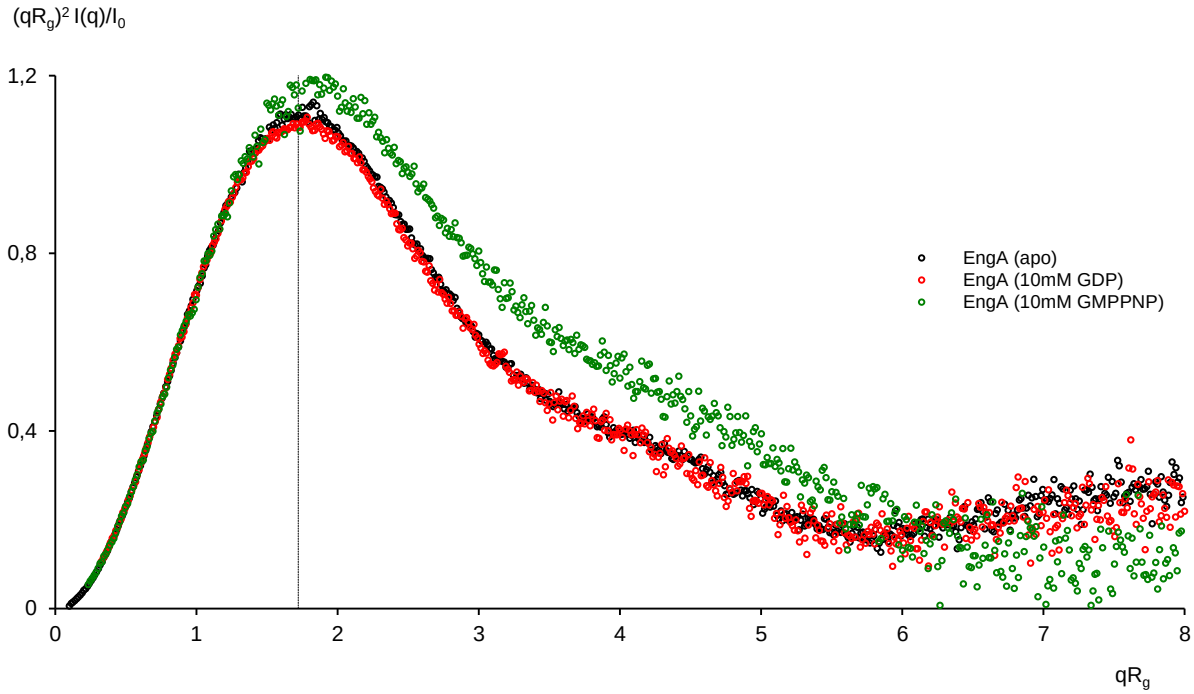

Fig. 53 Normalised Kratky plot [$(qR_g)^2 I(q)/I_0$ vs. qR_g] from the experimental data of EngA in the absence (black) and presence of 10 mM GDP (red) and GMPPNP (green). EngA presents a typical profile of a folded compact protein. The small tail at high qR_g suggests the presence of unstructured regions, possibly corresponding to loops from the nucleotide-binding site and linkers connecting domains. The presence of nucleotides does not seem to have a big impact on the flexibility of EngA.

Table 7 Overall parameters obtained by SAXS for EngA in the absence and presence of different concentrations of nucleotides (lines 1-4): *Conc* is the protein sample concentration (mg/mL), R_g is the radius of gyration (nm) obtained from the Guinier approximation, D_{max} is the maximum particle dimensions (nm) and *MW* is the molecular mass obtained from the excluded volume V_p (kDa). Fitting analysis with three models computed from PDB structures 4DCU, 1MKY and 3J8G (lines 5-7): χ^2 represents the chi-square values and R_{gE} and R_{gT} the experimental and theoretical radius of gyration (nm), respectively, obtained for the fits. The three last lines (8-10) indicate the χ^2 for the fit between models for statistical validation.

		APO	1 mM GDP	10 mM GDP	1 mM GMPPNP	10 mM GMPPNP	10 mM GMPPCP
***Conc.* (mg/mL)**		3.9	2.8	2.8	2.6	1.0	2.8
R_g (nm)		2.71	2.71	2.71	2.69	2.93	2.87
D_{max} (nm)		9.30	9.13	9.30	8.40	12.0	11.4
***MW* (kDa)**		53.2	52.4	52.4	52.1	53.4	52.3
Fit 4DCU	χ^2	3.16	2.78	2.22	2.65	3.32	5.84
	R_{gE}	2.71	2.69	2.74	2.69	2.84	2.79
	R_{gT}	2.68	2.67	2.69	2.67	2.69	2.69
Fit 1MKY	χ^2	32.14	29.65	8.30	24.31	2.01	3.74
	R_{gE}	2.71	2.69	2.74	2.69	2.88	2.79
	R_{gT}	2.80	2.79	2.81	2.79	2.85	2.85
Fit 3J8G	χ^2	45.32	43.18	12.12	32.30	4.02	6.25
	R_{gE}	2.71	2.69	2.74	2.69	2.88	2.78
	R_{gT}	2.88	2.86	2.88	2.87	2.92	2.93
χ^2 4DCU / 1MKY		38.33	33.28	12.05	26.51	4.28	11.98
χ^2 4DCU / 3J8G		56.41	50.21	18.02	37.05	4.93	14.87
χ^2 1MKY / 3J8G		9.60	8.37	2.99	6.03	1.58	3.59

2.1.2. Limited proteolysis by trypsin

Limited proteolysis was employed to further examine the nucleotides' effect on the conformation of $EngA_{WT}$ in solution. Limited proteolysis is based on the susceptibility of certain regions of a protein structure (due to flexibility and/or accessibility) to the action of proteases. Different conformations of one protein may exhibit differences in the regions that are exposed (due to global conformational changes on the tertiary structure) or flexible (due to local conformational changes on the secondary structure).

We aimed to use limited proteolysis to probe conformational changes of EngA in different nucleotide-bound states. Our observations from our SAXS data revealed a conformational change in solution when 10 mM GTP analogues are present. Thus, we speculated that the different conformations observed in solution would have different responses upon proteolytic digestion.

We initially hypothesised that binding of a GTP analogue would induce a conformational change (as observed by SAXS), which would generate a different fragmentation pattern upon proteolysis. The apo and the GDP-bound states, on the contrary, would present similar profiles, as the same conformation is adopted. Local conformational changes potentially due to the absence and presence of a bound nucleotide (independently of its nature) had also to be taken into consideration when analysing our results.

Trypsin is a serine protease that cleaves proteins on the C-terminal side of Arg and Lys residues. The proteolysis assay was conducted with a trypsin to EngA ratio of 1:600 at 20°C. Based in the observations from SAXS, we used a nucleotide concentration of 10 mM to be able to trigger conformational changes. Aliquots were collected at time intervals and the fragmentation pattern was visualised by tricine-SDS-PAGE analysis (Fig. 54).

In opposition to what we initially expected, the digestion pattern remains approximately the same for the three conditions. However, a change was observed in the kinetics of proteolysis. In the presence of 10 mM GMPPCP, EngA shows an increase in the digestion rate, indicating a change in conformation towards a state more susceptible to proteolytic cleavage. The apo and the GDP bound-state show a similar rate, suggesting a similar conformation.

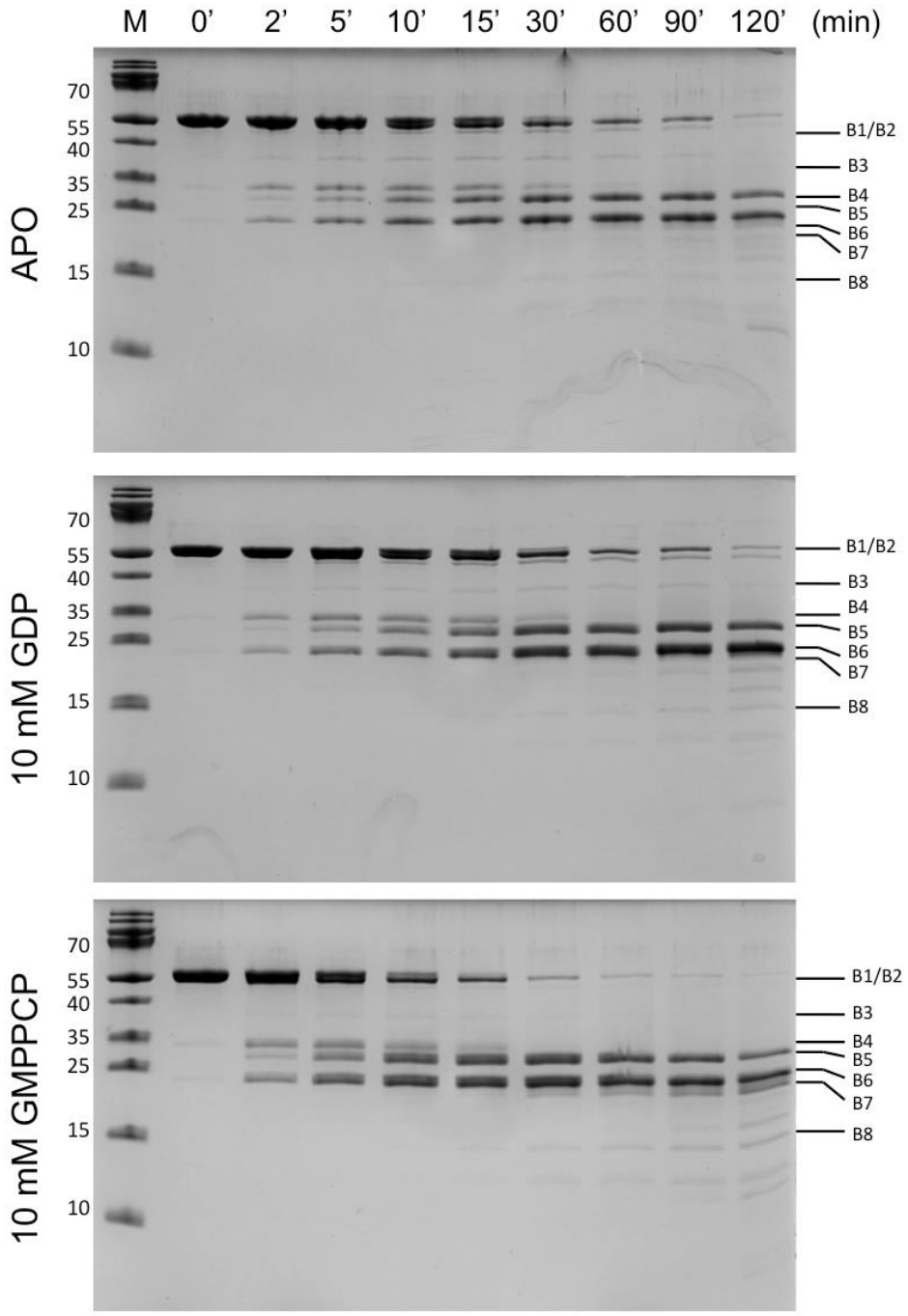

Fig. 54 SDS-PAGE analysis of the tryptic digestion of His-EngA_{WT} in different nucleotide bound-states. (M) molecular weight marker. Numbers on the top represent the time interval of the digestion in minutes. Each band generated upon fragmentation is labeled on the right (arrows and numbers).

To quantitatively analyse the proteolysis rate, the intensities of the bands corresponding to the full-length protein were measured and plotted against time (Fig. 55). The time at which the amount of full-length EngA in solution is reduced by half (t_{50}) was determined. In the apo-state and in the present of GDP, the t_{50} corresponds to 23.5 and 25.5 minutes, respectively. In the presence of a GTP analogue, however, no more than 50% full-length EngA are present after 10.5 minutes of digestion.

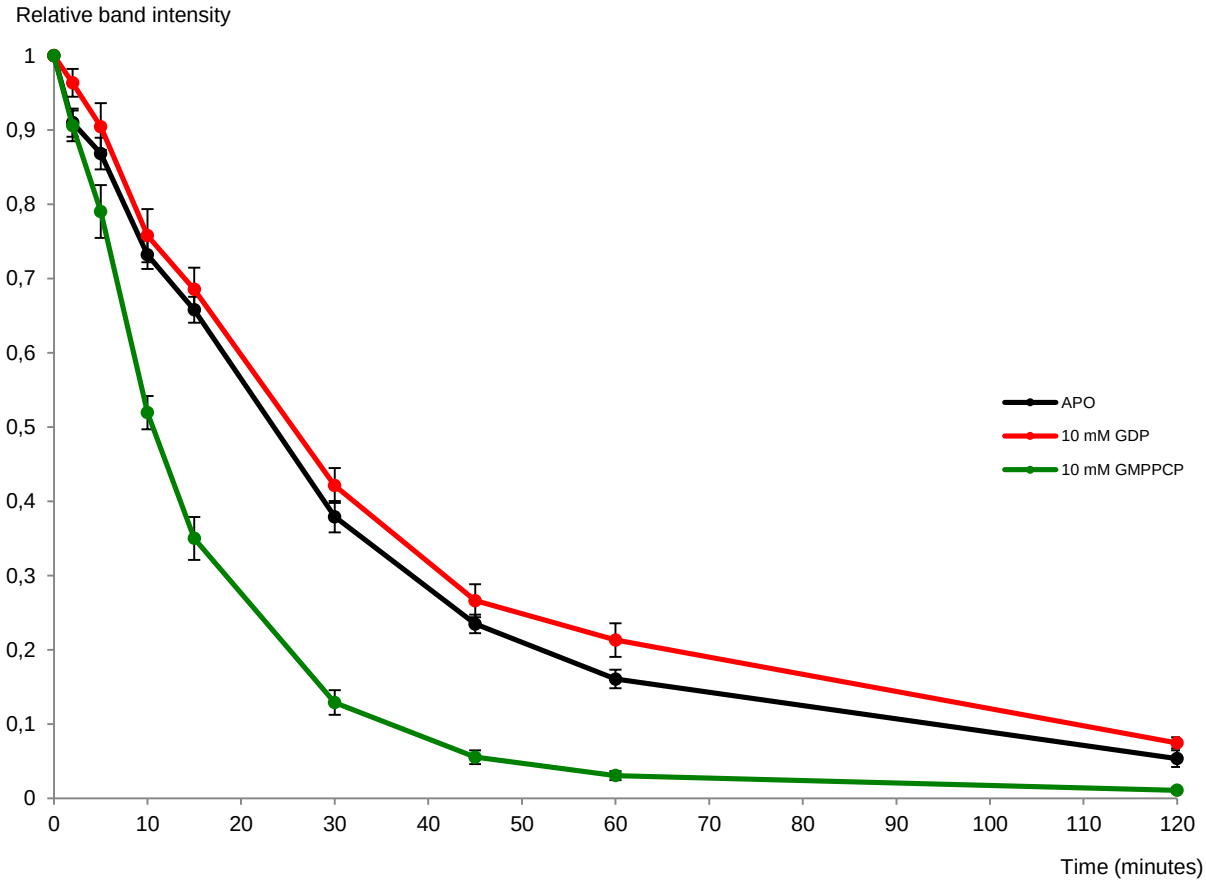

Fig. 55 Effect of nucleotides on the proteolytic rate of EngA by trypsin. The intensities of the bands on the SDS-PAGE were measured relative to time zero (100%), using ImageJ and ImageLab softwares. These values were plotted against time (minutes) and the t_{50} was determined. Each point corresponds to the mean of independent sets of results and error bars represent the standard error, where *n* varies between 3 and 5.

These observations suggest that EngA adopts the same conformation in the apo and GDP-bound states, undergoing conformational changes in the presence of a GTP analogue.

2.1.3. N-terminal sequencing and mass spectrometry analysis

In order to obtain some insights into the location of the regions affected by the GTP-induced conformational change, N-terminal sequencing and mass spectrometry (MS) were employed.

After proteolysis, aliquots of EngA were collected and subjected to N-terminal sequence analysis. The regions sensitive to proteolysis were identified as described in Table 8.

Table 8 Trypsin-sensitive regions detected by N-terminal sequencing analysis of a digestion mix of His-EngA_{WT} (apo) after 15 minutes of digestion. Please note that the additional amino acids introduced by the tag at the N-terminus of EngA are numbered from -19 to 0; the real sequence of EngA starts at +1.

Trypsin-sensitive regions	EngA regions
$^{-3}$R / GSH^{1}MGKPV	Cleavage of the His_6 tag.
^{127}R / ANIYDFYS	Cleavage at G4 motif of GD1.
^{236}K / GKVYETTE ^{238}K / VYETTEKY	Cleavage at switch II region of GD2.
^{345}K / ASENHSLRV	Cleavage at α6-helix of GD2.

To be able to assign each band observed on the SDS-PAGE analysis to a proteolysis-sensitive region identified by N-terminal sequencing, we used mass spectrometry that allows accurate determination of the fragments' molecular masses.

EngA in the absence of nucleotides and in the presence of 10 mM GDP or GMPPCP was analysed by mass spectrometry after tryptic digestion at 15 and 120 minutes. A comparative analysis was done between samples, in order to detect differences in peaks, either in their nature or in their intensities. The identified peptides at 15' and 2h of digestion in the absence and presence of 10 mM GMPPCP are listed in Table 9.

Table 9 Fragments originated after 15 and 120 minutes of tryptic digestion of apo and 10 mM GMPPCP His-EngA_{WT} detected by mass spectrometry. Please note that the additional amino acids introduced by the tag at the N-terminus of EngA are numbered from -19 to 0; the real sequence of EngA starts at +1.

Time (min)	EngA	Fragments	Molecular mass (kDa)	Cleavage sites	Gel band
15	Apo	$^{-18}$G – ^{436}K	50.805	------ (Full-length)	B1
		$^{-2}$G – ^{436}K	49.051	Tag	B2
		$^{-18}$G – ^{238}K	28.214	Tag / GD2 sw.II (N-ter)	B4
		$^{-2}$G – ^{236}K / $^{-2}$G – ^{238}K	26.278 / 26.463	GD2 sw.II (N-ter)	B5
		^{237}G – ^{436}K	22.791	GD2 sw.II (C-ter)	B6
	GMPPCP	$^{-2}$G – ^{436}K	49.051	Tag	B2
		$^{-2}$G – ^{236}K / $^{-2}$G – ^{238}K	26.278 / 26.463	GD2 sw.II (N-ter)	B5
		^{239}V – ^{436}K	22.606	GD2 sw.II (C-ter)	B6
120	Apo	$^{-2}$G – ^{188}K	21.046	GD2 P-loop (N-ter)	B7
		$^{-2}$G – ^{236}K / $^{-2}$G – ^{238}K	26.278 / 26.463	GD2 sw.II (N-ter)	B5
		^{237}G – ^{436}K / ^{239}V – ^{436}K	22.791 / 22.606	GD2 sw.II (C-ter)	B6
	GMPPCP	$^{-2}$G – ^{127}R	14.335	GD1 G4 (N-ter)	B8
		$^{-2}$G – ^{188}K	21.046	GD2 P-loop (N-ter)	B7
		$^{-2}$G – ^{236}K	26.278	GD2 sw.II (N-ter)	B5
		^{237}G – ^{436}K / ^{239}V – ^{436}K	22.792 / 22.606	GD2 sw.II (C-ter)	B6

Interpretation of the results was done by combining data from SDS-PAGE, N-terminal sequencing and mass spectrometry. Inspection of the SDS-PAGE allows identification of bands corresponding to the proteolytic fragments and comparison of the kinetics of fragment generation. Mass spectrometry is able to detect the generated fragments and measure their molecular mass, from which they can be assigned to the bands on the gel. Here, however, fragments present in low amounts can be hidden by high intensity peaks of other fragments eluted at a close retention time. Cleavage sites corresponding to fragments (including the omitted ones, as for its sensitivity) can be detected by N-terminal sequencing. All the obtained cleavage sites, molecular weights and SDS-PAGE bands can hence be combined to extract information on the structural features of different nucleotide-bound states of EngA.

For an easier understanding of the following analysis, Fig. 56 shows the bands observed in the gel that were assigned to a cleavage fragment:

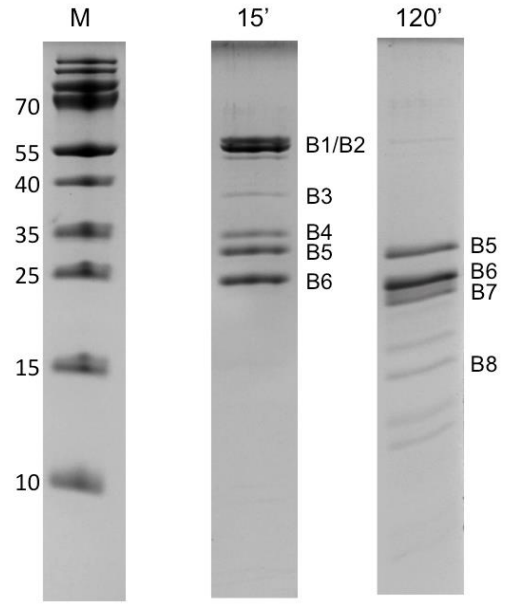

Fig. 56 SDS-PAGE from the tryptic digestion of His-EngA_{WT} at times 15 and 120 minutes. Each band was assigned to a fragment detected by N-terminal sequencing and mass spectrometry. Fragments are labelled from B1-8. M: molecular weight marker.

Mass spectrometry showed that the histidine tag of His-EngA_{WT} is actively cleft by trypsin, originating a fragment of 49.051 kDa. The full-length protein was assigned to band B1 and the tagless protein to band B2.

At time 15 minutes, no full-length protein remains in solution with 10 mM GMPPCP, with the His-tag being completely cleft. This is in line with the fast digestion observed in the limited proteolysis assay for EngA in the presence of GMPPCP. In the apo-state, as well as for the GDP-bound protein, the full-length protein and smaller peptides containing the tag are still present, which agrees with a slower rate of digestion.

As revealed by SAXS, the apo and GDP-bound proteins adopt the same conformation. The same behaviour observed for these two proteins, different from the GMPPCP-bound protein, suggests a change in the global structure of EngA in the GTP-bound state to a conformation where this region is more accessible to the action of the protease. Although the crystal structure of EngA in the GDP-bound state (PDB 4DCU) does not contain the tag, inspection of the N-terminal residues shows they are turned towards the acidic linker and the GD2 (Fig. 57), possibly hampering the access for trypsin. This suggests that in the new GTP-bound conformation the tag may be more exposed to the solvent. It is interesting to notice that the N-terminal region in the 1MKY structure – proposed to mimic the GTP-bound state –, seems to be inaccessible to the action of a protease, as it is hidden in a pocket formed by the GD1, the acidic linker and the KH-domain (Fig. 57). As we have previously seen from our SAXS results, these observations further suggest that 1MKY may not be a real mimic of the GTP-bound structure of EngA. It is difficult to draw conclusions from the observation of the 3J8G structure of EngA due to the low resolution and the absence of a resolved linker, located close to the N-terminal residues.

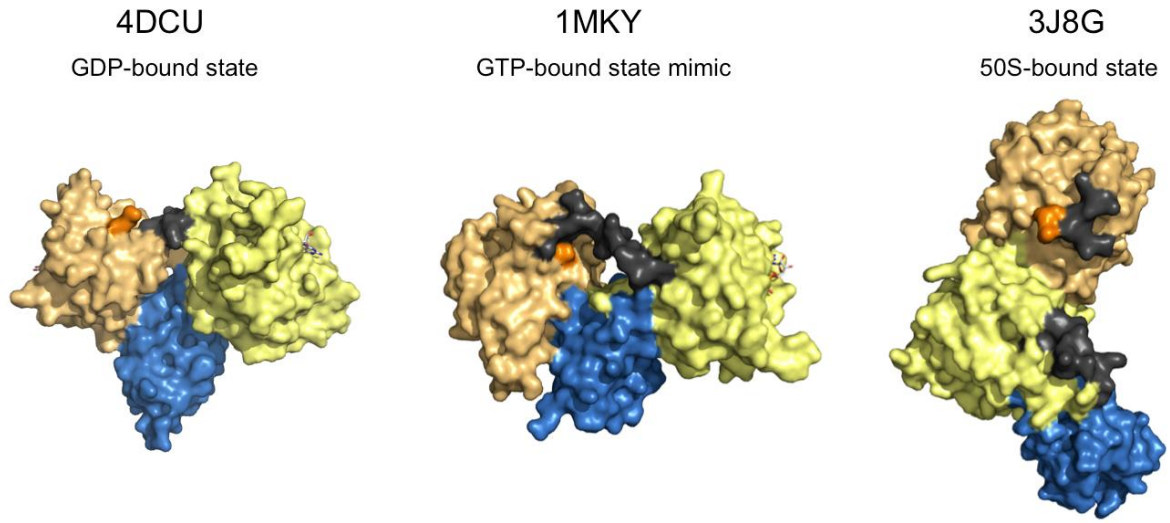

Fig. 57 Surface representation of the 3D structures of EngA from the PDB files 4DCU (left; GDP-bound conformation), 1MKY (middle; pseudo GTP-bound conformation) and 3J8G (right; 50S-bound conformation). GD1 is coloured in light orange, GD2 in yellow, KH-domain in blue and the acidic linker in black. The N-terminus is represented in dark orange. The linker in the 3J8G structure is not structured, due to the low resolution of the cryo-EM structure, and it is present as two fragments at the C-terminus of GD1 and N-terminus of GD2.

Another region that is found cloven at time 15 minutes is the switch II region of the GD2 (at residues 236-239). Cleavage at this region generates two peptides: from N-terminal to the cloven residue (~28.2 kDa or ~26.4 kDa, depending on the presence or absence of the His-tag) and from this residue to the C-terminus (~22.7 kDa). The two N-terminal peptides were assigned to bands B4 and B5 and the C-terminal peptide to band B6. Comparative analysis

of the corresponding peaks in the MS spectra reveals a significant different in their intensities (Fig. 58).

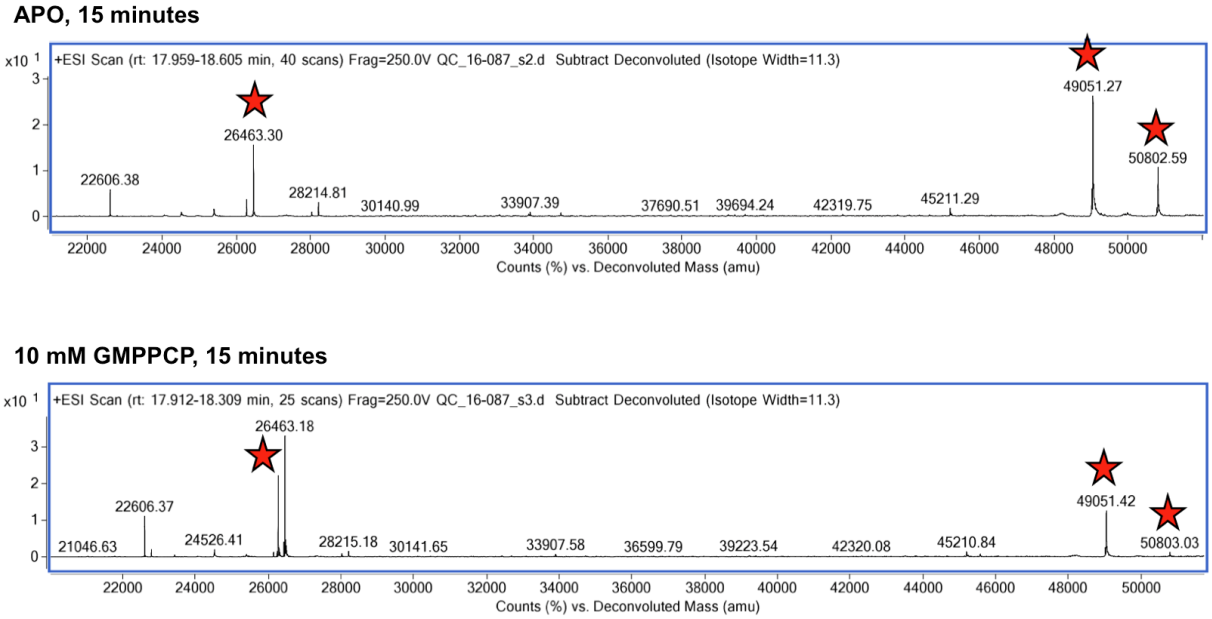

Fig. 58 ESI-TOF MS deconvoluted spectra of the tryptic digestion (time 15 minutes) of His-EngA_{WT} in the absence of nucleotide (top) and in the presence of 10 mM GMPPCP (bottom). Red stars highlight peaks that show significant differences in intensities for both samples. Two such peaks are clearly identified: the one corresponding to cleavage at the His-tag, and the one corresponding to cleavage at the GD2-switch II.

The N-terminal peptide ($^{-2}$G – ^{238}K, 26.463 kDa) is present in higher amounts in the sample with 10 mM GMPPCP, suggesting this region is more susceptible to cleavage when a GTP analogue is bound. This agrees with the observations from the SDS-PAGE, where band B5 is more evident when GMPPCP is in solution. The N-terminal peptide containing the tag (band B4) is also nearly absent from the GMPPCP sample at time 15', while a clear band can be observed in the apo sample (Fig. 54), agreeing with the fast cleavage of the His-tag from the GMPPCP-bound EngA.

Susceptibility to proteolysis requires the backbone chain to be both accessible and flexible enough to adapt and interact with the protease catalytic site (Fontana et al., 2004). According to our SAXS data, the apo-EngA adopts the same conformation than the GDP-bound state. Inspection of the crystal structures of GDP-bound EngA shows that the nucleotide-binding site is well oriented towards the solvent and apparently accessible to a potential protease. Thus, in what concerns accessibility, the apo and GDP-bound global conformation of EngA should not hamper cleavage at this site. This may suggest that the difference in cleavage at the GD2-switch II may arise from a local change in conformation of this region that affects flexibility, rather than a global change of conformation of EngA. Binding of a GTP molecule seems to induce a change in the secondary structure of the GD2-switch II that alters its flexibility and hence its susceptibility to the action of a protease,

enhancing proteolytic cleavage. Indeed, the GD2-switch II is well structured in all *B. subtilis* EngA crystal structures, indicating a lower degree of flexibility. Analysis of the GDP-bound EngA crystal structures shows that this stabilisation occurs, at least partially, by contacts between the K236 and the β-phosphate of the nucleotide. Concurrently with our interpretation, analysis of several crystal structures obtained during our work reveals a destabilisation of the GD2-switch II region upon binding of GTP analogues (discussed afterwards in *Results, section 3.5.3*). This destabilisation in the GTP-bound protein may promote the action of trypsin on this loop. The similar behaviour of the apo and the GDP-bound proteins suggests that the switch II is stabilised in both cases. Hence, similarly to what we already have speculated in the interpretation of our SAXS data, we may think that a GDP is bound to the GD2 in the apo-protein.

Based on the data derived from the gels, MS and N-terminal sequencing, both the N-terminus and the GD2-switch II region seem to be the major cleavage sites responsible for the differences observed between the apo/GDP and GMPPCP samples. We are hence able to confirm the change in the tertiary structure observed by SAXS, and to further reveal the presence of local conformational changes induced by the presence of di- or tri-phosphate nucleotides.

Further analysis of our data revealed additional dissimilarities in the susceptibility of other cleavage sites. N-terminal sequencing detected a cleavage site in the apo sample at 15 minutes that corresponds to residue ^{345}K. Cleavage of the intact protein would generate two fragments: from N-terminal to ^{345}K (~38.7 or ~40.4 kDa, depending on the presence of the His_6 tag) and from ^{346}A to C-terminal (~10.4 kDa). Inspection of the SDS-PAGE from EngA-apo reveals a faint band migrating at a molecular mass close to 40 kDa (band B3), which was assigned to the N-terminal fragment. The weak intensity of this band possibly explains why this fragment could not be detected by mass spectrometry. The same band in SDS-PAGE can be observed for the GDP sample. Indeed, a peak corresponding to the C-terminal fragment from this cleavage (^{346}A-^{436}K, 10.413 kDa) was detected in the GDP sample at 120 minutes by mass spectrometry (not shown). A close look at the SDS-PAGE from EngA-GMPPCP shows an almost undetectable band. This suggests that the cleavage site ^{345}K is less susceptible to tryptic digestion in the GMPPCP-bound protein. This residue is located in the α6-helix of GD2, the last α-helix of this domain, close to the linker connecting the KH-domain. This region is not involved in nucleotide binding, excluding the hypothesis of a potential increased stability in the EngA-GMPPCP due to the presence of a triphosphate nucleotide. Furthermore, as seen from the crystal structure 4DCU of EngA, it corresponds to

a well structured region, which suggests that enhanced cleavage in the EngA-apo and EngA-GDP is not due to a larger flexibility in these proteins. Hence, the slow or absent proteolysis on the GTP-bound protein may be due to a protection of this region, that may become less accessible for a protease. Inspection of the 4DCU structure shows that in the conformation adopted by the apo and GDP-bound EngA this helix is exposed to the solvent (Fig. 59). This may suggest that the conformational change triggered by the triphosphate nucleotide that renders the N-terminus more accessible to proteolysis, could also hide this surface of GD2, protecting it from cleavage.

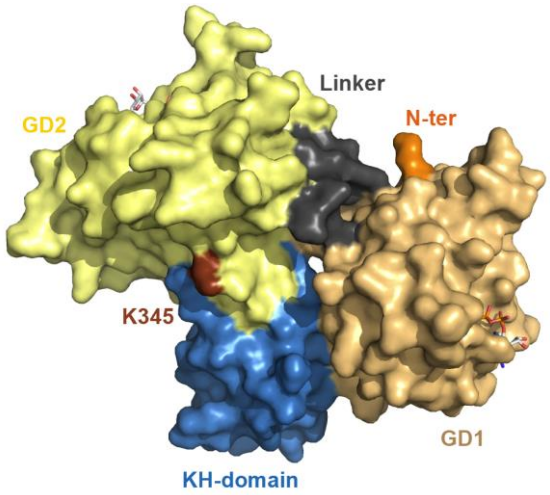

Fig. 59 3D structure of GDP-bound EngA. Two regions are possibly associated to changes in the protein global conformation triggered by a triphosphate nucleotide. Cleavage at these regions, identified as the N-terminus (N-ter, represented in orange) and the α6-helix of GD2 (K345, represented in dark brown), show different rates of fragment generation in different nucleotide bound-states of EngA. In the GMPPCP-bound form, the N-terminal appears to be more susceptible to proteolysis, while the GD2 α6-helix seems to be more protected. A possible movement of one G-domain upon binding of a GTP analogue may alter the accessibility of these two regions to a protease. Image edited in PyMol from PDB 4DCU.

Another cleavage site detected by N-terminal sequencing for the apo-protein at 15 minutes corresponds to residue ^{127}R, close to the GD1-G4 motif. Cleavage at this site generates an N-terminal fragment ($^{-2}$G-^{127}R) of 14.334 kDa, which was detected for the GMPPCP sample at 120 minutes. This fragment was assigned to band B8, which starts to be visible at 15-30 minutes. However, the profile of this band in the SDS-PAGE from the three samples is very difficult to analyse due to the weak intensity of the bands. Based on N-terminal sequencing and MS, this region seems to be cloven both in the apo and GMPPCP sample, suggesting that GDP may stabilise it.

Interpretation of the results of digestion at 120 minutes requires particular attention to an important aspect. While probing global conformational changes, the extent of cleavage that the protein undergoes for longer digestion times and how if affects its tertiary structure is a factor to take into consideration. To analyse changes in the global protein conformation, the

tertiary structure must remain intact. If this organisation is lost, we can assume that the generated fragments will no longer be related to a specific nucleotide-induced conformation, but simply to an extended cleavage of peptides in solution.

Mass spectrometry analysis of samples at time 120 minutes identified one region as being subjected to tryptic cleavage in a later step of the digestion. This region corresponds to the P-loop of GD2 (cleavage at ^{188}K) (Table 9). The N-terminal fragment generated by cleavage at ^{188}K has a molecular mass of 21.046 kDa. This peptide was assigned to band B7 on the SDS-PAGE. Comparison of the MS spectra from the apo and GMPPCP samples (Fig. 60) shows that this peptide is present in high concentrations in the GMPPCP sample, while it is barely present in the apo sample. This matches the observations from the SDS-PAGE, where band B7 is clearly present for the GMPPCP sample, but not for the apo.

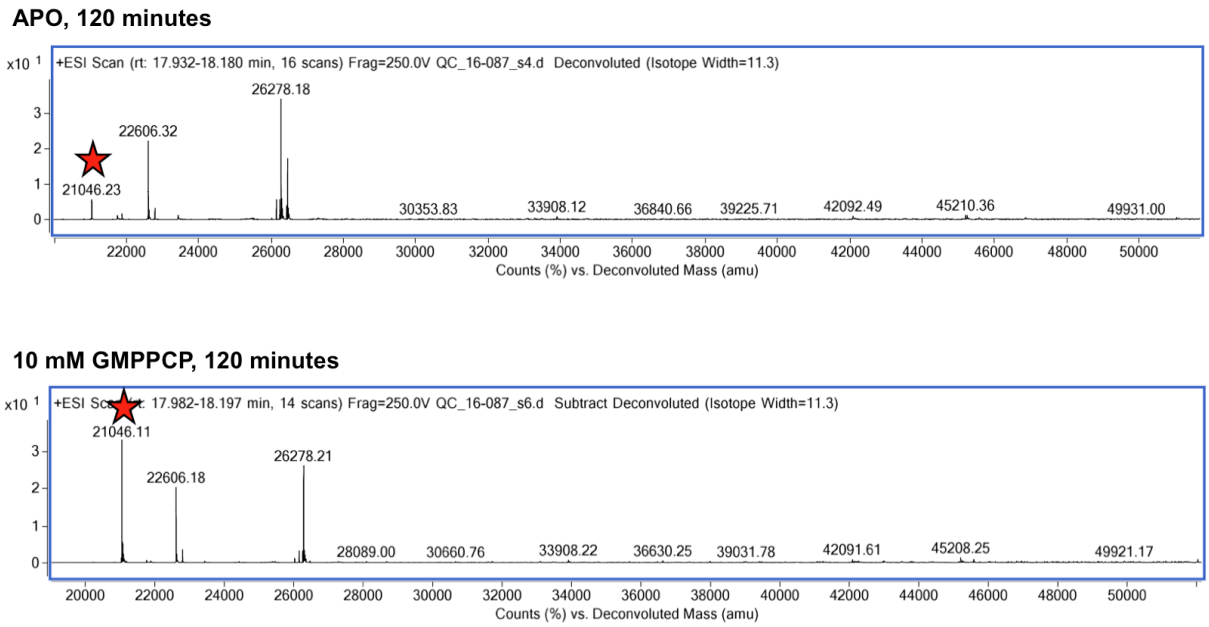

Fig. 60 ESI-TOF MS deconvoluted spectra of the tryptic digestion (time 120 minutes) of His-EngA_{WT} in the absence of nucleotide (top) and in the presence of 10 mM GMPPCP (bottom). Red stars highlight peaks that show significant differences in intensities for both samples.

Increased cleavage of this region in the GMPPCP sample indicates that it is more susceptible to proteolysis when the triphosphate nucleotide is bound. Inspection of the gel from the GDP sample shows a profile closer to the apo, with band B7 basically absent, indicating that the P-loop is also protected here. Thus, it seems that the P-loop is less susceptible to proteolysis in the apo and the GDP-bound protein than in the GMPPCP-bound protein. One possibility for the decreased cleavage of GD2-G1 in the apo/GDP samples may be the orientation of K188, which stabilises the nucleotides' γ-phosphate. The absence of such contacts in the apo/GDP sample may induce a different conformation of the Lys side-

chain that may protect the main-chain from cleavage. Another possible reason for the increased digestion of the GD2-P-loop in the GMPPCP sample may be the fast digestion of the GD2-switch II in this same sample. The presence of the switch II loop may possibly hamper the access to the P-loop to the protease. Once the switch II is cleft, the P-loop may become more accessible, promoting its cleavage. The similar profiles of the apo and the GDP samples seem to once again suggest the presence of a GDP bound to the GD2 of the apo protein. The apparent higher affinity of GD2 over GDP may also protect the P-loop from proteolysis, while a lower affinity for GMPPCP may possibly create an equilibrium between events of binding and unbinding, allowing the action of trypsin.

The fact that differences in proteolytic fragments can be seen between different nucleotide-bound samples suggests that the integrity of the tertiary structure of EngA is maintained at 2h of digestion. This seems to be further suggested by the fact that the two main fragments formed by the N- and C-terminus after cleavage at the switch II (bands B5 and B6, respectively) are still present at 2h, while loss of the tertiary structure would most probably generate many small fragments derived from extensive cleavage. Hence, the information extracted from the analysis of proteolysis at 2h and its correlation with ligand-induced conformational changes seems to be reliable.

Based on the interpretation of our results from limited proteolysis, we were able to create a map of the fragmentation of EngA:

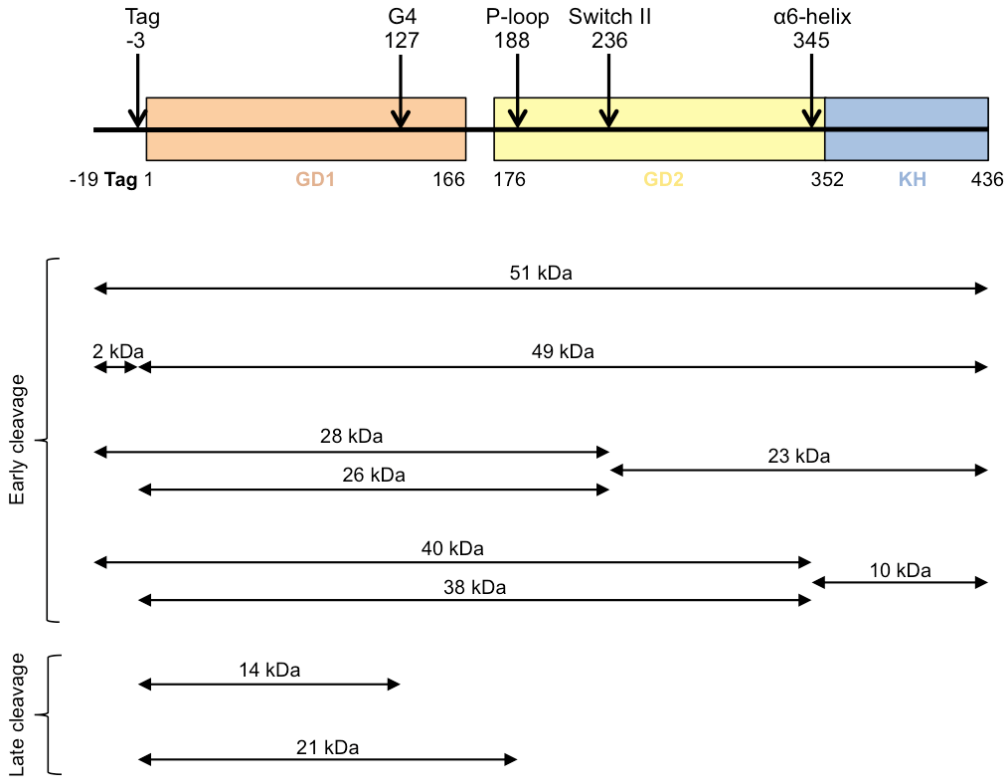

Fig. 61 Fragmentation map of tryptic cleavage of His-EngA_{WT}.

2.2. Mutants trapping different conformational states

The available structures of EngA orthologs show three different conformations that have been associated with different ligand-bound states. One of these conformations, represented by the 4DCU crystal structure, clearly demonstrates a GDP molecule bound to each G-domain and its occurrence in solution was confirmed by our SAXS analysis. Thus, this conformation seems to clearly represent the GDP-bound off-state of EngA. The biological implication of the conformation represented by the 1MKY crystal structure, however, is unclear. Although this conformation has since been considered the GTP-bound on-state of EngA, a detailed observation of the nucleotide-binding site structural features may contest this assumption. Furthermore, the conformational change that we were able to observe by SAXS for EngA in the presence of GTP analogues did not match this structure. The third conformation obtained for EngA from cryo-EM images of the complex 50S:EngA suggests that this structure represents the ribosome-bound state. Despite the high overall resolution of this structure, the EngA binding-site has a rather low resolution and neither the nucleotide-binding sites nor the atomic structural details of EngA structure can be observed.

With the aim of further studying conformational changes of EngA, two mutants were generated to trap the protein in different conformations. These mutants intend to block the protein in the 4DCU or 1MKY conformations, suggested to represent the GDP-bound and the GTP-bound states, respectively. Because the cryo-EM structure, more recently resolved, was not available at the time of our study, a EngA variant for this structure was not created. By being able to block EngA in the desired conformations, we can study their biochemical and structural properties and perform interaction studies, eventually assessing their biological role.

2.2.1. Generation of mutants

Two double-cysteine mutants were designed that contain two substitutions in which the residues were replaced by cysteines. The positions of the substituted residues were chosen in order to block the protein in the two conformations observed in the crystals. These two conformations correspond to PDB structures 1MKY and 4DCU.

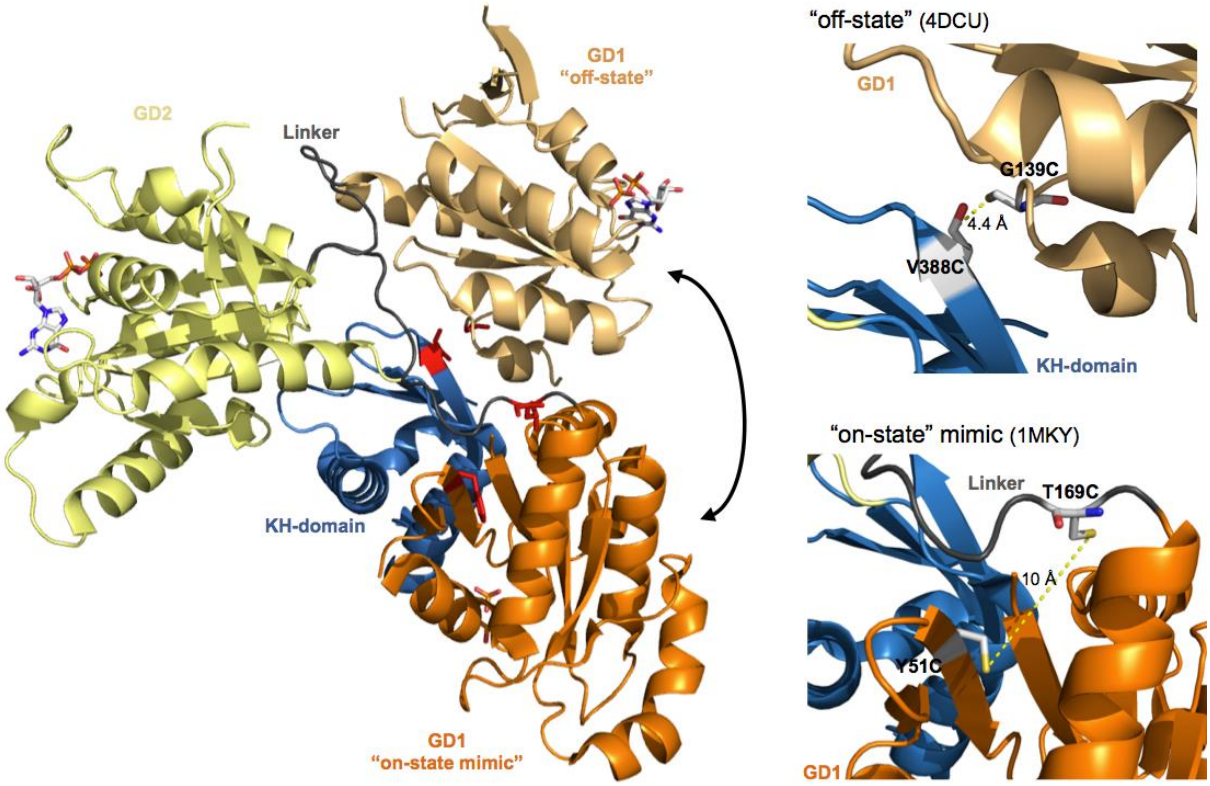

Fig. 62 Superposition of the crystal structures of the GDP-bound EngA ("off-state", PDB 4DCU) and phosphate-bound EngA ("on-state" mimic, PDB 1MKY). Both structures are superposed by their GD2 (yellow) and KH (blue) domains. The GD1 has a different position relative to GD2-KH depending on the presence of a GDP (represented in light orange) or a GTP mimic (represented in dark orange). In order to block these two conformations, specific residues were substituted by two cysteines that can crosslink by forming disulfide bonds. Mutated residues are represented in red. The two insets on the right represent close views of the position of the mutated cysteines aimed to form disulfide bonds in each conformation. The cysteines in the mutant blocking 4DCU are located in the GD1 and KH-domain (top); in the mutant blocking 1MKY, the cysteines are located in the GD2 and the acidic linker connecting the two G-domains.

To mimic the "off-state" 4DCU conformation, residues G139 and V388 were chosen to generate the mutant $EngA_{G139C/V388C}$. In this conformation, these two residues, located in the GD1 and KH-domain respectively, are close enough to form a disulfide bond if cysteines are present instead (Fig. 62). This mutant will be from now on called GV.

The same rational was applied for the "on-state mimic" 1MKY conformation. The residues chosen were Y51 in the GD1 and T169 in the acidic linker connecting G-domains, generating the $EngA_{Y51C/T169C}$ mutant (Fig. 62). This mutant will be from now on called YT.

2.2.2. Purification of EngA mutants

The two variants His-$EngA_{G139C/V388C}$ and His-$EngAT_{Y51C/T169C}$ were generated by site-directed mutagenesis and the mutations were confirmed by DNA sequencing.

The expression and purification of these mutants was done similarly to the wild-type protein, with similar elution peaks and times. After IMAC and SEC, we recovered 120 mg of $EngA_{Y51C/T169C}$ and 25 mg of $EngA_{G139C/V388C}$ for 1 L culture.

The mutations on the constructs do not affect significantly the catalytic activity ($EngA_{Y51C/T169C}$: 6.4 / 71.8 $nmol.min^{-1}.mg^{-1}$ (0 / 300 mM KCl); $EngA_{G139C/V388C}$: 8.5 / 66.3 $nmol.min^{-1}.mg^{-1}$ (0 / 300 mM KCl)).

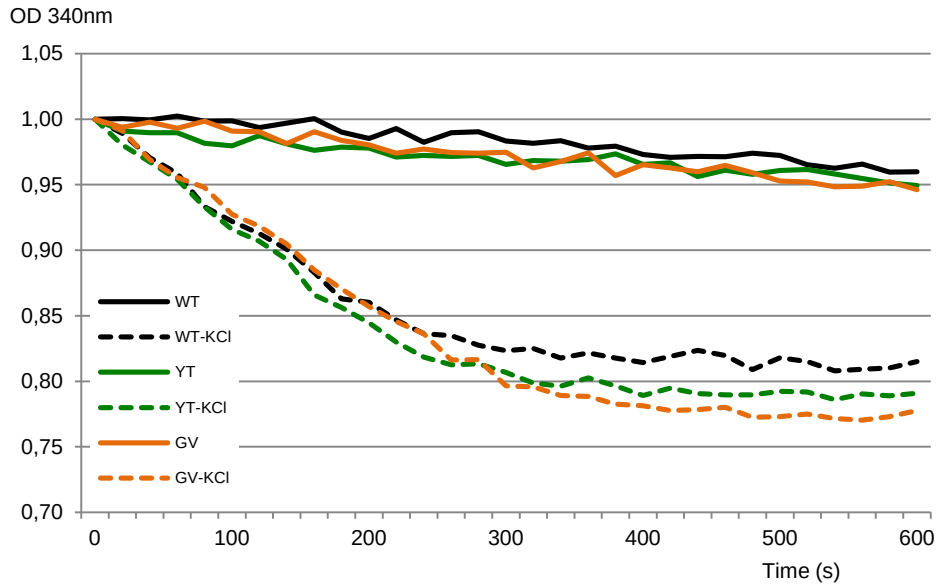

Fig. 63 GTPase activity assay for double-cysteine mutants $EngA_{Y51C/T169C}$ (YT, green) and $EngA_{G139C/V388C}$ (GV, orange). $EngA_{WT}$ (WT, black) is represented for comparison. The curves were corrected by a control where no GTP was used and normalised. The enzymatic activity was tested in the absence of KCl (continuous lines) and in the presence of 300 mM KCl (dashed lines). The specific enzymatic activities were computed from the linear region of the curves.

2.2.3. Cross-linking assay

To induce formation of disulfide bonds between cysteines and block the desired conformations, we performed a cross-linking assay. The two double Cys mutants were treated in oxidising conditions and subjected to SDS-PAGE analysis under reducing and non-reducing conditions. The formation of an intramolecular disulfide bond results in a more compact particle and thus changes the protein mobility in SDS-PAGE in non-reducing conditions.

For each mutant, the cross-linking assay was performed in the absence or presence of 10 mM nucleotides. We speculated that the GV mutant should be able to form intramolecular disulfide bonds in the presence of GDP. Based on our SAXS data that indicates the same conformation for the apo and GDP-bound sates, this variant should also cross-link, at least partially, in the absence of nucleotides. In the presence of a GTP analogue, intramolecular disulfide bonds should not be formed to a big extent, as a change in conformation would lead

to a larger distance between cysteines. Accordingly, if the 1MKY conformation truly represents the GTP-bound state of EngA, we would speculate that the YT variant should form disulfide bonds in the presence of a GTP analogue, but not with GDP or in the absence of any nucleotide.

The results of the SDS-PAGE analysis are shown below:

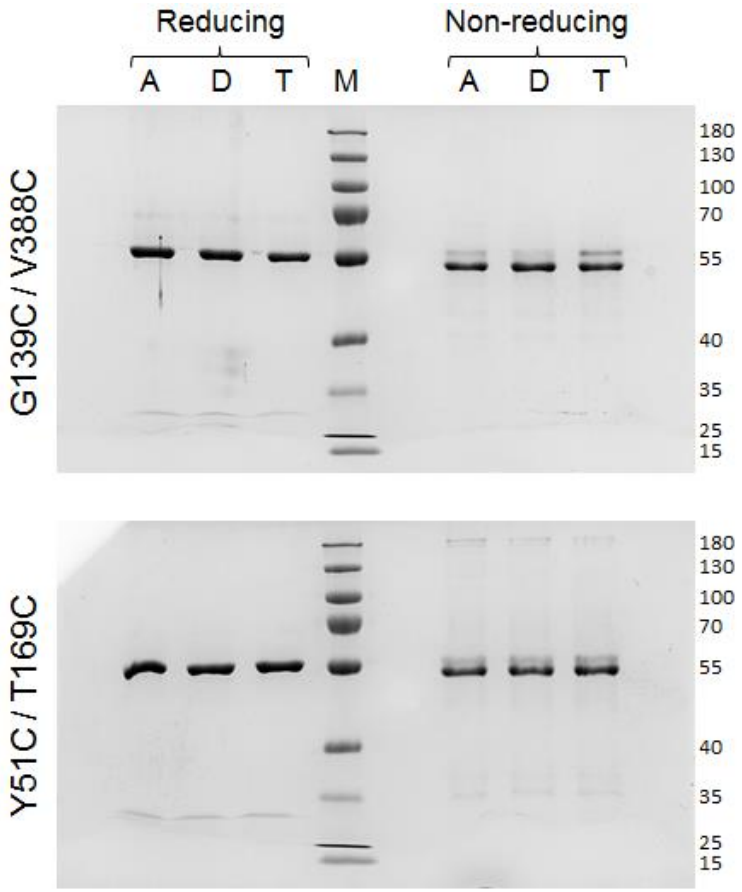

Fig. 64 SDS-PAGE analysis of the cross-linking assay for double Cys mutants $EngA_{G139C/V388C}$ (top) and $EngA_{Y51C/T169C}$ (bottom). (A) apo, no nucleotides; (D) 10 mM GDP; (T) 10 mM GMPPCP; (M) molecular weight marker.

For both mutants, EngA treated under oxidising conditions migrated as two bands of different mobility, with a slower component representing the non-cross-linked protein and the faster component representing the cross-linked protein. This was confirmed by addition of a reducing agent that converted all sample to a single slow migrating band.

As we expected, the intramolecular disulfide bond was formed in almost 100% of the GV variant in the presence of 10 mM GDP. This is in line with the conformation represented by the PDB 4DCU structure being the real GDP-bound state of EngA. In the absence of nucleotides, although to a smaller degree, the cross-link was also formed, with a faint band corresponding to the non-cross-linked component. GMPPCP seems to slow down the

formation of cross-link. Although the protein is partially able to form a disulfide bridge, a clear band corresponding to the slower component is visible.

The YT variant shows a similar pattern in the three tested conditions. The formation of an intramolecular disulfide bond indicates that a conformation that brings the two Cys residues close together (potentially that of PDB 1MKY) is adopted in solution. The occurrence of S–S bridges in the absence of nucleotides suggests that there is a dynamical equilibrium of different conformations in solution. Despite the SAXS data for the apo-EngA clearly revealed the 4DCU structure, the presence of additional conformation(s) for a small fraction of the sample would not affect the SAXS-derived protein envelope but could allow the accumulation of cross-linked proteins over the time of the assay. Unexpectedly, GDP does not seem to avoid the formation of cysteines cross-link. One explanation for this may reside in the position of the T169C mutation, that is located in the flexible linker between both G-domains. High flexibility in this region may eventually allow the two Cys residues to come closer, even if the G-domains are apart.

2.2.4. Small-angle X-ray scattering (SAXS)

After obtaining the cross-linked EngA variants, we were interested in confirming the adopted structure and in studying the stability of these blocked conformations upon addition of nucleotides. To do so, we employed SAXS following the same analysis as for the wild-type protein.

The GV mutant intended to adopt the conformation of the 4DCU crystal structure. The fit of the experimental scattering curve of the cross-linked GV mutant and the computed scattering curves from the PDB coordinates of 4DCU is represented in Fig. 65.

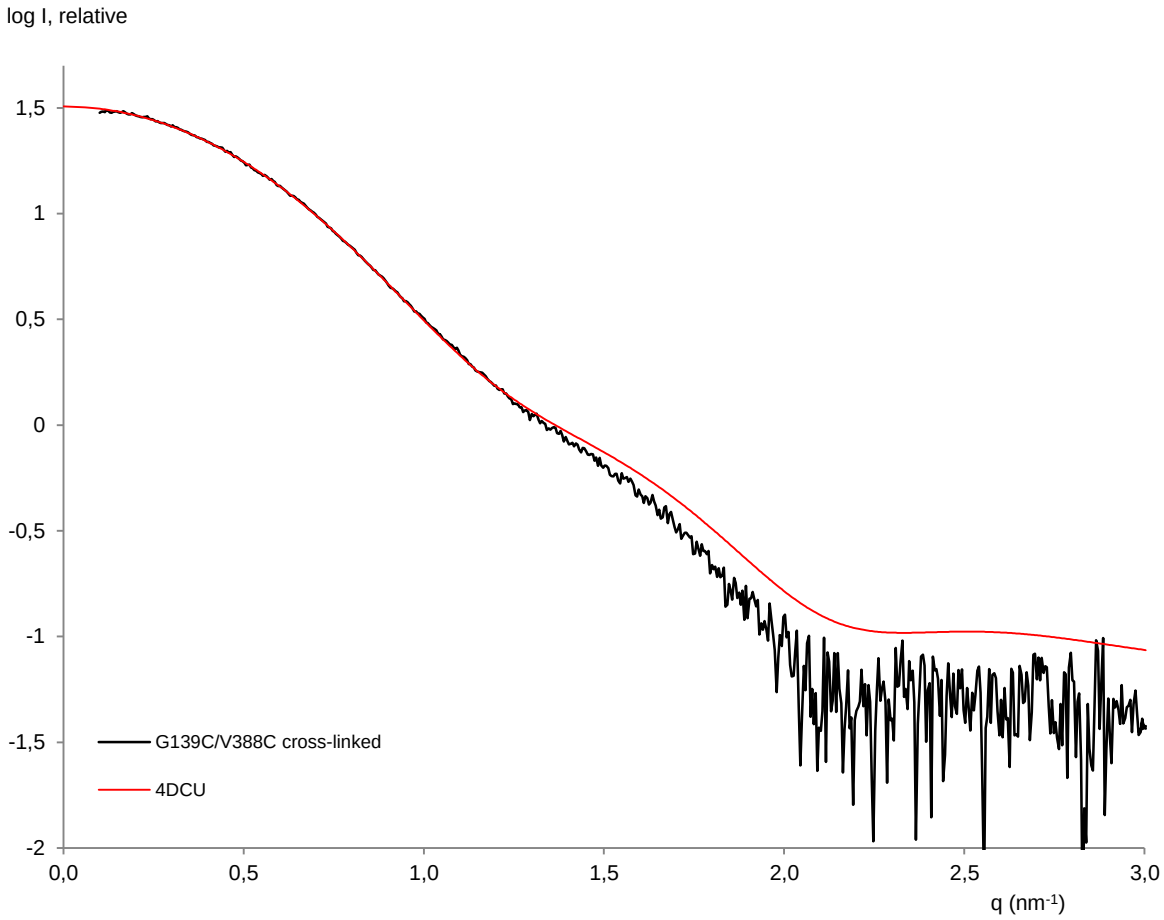

Fig. 65 Logarithmic plot of the scattering curve [log I(q) vs. q] for cross-linked His-Eng$A_{G139C/V388C}$. The black line denotes the experimental data. The red line corresponds to the fit with the crystal structure PDB 4DCU.

The χ^2 value for the fit is 7.39. Comparison of the two curves shows a good match in the Guinier region, with identical R_g. Slight differences start to appear in the end of the Porod region, suggesting that the conformation induced by cross-link of the GV mutant may not completely match the one from the apo and GDP-bound proteins. Despite these differences, these two conformations seem to be similar.

Several reasons led us to anticipate that a similar conformation to that of the crystal structure would be easily induced by cross-linking. The GV mutant was designed from the 4DCU structure, which belongs to the *B. subtilis* protein and hence the sequence/structural analysis and determination of the residues to mutate could be done more easily. Both cysteines are located in well-defined structured regions of EngA (GD1 and KH-domain), where large movements are not expected. Furthermore, this conformation revealed to be a stable conformation adopted by the apo and GDP-bound proteins. Small changes are however reasonably expected, as a single S–S bond may not be enough to perfectly block a desired conformation. The fact that small changes can be observed confirms that the cross-link between both cysteines was induced.

The cross-link in the YT mutant intended to block the conformation of the 1MKY crystal structure. The same fitting analysis was made for this mutant. The fit of the scattering curves (experimental and models) is represented in Fig. 66.

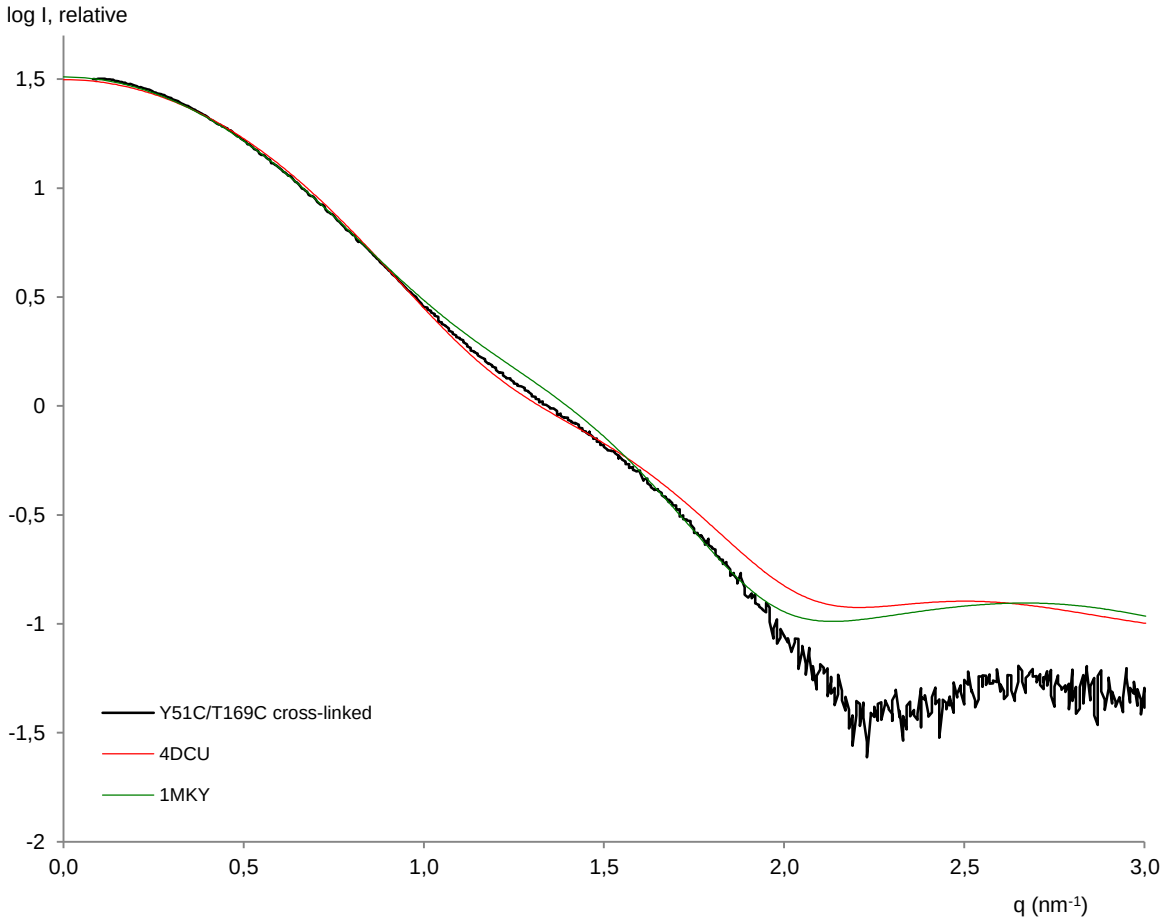

Fig. 66 Logarithmic plot of the scattering curve [log I(q) vs. q] for cross-linked His-EngA$_{Y51C/T169C}$. The black line denotes the experimental data. The red line corresponds to the fit with the crystal structure PDB 4DCU and the green line to the fit with the crystal structure PDB 1MKY.

The cross-link on the YT mutant has disrupted the 4DCU conformation, as seen by the fit with the PDB model, which as a χ^2 value of 94.24. The R_g increased to 2.88 ± 0.01 nm, approaching that of 1MKY and significant differences in the Porod region are observed, indicating that the cross-link was formed and the apo-conformation disrupted. However, comparison with the 1MKY structure reveals that the conformation triggered by the disulfide bond does not correspond to the intended one. Although there is a good agreement within the Guinier region, significant differences are present in the scattering curve and the χ^2 value for the fit with 1MKY is 78.84. This indicates that the cross-link between residues C51 and C169 is not able to block the 1MKY conformation. Which conformation is instead adopted cannot be deduced by SAXS analysis.

This outcome was not completely unexpected. The 1MKY structure belongs to the *T. maritima* orthologue, and hence analysis of this structure and identification of the residues to mutate in the *B. subtilis* protein was not straightforward. Furthermore, one of the mutated cysteines is located in the acidic linker between G-domains, which, being a flexible region, may allow movement of the G-domains into different positions despite the disulfide bond.

While a cellular functional meaning cannot be inferred for the YT variant, the cross-linked $EngA_{G139C/V388C}$ protein seems to be an interesting approach to study biochemical properties and ribosome-binding capability of the GDP-bound EngA.

3. Crystallographic studies

X-ray protein crystallography allows us to visualise the atomic details of proteins and study their three-dimensional structure. Different crystal structures of EngA orthologs are available in the PDB. Here, we have performed several crystallographic studies on *B. subtilis* EngA in order to obtain structural information that could help in our attempt of targeting EngA for the design of antibiotics. More specifically, we aimed at identifying the mode of binding of protein partners or inhibitors to EngA and characterising different conformational states.

3.1. Crystallisation goals

In search of the GTP-bound conformation

SAXS analysis revealed that 10 mM GTP analogues trigger a change in conformation on EngA. The GTP-bound conformation has been assigned to the crystal structure from *T. maritima* EngA (PDB 1MKY), where GD1 has two phosphate ions bound in the nucleotide-binding site. However, comparison of the structural properties from the conformation we observed in solution and the 1MKY structure shows significant differences. This suggests that the GTP-bound conformation of *B. subtilis* EngA is still to be determined. Understanding how the switch mechanism occurs in EngA is an important step to understand its interactions with effector molecules and determine its function in cells. We have done crystallisation trials in order to crystallise EngA in its GTP-bound conformation and determine its structure.

Assessing binding to ribosomal protein S7

Several studies have demonstrated that EngA can bind both ribosomal subunits. While the *E. coli* EngA:50S complex has recently been solved by cryo-EM (Zhang et al., 2014), the structural aspects of the interaction with the 30S subunit are not known. The S7 protein from the small subunit was identified as being an interaction partner of EngA (Lamb et al., 2007). With the aim of obtaining the structure of the EngA:S7 complex and further assess interaction mechanisms with the small subunit, we have performed co-crystallisation trials with these two proteins.

Structure-based drug design: defining ligand-binding sites

Knowing the three-dimensional structure of biological macromolecules targeted for drug discovery can be an effective tool for determination of ligand-binding sites and leads optimisation. Bharat and coworkers (Bharat et al., 2013) have developed a high-throughput screen towards inhibitors of the GTPase activity of EngA, which identified four hit compounds. Two of these compounds plus a structural analogue (see Fig. 34) were synthesised by Ahcène Boumendjel (UGA, Grenoble) and used in crystallisation trials, with the aim of defining their binding-mode and potentially understand a rational basis for the design of more specific EngA-directed drugs.

3.2. Crystallisation trials

Obtaining well-ordered protein crystals is a critical step of X-ray crystallography. For proteins to crystallise, the protein solution must be brought to a supersaturation state. Different factors – such as temperature, organic solvents, ionic strength, pH, the nature of the buffer and salts, among others – affect protein solubility. These factors can be exploited by adjusting each component in incremental amounts, in order to achieve supersaturation, favour protein-protein interactions and obtain a crystalline form (DeLucas et al., 2003; McPherson, 2004).

The first step to obtain protein crystals is to identify conditions that yield crystalline material. We have started our crystallisation trials with a high-throughput screening at the HTX Lab platform (EMBL, Grenoble). For co-crystallisation of EngA with GTP analogues, S7 protein or CEB-II-5x compounds, six commercial standard screen plates were initially tested at 20°C (*Classics Suite* and *PEGs-I* from Qiagen; *JCSG+* and *PACT* from Molecular Dimensions; *Salt Grid* from Hampton Research; *Wizard I&II* from Rigaku Reagents). These screens are formulated based on the analysis of successful protein crystallisation conditions and allowed us to cover a wide range of salts, precipitants, buffers and pH range. About 1/5 of the screened conditions provided crystals, with different shapes and sizes (Fig. 67). Most of the conditions prone to produce crystals were composed of phosphate and sulfate salts, with a pH range of 6.5–8.5 and PEG 3350–8000 between 10–25% as a precipitating agent.

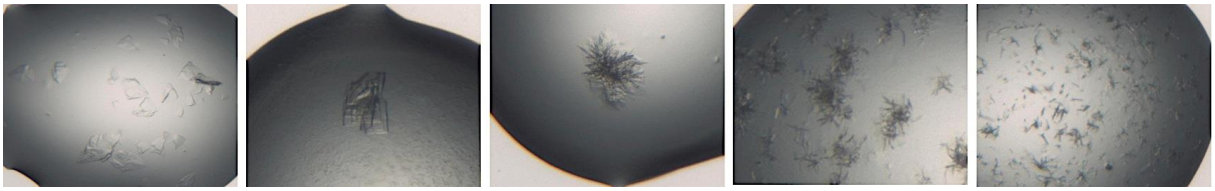

Fig. 67 Crystals obtained from the first crystallisation screening of EngA with GTP analogues. Similar shapes were obtained in crystallisation trials with other ligands (S7, CEB-II-5x). Crystals varied between sea urchins, needles (most of the time growing from a single nucleation centre) and clustered plates. These photographs were obtained from the Imaging Robot at the HTX Lab.

At this point, the first diffraction tests were done at the ESRF to check for the nature of the crystals (either protein or salt crystals) and their diffracting power.

After selecting promising conditions from the initial coarse screening trials, different factors that can affect crystallisation were refined in order to optimise crystal growth. We prepared 96-well plates at the HTX for a fine screen of salt and precipitant concentrations. Crystal growth was controlled at 20°C and 4°C to test for the effect of temperature. Based on these refinements, we prepared hand-made drops in 15-well plates in order to obtain larger crystals. In parallel with fine refinement of hit conditions, salts, buffers and precipitants from these positive hits were used in a combinatorial analysis. This was done with the program SAmBA (Audic et al., 1997) that computes an optimal set of experiments derived from the original variables.

As most of our crystals were rather small (20–100 µm), two-dimensional (plates or needles) and growing as clusters, additional strategies were used to improve crystal morphology and size. Additives from a commercial screen (Hampton Research) were tested. Microseeding was also employed in an attempt to separate nucleation from crystal growth.

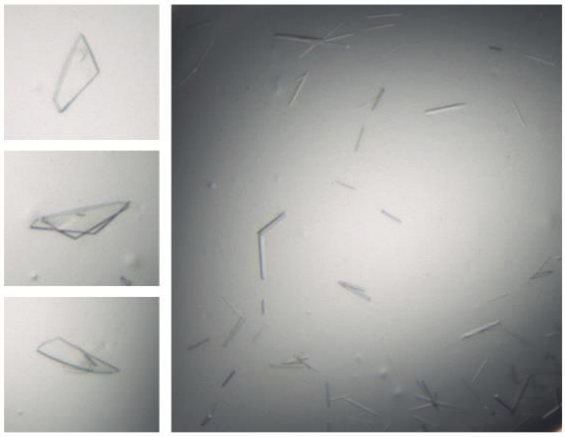

Fig. 68 Crystals obtained after steps of optimisation of the crystallisation conditions of EngA. Crystals here grow as single needles or plates. These photographs were obtained from the Imaging Robot at the HTX Lab.

Additional strategies besides co-crystallisation were employed for crystallising EngA in complex with CEB-II-5x inhibitors. Crystals of apo-EngA were grown in optimised conditions and soaked in a liquor containing these compounds. Soaking allows ligands to diffuse into the crystal and to bind to their binding-site, if this is accessible in the crystal packing. Both in co-crystallisation and soaking, the low solubility of these compounds limited the ligand concentration used in solution. Hence, to avoid this limitation but also the use of DMSO (that is required for solubilising CEB-II-5x, but that may be harmful for proteins or protein crystals), we employed a third strategy. Trials of dry co-crystallisation were done at the HTX and on hand-made drops, where crystallisation wells were pre-coated with dried ligands.

Along this process, several crystals were tested at the ESRF. Some crystals diffracted well enough to obtain an atomic structure of EngA. While refining these structures, we realised that phosphate and sulfate ions were competing with GTP analogues and were most of the time present in the nucleotide-binding site. Based on this observation, and in order to search for a GTP-bound conformation, we re-analysed the hit conditions from the initial coarse screening phase and selected those that did not contain neither phosphate nor sulfate salts. We worked again on the refinement and optimisation of this set of conditions, as described previously. Additional parameters were tested. Crystals were grown at 4°C, 20°C and 30°C to test the effect of temperature on the rate of crystal growth. Motivated by our SAXS results, different concentrations of GTP analogues were used, starting from 1 mM to 20 mM. The effect of potassium was also assessed by addition of 700 mM KCl. Five additional commercial screens from the HTX platform were tested (*Morpheus*, from Molecular Dimensions; *JBScreen Classic 1* from Jena Bioscience; *MPD* and *Cations* from Qiagen; *Polymer Grid* from Hampton Research). These screens were chosen to cover particular aspects in the crystallisation of a new form of EngA, such as: screening of new additives; screening of PEGs (which had been given good hits for EngA); or screening of several salts not including PO_4^{3-}/SO_4^{2-}. Finally, one promising crystallisation condition was optimised using two commercial screens from the HTX platform (*Opti-Salts* from Qiagen and *Additive Screen* from Hampton Research) that screen salt additives, small molecules and pH.

Unfortunately, after approximately 7300 crystallisation trials and 325 crystals tested at the ESRF, we were not able to obtain EngA in complex with S7 or CEB-II-5x, nor in a new conformation.

Nevertheless, seven crystals diffracted up to ~3 Å, which allowed us to refine their structure. We were thus able to inspect the nucleotide-binding site and make a comparative analysis of these structures.

3.3. Determination of EngA structures

Sixteen crystals tested at the ESRF gave useful data for molecular replacement with Phaser. Molecular replacement was possible as there are structures of EngA orthologs that can be used as models for phasing. Three of these structures were used as models: 4DCU (*B. subtilis*, GDP-bound), 1MKY (*T. maritima*, phosphate-bound) and 3J8G (E. coli, 50S-bound). Either the entire structures of the individual GD1, GD2 and KH domains were searched.

The search for a solution is done by first determining the model's orientation by solving the rotation function (i.e. finding the model's orientation for which the model and the experimental Patterson functions correlate the most) and then determining the position of the oriented model by solving the translation function (i.e. the position at which the calculated and experimental structure factors have the best match). Inspection of several parameters allowed us to know whether a good solution was found. The presence of one strong peak tells that the correct solution was most probably found, as it indicates a clear discrimination between the rotation/translation functions of this particular solution and the others. The LLG (log-likelihood gain) tells how better the experimental data can be predicted by the model in comparison to a random distribution of its atoms, and a high value suggests that the model used was correct. A high LLG should also be accompanied by a high Z-score, which measures the signal-to-noise ratio. Phaser also checks that no bad contacts are present in the solution by eliminating solutions with too many steric clashes. Finally, the solution generated was inspected in Coot to confirm that no bad contacts were present between the three searched domains and check that the packing was reasonable and that the electron density was in good agreement with the model.

These several diffraction data sets corresponded to crystals from crystallisation trials of EngA with S7, nucleotides and inhibitors. Processing of these data revealed a good match with the 4DCU structure for all crystals. Such outcome from co-crystallisation drops of EngA and S7 means that, despite the efforts of obtaining the complex, EngA crystallised alone. Likewise, this observation for co-crystallisation drops of EngA and GTP analogues indicates that a new conformation was not obtained, despite the use of 20 mM nucleotides. Due to limitations of resolution or completeness of the data, the content of the nucleotide-binding site could not be investigated in all structures to assess the presence of a GTP analogue or a CEB-II-5x inhibitor.

From these sixteen crystals, however, seven diffracted to a resolution good enough to refine the structure. The statistics of the diffraction data processing and refinement for these seven structures are presented in Table 10. Table 11 contains the crystallisation conditions and

ligand occupancy of G-domains. A brief description of these structures is presented next, followed by a more detailed analysis of the crystal packing and structural changes.

Table 10 Statistics of diffraction data processing and refinement from seven His-EngA_{WT} crystal structures (Crystals 1–7). The last column referring to Crystal C2 corresponds to a crystal that was obtained in a different space group (*C*2), although refinement was not possible due to low completeness. The values in brackets for resolution correspond to the lowest resolution limit. The values in brackets for completeness, Rsym, CC(1/2) and I/sigma(I) correspond to the highest resolution shell.

	Crystal 1	**Crystal 2**	**Crystal 3**	**Crystal 4**	**Crystal 5**	**Crystal 6**	**Crystal 7**	**Crystal C2**
Beamline (ESFR)	ID23-2	ID23-1	BM30	ID23-1	BM30	ID30-B	ID30-B	ID23-2
Collection date	2015/02/01	2015/06/15	2015/07/08	2015/09/11	2016/02/15	2016/04/28	2016/04/28	2015/02/01
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	*P*1	*P*1	*C*2
Unit cell parameters (a, b, c) (Å)	50.93 63.88 62.45	50.92 64.15 62.01	50.05 65.12 62.40	51.28 65.84 62.62	53.58 64.80 62.47	57.27 63.17 65.91	57.21 63.19 65.95	141.99 64.66 62.60
Unit cell parameters (α, β, γ) (°)	90.0 97.3 90.0	90.0 96.4 90.0	90.0 97.6 90.0	90.0 96.7 90.0	90.0 97.2 90.0	90.0 109.5 97.7	89.9 109.5 97.7	90.0 106.8 90.0
Resolution limit (Å)	3.15 (61.94)	3.20 (61.62)	3.10 (14.1)	3.00 (13.4)	3.00 (13.6)	2.90 (13.2)	2.90 (13.0)	4.00 (18.0)
Completen. (%)	88.5 (92.1)	98.5 (96.5)	99.2 (100)	99.3 (98.9)	95.5 (92.7)	85.1 (87.6)	85.4 (89.2)	49.3 (46.9)
Rsym (%)	11.4 (44.2)	9.3 (74.4)	15.6 (64.9)	9.6 (50.9)	14.5 (119.7)	13.7 (64.4)	13.0 (59.7)	18.9 (24.2)
I/sigma(I)	7.0 (2.3)	9.7 (1.7)	8.8 (2.5)	9.8 (2.4)	6.0 (0.9)	4.4 (1.3)	5.1 (1.3)	3.43 (2.4)
R-factor	0.221	0.222	0.204	0.195	0.245	0.279	0.271	-
R-free	0.318	0.363	0.350	0.305	0.368	0.390	0.380	-

Table 11 Sample and crystallisation conditions from each crystal described in Table 10.

	Sample	Crystallisation condition	Nucleotide-binding site content	
Crystal 1	His-EngA_{WT} 3.7 mg/mL S7-His 1.65 mg/mL 1 mM GMPPNP 2.5 mM EDTA	0.2 M Na/KPO_4 0.1 M Tris pH 6.5 16% PEG 3350	GD1: 2 PO_4^{3-}	GD2: GMPPNP
Crystal 2	His-EngA_{WT} 4.5 mg/mL	0.5 M Na_2SO_4 0.1 Bis-tris propane pH 6.5 16% PEG 3350 Soaking solution: 0.1 mM CEB-II-54 10 % DMSO 0.05 M CAPS pH 10	GD1: SO_4^{2-}	GD2: GDP
Crystal 3	His-EngA_{WT} 4.5 mg/mL	0.5 M Na_2SO_4 0.1 HEPES pH 7.0 8% PEG 3350 Soaking solution: 0.2 mM CEB-II-56 10 % DMSO 0.05 M CAPS pH 10	GD1: SO_4^{2-}	GD2: SO_4^{2-}
Crystal 4	His-EngA_{WT} 4.5 mg/mL 10 mM GMPPNP	0.5 M Na_2SO_4 0.1 MES pH 6.5 10% PEG 3350	GD1: SO_4^{2-}	GD2: GMPPNP
Crystal 5	His-EngA_{WT} 3.6 mg/mL 20 mM GMPPCP 20 mM $MgCl_2$	0.2 M Na_2SO_4 0.1 Bis-tris propane pH 7.5 18% PEG 3350	GD1: SO_4^{2-}	GD2: GMPPCP
Crystal 6	His-EngA_{WT} 5.4 mg/mL 10 mM GMPPCP 10 mM $MgCl_2$	0.2 M KSCN 18% PEG 3350 2.4M NaSCN	Chain A: GD1: (GMP)PCP Chain B: GD1: (GMPPCP)	Chain A: GD2: GMPPCP Chain B: GD2: GMPPCP
Crystal 7	His-EngA_{WT} 5.4 mg/mL 10 mM GMPPCP 10 mM $MgCl_2$	0.2 M KSCN 18% PEG 3350 2.4M NaSCN	Chain A: GD1: (GMP)PCP Chain B: GD1: apo	Chain A: GD2: GMPPCP Chain B: GD2: GMPPCP
Crystal C2	His-EngA_{WT} 3.7 mg/mL S7-His 1.65 mg/mL 1 mM GMPPNP 2.5 mM EDTA	0.1 M Tris pH 8.5 16 % PEG 8000	-	-

3.3.1. Crystallisation with CEB-II-5x inhibitors

Our efforts to obtain the CEB-II-5x inhibitors bound to EngA have resulted in two crystal structures (Crystals 2 and 3). These crystals were grown in similar conditions in the apo-form and soaked with ligands (see Table 11), and belong to space group $P2_1$. However, inspection of the nucleotide-binding site revealed that the inhibitors were not bound to EngA. Instead, sulfate ions occupy the place of the β-phosphate of the nucleotide (Fig. 69A). These crystals were obtained from drops in which Na_2SO_4 was present, which we now know to compete with other ligands in binding to the nucleotide-binding sites, as observed in crystallisation trials of EngA.

When refining the structure of Crystal 2, a positive density in the residual map was observed in the GD2 that was not satisfied by a SO_4^{2-} ion. Attempts of placing a CEB-II-54 molecule were neither successful. Finally, a GDP molecule was modelled in the density. While the phosphate moiety of GDP is correctly positioned, the ribose and guanine base show a rotation that places the nucleobase at distances of 9–15 Å from G4 and G5 motifs (Fig. 69B). Indeed, the occupancy of the nucleobase is low, indicating it is not stabilised and is free to move.

GDP had not been added in the crystallisation conditions. This is not the first evidence we obtain about the presence of a GDP on GD2 in the purified apo-protein. It is surprising though how this GDP molecule could remain bound to GD2 given the crystallisation conditions. In Crystal 3, the presence of 0.5 M Na_2SO_4 was enough to replace GDP. The presence of GDP is even more intriguing given the lack of stability of the nucleobase, which would suggest that the nucleotide is not strongly bound. We can think of the possibility of the CEB-II-54 being able to displace the nucleobase but not the phosphates. However, if this is the case, we would expect that 0.5 M SO_4^{2-} would easily displace the nucleotide phosphates afterwards. We may also have a heterogeneous content of this GD2 nucleotide-binding site within the crystal, which renders the interpretation of the electron density quite difficult.

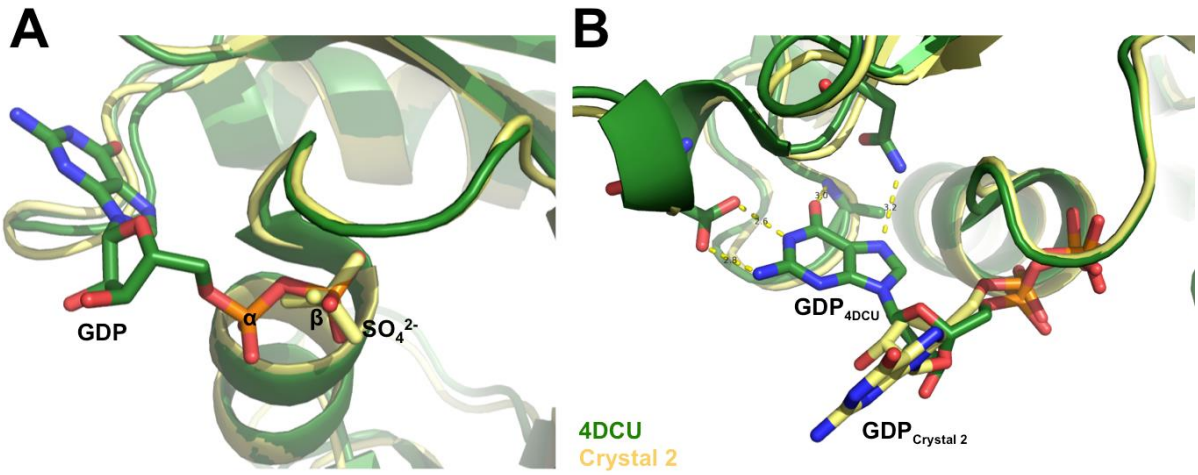

Fig. 69 Details of the nucleotide-binding site of EngA from Crystal 2. The 3D structure of Crystal 2 is represented in yellow. Crystal 3 is not shown as the details of both nucleotide-binding sites are identical to that of Crystal 2 in image A. In green, 4DCU is represented for comparison. **(A)** The GD1s from Crystals 2 and 3 and the GD2 from Crystal 3 have a sulfate ion occupying the position of the β-phosphate of the nucleotide. **(B)** In Crystal 2, GD2 has a GDP molecule bound. The nucleobase adopts a different position, where it cannot form contacts with residues from G4 and G5.

3.3.2. Crystallisation with S7

Co-crystallisation of EngA and S7 has resulted in one crystal structure (Crystal 1). However, in place of the EngA:S7 complex, EngA crystallised alone. This crystal was grown in similar conditions to Crystals 2 and 3 (see Table 11), and belongs to the same space group $P2_1$,

with identical unit cell parameters. Inspection of the nucleotide-binding site revealed the presence of a GMPPNP in GD2 and a phosphate ion in GD1. Like the previous structure, the phosphate ion is occupying the position of the β-phosphate of the nucleotide. The triphosphate nucleotide bound to GD2 did not trigger any change in the global conformation of EngA (Fig. 70). This had already been observed in one crystal structure of *B. subtilis* (PDB 4DCV; see *Introduction, section 5.4*) that adopts the 4DCU conformation despite the presence of a GMPPCP in GD2.

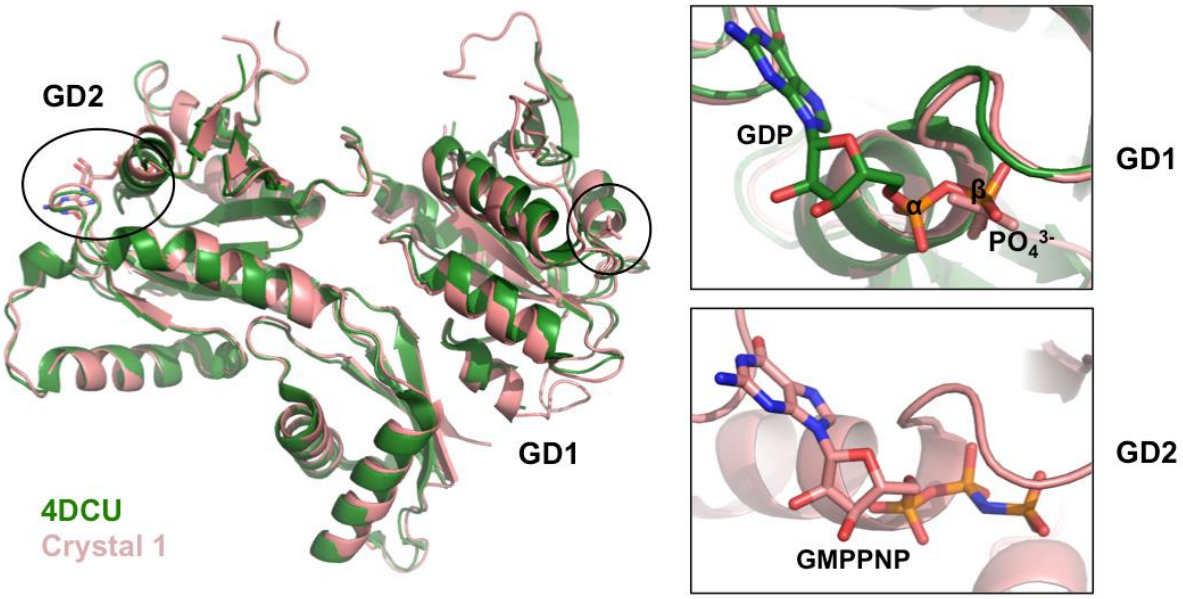

Fig. 70 (Left) EngA 3D structure obtained from Crystal 1 (pink) superposed with 4DCU (green). This crystal was obtained from a co-crystallisation trial of EngA and S7, however EngA was present alone in the crystal. Circles represent the nucleotide-binding sites of G-domains. **(Right)** Close view of the nucleotide-binding sites of GD1 and GD2. GD1 has a bound phosphate ion in the position of the nucleotide β-phosphate. GD2 has a GMPPNP molecule bound in the nucleotide-binding site. However, the global conformation of EngA did not change.

3.3.3. Crystallisation with GTP analogues

Several attempts were made in order to obtain a new conformation for the GTP-bound EngA. From these crystallisation assays, four crystal structures were refined (Crystals 4, 5, 6 and 7).

Our initial co-crystallisation trials were done with 1 mM nucleotide. However, SAXS indicated that we need 10 mM GTP analogues to trigger a change in conformation. Hence, in order to maximise our chances of having the triphosphate nucleotide bound in both domains and obtain a new structure, the subsequent trials were done with 10 and 20 mM GMPPC/NP.

Two crystals (Crystals 4 and 5) were obtained from sulfate-containing drops. Similarly to Crystals 1, 2 and 3, they belong to space group $P2_1$ and have identical unit cell parameters. The content of the nucleotide-binding sites does not differ either. Both structures have a sulfate ion in the same position in GD1, and a GTP analogue in GD2 (as in Fig. 70, for Crystal 1).

Although the high sulfate concentration allows the binding of a nucleotide in GD2, this is never the case in GD1, suggesting a lower affinity of this domain towards nucleotides. This was seen for Crystals 1, 2, 3, 4 and 5, where sulfate/phosphate competed with the nucleotides in solution. In addition, for Crystal 3, sulfate was even able to replace the intrinsically bound GDP on GD2. From this point, all crystallisation conditions including sulfate or phosphate salts were rejected from our assays.

Two other diffraction datasets were more recently obtained from co-crystallisation of EngA and GMPPCP (Crystals 6 and 7). They were collected from the same crystal, by hitting two different regions of the crystal. However, when trying to merge the two datasets, the statistics for the scaled data became much worse, with increased R-sym and R-meas values. These two datasets where then processed separately. This crystal was grown in different conditions from all previous (see Table 11), and belongs to space group *P*1, with two molecules in the asymmetric unit. The four molecules in these two structures have the same 4DCU conformation. However, some differences on the content of the nucleotide-binding sites can be seen.

In both structures, one of the molecules (chain A) has a GMPPCP bound in GD2 (Fig. 71B). When inspecting GD1, we saw electron density corresponding to the binding-sites of the α- and β-phosphates. However, there was no density for the ribose, nucleobase or γ-phosphate. The present electron density was too large to be consistent with any ion from the crystallisation solution. It was also unlikely that a GDP would be present (from hydrolysis of GMPPCP, since no GDP was added), as in this case we would expect the nucleobase to be visible in the electron density. We interpreted this observation as the β- and γ-phosphates of GMPPCP bound to the positions of the α- and β-phosphates (Fig. 71A). Possibly because of this shift, the guanine base cannot be well stabilised by G4/G5 motifs and hence we cannot see its electron density.

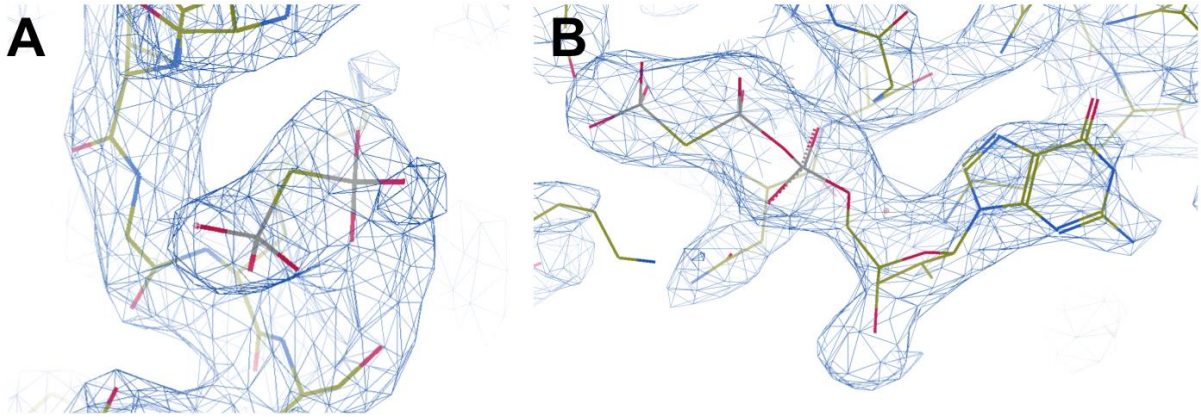

Fig. 71 Electron density map of molecule A from the asymmetric unit of Crystal 6. Crystal 7A is not represented, as it is identical. **(A)** Nucleotide-binding site of GD1, where only two phosphates from the nucleotide seem to be well stabilised. **(B)** Nucleotide-binding site of GD2, showing full occupancy of a GMPPCP. The electron density map represented in this image was observed with Coot..

In the second molecule in the asymmetric unit (chain B), the same full occupancy for GMPPCP was observed in GD2 for both Crystals 6 and 7 (as in Fig. 71B). However, the content of GD1 differs from the observed in chain A, and also between Crystals 6 and 7. This probably contributed for the bad correlation between datasets when merging them. In Crystal 7B, GD1 is in the apo state, without any ligands bound (Fig. 72A). In Crystal 6B, a GMPPCP molecule seems to be present, although the occupancy level is quite low and the B-factor high (Fig. 72B).

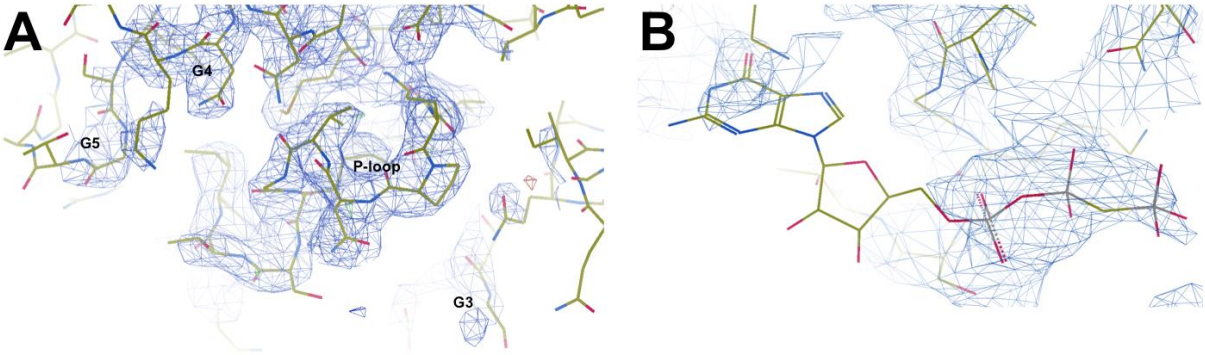

Fig. 72 (A) Electron density map of the GD1 nucleotide-binding site from Crystal 7, chain B. No electron density is present for a potential ligand. **(B)** Electron density map of the GD1 nucleotide-binding site of Crystal 6, chain B. A GMPPCP is present here, although the density is not perfect for a full-occupancy. The electron density maps represented were observed with Coot.

Molecule B from the asymmetric unit of Crystal 6 hence bears a GMPPCP in both G-domains. However, no change in the conformation was observed. So far, the 4DCU conformation was obtained in four different crystal forms: $P2_12_12_1$ (4DCU), $P2_1$ (Crystals 1-5) $P1$ (Crystals 6-7) and $C2$ – a crystal whose structure we were not able to refine due to poor resolution and completeness, but where molecular replacement gave a solution matching 4DCU. Hence, this conformation seems to be quite stable and prone to crystallise at the expense of other alternative conformations. We have analysed the crystal packing of EngA structures in each space group to search for crystal contacts that may be conserved.

3.4. Crystal packing and contacts analysis

We used the program PISA (Krissinel and Henrick, 2007) to examine the interfaces and protein-protein contacts between molecules in each crystal packing. As EngA has always been observed as a monomeric protein in solution, all contacts in the crystal are due to its packing and should not represent physiologically relevant, specific protein-protein interactions. We searched for large crystal contacts with an interface area above 800 $Å^2$.

We have started by analysing the packing of PDB 4DCU. 4DCU crystal ($P2_12_12_1$) contains one molecule in the asymmetric unit. Molecules in this crystal are related by symmetry operations and have the same packing environment. The largest packing contact has an interface area of 1233.2 Å^2. Fig. 73 represents the structures involved in this interface:

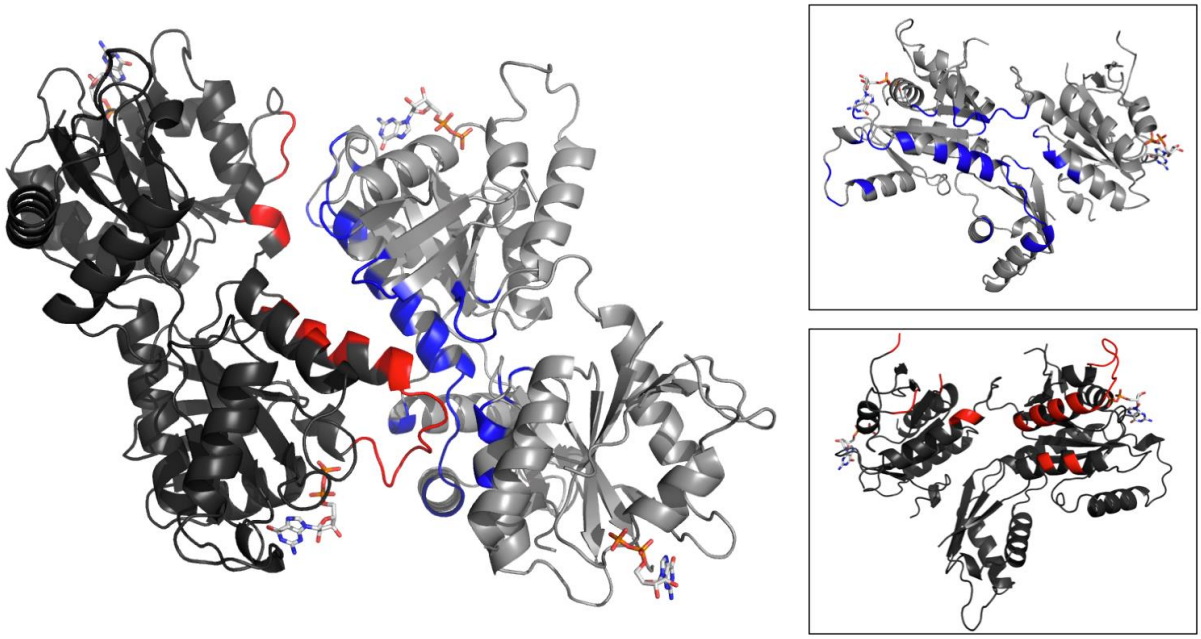

Fig. 73 Interface between two molecules of 4DCU related by crystallographic symmetry. The residues involved in this interface are coloured in red and blue. The two insets on the right present the interface region on both molecules towards the reader.

The nucleotide-binding sites are not involved in these contacts. In the reference molecule (represented in light grey/blue in Fig. 73), the interface involves mainly residues from the α6-helix of GD2 and the linker connecting the KH-domain (^{332}T-^{358}V). In the symmetry related molecule (dark grey/red), it involves mainly the loop that comprises the GD2-switch II and the following α3-helix (^{235}K-^{256}D). The contacts between these regions account for more than 40% of all contacts of the interface. The complete list of contacts in the interface is presented in Table 12.

The same crystal packing analysis was done for crystals in space groups $P2_1$ and $C2$ (Table 12). For all structures, the interface involves the same secondary structures of EngA as in 4DCU and the hydrogen/salt bonds are maintained with minor variations.

Table 12 Hydrogen (H) and salt (S) bonds involved in the interface of each crystal packing. (IA) is the interface area in Å^2. PISA analysis results shown here correspond to crystals 4DCU ($P2_12_12_1$), Crystal 4 ($P2_1$), Crystal 6 ($P1$) and one crystal in space group $C2$ where only molecular replacement was possible. (Ref) refers to the molecule taken as a reference and (Sym) to a symmetry related molecule. (*) indicates contacts from the GD2-KH region.

	***P*2$_1$2$_1$2$_1$**		***P*2$_1$**		***C*2**		***P*1**	
Bond	**Ref.**	**Sym.**	**Ref.**	**Sym.**	**Ref.**	**Sym.**	**Chain A**	**Chain B**
H	R109	E241	R109	E241	R109	E241	R109	E241
H			K111	Y246				
H	E173	K245	E173	K245	E173	K245		
H	E174	K245	E174	K245	E174	K245	E174	K245
H			E174	T243	E174	T243	E174	T243
H					V175	K245		
H	R200	D256	R200	D256	R200	D256	R200	D256
H	R200	Y281	R200	Y281	R200	Y281	R200	Y281
H	N221	R277	N221	R277	N221	R277	N221	R277
H	Q222	R277	Q222	R277	Q222	R277		
H	E259	Y246	E259	Y246	E259	Y246		
H*	K334	D52			K334	D52		
H*	S347	Y246	S347	Y246	S347	Y246	S347	Y246
H*							E348	R250
H*	S351	Y246	S351	Y246	S351	Y246	S351	Y246
H*	R353	Y240	R353	Y240	R353	Y240	R353	Y240
H*	Q355	K238	Q355	K238	Q355	K238	Q355	K238
H*	Q355	V239			Q355	V239		
H*			T356	Y240				
H	A420	K238			A420	K238	A420	K238
S							R109	E241
S							E174	K245
S					E199	K164		
S	R200	D256	R200	D256	R200	D256	R200	D256
S*					K334	D52	K334	D52
S*	E348	K235			E348	K235		
S*							E348	R250
IA (Å^2)	1233.2		1201.4		1278.6		1103.6	

Crystals 6 and 7, in space group *P*1, have two molecules in the asymmetric unit. A similar analysis was done to check for the interface between both molecules. As seen in Table 12, this interface corresponds to the same observed in the previous crystals. Fig. 74 shows a comparison of the content of the asymmetric unit of these crystals with two symmetry related molecules from 4DCU. As expected, the same large interface is present. The good superposition of these molecules suggests that small distortions in the crystal packing may have disrupted the symmetry found in 4DCU. However, these changes seem subtle.

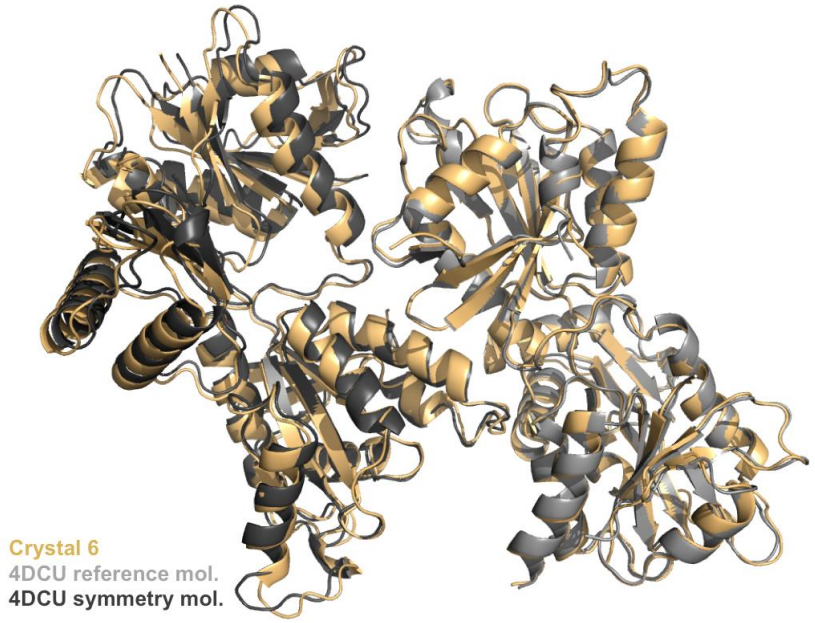

Fig. 74 Comparison of EngA structures from Crystal 6 (yellow) and two symmetry related molecules from 4DCU (dark/light grey). Superposition was done on one molecule of the asymmetric unit of Crystal 6 and the reference molecule from 4DCU.

This large interface and the interactions involved seem to be conserved in all crystals obtained so far in four different space groups. This may indicate that the formation of these contacts is favourable for crystallisation of EngA. Thus, the GDP-bound conformation may be selected during nucleation and crystal growth.

One interesting feature of this interface is the contribution of residues from α6-helix of GD2. This helix was detected as containing a tryptic cleavage site in limited proteolysis. This region seemed to be more protected from proteolysis in the GTP-bound state, which may suggest that it would not be available for forming crystal contacts with a neighbouring molecule. This may be one possible reason why the GTP-bound conformation is more difficult to crystallise.

3.5. Analysis of structural changes

3.5.1. Structures alignment

To compare the different EngA structures and investigate global or local conformational changes, we used the program ALIGN (Satow et al., 1986) that superposes structures by aligning the Cα atoms.

All structures from Crystals 1–7 were superposed, including the two molecules present in the asymmetric unit of Crystals 6 and 7. The PDB structure 4DCV was also included, which is

the only available EngA PDB entry with a triphosphate nucleotide in GD2 (here, GD1 is empty). PDB entry 4DCU was used as the reference molecule for the alignment.

For alignment of G-domains, the flexible loops (P-loop, switch I, switch II, G4 and G5) were removed so that only the structurally conserved regions are taken into account.

3.5.2. Overall conformation

SAXS and limited proteolysis show a nucleotide-induced change in the global conformation of EngA. This change has not been observed in any of the obtained structures of EngA, even though the content of the nucleotide-binding site varies between GDP, GTP analogues, sulfate/phosphate ions or the apo-form. Phasing by molecular replacement indicated a good match with 4DCU, i.e., the GDP-bound state. We have superposed all structures by their GD2 domain. The alignment is represented in Fig. 75, where an identical tertiary structure is observed for all structures.

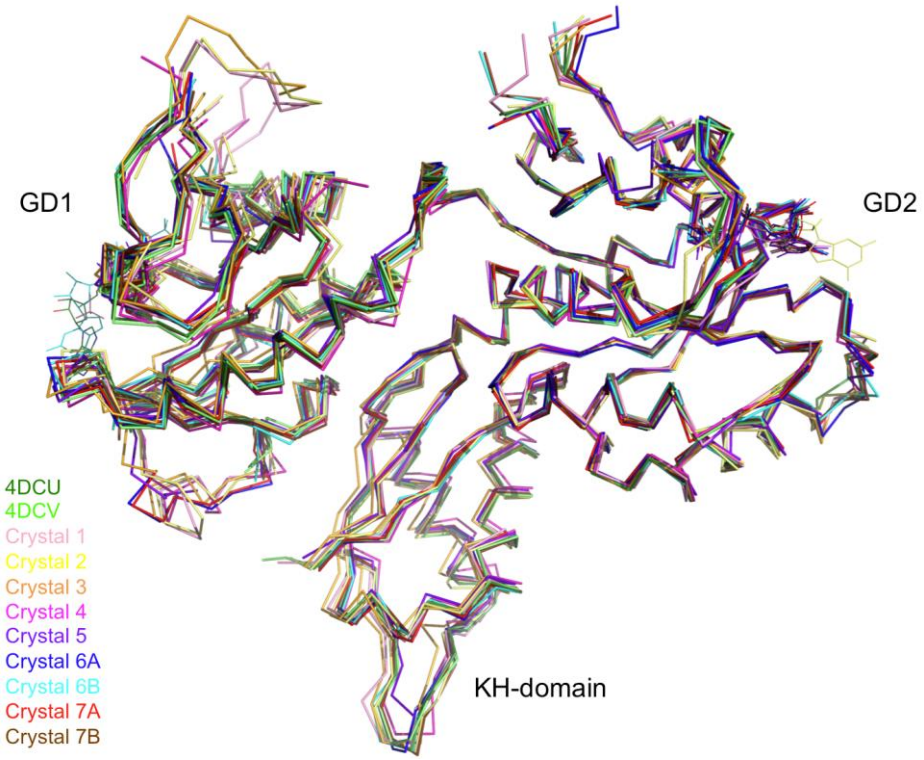

Fig. 75 Superposition of EngA structures 4DCU and 4DCV from the Protein Data Bank and our refined structures Crystals 1–7. Alignement was done on GD2.

We computed the root mean square deviation (rmsd) between atomic coordinates of the sequence-aligned structures in order to have a quantitative measure of their similarity. The rmsd were calculated on 382 Cα and are presented in Table 13.

Table 13 Rmsd (Å) values of the eleven structures of EngA after superposition on the GD2.

	4DCV	C1	C2	C3	C4	C5	C6A	C6B	C7A	C7B
4DCU	0.435	0.828	0.946	0.963	0.965	0.790	0.680	0.674	0.632	0.624
4DCV		1.026	1.116	1.003	1.104	0.982	0.913	0.855	0.847	0.798
C1			0.712	1.080	1.087	0.920	0.896	0.900	0.831	0.854
C2				1.082	1.216	1.056	1.013	0.970	0.961	0.977
C3					1.364	0.968	1.065	0.973	1.002	0.982
C4						1.190	0.911	1.077	0.956	1.040
C5							0.735	0.826	0.731	0.806
C6A								0.671	0.434	0.674
C6B									0.639	0.499
C7A										0.590

For analysis of similarities, a large rmsd value indicates dissimilarity and zero indicates identical conformations. The lowest rmds obtained was 0.435 Å (4DCU / 4DCV), while the highest value was 1.364 Å (Crystals 3 / 4). Comparison of a large number of structure pairs obtained from identical proteins in the PDB shows that rmsd values fluctuate between 0–1.2 Å due to inherent protein flexibility and experimental resolution limits (Kufareva and Abagyan, 2012). Hence, our values are within the acceptable range for equivalent atom pairs and no significant differences seem to be related to the nucleotide content. Alignment of the G-domain of Era in the GDP- and GTP-bound forms, for example, shows a rmsd of 11.1 Å between both structures, with the largest movements seen in switch I and II loops. The maximum displacement seen in GD1 and KH domains when the GD2s from our structures are superposed is <2 Å, confirming that this is a very rigid conformation, with only minor movements induced by the crystal packings.

3.5.3. Structural analysis: GD2

Whereas the global conformation of EngA remains the same, some regions show differences, in particular flexible loops. The several structures of EngA aligned by their GD2 were compared in order to understand if these differences are related to the ligand.

4DCU and Crystal 2 both have a GD2:GDP, while Crystal 3 has a SO_4^{2-} ion in the position of the β-phosphate. The backbone of the P-loop shows no major deviations for these three structures, except for P184 and G187 (^{182}GRPNVGKS^{189}) (Fig. 76). G187, which forms contacts with the α-/β-phosphates, shows a movement of 0.8 Å between 4DCU and Crystal 2, which is probably due to the movement of the α-phosphate of the GDP molecule in the same direction (0.7 Å). The movement of 1.0 Å of P184 in Crystal 3 follows the main-chain NH group of N185 whose orientation adapts to the position of SO_4^{2-} (which slightly precedes the GDP β-phosphate) to form contacts with the ion.

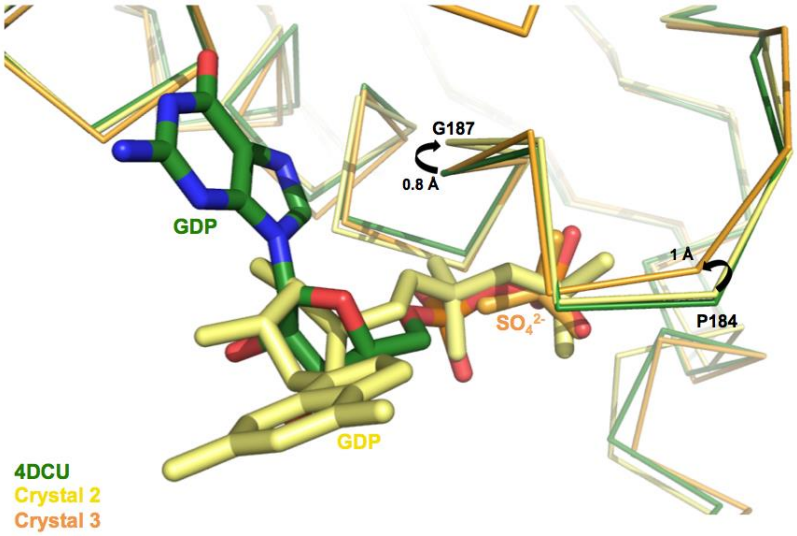

Fig. 76 Close view of the GD2 P-loop from structures 4DCU (green; GDP-bound), Crystal 2 (yellow; GDP-bound) and Crystal 3 (orange; SO_4^{2-}-bound). The presence of a GDP in a non-classical position in Crystal 2 induces some changes in glycine 187 that moves to create place for the α-phosphate. In Crystal 3, the sulfate ion, placed near the β-phosphate, induces movement of the proline 184 for formation of contacts, where the loop adopts a more closed conformation.

Comparison of 4DCU and 4DCV, where GD2 has a bound GDP and GMPPCP, respectively, shows only minor differences in the backbone of the P-loop (0.1–0.3 Å) (Fig. 77A). The presence of a γ-phosphate in 4DCV does not seem to trigger any change in the P-loop. On the other hand, superposition of Crystals 1, 4, 5, 6A, 6B, 7A and 7B, which all have a triphosphate nucleotide in GD2, shows significant deviations (Fig. 77B).

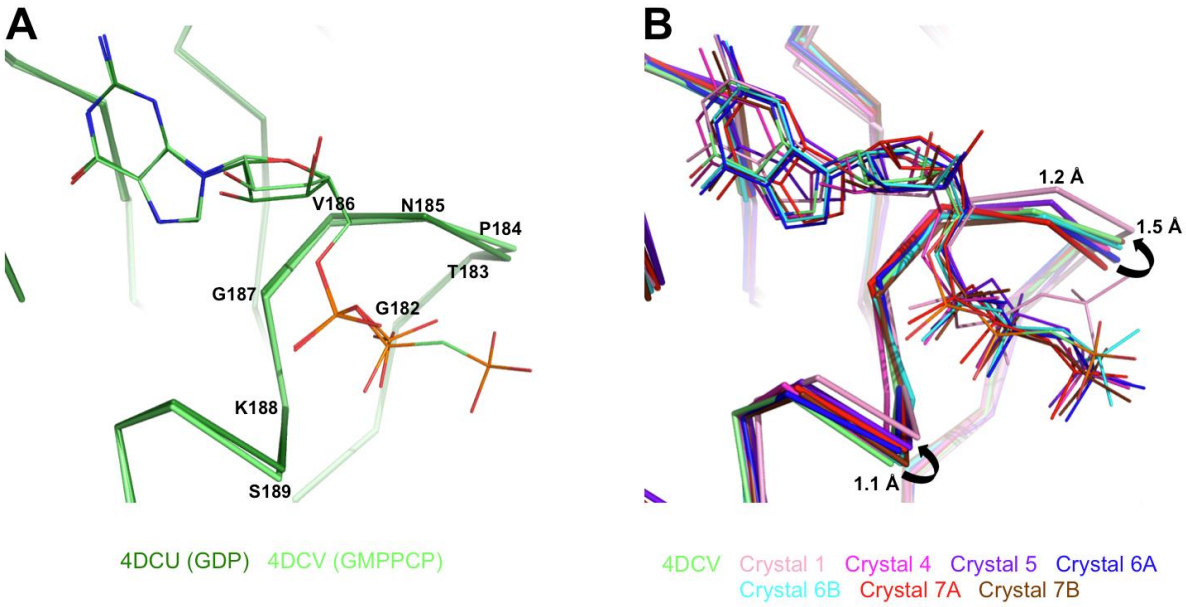

Fig. 77 (A) Alignment of the GD2 of PDB entries 4DCU (dark green) and 4DCV (light green). The presence of a GDP or a GTP analogue does not induce major changes in the P-loop. **(B)** Comparison of the different GMPPN/CP-bound structures. The γ-phosphate adopts different relative positions, inducing movements of the P-loop.

Residues from the P-loop display variations of 0.8–1.5 Å between each structure. The largest deviations are seen in residues P184, N185 and S189, which move 1.5, 1.2, 1.1 Å, respectively. Inspection of the γ-phosphate in each structure reveals that the position adopted has significant variations. These tend to diminish from the β-phosphate until the guanine base. The movements of the γ-phosphate are accompanied by the displacement of S189, which contacts the β-phosphate, and P184/N185, whose movements create place for the β-/γ-phosphates. In particular, the γ-phosphate of GMPPNP in Crystal 1 (Fig. 77, light pink) is at 2.5 Å distance from the other structures, and presents the largest movement of the P-loop.

These different orientations of the γ-phosphate also create deviations in the switch II region. The loop from ^{229}DTAGMRKKGKVYET^{242} seems to become destabilised by the presence of the γ-phosphate, showing deviations between 2–3 Å between structures (Fig. 78), while variations in the GDP/SO_4^{2-}-bound structures are around 1–1.5 Å. K236 hydrogen-bonds with the γ-phosphate in 4DCV, although in all other structures it is positioned at a distance of 4.4–5.9 Å away from the nucleotide (inset of Fig. 78). Also, G232 from the G3 motif, which in other GTPases is seen to stabilise the γ-phosphate and to participate in the hydrolysis, is not in a position to form contacts with the nucleotide in any of the structures (inset of Fig. 78).

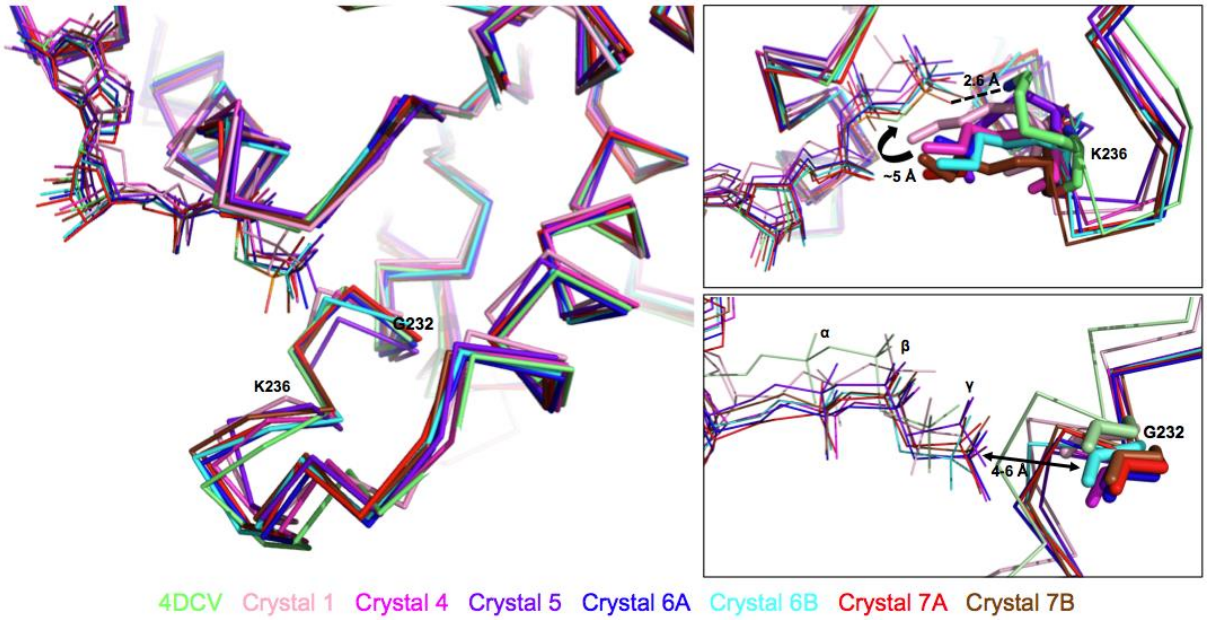

Fig. 78 Close view of the GD2-switch II region of the several GMPPN/CP-bound structures. This region shows larger displacements (2–3 Å) when compared with ion-/GDP-bound structures (1.0–1.5 Å; not shown).

The G4 motif in GD2 is present in all EngA crystal structures and is structurally well-conserved, independently of the ligand bound in the nucleotide-binding site. Although slight deviations can be observed in this loop (up to 0.7 Å), both N294 and D297 point towards a

position that would stabilise the nucleotide base (Fig. 79). This position is even maintained in Crystal 2, where the GDP base is not on its conserved binding site; and in Crystal 3, where a sulfate ion is present in the nucleotide-binding site.

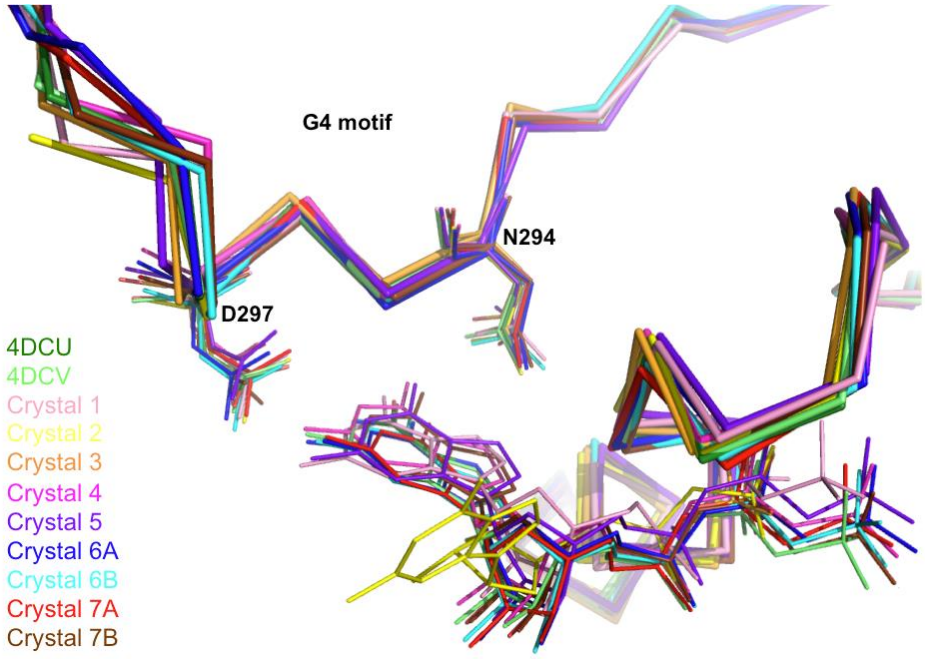

Fig. 79 Close view of the GD2-G4 motif. In all our obtained crystal structures this loop is well structured and maintains the same conformation towards the stabilisation of the nucleobase.

3.5.4. Structural analysis: GD1

The same alignment was done on GD1s of EngA structures. Comparison of the three structures GDP-bound (4DCU) and apo (Crystal 7B and 4DCV) shows no major differences in the P-loop (^{10}GRPNVGKS^{17}), with the exception of G15, which adapts to the absence/presence of the β-phosphate to form hydrogen-bonds (Fig. 80A). The two structures containing a pyrophosphate (Crystals 6A and 7A) also show a good superposition with 4DCU. The largest differences are present on residues S17 and N13, which adjust to form contacts with the β-phosphates that are placed in slightly different positions in the three structures.

The structures containing PO_4^{3-}/SO_4^{2-} in GD1 show large variations on the position of the ions. Despite their close position to the β-phosphate, in some structures we can observe a displacement of up to 0.6 Å (Crystals 1 and 5). Deviations in the P-loop between 0.8–1.3 Å follow the positions of the ions (Fig. 80B).

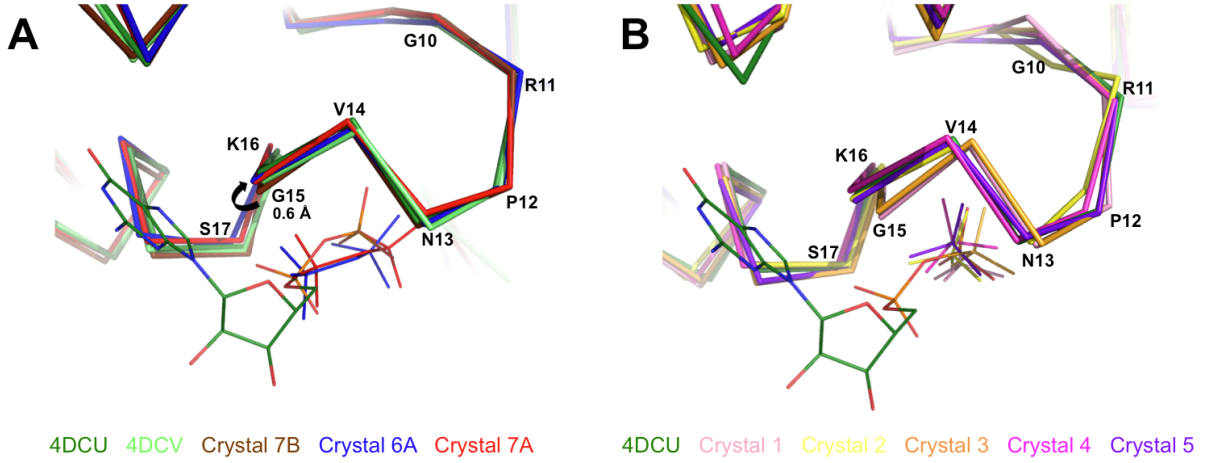

Fig. 80 Close view of the GD1 P-loop. **(A)** The conformation of the P-loop does not change significantly when the GD1 is either in the apo, GDP-bound or pyrophosphate-bound forms. **(B)** The presence of sulfate/phosphate ions around the position of the β-phosphate induces movements of the P-loop.

Due to its high flexibility, the GD1-switch II region is not seen in most *B. subtilis* EngA structures, with the exception of PDB entry 4KYU (corresponding to the structure of GD1 alone). In GTPases, the conserved Asp residue in the G3 motif stabilises the Mg^{2+} ion, while the conserved Gly residue stabilises the γ-phosphate. During refinement of Crystal 1, we were able to observe a well-defined loop corresponding to the switch II region. However, superposition of GD1 from Crystal 1 with 4KYU and 1WQ1 (GTPase Ras containing GDP-AlFx-Mg in the nucleotide-binding site) showed that the conserved D57 is not in a position towards the stabilisation of the magnesium (Fig. 81). Additionally, the G60 is not structurally conserved in EngA structures, presenting a large mobility in different structures.

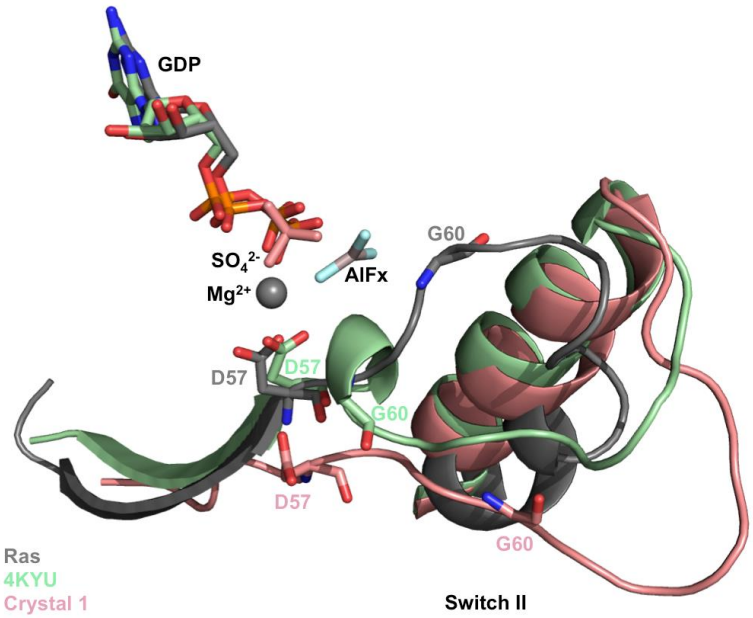

Fig. 81 Close view of the switch II region of GD1. Crystal 1 (light pink) is the only structure where we were able to built a well-defined switch II region. When comparing to other GTPase structures where G3 motif is structured, a deviation in this loop can be observed for our structure, where contacts with the nucleotide/magnesium ion are missing.

In EngA, the GD1-G4 loop is very flexible and is not present in most structures. The only PDB available structure showing the G4 motif corresponds to 4KYU. In Crystals 1, 2, 4 and 5 we were able to obtain a structured G4 loop. Although this loop is structurally conserved in all four crystal structures, with no major deviations, its position does not seem to represent a stabilised form of this region. The conserved Asp residue (D122), which in 4KYU is pointing towards the nucleobase and is stabilising it, is at 10–11 Å away from the guanine base in our crystal structures (Fig. 82).

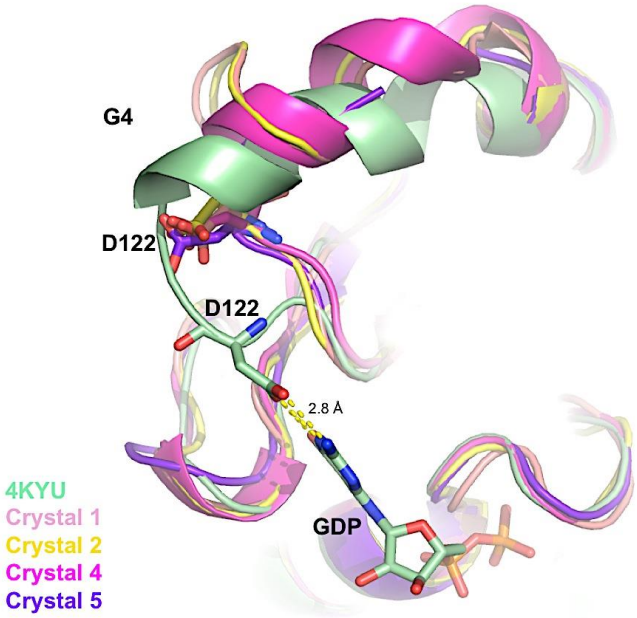

Fig. 82 Close view of the GD1-G4 motif. This loop presents high flexibility and is not visible in the electron density of most EngA structures, with the exception of PDB entry 4KYU. In four of our refined crystals (1, 2, 4 and 5) we were able to observe the electron density and build the G4 loop. In these four crystal structures, the G4 is structurally conserved. However, analysis of contact shows that its conformation (namely that of the D122 residue) is not in a position towards stabilisation of a nucleotide.

The GD1-G5 motif (^{145}SGT^{147}) is also destabilised in the apo-/ion-/pyrophosphate-bound structures, possibly due to the lack of contacts with the nucleobase. Residues S145, G146 and T147 show distances of 3.0, 2.7 and 2.9 Å, respectively, between different structures. Comparison of 4DCU (GDP) with Crystal 6B (GMPPCP) shows a shift in the relative position of the nucleotides (Fig. 83). Despite this shift, the main contacts between the guanine base and the G5 motif do not change.

This is the first time that we see a GTP analogue bound to the GD1 of EngA. In this structure, the γ-phosphate is stabilised by residue Q75 from the switch II loop (Fig. 83). This observation corroborates activity measurements performed by Anne-Emmanuelle Foucher (IBS, Grenoble) in EngA mutants, where a variant Q75A showed half of the activity of the wild-type protein. This may suggest a catalytic role for this glutamine in EngA.

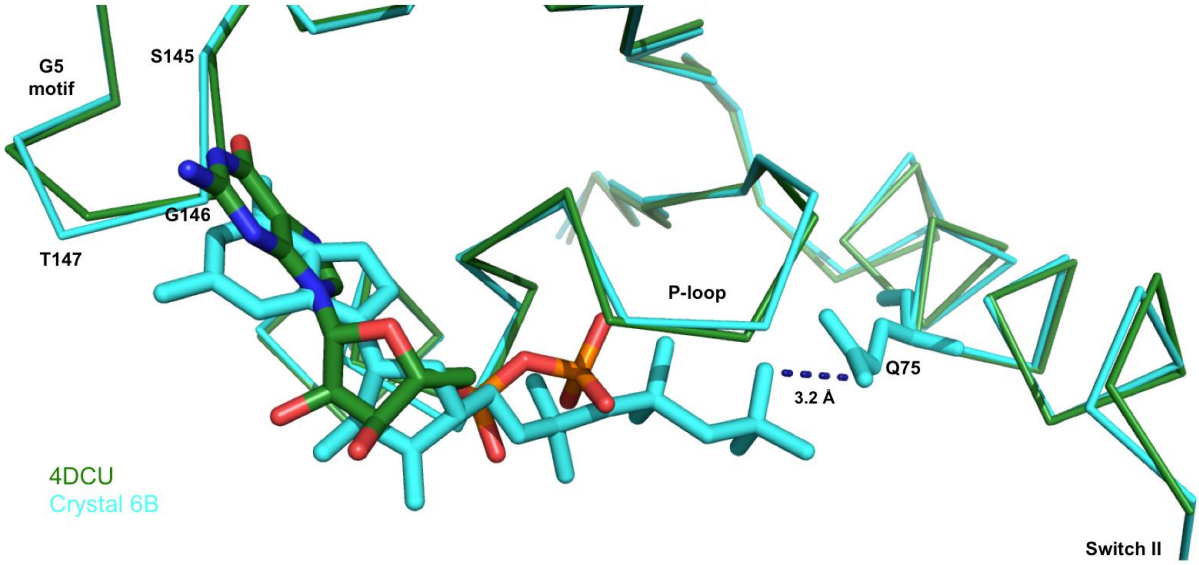

Fig. 83 Alignment of the GD1 from 4DCU (green; GDP-bound) and Crystal 6B (cyan; GMPPCP-bound).

The small changes observed in the nucleotide-binding site seem to be associated with the presence of a ligand in the nucleotide-binding site. These ligands induce small movements in flexible regions for the peptide chain to adapt to their presence. However, with the exception of GD2-switch II, these structural changes do not seem to be correlated with the nature of the ligand (either di-, tri-phosphate or ions).

4. Interaction studies

While the exact function of EngA in ribosome biogenesis has not yet been determined, different studies have demonstrated interactions with the 70S and individual 50S and 30S subunits in cells (Agarwal et al., 2012; Bharat et al., 2006; Hwang and Inouye, 2006; Tomar et al., 2009) and *in vitro* (Agarwal et al., 2012; Schaefer et al., 2006; Tomar et al., 2009; Zhang et al., 2014). These studies indicated that binding of EngA to the ribosome occurs in a nucleotide-specific manner. Distinct affinities over the ribosome triggered by binding of a nucleotide have been suggested to arise from conformational changes that regulate the affinity for the effector.

These previous interaction studies were carried out using 1 mM nucleotides. Surprisingly, our data on the structural analysis of EngA have shown that a concentration of 10 mM GTP analogues is necessary to trigger a switch to the active GTP-bound conformation.

We have thus performed interaction studies to further assess how conformational changes and ribosome-binding events are related. We speculated that the same conditions that induce a change in conformation should enhance binding to the ribosome. To confirm this, the information obtained from our analysis of conformational changes on EngA was applied in interaction studies.

Different strategies to study the interaction between EngA and ribosomal elements were tested. These different set-ups were also optimised with the aim of establishing an assay for screening inhibitor molecules of the interaction EngA/bacterial ribosome.

4.1 EngA and the bacterial ribosome

4.1.1 ELISA assay

An ELISA assay was setup in order to have a reliable assay to detect EngA interactions with the bacterial ribosome. Ribosomes were used to coat a 96-well plate and His-EngA_{WT} was added and probed with a conjugated anti-polyhistidine antibody.

To examine the nucleotide-dependent binding of EngA to the ribosome, the interaction was probed in the absence and presence of GDP and GTP analogues. Based on our results from SAXS, where a change in conformation was observed at 10 mM GTP analogues, different nucleotide concentrations were tested in the ELISA assay. According to the literature, the GTP-bound state of EngA is prone to bind the 70S ribosome. We speculated that, if a

change in conformation does relate with EngA binding ability, an increased affinity would be observed in the presence of 10 mM GTP analogues.

The interaction curves represented in Fig. 84 show a nucleotide-dependent binding of EngA to the 70S ribosome and confirm the effect of the conformational change observed by SAXS and limited proteolysis on the binding ability.

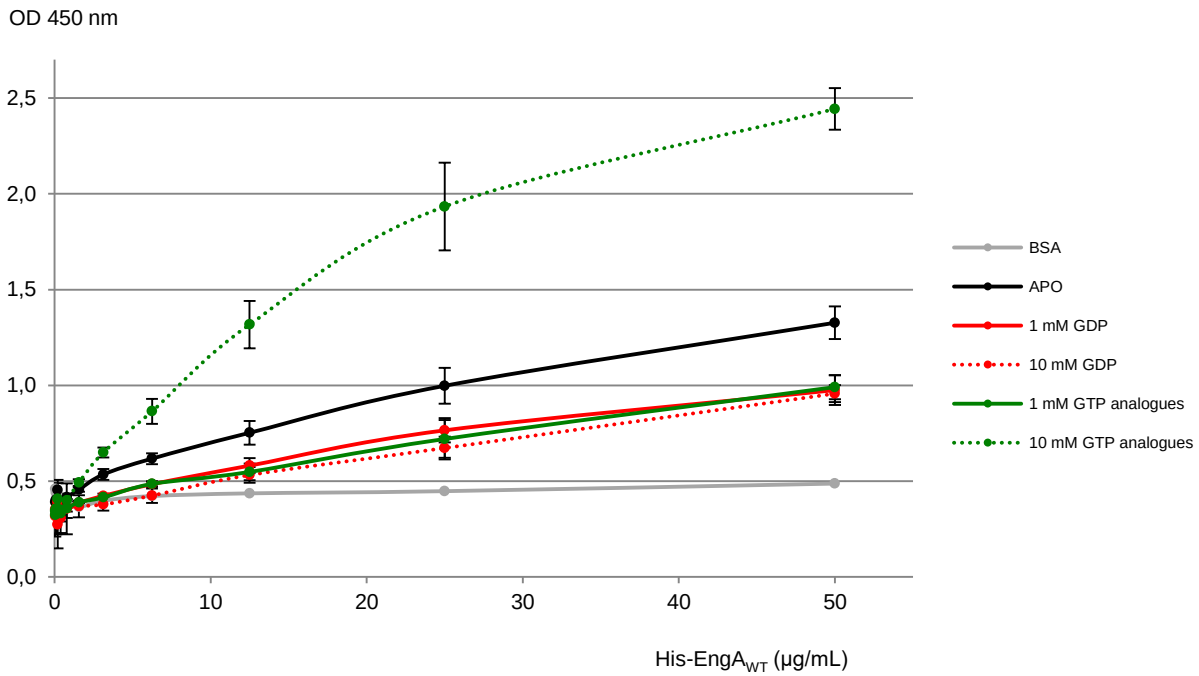

Fig. 84 Interaction between His-EngA and 70S ribosome measured by ELISA assay. The purified *B. subtilis* 70S ribosome was used to coat a 96-well plate. EngA was incubated with different nucleotides prior to addition to the coated wells. "GTP analogues" refers to samples tested with GMPPNP and GMPPCP, which gave identical results and thus were analysed as one. OD_{450} values were plotted against EngA concentration (µg/mL). Each point corresponds to the mean of independent sets of results and error bars represent the standard error where *n* varies between 2 and 6 between the different samples.

According to the obtained results, EngA is able to bind the 70S ribosome in the apo-state and the interaction is slightly destabilised by 1 mM GDP. This inhibitory effect of GDP on the interaction between EngA and the ribosome had already been observed in previous studies (Hwang and Inouye, 2006; Schaefer et al., 2006; Tomar et al., 2009).

The presence of 1 mM GTP analogue provides a similar interaction curve to that of GDP. This suggests that 1 mM triphosphate nucleotide is not enough to induce a change in the affinity of EngA over the ribosome. If we expect a conformational change to be related to an increased affinity for the ribosome, these results seem to be consistent with SAXS, where 1 mM GTP analogues did not trigger any conformational change. To confirm a relation between a GTP-induced conformation and ribosome-binding ability, we tested the interaction in the presence of 10 mM nucleotides.

Increasing GTP analogues concentration to 10 mM greatly enhances the binding of EngA to the ribosome. This confirms that the conformational change observed by SAXS and limited proteolysis for EngA is associated with its ribosome binding capability. Varying the concentration of GDP between 1 and 10 mM does not seem to affect the affinity for the ribosome. This suggests that the increase in binding observed with 10 mM GTP analogues is specific of the triphosphate nucleotide and not due to a side-effect of high nucleotide concentrations in solution.

We have thus confirmed the need of 10 mM GTP analogues to induce a change in conformation and an effect in binding ability.

The development of this ELISA assay also provides a consistent tool for the screening of inhibitors. The experiments carried within this scope are presented in *Results, section 5*.

4.2 EngA and individual ribosomal subunits

4.2.1 Sucrose density gradient ultracentrifugation

Following our interaction studies with the 70S ribosome, we have decided to disrupt the association of ribosomal subunits and assess interactions with individual 50S and 30S. The interaction of EngA with both ribosomal subunits had been previously shown by sucrose gradient ultracentrifugation experiments (Agarwal et al., 2012; Bharat et al., 2006; Hwang and Inouye, 2006; Tomar et al., 2009).

The western-blotting represented in Fig. 85 shows that in the absence of nucleotides, the apo-EngA cofractionates with the 50S subunit. A band is visible in the peaks corresponding to the large ribosomal subunit, indicating association of EngA. As EngA was used in large excess (see *Material and Methods*, *section 6.2*), high amounts of unbound EngA remain on top of the gradient. A band corresponding to the peak of the 30S subunit is also visible in the membrane. However, this band may correspond to unbound EngA that seems to diffuse from the top fractions. Thus, an interaction with the small ribosomal subunit could not be established with confidence.

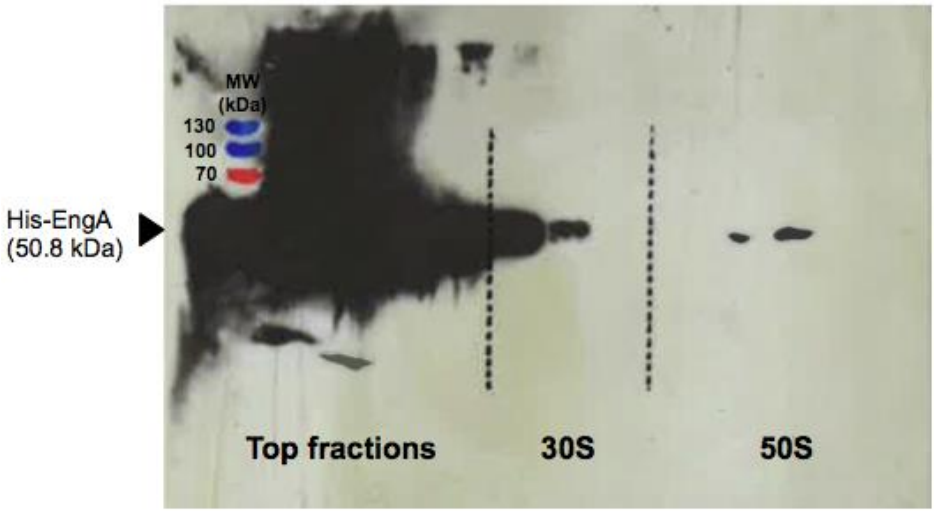

Fig. 85 Western-blot analysis of the fractions obtained from centrifugation on a 5–20% sucrose gradient of a mixture of His-$EngA_{WT}$ (apo) and dissociated 70S ribosome. Bands from the molecular weight marker are represented in blue/red. The fractions were collected from the top to the bottom of the cushion and applied onto an SDS-PAGE. EngA was detected by immunoblotting using an anti-polyhistidine antibody. EngA that is not associated to the ribosome is present in the top fractions. EngA was detected in the peak corresponding to the 50S subunit, indicating it cofractionates with the large ribosomal subunit. Although a band was also detected in the peaks corresponding to the 30S subunit, the high amounts of unbound EngA present in the top fractions that seem to diffuse in the top of the gradient do not allow us to conclude whether it corresponds to dissociated or 30S-associated EngA.

Unfortunately, this interaction could not be probed with 10 mM nucleotides, as preparation of such gradients would require infeasible amounts of nucleotides (~60 mg/ gradient). Due to this limitation, other methodologies were instead applied to assess interactions of EngA with individual ribosomal subunits. The following results represent preliminary data on these interactions.

4.2.2 ELISA assay

The same ELISA assay applied for probing the interaction EngA:70S was employed for the 50S and 30S ribosomal subunits. We have done a preliminary trial using the apo-protein. Different concentrations of coated subunits and of EngA as the ligand were tested. The interaction curves in Fig. 86 show a specific binding of EngA to the 50S and the 30S in a concentration-dependent manner.

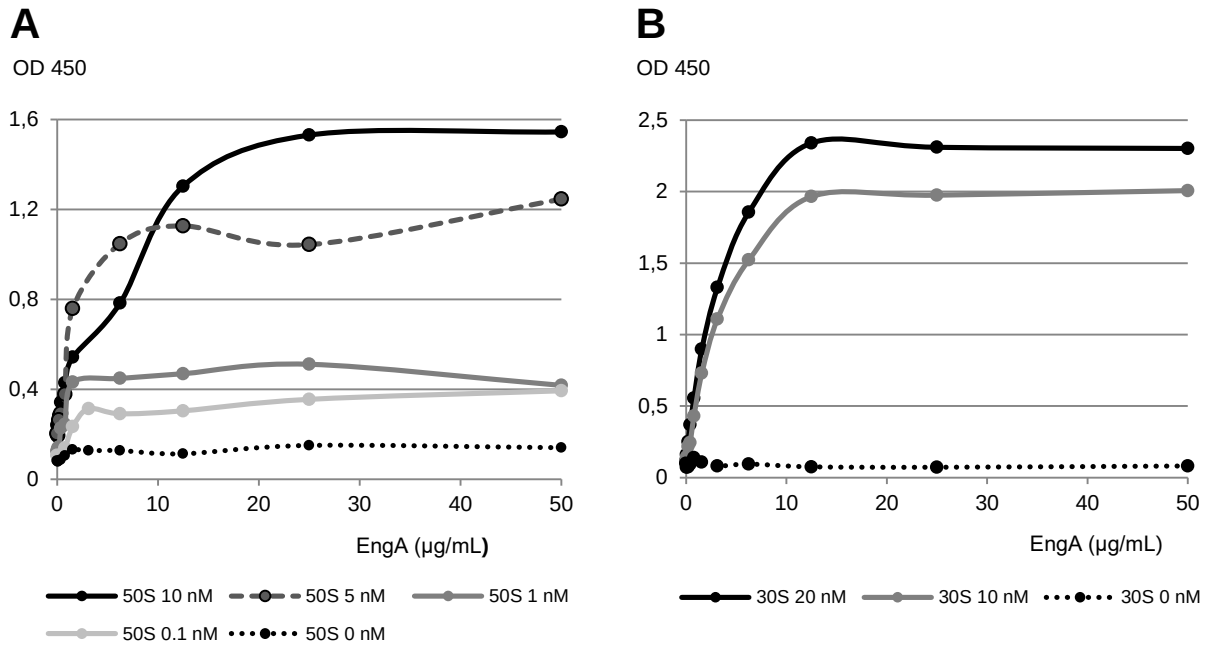

Fig. 86 Interaction curves of His-EngA_{WT} (apo) with the **(A)** 50S subunit and the **(B)** 30S subunit. The purified *B. subtilis* ribosome subunits were used to coat a 96-well plate and associated His-EngA_{WT} was probed with an anti-polyhistidine antibody. OD_{450} values were plotted against EngA concentration (µg/mL). The antibody was confirmed not to bind ribosomal subunit. Each point corresponds to a single measurement from a preliminary experiment.

Although this observation is based on preliminary experiments, the ELISA assay seems to be a reliable tool for assessing the formation of EngA:50S and EngA:30S complexes. Several set-ups are now being implemented:

- The binding of EngA with 10 mM nucleotides is being tested in order to assess if the same effect observed for the ribosome is also observed for the individual subunits.

- The binding of the cross-linked mutants Eng$A_{G139C/V388C}$ and Eng$A_{Y51C/T169C}$ to the individual subunits is being assessed. The GV variant blocks the protein in the GDP-bound state. As GDP has an inhibitory effect on the binding of EngA to the ribosome, we speculate that the interaction should be abolished. Although the presence of GDP seems to allow a certain level of binding events to still occur (from observations from the ELISA assay 70S:EngA), these interactions may be due to an intrinsic dynamic switch between different conformations of EngA in solution. Having all particles blocked in the same conformation should clearly indicate whether the GDP-bound conformation of EngA can bind the ribosome or one of its subunits. Since the structure of the YT variant in solution cannot be determined and differs significantly from the observed in 1MKY crystal structure, no reliable prediction can be made on the binding to the ribosome. However, it may still be interesting to test its binding to the ribosome in order to assess the impact of blocking EngA in another conformation.

4.2.3 Bio-layer interferometry (BLI)

The Octet Red96 system (ForteBio) was used to assess the interaction of EngA and the large subunit and possibly determine kinetic parameters. This technique requires that one interaction partner is captured at a surface, while the other is free to bind in solution.

The kinetics workflow of a BLI experiment is depicted in Fig. 87. After capturing the molecule on the biosensor surface, the analyte is added in solution. During a measurement, an association phase corresponds to binding of the analyte, while in the dissociation phase the analyte is washed away from the surface. Kinetic constants can be obtained from a titration series of such analyte measured against an immobilised ligand.

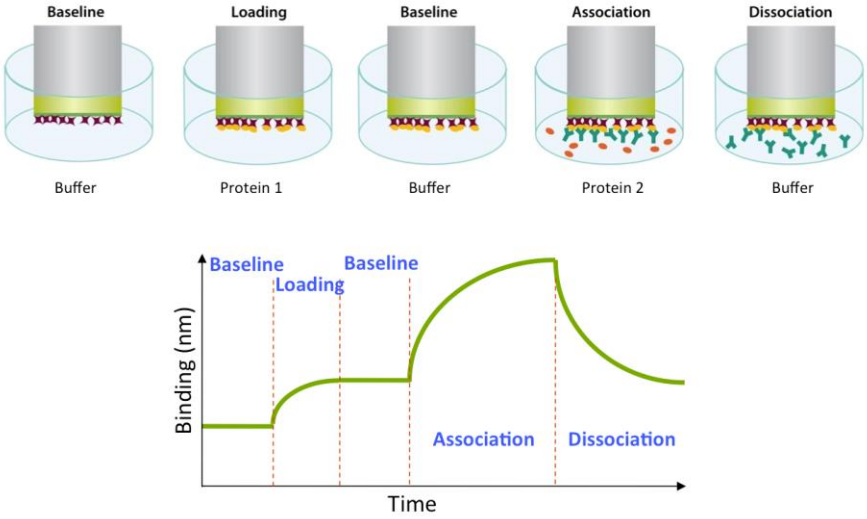

Fig. 87 Kinetics workflow of a BLI experiment. An initial step of equilibration in buffer is done to set the baseline. The first molecule – the ligand – is immobilised on the biosensor (loading). A new baseline is set with the buffer from the next sample. Addition of the second molecule – the analyte – leads to an association phase, where the analyte binds to the ligand. Transfer of the biosensor to an analyte-free solution will cause dissociation from the ligand. Adapted from http://www.fortebio.com/bli-technology.html.

We have tested two set-ups, immobilising either EngA or the 50S.

For immobilisation of EngA, two biosensors capturing the His_6 tag were tested, Ni-NTA and HIS1K. Capture with the HIS1K biosensor was not stable and a strong release of EngA occurred during the second baseline, association and dissociation steps (Fig. 88). In comparison with HIS1K, the Ni-NTA biosensor allowed a strong capture of EngA. However, when adding a solution of 50S, no interaction could be detected (not shown).

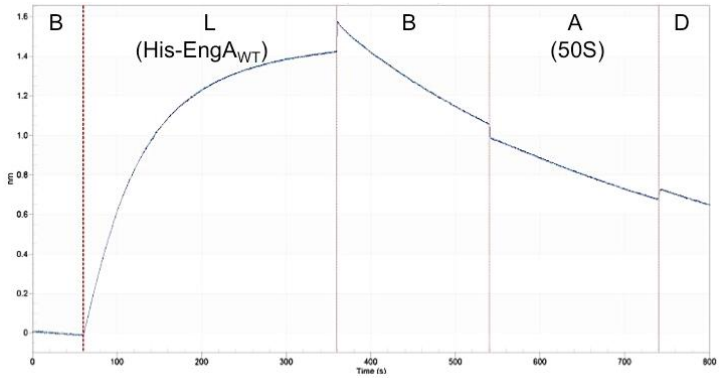

Fig. 88 Sensorgram obtained from immobillisation of His-EngA$_{WT}$ on a HIS1K biosensor. After the loading step (L), a strong release of EngA occurs during baseline (B), association (A) and dissociation (D).

One possible reason for the absence of binding could be that the concentration of 50S used in this assay was not enough for the affinity of this interaction. Higher ribosome concentrations, as well as the presence of a GTP analogue, that increases EngA affinity for the ribosome, should have been tested. As we were limited by the 50S concentration, we decided to invert the set-up and immobilise the large subunit.

As no tag was present in the ribosome, an initial step of biotinylation had to be done. Biotinylated-50S were immobilised onto a streptavidin biosensor and a range of concentrations of EngA was injected. The capture of 50S in the biosensor was very low. Different reasons may have contributed to low loading, such as an incomplete biotinylation process, low ribosome concentrations or an inadequate loading buffer. These parameters should be inspected and optimised in the future in order to increase immobilisation of 50S on the sensor. Addition of EngA showed a weak, dose dependent signal. However, based on the low immobilisation of 50S, this may possibly be due to a non-specific binding of EngA to the biosensor.

Due to limitations in time, as this experiment was part of a ForteBio demonstration of the BLI technology, we were not able to optimise this assay. Nevertheless, this can be a promising approach to perform a kinetic characterisation of the interaction 50S:EngA and afterwards screen inhibitors.

4.2.4 Cryo-electron microscopy (cryo-EM)

The structure of *E. coli* EngA associated to the 50S subunit was recently solved by cryo-EM (Zhang et al., 2014). This structure shows EngA bound to the PTC in a different conformation from the ones previously obtained by crystallography for the GDP-bound and the pseudo GTP-bound states.

Our SAXS data indicates that the GTP-bound conformation of *B. subtilis* EngA does not match the one determined by this structure (PDB 3J8G). Different reasons have led us to try to investigate this observation. A possibility exists that the GTP-induced conformation and the ribosome-bound conformation are not the same. While we have seen that the binding of GTP triggers a conformational change in EngA increasing the affinity for the ribosome, we cannot exclude the hypothesis of another conformation being adopted upon binding. Another aspect of this structure to consider is that despite the good global resolution of the cryo-EM structure, the local resolution achieved in the EngA binding-site (around 15 Å) does not allow the inspection of the details of the EngA at the atomic level nor the nucleotide content. Lastly, this structure corresponds to the *E. coli* EngA, whereas we study the *B. subtilis* ortholog, and differences between both proteins may exist.

Hence, having identified the right conditions to trigger a conformational change and maximise binding of EngA to the ribosome, we were interested in investigating the structure of the EngA:50S complex of *B. subtilis*.

Quality control analysis of the 50S subunit was done by negative staining. The images represented in Fig. 89 show that, despite the presence of few aggregates, the preparation of the large ribosomal subunit is rather homogeneous and can be used for further EM analysis.

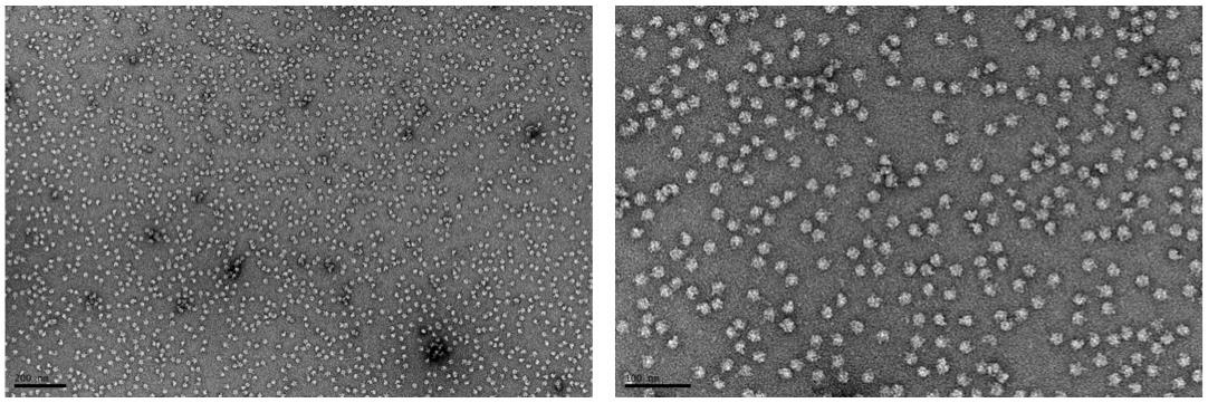

Fig. 89 Negative staining images of the 50S subunits at a magnification of 11000x (left) and 30000x (right).

After determining the adequate sample concentration for cryo-EM analysis, the complex 50S:EngA was prepared and several grids were analysed (see *Material and Methods, section 5.7*).

A first analysis performed with images recorded on the Tecnai Polara microscope revealed a highly preferred orientation of the 50S particles (Fig. 90). This distribution led to a lack of information in several directions, which limited the maximum achievable resolution in these directions, and hence did not allow us to determine whether EngA was bound.

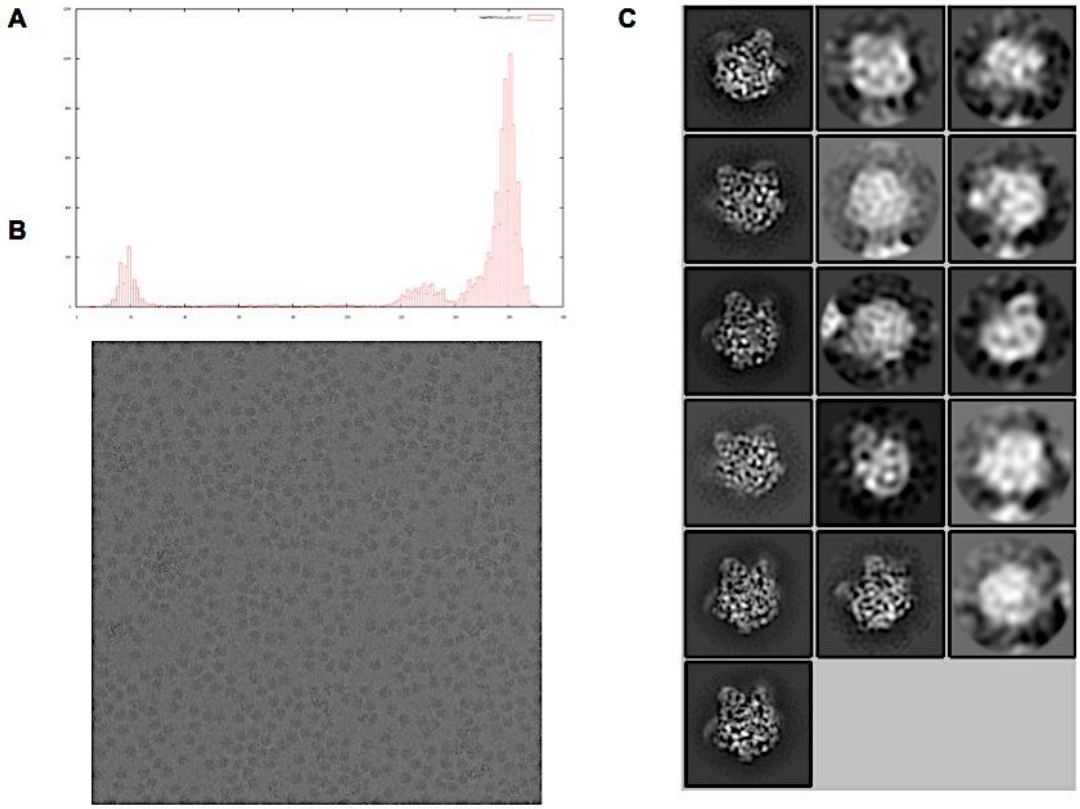

Fig. 90 (A) Distribution of the particles' orientation. In these images, the 50S:EngA particles show a preferential orientation. **(B)** One of the analysed images of the 50S:EngA complex obtained at 20000x magnification. **(C)** Classification of particles into different classes. Some clusters show a lot of details (meaning a good resolution) due to a great number of particles adopting the same orientation. However, for the missed orientations the resolution is very low.

Two additional grids were observed and the corresponding recorded images have been analysed following the same procedure. Although these grids were prepared concomitantly to the previous one, no problem of particles orientation has been observed this time. 140 images were collected and a total of 57545 particles were selected. A 2D *ab initio* classification grouped 54198 of these particles into 40 classes (Fig. 91).

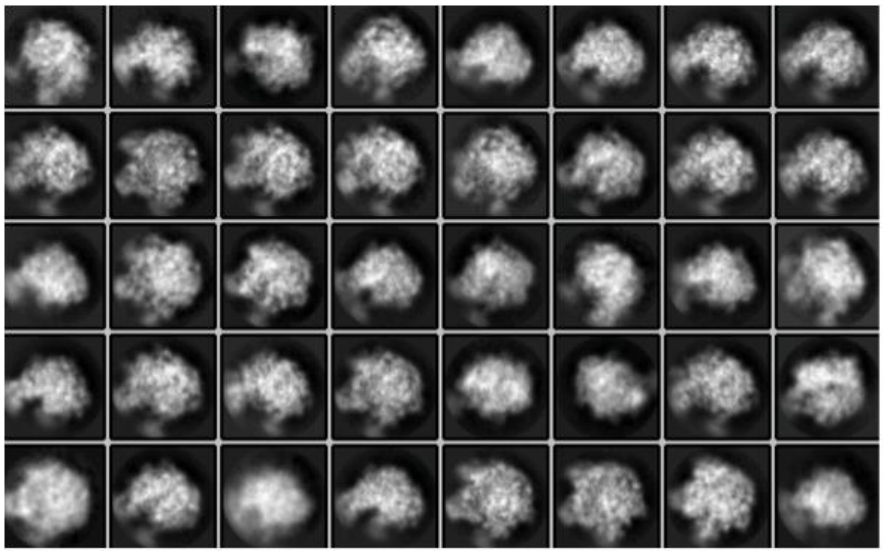

Fig. 91 40 out of 100 classes obtained after 2D *ab initio* classification of the 50S:EngA particles. These 40 classes correspond to 54198 particles out of 57545 initially analysed.

To perform the 3D reconstruction, the structure of the *B. subtilis* 50S subunit (PDB entry 3J3V) filtered at 30 Å resolution has been used as reference model. An additional 3D classification allowed us to separate our data set into two groups: one where no additional density is visible, and one where an additional density is visible that may possibly correspond to EngA (Fig. 92).

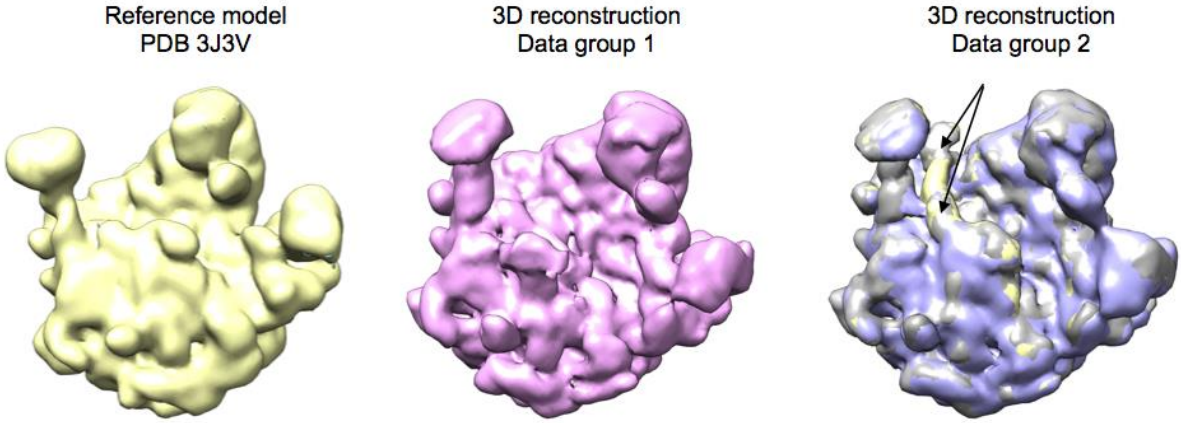

Fig. 92 3D reconstructions of the 50:EngA complex. In yellow, the reference model used to perform the 3D classification and the 3D refinement (PDB entry 3J3V) filtered to 30 Å resolution. The 3D reconstruction shown in pink corresponds to the dataset obtained from one subset of particles where no additional density is observed, when compared to the reference model. The reconstruction on the right corresponds to the superposition of three independent reconstructions (yellow, pink and purple) and shows an extra density (highlighted by black arrows), giving a second subset of particles.

This second data set, exhibiting an additional density, corresponds to 44584 particles (about 82% of the total data). A refinement of this reconstruction has then been performed, resulting in a 3D structure at a resolution between 6.40–4.86 Å (Fig. 93).

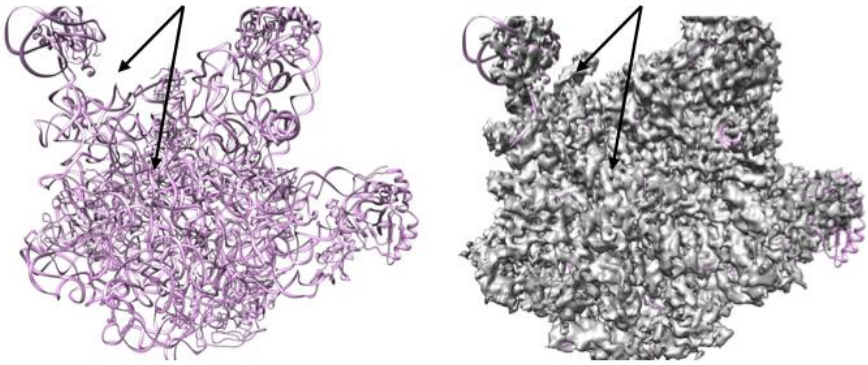

Fig. 93 Refined 3D reconstructions of the 50:EngA complex. The model (PDB entry 3J3V) used as a reference is shown in pink (left). The refined 3D reconstruction from data set group 2, shown in grey, exhibits an additional density (highlighted by black arrows) when superposed with the 3J3V structure (right).

Further analysis of these data is being carried out to understand whether the additional density might correspond to EngA bound on the 50S subunit.

4.3 EngA and the small ribosomal subunit protein S7

S7 is a protein from the small ribosomal subunit. An interaction between EngA and S7 was shown by another group using ITC (Lamb et al., 2007). However, problems of protein precipitation did not allow them to quantitatively assess this interaction. We have tried to characterise the binding of EngA to S7 by ELISA assay, SPR and BLI. Attempts of co-crystallisation were also made but, as described in the previous chapter, we were not able to obtain any crystals containing the EngA-S7 complex.

4.3.1 ELISA assay

The set-up of our ELISA assay for determining an interaction between EngA and S7 consisted on using GST-S7 to coat wells and probe binding of His-EngA_{WT} with an anti-histidine antibody. However, the initial assays revealed an unspecific interaction between EngA and GST. Hence, we defined a protocol to remove the GST-tag of S7 and to use tag less S7 instead for the coating step.

The results from this assay were not conclusive about a specific interaction between EngA and S7. The histogram in Fig. 94 shows measurements of the interaction EngA:S7 using an increasing range of EngA concentration. No major differences are seen for the initial points. The slight increase in response between 0.25–1 mg/mL may indicate that we did not use the EngA concentration range required to detect an interaction. Moreover, no differences were seen between different nucleotide-bound samples. We now know that this is certainly due to the low nucleotide concentration used at that time. As these assays were performed in the beginning of our project, we were using final concentrations of 1 mM GDP/GMPPNP, as often used in works published by other groups. Hence, different ranges of EngA and/or S7 concentration and 10 mM nucleotides should have been tested to assess the binding of EngA to S7. Because the assay using 70S provided promising results in the initial trials, we focused on this interaction and were not yet able to optimise the EngA-S7 set-up.

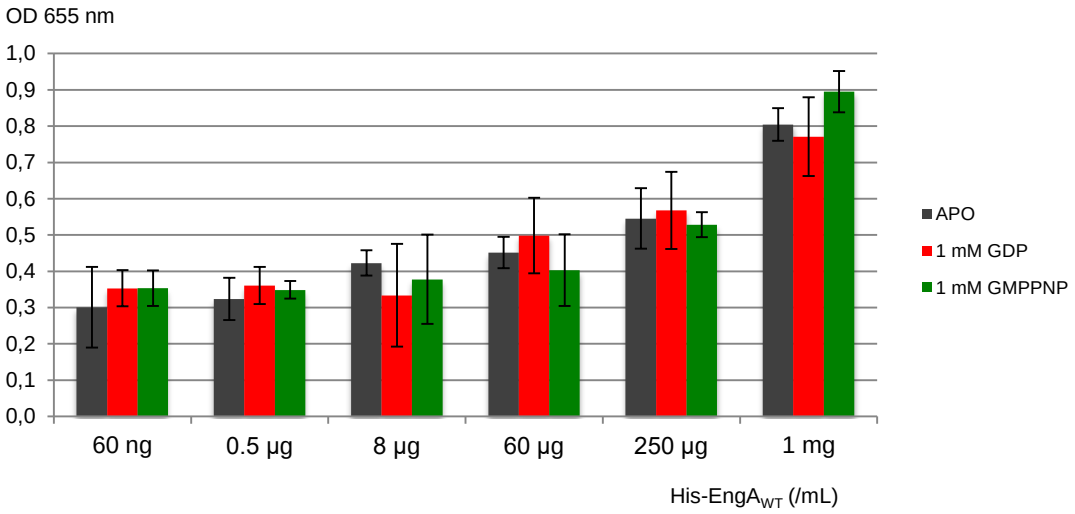

Fig. 94 Interaction between His-EngA$_{WT}$ and S7 measured by ELISA assay. The purified *S7* protein was used to coat a 96-well plate. EngA was added to the coated wells in the absence or presence of 1 mM nucleotides. OD_{655} values were plotted against EngA concentration. Each bar corresponds to the mean of independent sets of results and error bars represent the standard error where *n* = 3. No significant differences are seen between different samples. An increase in response is only observed in the last points, suggesting that the protein concentrations used are not in the optimal range for this interaction.

4.3.2 SPR

Surface plasmon resonance was employed to characterise the kinetics of the interaction EngA:S7. In the single-cycle kinetics approach, the analyte (in our case, His-EngA$_{WT}$) is injected in increasing concentration in a single cycle, without the need to regenerate the surface between injections. We have used two different set-ups, using either S7-His or GST-S7 as ligands. Coating the chip surface with GST-S7 via an anti-GST antibody has the advantage of having all S7 molecules with the same orientation. However, if the binding-site is hidden by this orientation, the interaction may not be detected. Also, unspecific interactions of EngA with the GST (previously seen in the ELISA) may interfere with the assay. Using S7-His as the analyte immobilised by amine coupling eliminates these unspecific interactions but has the disadvantage of having a mixture of ligand molecules with different points of attachment. Furthermore, covalent immobilisation to the surface induces chemical modifications on the ligand that may affect the interaction with the analyte.

Both set-ups gave very similar results. Fig. 95 shows the sensorgram obtained for the interaction His-EngA$_{WT}$ (apo) and S7-His after subtraction of the reference curve. A specific interaction between EngA and S7 seems to be present. However, the analysis of the data could not be done by fitting with the expected “1:1 binding” model. The best fit was obtained with the “heterogeneous ligand” model. This binding model describes an interaction of the analyte with two independent binding-sites on the immobilised ligand. According to this model and to the kinetic parameters obtained, a fast association ($ka_1 \approx 4000\ M^{-1}.s^{-1}$) with a K_D in the micromolar range ($K_{D1} \approx 1\ \mu M$), and a slow association ($ka_2 \approx 200\ M^{-1}.s^{-1}$) with a K_D

in the picomolar range (K_{D2} ≈ 3 pM) are present. In the presence of 1 mM GDP or 1 mM GTP/EDTA, a similar binding profile and similar kinetic parameters were obtained.

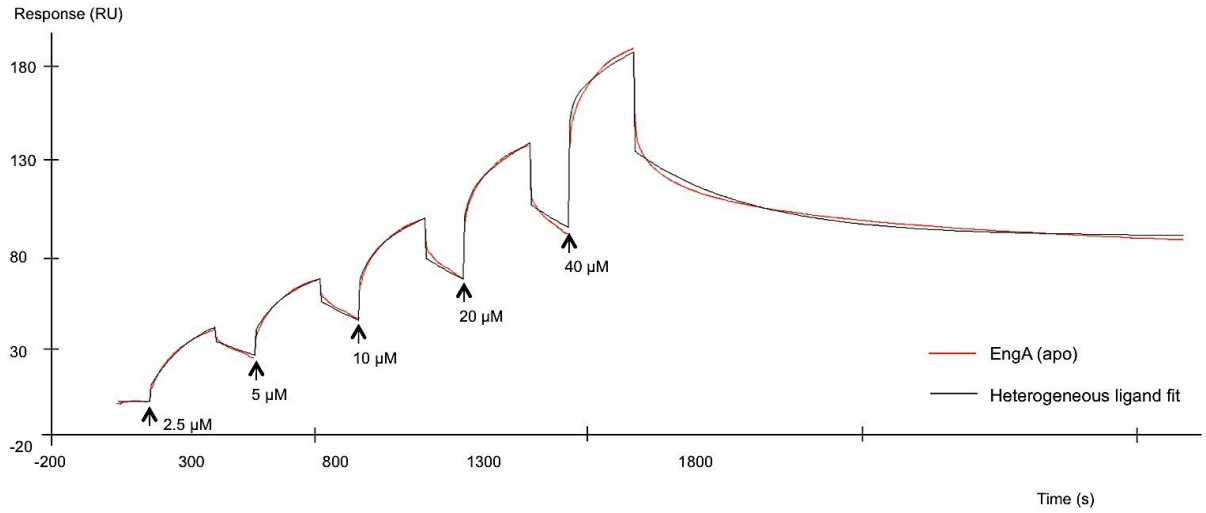

Fig. 95 Sensorgram obtained for binding of His-EngA_{WT} (apo) to S7-His immobilised in a CM5 chip via amine coupling. Arrows and numbers indicate the concentration of EngA solution injected at each time. The experimental curve is represented in red. The black line corresponds to the "heterogeneous ligand" fit model. A fast association was described with a ka_1 = 3957 $M^{-1}.s^{-1}$ and a K_{D1} = 1.2 µM; a slow association shows a ka_2 = 191 $M^{-1}.s^{-1}$ and a K_{D2} = 2.8 pM.

From the "heterogeneous ligand" binding model, as two interactions are described, it is difficult to figure out whether one or both interactions are biologically relevant. While one set of parameters may describe the expected interaction EngA:S7, the other association may correspond to unspecific binding events of EngA. As seen from the ELISA assay, EngA may be able to bind GST, which is present in one set-up. The different orientations of S7 immobilised by amine coupling may also result in defective or unspecific binding of EngA to miss-oriented S7 proteins. As we expect EngA to be able to bind the ribosome and dissociate once its role is completed, accomplishing its GTPase cycle, we do not expect a very strong interaction. Hence, we think that the parameters corresponding to the "interaction 1" (fast association and micromolar affinity) may represent the EngA-S7 interaction.

4.3.3 BLI

A similar approach was done using bio-layer interferometry. In this case, the set-up was done by immobilising His-EngA_{WT} (apo) on the biosensor and binding of tagless S7 was detected. So far, as an initial trial, one single concentration was tested. Fig. 96 shows the obtained sensorgram, which suggests a specific interaction between EngA and S7. A negative control (by recording the association of S7 to a non-immobilised sensor) confirmed the absence of unspecific binding of S7 to the biosensor.

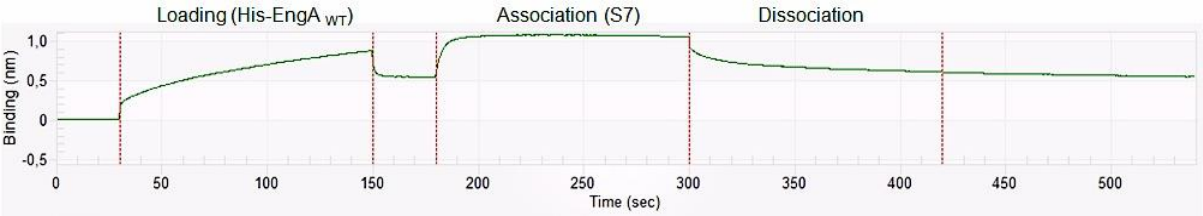

Fig. 96 Sensorgram obtained for binding of S7 to His-EngA_{WT} (apo) immobilised in a Ni-NTA biosensor.

Fitting of the experimental data with the "1:1 binding" model showed a good agreement (Fig. 97). The kinetic parameters obtained from this curve were approximately ka ≈ 34000 $M^{-1}.s^{-1}$ and K_D ≈ 0.1 μM. Although these values were computed from a single curve, they seem to be closer to the ones obtained by SPR on "interaction 1".

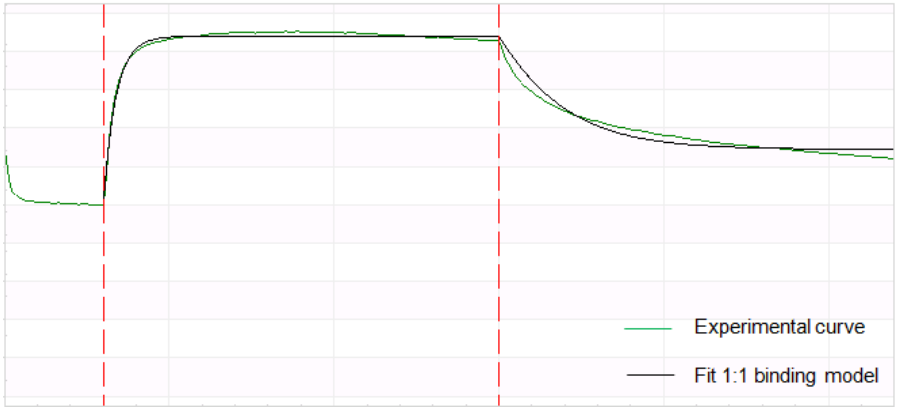

Fig. 97 Fitting of the experimental data of the interaction EngA:S7 (green line) with the 1:1 binding model (black line).

The BLI experiment is now being repeated, by using a multi-cycle kinetic set-up, where different concentrations of analyte are injected. This will allow us to obtain more accurate kinetic parameters for this interaction.

5. Screening of inhibitors

Eventually, the final aim for studying EngA is to identify inhibitors of the ribosome assembly process. In order to attain such a goal, one can either develop methods to set up reliable screening experiments or investigate already identified EngA inhibitors.

We have tested two compounds identified by Bharat and co-workers (Bharat and Brown, 2014) as inhibitors of the GTPase activity of EngA, plus an analogue (see Fig. 34). As these compounds were described as lacking specificity for EngA, the objective was to carry out structural studies in order to investigate whether they could be optimised and be made more specific to the EngA family. Crystallisation assays and crystallographic data collection (described in *Results, section 3*) were performed on EngA in the presence of either one of these three compounds in order to determine their binding-site. However, we were not able to obtain the structure of EngA in complex with either one of these inhibitors.

In order to confirm the binding of these compounds to EngA and their possible impact on EngA-ribosome interaction, thermal shift and ELISA assays were employed.

5.1 Thermal shift assay

Previous screening assays were done at the laboratory by Anne-Emmanuelle Foucher and Charlotte Caneiro (IBS, Grenoble), using thermal shift assay. This assay allows identification of ligands that either stabilise or destabilise a target protein, by following its thermal denaturation and determining the temperature at which the protein is no longer folded. These trials were performed using a small library of compounds synthesised by Ahcène Boumendjel (UGA, Grenoble) that aimed at competing with nucleotides and targeting the active site. However, an accurate assay for inhibitors screening could not be developed as these compounds interfered with the detection method of TSA.

We have tried to test the three CEB-II-5x compounds by TSA, but a similar interference was observed in samples containing these molecules. Fig. 98 shows the TSA profile obtained for EngA in the presence of these inhibitors.

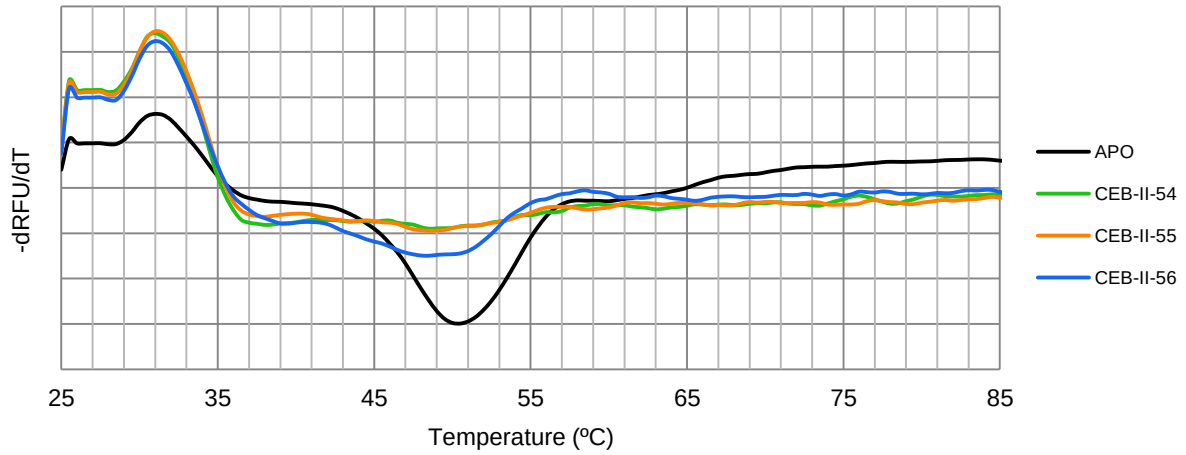

Fig. 98 TSA profile of His-EngA_{WT}. EngA was tested in the absence of ligands (APO) and in the presence of the three CEB-II-5x inhibitors. The Tm for EngA-apo was determined as 50.4ºC. However, for EngA in the presence of inhibitors, the Tm could not reliably determined.

The plot here is represented as the derivative of the fluorescence as the function of temperature. In a TSA experiment, proteins in solution are subjected to thermal denaturation. While proteins are being denatured, hydrophobic residues become exposed, which increases the emission of fluorescence by the fluorescent dye. The resulting plot of the fluorescence (RFU) versus temperature is a sigmoidal curve whose inflexion point corresponds to the melting temperature (Tm), i.e. the temperature at which the concentration of native protein equals the concentration of denatured protein. The inflexion point corresponds to the minimum of the symmetric of the derivative, which can then be calculated to determine the Tm.

The Tm for the apo-protein was determined as 50.4°C. In the case these inhibitors stabilise the protein, the Tm should increase, since an increase in Tm reflects a denaturation process occurring at higher temperatures. Hence, in such situation, we would expect a shift of the curve towards the right. The reverse would be expected for ligands that destabilise the protein. When putative inhibitors were added to *B. subtilis* EngA, no fluorescence response was observed and so the Tm could not be reliably determined. As previously observed for other inhibitors, these compounds must interfere with the fluorescent dye. Hence, TSA does not seem to be the best method to screen inhibitors for EngA. A different approach was then established.

5.2 ELISA assay

The ELISA assay was selected as the method of choice for screening compounds. Having developed an ELISA set-up to accurately detect interactions between EngA and the bacterial

ribosome, we started testing the effect of CEB-II-5x inhibitors on this interaction. However, due to the low solubility of these compounds, we encountered several problems while setting up the screening assay.

The three tested compounds are soluble in DMSO and precipitate when diluted in water. We started by testing these inhibitors at 100 µM in assay medium, however precipitates immediately started to form. The solubility of these compounds over time was then investigated using different concentrations of DMSO. Eventually, a concentration of 100 µM with 2% vol/vol DMSO was chosen, as it appeared to remain soluble during the duration time of an ELISA assay.

The initial ELISA assays performed in such conditions showed a very fast development of the blue colour upon addition of the TMB solution. TMB is a chromogenic substrate that results in a blue reaction product upon oxidation by peroxidase. The assays performed for detection of the interaction EngA:70S required incubation of TMB during the reasonable amount of time indicated by the supplier. However, the assays including inhibitors developed a blue colour that became saturated in a few seconds. Such a short time was not compatible with the time needed for pipetting and measurement of the absorbance. We have thus checked several controls and concluded that the DMSO was responsible for this effect. The washing steps in the ELISA were improved in order to remove any remaining DMSO after incubation of EngA/inhibitors. Although the effect of DMSO on the detection of the product was diminished, it was difficult to obtain reproducible results, suggesting that some interference may still be present.

The interaction curves shown in Fig. 99 represent the average of independent experiments of inhibitor screening using the assay described above. The variations in the response from different assays do not allow us to conclude about the effect of CEB-II-54 and -55 on the binding of EngA to the ribosome. These ligands generate interaction curves that are similar to the apo protein, and fluctuations observed in independent assays result either in a weaker or stronger interaction. On the other hand, CEB-II-56 always results in a significant increase in response, suggesting a stronger binding of EngA to the ribosome. This may suggest that the inhibition of the GTPase activity in EngA does not necessarily block its binding to the ribosome. However, the ELISA set-up for drug screening should be improved to eliminate all interferences and obtain consistent data on the effect of the tested compounds.

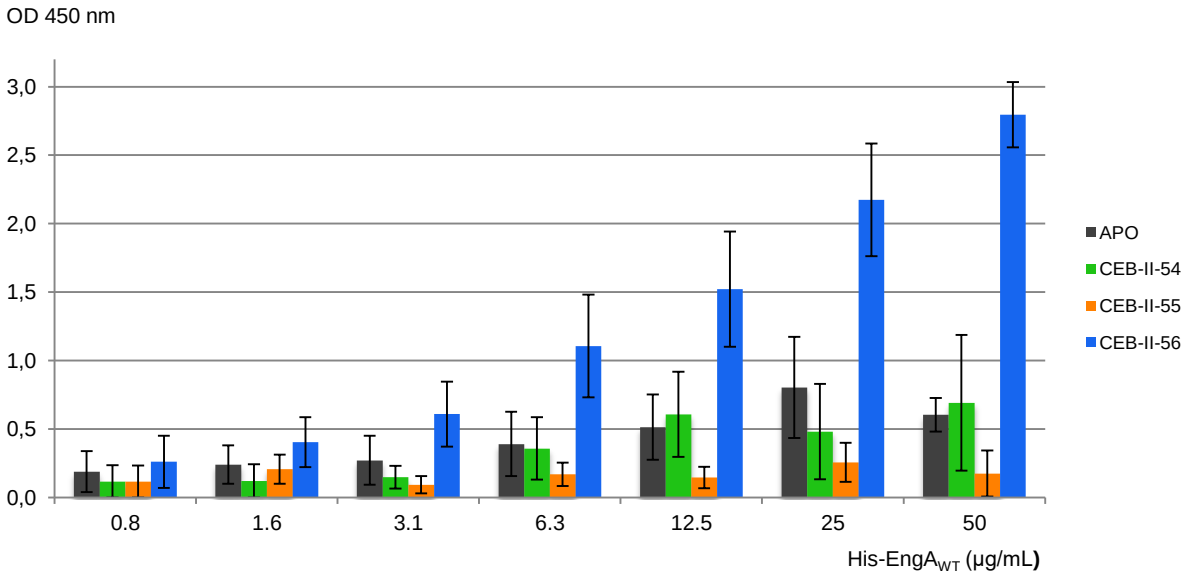

Fig. 99 Interaction between His-EngA$_{WT}$ and 70S ribosome by ELISA assay in the presence of CEB-II-5x inhibitors of the EngA GTPase activity. Each point corresponds to the mean of three independent sets of results and error bars represent the standard error (where *n* = 3). Compounds CEB-II-54 and -55 show large variation in response and we are not able to draw conclusions about their inhibitory effect on the 70-EngA interaction. CEB-II-56 on the other hand seems to enhance this interaction.

We have tried using chemiluminescent detection to investigate whether this method was less sensitive to the presence of DMSO. Although no obvious differences were seen in controls of EngA:70S with and without DMSO, we observed large variations between triplicates and obtained response curves that appear not to be related to protein concentration. These variations may suggest inconsistencies while pipetting DMSO. As for other organic solvents, which have different physical properties than aqueous solutions, DMSO may require particular caution when handling pipettes. Despite these variations, chemiluminescence may be a good option for screening DMSO-soluble compounds via ELISA assay, as DMSO seems not to interfere with this detection.

Other strategies were employed to solubilise compounds without the need of DMSO. We tested the effect of pH, other organic solvents and detergents. A few conditions were identified for solubilisation of the CEB-II-5x compounds, such as the presence of LDAO, glycerol or pH 10. The effect of detergents or basic pH on the stability of EngA and the ribosome should however be confirmed prior to defining these conditions for the screening assay.

The developed ELISA set-up is a reliable tool for detecting and probing interactions between EngA and the ribosome. However, for ligands such as the CEB-II-5x whose chemical properties interfere with the assay, some optimisation is still needed before implementing large scale screening trials.

DISCUSSION

1. High concentrations of GTP analogues trigger conformational changes in EngA and promote ribosome binding

We have been able to assess nucleotide-dependent conformational changes of EngA in solution by SAXS and limited proteolysis. Conformational changes are part of the catalytic cycle of GTPases. These enzymes switch between different states depending on the nucleotide content of the G-domain. The presence of two G-domains renders the switch mechanism more complex and unique in EngA, as tertiary structure changes can be expected in addition to modifications in the switch regions of individual G-domains. Crystallographic studies with EngA orthologs have determined different conformations that were associated with different states of the catalytic cycle. However, a close observation of these structures poses several questions and suggests that the different conformations that EngA can adopt through the GTPase cycle still need to be clarified. Because of these questions, and to avoid any bias of crystal packing on protein structure, we chose to study EngA in solution. Our SAXS experiments confirmed that the GDP-bound crystal structure of *B. subtilis* EngA, as represented by the 4DCU PDB entry, is indeed predominantly present in solution. Furthermore, we have shown that this conformation is also the adopted one by the apo-protein. Thus, our data clearly indicate that the EngA conformation observed in the 4DCU crystal structure is biologically relevant and corresponds to the "off-state" of the enzyme. With the aim of triggering a change of conformation, SAXS experiments were carried out in the presence of GTP analogues. As indicated in *Results, section 2.2.1*, no significant change is observed in the presence of 1 mM GTP analogues, while a different conformation for the EngA GTP-bound form is observed upon addition of 10 mM GTP analogues. Such a conformational change has been confirmed by limited proteolysis. Since the different nucleotide-bound states of EngA are related with its affinity for the ribosome, we assessed the impact of GTP analogues on ribosome interactions. Interaction studies confirmed that 10 mM GTP analogues were required to observe a significantly improved binding of EngA to the bacterial ribosome when compared to the apo or the GDP-bound protein.

The need of 10 mM nucleotides to induce a change in conformation may seem unexpected given the effect of nucleotides on ribosome binding observed by other groups using lower concentrations (Agarwal et al., 2012; Foucher et al., 2012; Hwang and Inouye, 2006; Schaefer et al., 2006; Tomar et al., 2009). However, the bacterial nucleotide intracellular concentrations seem not to be far from such range. Indeed, the GTP pool has been determined in *E. coli* as being in the order of the millimolar range (~1.7 mM) (Buckstein et al., 2008). The GTP concentration in cells may suffer great variations in response to different stimuli: in this same study, when cells entered the stationary phase, a 4-fold decrease in

GTP concentration was observed. In cells where GTP homeostasis was compromised by gene deletion, intracellular GTP levels reached >10 mM (Kriel et al., 2012). Taking together, such values and our observations on EngA conformational changes and ribosome binding suggest that EngA may be sensitive to these variations in GTP intracellular concentration and adapt its interactions with the ribosome according to it. Indeed, previous observations had suggested that EngA, as well as other RA-GTPases, could be involved in sensing nucleotide concentrations to regulate ribosome synthesis according to the cell requirements (Britton, 2009; Karbstein, 2007). Cell growth and ribosome defects caused by a mutation in EngA-GD2 (N321D) were rescued by accumulation of (p)ppGpp (Hwang and Inouye, 2008), which has been shown to bind to EngA and to inhibit its GTPase activity (Bharat and Brown, 2014; Hwang and Inouye, 2008). TSA experiments performed by Anne-Emmanuelle Foucher (IBS, Grenoble) also showed that (p)ppGpp can bind to EngA and stabilise it against thermal denaturation. (p)ppGpp are modified guanine nucleotides and elevated levels are associated with the stringent response. The stringent response is initiated in cells enduring stress in order to drive their metabolism towards a “survival mode”. Upon accumulation of uncharged tRNA in the ribosome, following amino acid depletion, (p)ppGpp synthesis is enhanced, which alters gene transcription. Promoters involved in cell growth are down-regulated (including tRNA and rRNA production), while promoters involved in stress response are up-regulated (Corrigan et al., 2016; Hauryliuk et al., 2015). The mechanisms for transcription regulation by (p)ppGpp differ between bacterial species. In *B. subtilis*, this regulation is achieved by controlling intracellular GTP levels. During the stringent response, (p)ppGpp decreases GTP levels by direct inhibition of its biosynthesis pathways (Kriel et al., 2012). Interestingly, even in the absence of stress, basal levels of (p)ppGpp are necessary to maintain the homeostasis of GTP (Kriel et al., 2012). In the study by Buckstein and co-workers, the 4-fold decrease in GTP level during slow growth-rate occurred concomitantly with an increase in (p)ppGpp level, a behaviour similar to a stringent response. Consistent with the bacterial GTP pools range, (p)ppGpp levels can increase up to 1–2 mM during stress response (Corrigan et al., 2016). Hence, it is possible that in cells EngA binds GTP, GDP or (p)ppGpp depending on their intracellular concentrations and exerts different effects on the ribosome. It would be interesting to perform conformational analyses and interaction studies on EngA in the presence of (p)ppGpp, similarly to what we have done with GDP and GTP analogues. This would allow us to understand how conformational changes of EngA can be regulated by different guanine nucleotides and how they are connected with EngA-ribosome interactions.

2. Contrasting affinities of EngA G-domains for nucleotides

The affinity of *B. subtilis* EngA for nucleotides was measured by fluorescence on truncated constructs of EngA (GD1 and GD2-KH) (Foucher et al., 2012). This set-up allowed determination of the affinity of each individual G-domain. Both GD1 and GD2 showed a similar affinity in the order of the micromolar range. However, several observations from our experiments done on the whole protein suggest a different behaviour.

Both SAXS and crystallographic studies indicate the presence of GDP bound to EngA even if the protein is purified in the absence of nucleotide (apo-protein). We observed a significant divergence between the known molecular mass of His-$EngA_{WT}$ (50.8 kDa) and the one obtained from the forward scattering intensity I_0 (37.6 kDa) based on OD_{280} measurements and theoretical ε_{EngA}. When the theoretical $\varepsilon_{EngA:GDP}$ is applied instead, a much better agreement is obtained, suggesting that about one GDP molecule remains bound to EngA. Indeed, one crystal structure (Crystal 2) shows EngA-GD2 with a GDP molecule bound to it, while the crystallisation conditions did not include any nucleotides. Such observation of a GDP bound to the GD2 domain in a crystal structure while it was not present either in the crystallisation condition or in the protein buffer was also made on *T. maritima* EngA (Robinson et al., 2002). Hence, the affinity of GD2 for GDP is high enough so that the nucleotide remains bound after all purification steps, and thus much higher than that of the GD1.

Furthermore, a change in conformation was observed only by increasing the concentration of GTP analogues up to 10 mM. This suggests that such high GTP concentrations are required to have both G-domains in the GTP-bound state: either GD1 has a very low affinity, either GDP is strongly bound to GD2 and needs high amounts of GTP to displace it, or both. In our crystal structures, we were able to obtain GD2 bound to GTP analogues using 1–20 mM nucleotides. However, GD1 was never seen bound to GTP analogues with full occupancy: the only structure obtained having a low occupancy GTP analogue in GD1 domain (occupancy ≈ 0.5) was obtained with 20 mM GMPPCP in the crystallisation drop. All together, this suggests that GD1 may have a low affinity for nucleotides, in the millimolar range, *in vitro*.

Observation of the nucleotide-binding site in GD1 and GD2 from the 4DCU structure (GDP-bound) shows some differences in the contacts formed between the nucleotide and the protein. Although no big differences are seen at the level of the phosphate moiety and ribose, the nucleobase is lacking some contacts in the GD1. In particular, all three residues ^{329}SAL^{331} from G5 motif are forming hydrogen bonds with the nucleobase in GD2, while in GD1 only G146 (^{145}SGT^{147}) is forming contacts (Fig. 100). In the first G-domain, S145 is pointing in the

opposite direction of the nucleotide (at 5.0 Å distance from the O6 of the base, versus 3.2 Å in GD2) and a slight movement of the backbone of this loop increases the distance of T147 from the nucleotide to 3.6 Å (versus 3.2 Å in GD2-L331). Such contacts in the nucleotide-binding site lacking in the GD1 could contribute to a lower affinity. However, the absence of some structured features in GD1 (such as the G4 motif) make this analysis incomplete. The G4 motif seems to be very mobile in GD1, as it is not present in most of the EngA crystal structures (with only the PDB entry 4KYU, a *B. subtilis* EngA GD1 domain crystal structure, showing a well-structured loop). In four of our refined structures, the G4 is well defined but pointing towards the opposite direction of stabilisation of the guanine base. On the contrary, in GD2 this motif is visible in all structures and structurally well conserved, pointing to the guanine base. This may further contribute to a higher affinity of GD2 and a lower affinity of GD1 for nucleotides.

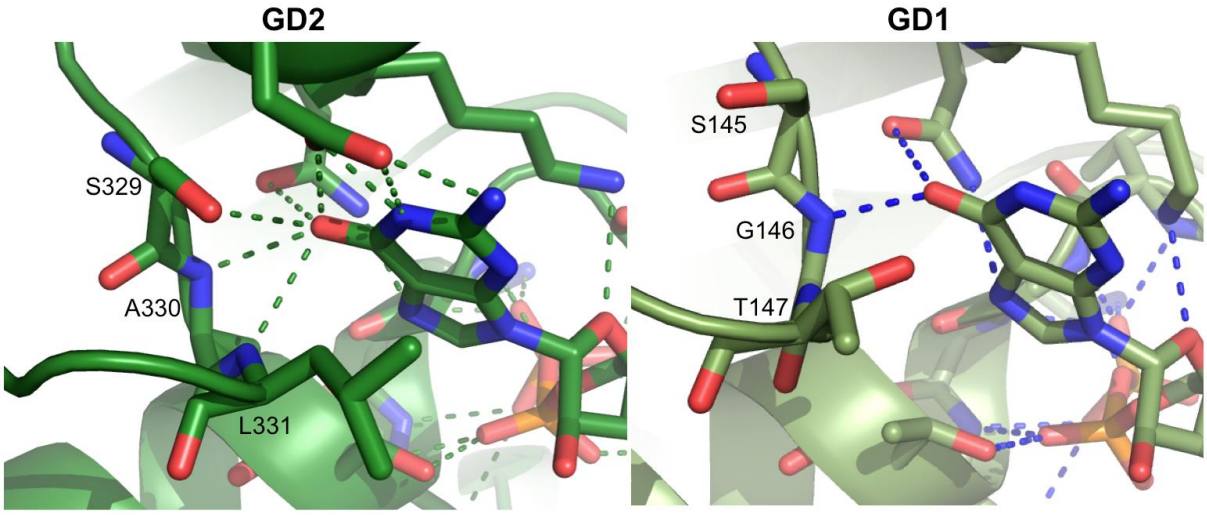

Fig. 100 Contacts in the nucleotide-binding site in GD1 (light green, right) and GD2 (dark green, left) of *B. subtilis* GDP-bound EngA. The G5 motif forms fewer contacts in GD1 than in GD2. While the OH- group from GD2-S329 is stabilising the nucleobase, GD1-S145 is pointing in the opposite direction, at a distance of 5 Å from the nucleobase. GD2-L331 is further stabilising the nucleotide, while a shift of the backbone in this loop in GD1 positions the T147 at a larger distance from the nucleotide (3.6 Å vs 3.2 Å in GD1 and GD2, respectively). In the 4DCU structure, as well as in most EngA structures, the GD1-G4 motif is not structured and the conserved Asp residue is not seen stabilising the guanine base, in opposition to what is observed in GD2. This image was edited using PyMol from PDB 4DCU.

The contrasting affinities measured on either the isolated domains (GD1 and GD2-KH) or the whole protein may be explained by a cooperative mechanism between GD1 and GD2. Previous studies have shown that both G-domains are crucial for GTPase activity and ribosome binding, and that they have a cooperative role in regulating EngA function (Bharat et al., 2006; Tomar et al., 2009). Hence, it is possible that the affinities of G-domains may be different in the context of the full-length protein, whereas this cooperativity is disrupted when G-domains are separated (as measured for *B. subtilis*). Indeed, studies in *S. typhimurium* EngA by ITC (Lamb et al., 2007) allowed modelling of the data with two sites with distinct affinities for GTP, one with low (K_D > 100 µM) and one with high (K_D = 12.7 µM) affinity. In *B.*

subtilis, fluorescence experiments on the full-length protein were not able to discriminate two distinct binding-sites (Foucher et al., 2012). Possibly, the affinity of the *B. subtilis* EngA GD1 is too low to be assessed by this technique.

3. In search for the GTP-bound conformation: crystallisation strategies based on crystal packing analysis

By using the SAXS technique, we were able to observe a new conformation of *B. subtilis* EngA triggered by 10 mM GTP analogues. Comparison of the structural properties of the induced conformation with the crystal structure previously proposed to mimic the GTP-bound form (PDB 1MKY) shows significant differences. This seems to be supported by the MS analysis on the fragments generated after tryptic digestion. These data indicate that 1MKY does not correspond to the active-state of *B. subtilis* EngA. With the aim of determining the structure of the GTP-bound EngA, we performed numerous crystallisation trials. Despite all the efforts, and although we were able to grow crystals that diffracted to a reasonable resolution, we always obtained the GDP-bound conformation. So far, this same conformation has been obtained in four different space groups, $P2_12_12_1$, $P2_1$, $P1$ and $C2$, with either one or two molecules in the asymmetric unit. This seems to confirm that the GDP-bound form is a very stable conformation in solution and not much affected by crystal packing. Analysis of the crystal packing of all these crystal forms shows that the largest interface (~1200 $Å^2$) identified with PISA (Krissinel and Henrick, 2007) is basically conserved in all crystal forms (Fig. 101). This suggests that the nucleation process might be initiated by the formation of this dimer.

The presence of this conserved dimer in all crystal forms is surprising, as EngA was never seen as a dimer in solution. In gel filtration and SAXS, EngA was always present as a monomer. Experiments of analytical ultracentrifugation carried out by Anne-Emmanuelle Foucher (IBS, Grenoble) showed the same monomeric behaviour for EngA. Dimerisation of G-domains occurs for some GTPases, such as MnmE, where a homodimer is formed across the nucleotide-binding site of both G-domains, facilitating the catalysis (Scrima and Wittinghofer, 2006). Observation of the EngA dimer from the crystal structures shows the nucleotide-binding sites of all four G-domains from the two EngA molecules accessible to the solvent. No evident effect on the stabilisation of the nucleotide-binding sites seems to arise from dimerisation. Together with the fact that no dimer/multimer was ever seen in solution, this does not seem to point to a biological significance for the dimeric form of EngA.

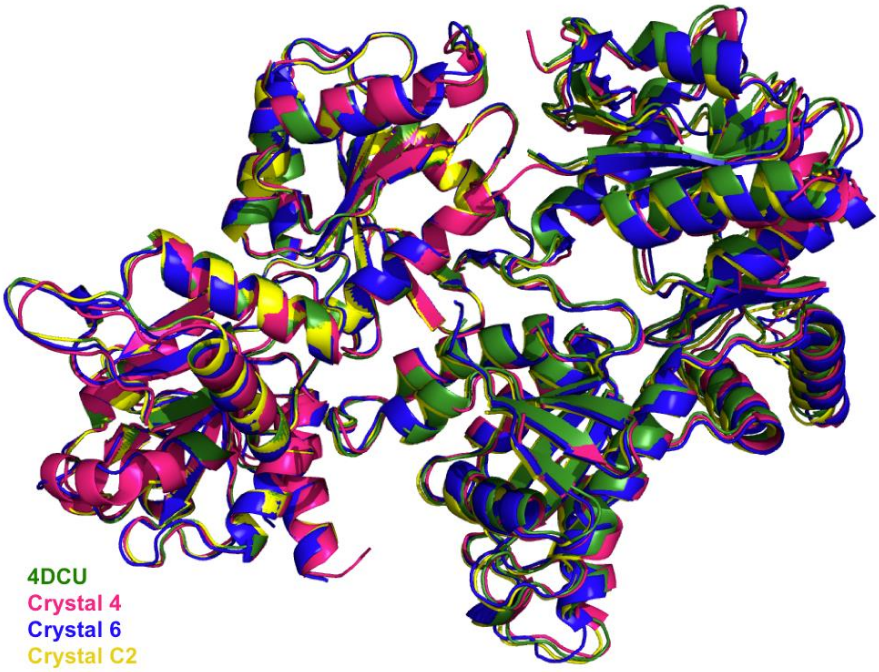

Fig. 101 Superposition of the 3D structures of EngA dimers in four crystal forms. PDB entry 4DCU (green; space group $P2_12_12_1$) and Crystal 4 (magenta; space group $P2_1$) are represented by one molecule per asymmetric unit and a symmetry related molecule). Crystal 6 (blue; space group $P1$) has two molecules in the asymmetric unit. The EngA structure represented in yellow corresponds to the crystal in space group $C2$, with one molecule in the asymmetric unit and a symmetry related molecule. The four structures are superposed in one monomer. The dimer is maintained by the same interface, which is stabilised by conserved hydrogen bonds and salt bridges between residues of both molecules.

The interface formed by this dimer is only possible if EngA adopts the 4DCU conformation. Crystal contacts formed in this interface may possibly be favoured during crystallogenesis. EngA in the GDP-bound conformation may be selected during crystallisation at the expense of other alternative conformations (from which possibly the GTP-bound one) that stay in solution. One possible strategy could be to destabilise this interface by mutagenesis of the surface residues. A study by Charron and co-workers (Charron et al., 2002) has shown that modification of the crystal contacts by point mutations that alter the surface charge, hydrophobicity and/or H-bonds can affect protein crystallisation. A correlation was found between disruption of packing contacts and prevention of crystallisation; and between addition of potential contacts and favoured crystallisation. Several examples have shown that site-directed mutagenesis can be successfully employed to disrupt crystal contacts. The crystallisation of the replication protein A 70 subunit with inhibitors was only possible after directed mutagenesis (Feldkamp et al., 2013). Analysis of crystal packing revealed that the binding site was occluded in the crystal lattice, preventing binding of ligands in the wild-type protein. Disruption of crystal lattice contacts by mutagenesis led to crystallisation in an alternate crystal packing where the ligand-binding site was available. Similarly, substitution of Glu residues involved in a hydrogen-bonding network from the crystal packing of the human Keap1 protein allowed disruption of crystal contacts (Hörer et al., 2013). The mutated protein

crystallised in a different space group from the wild-type, and the occurrence of non-crystallographic symmetry led to exposure of a substrate binding-site in one of the molecules in the asymmetric unit. Hence, we may think of residues that form crystal contacts involved in the conserved interface in EngA crystals that could be substituted in order to destabilise salt bridges or hydrogen bonds. E348, which forms a salt bridge, could be replaced by a reverse charge residue such as an arginine. R353 and Q355, which form hydrogen bonds, could be substituted by short side-chains, such as alanines, to avoid contacts. Similarly, substitution by bulky residues, such as tryptophans, could engender steric hindrance that would also prevent formation of crystal contacts. However, the approaches for crystal engineering cannot be generalised for all proteins and the results of point mutations may be difficult to predict in the context of the entire protein. In a study by Dale and co-workers, a rational prediction of crystal contacts was applied to a wild-type and a trimethoprim-resistant dihydrofolate reductase from *S. aureus*. However, the same residues that generated close crystal interactions in the wild-type protein in a hexagonal crystal form, after insertion in the trimethoprim-resistant protein generated a tetragonal space group with different crystal contacts (Dale et al., 2003). Therefore, this approach has to be adapted to each case. Particular attention may be needed in the search for a new conformation of EngA, as mutations intended to disrupt crystal contacts may also affect conformational changes. Namely, the main crystal interface that occurs in our EngA crystals is largely formed by residues from the α6-helix of GD2 that, according to the interpretation of our limited proteolysis assay, may have different surface environments in the GDP- or GTP-bound forms. If specific protein-protein interactions are disrupted upon site-directed mutagenesis, we may still not be able to obtain the biologically relevant GTP-bound conformation.

4. Structural features of EngA G-domains: implications in nucleotide binding and hydrolysis

The several EngA structures obtained from our crystallographic experiments show different nucleotide-binding site contents and different conformations for some flexible regions. Analysis of these structures allowed us to gather other interesting structural information about EngA.

Local changes in the nucleotide-binding site can be observed comprising the conserved regions involved in nucleotide binding and hydrolysis. Most of these changes seem to arise from the large flexibility of loops within this region, and do not correlate with specific nucleotide-bound states of the catalytic site. An exception seems to be the P-loop and switch

II regions of GD2, that become more destabilised in GTP-bound proteins by the presence of the γ-phosphate. Comparison of six *B. subtilis* EngA structures bearing a GD2:GTP analogues (PDB 4DCV and Crystals 1, 4, 5, 6 and 7) revealed large deviations in the position of the γ-phosphate (up to >2 Å) that induce movements of the P-loop and switch II in order to adapt to the presence of the nucleotide.

Differences in the flexibility of the G4 motif in EngA G-domains were also observed. The large flexibility of GD1-G4 in comparison to GD2-G4 seems to explain in part the differences seen in the affinity of both G-domains for nucleotides (see *Discussion, section 2*).

Furthermore, in almost all crystal structures of *B. subtilis* EngA, the GD1-switch II is not structured due to its high flexibility. In one of our crystal structures (Crystal 1) we were able to observe this loop. However, the conserved D57 and G60 from the GD1-G3 motif adopt a position where stabilisation of the magnesium ion and the γ-phosphate is not possible. These residues are known to be catalytically important and their conserved position in GTPases characterises the GTP-bound state. Another clue on one of the catalytic residues of EngA came from the only structure so far obtained with GD1 bound to a triphosphate nucleotide (Crystal 6B). In this crystal structure, residue Q75 following the GD1-switch II motif is stabilising the γ-phosphate. Previous studies carried out by Anne-Emmanuelle Foucher (IBS, Grenoble) showed that mutating this glutamine to alanine in the full-length EngA affects enzymatic activity, decreasing it to half of the wild-type protein. These observations raise interesting questions regarding the hydrolysis mechanism present for EngA. In Ras-like GTPases, a catalytic glutamine (Gln^{cat}) follows the G3 motif (DxxGQ) and is involved in the canonical GTP hydrolysis mechanism. In contrast, in HAS-GTPases a hydrophobic residue is replacing the Gln^{cat}, suggesting the presence of an alternative hydrolysis mechanism. Indeed, in HAS-GTPases the switch II is suggested to be retracted and the hydrophobic residue to be pointing towards a hydrophobic pocket opposite to the catalytic site, while a catalytic residue is provided by neighbouring domains or by another interacting protein (Mishra et al., 2005). EngA has been classified as a HAS-GTPase due to the presence of such hydrophobic residues following the G3 motif (namely an Ile in GD1 (DTGGI) and a Met in GD2 (DTAGM)). However, this classification has been done based on sequence analysis. In the context of the protein three-dimensional structure, a Gln residue up- or downstream the G3 motif in the protein sequence may be close to the catalytic site. Our observations seem to suggest that Q75, located 14 residues downstream the canonical position of the Gln^{cat}, may participate in catalysis, playing a similar role to the Gln^{cat} in Ras. Hence, as far as the GD1 is concerned, EngA may possibly not be a genuine HAS-GTPase. To confirm this hypothesis, the effect of the Q75A mutation on the enzymatic activity should be checked in a truncated construct of

EngA-GD1. This EngA mutant may thus represent a good starting-point for activity assays and to eventually understand the catalytic mechanism present in GD1. The presence of a Gln^{cat} may also in part explain the higher activity of GD1 in comparison to GD2. Different catalytic mechanisms may be present for GD1 and GD2. Indeed, studies have suggested a cooperative action between both G-domains, with one of the domains possibly playing a regulatory role. Further investigation on the sequence, structure and biochemistry of EngA are still necessary to identify the attributes of each G-domain.

5. Mutants trapping conformations: insights into the on-/off-states of EngA

Two EngA variants, $EngA_{G139C/V388C}$ and $EngA_{Y51C/T169C}$, were generated to block the EngA conformation in either the one observed in 4DCU PDB entry (GDP-bound conformation) or the one observed in 1MKY PDB entry (pseudo GTP-bound). These conformations were blocked by formation of a disulfide bond between substituted cysteines.

SAXS analysis of the $EngA_{G139C/V388C}$ mutant indicated that the intended 4DCU conformation, or at least a very similar one, was obtained. The $EngA_{G139C/V388C}$ mutant may be a valuable tool for studying the GDP-bound state of EngA and further understand the role of this stable conformation of EngA in ribosome assembly. Having all proteins blocked in the same conformation will allow us to have a homogeneous sample (regarding the tertiary structure) and study the properties of the off-state of EngA, avoiding the dynamic interchange between conformations in solution that may occur even in the presence of GDP. Interaction studies between $EngA_{G139C/V388C}$ and the ribosome (either the 70S or individual subunits) have now been initiated. Based on our results and on what was already known from the literature, we speculate that this mutant should not bind the 70S ribosome or the 50S subunit, and should eventually bind the 30S subunit. It may also be interesting to determine whether the enzymatic activity of this cross-linked mutant is compromised after preventing conformational changes.

The analysis of the $EngA_{Y51C/T169C}$ variant SAXS data indicated that it adopts a conformation in solution that differs significantly from all known structures of EngA, represented by 4DCU, 1MKY or 3J8G PDB entries. Thus, this double mutant fails in blocking EngA in the conformation adopted in 1MKY structure. Since we cannot reliably predict the tertiary structure of the $EngA_{Y51C/T169C}$ variant from the SAXS data, we cannot reason any biological role for it. It cannot also be used to test whether the EngA conformation, observed in the

1MKY crystal structure and claimed by the authors to represent the GTP-bound form, is biologically relevant or whether it is the result of very specific crystallisation conditions.

SAXS analysis on the wild-type protein revealed that the 1MKY does not represent the real GTP-bound state of *B. subtilis* EngA. Previous observations from this structure had already suggested that 1MKY could possibly not be the GTP-bound conformation as claimed by the authors. The presence of two phosphate ions would hardly mimic a β-γ covalent bond due to steric hindrance and electrostatic repulsion. This is shown by the perfect overlap of one PO_4^{3-} on the β-phosphate while the second PO_4^{3-} stands at a larger distance from it than the γ-phosphate (see *Introduction, section 5.4*). This second phosphate ion, close to the position of the γ-phosphate, is also not stabilised in this structure by the conserved Gly from the switch II, which is pointing in the opposite direction from the one found in other GTP-bound GTPases. Also, comparison of the three EngA structures 4DCU, 1MKY and 3J8G (see Fig. 29) reveals that the pseudo GTP-bound and the 50S-bound states differ the most in the protein global conformation, inconsistent with the fact that GTP binding enhances ribosome interactions. Understanding the properties of the on-state of EngA would be a major step in understanding EngA function in ribosome assembly. As so, improvements in the crystallisation conditions are crucial to obtain the GTP-bound conformation of EngA. Having this structure would allow us to propose an EngA variant to block such conformation, and thus open the way for setting up efficient screening strategy to identify compounds interfering with the biological role of this on-state.

At the time this work was started, the 50S-bound structure of EngA was not available. Now that this conformation is known, it could be also interesting to try to generate a mutant to block this state. However, obtaining an EngA variant that could block this conformation could be difficult to achieve. In the 3J8G cryo-EM structure, EngA adopts an extended conformation where both G-domains are facing each other via their nucleotide-binding site (Fig. 102). Hence, the interface between GD1 and GD2 is mainly composed of flexible loops, which, as we have seen for the $EngA_{Y51C/T169C}$ mutant, may not be optimal to block the protein in the desired tertiary structure even when bridged by a disulfide bond. Introducing more than one disulfide bond could help to rigidify the inter-domain interface. However, this could also lead to the formation of different cross-linked species, by formation of different S–S bridges, which would be impossible to separate.

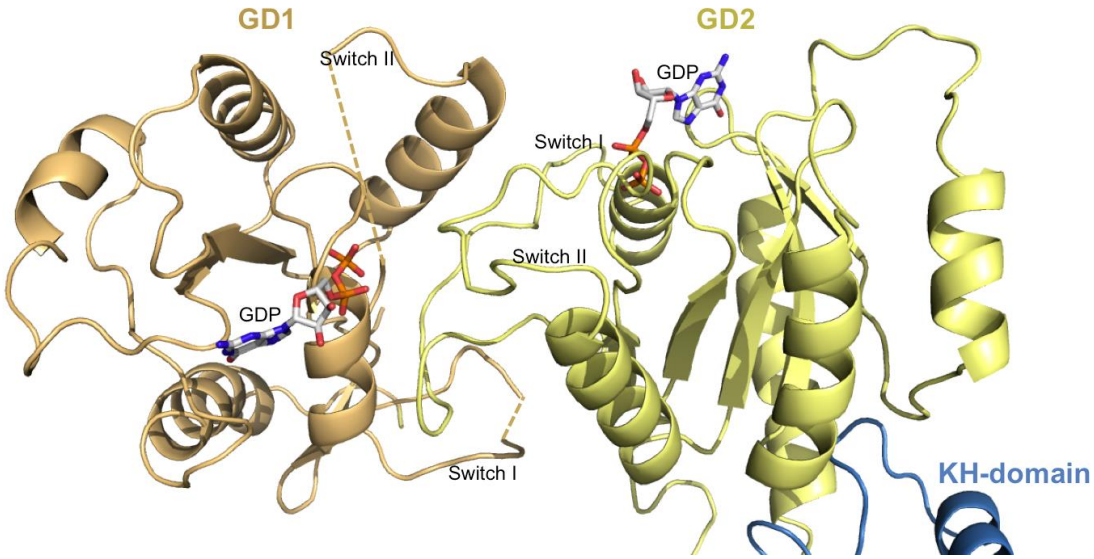

Fig. 102 Close view of the G-domains of *E. coli* EngA in the 50S-bound conformation. GD1 is represented in light orange and GD2 in yellow. In this conformation, the nucleotide-binding sites of the G-domains are facing each other, with the switch regions forming the interface. The nucleotides, which are not visible in the cryo-EM structure, were added for a better visualisation of the nucleotide-binding sites. This image was edited using PyMol from PDB entry 3J8G.

6. Inhibitors of ribosome assembly: targeting EngA function

6.1. Challenges in screening bioassays

We have developed an ELISA assay set-up to reliably detect interactions between EngA and the bacterial ribosome. This assay has started to be applied for screening small compounds in order to identify inhibitors of this interaction. The compound library was dissolved in DMSO to a final concentration of 10–50 mM and stored at 4°C. Aliquots of these stocks were then tested at final concentrations in the µM range. However, due to their hydrophobicity, precipitation occurs immediately upon addition of compounds to protein solution. The low water solubility of organic compounds is a big challenge in *in vitro* experiments. Not only the experimental conditions must be optimised, but also the screening results may not be reliable (Di and Kerns, 2006). In the particular case of drug discovery, solubility problems affect not only the screening assay, but also the subsequent formulation for *in vivo* delivery and the pharmacokinetic/pharmacodynamic behaviour of drugs (Kerns et al., 2008). In order to reach their biological targets, drugs must be soluble. Poor water-solubility results in poor bioavailability and sub-optimal drug delivery. Indeed, poor pharmacokinetic properties are one of the main factors leading to the rejection of drug candidates during development (Henchoz et al., 2009). Generally, compound concentrations used in screening bioassays may vary from very low concentrations up to 100–200 µM, depending on the nature and affinity of the compounds. One way to possibly avoid solubility problems may be to use lower

compound concentrations (for example, starting with 1 µM). However, low affinity compounds – whose design could afterwards be optimised for enhanced affinity – may be missed in such set-up. For compounds that show no inhibitory response at low concentrations, higher concentration can then be tested, limiting solubility problems to fewer compounds.

Precipitation of compounds occurred even in DMSO solutions after long storage and freeze-thaw cycles between assays. Precipitates (either in stock solutions or in screening-plates) can cause non-reproducible data by over- or under-estimation of the concentrations pipetted in the bioassay. Storing stock solutions in individual aliquots may be a useful strategy to avoid precipitation related to freeze-thaw cycles. For short-term use, storage at room temperature instead of 4°C may also prevent reduced solubility of compounds in DMSO solutions, and thus minimise precipitation. Preparing lower stock concentrations also improves solubility in DMSO solutions, but limits the maximum concentration to be screened. The manipulation of DMSO dilutions seems to also be a factor in the formation of precipitates. In the course of our experiments, for example, we observed different behaviours on the solubilisation of these organic compounds when performing up-and-down mixing, which more easily lead to precipitates, versus vortexing, which contributed to a better solubilisation possibly due to a faster homogenisation.

Prior to screening trials, different concentrations of compounds and buffer/DMSO ratios were tested over time to check for the presence of precipitation in the assay medium. The use of low DMSO concentrations in the bioassay helped us solubilising the compounds in the assay medium, but at the same time interfered with the assay itself. To avoid DMSO, we have tested the effect of pH and additives (such as alcohols and detergents). Some compounds can have a pH-dependent solubility and adjusting the pH may be a strategy to improve solubility. In drug formulation, pH modifiers are commonly used to deliver insoluble drugs. Intravenous formulations of the fluoroquinolone ciprofloxacin, for example, include lactic acid to increase solubilisation (Kalepu and Nekkanti, 2015). Co-solvents and surfactants are also widely used as formulation vehicles for non-polar drugs. Intramuscular formulations of the benzodiazepine lorazepam use up to 18% PEG 400 and 80% propylene glycol to increase solubility and palclitaxel, an antineoplastic agent, is delivered in oral formulations in the form of polymeric micelles (Kalepu and Nekkanti, 2015; Strickley, 2004). Another strategy to avoid DMSO may consist on using a dry film of compounds after evaporation of DMSO, similarly to the “dry co-crystallisation” strategy we have employed for protein crystallisation.

The washing steps from the ELISA assay were also adapted for compound screening. Other parameters can still be improved, such as buffer composition, temperature or the incubation time of the assay. When dealing with poor soluble compounds, all experimental conditions

must be adapted to each particular bioassay. While such a fine optimisation may be reasonable when dealing with a few compounds, it becomes impossible to do when screening large chemical libraries composed of thousands of inhibitors. To avoid the difficulties in optimisation of bioassays and the costs of developing drug formulation strategies, solubility can always be included as a criterion in the design of libraries.

As we have experienced, the problems derived from poor solubility may also affect efforts in structure-based lead optimisation. Knowing the binding-site and mechanism of action of inhibitors is an important step in optimising their design for improved specificity and affinity over the target. However, similarly to the screening assay, solubility issues affected our crystallisation trials, limiting the compound concentration used in crystallisation drops. Strategies used for screening assays and drug formulation can be tested. It could be promising to screen crystallisation conditions buffered at pH values around 10, which showed to be efficient in solubilising compounds. The use of additional solvents/additives, such as detergents, may also help to solubilise compounds, but their effect on crystallogenesis may be difficult to predict. Furthermore, we will have to check that EngA remains properly folded in these conditions.

6.2. Inhibitors of EngA function

In the search of inhibitors of EngA function on the ribosome assembly pathway, molecules with different mechanisms of action can be considered. Suppressing the GTPase activity in EngA can be one strategy to inhibit its function in ribosome biogenesis. EngA can be targeted by small molecules that bind to the nucleotide-binding pocket, displacing GTP. For proteins such as EngA that undergo large conformational changes, modulation of the enzymatic activity can also be done by molecules that bind to allosteric sites. Binding of an allosteric inhibitor can trigger EngA to switch to its inactive state, slowing or suppressing the reactions carried by the catalytic site. The latter approach presents the advantage of being more selective towards their target. Whereas allosteric sites may present structural diversity, the catalytic site is usually well conserved between members of the same family and so cross-inhibition may occur with competitive antagonists. However, as allosteric sites are unknown, drug design can be more challenging than for competitive inhibitors, which can be based on the structure of the endogenous ligand (Grover, 2013; Nussinov and Tsai, 2013). We have performed crystallisation assays with two GTPase activity inhibitors identified by Bharat and colleagues (Bharat et al., 2013) and a structural analogue, in order to identify their binding-site in EngA. Knowing where and how these compounds bind EngA would be an important step for structure-based drug design and optimisation. However, despite the

different crystallisation techniques implemented, so far we have not been able to obtain the structure of the EngA-ligand complexes. Further crystallisation conditions/techniques and additional strategies for solubilisation of these compounds (which mainly account for the difficulties in our crystallisation set-ups) should be tested.

Besides probing EngA as a single target, one can also think of targeting EngA-ribosome interactions. In the catalytic cycle of GTPases, binding of GTP allows interactions with an effector molecule and GTP hydrolysis promotes dissociation. We can therefore speculate that inhibitors of the GTPase activity will affect EngA-ribosome interactions. However, the exact mechanisms that interplay GTP hydrolysis, conformational changes and ribosome binding in EngA are not completely determined. Thus, the modulation of these events by inhibitors may be difficult to predict. We have tested the effect of the previously mentioned GTPase inhibitors on the interaction between EngA and the ribosome to understand their impact in protein-protein interactions. Interestingly, one GTPase inhibitor seems to strengthen this interaction. Although there is the possibility of this result not being reliable due to the experimental problems during the screening assay, it may be worth to consider such effect on EngA-ribosome interactions. Depending on how catalysis, conformational changes and association/dissociation events are correlated in the functional cycle of EngA, different scenarios may be expected. We can speculate that inhibition of GTPase activity can avoid EngA from interacting with the ribosome. Allosteric inhibitors, for instance, may induce a conformation that is not prone to bind the ribosome, or even mechanically block the interface for protein-protein interactions. Conversely, inhibitors of GTPase activity may allow this interaction but prevent dissociation of EngA from its binding-site in the ribosome, as GTP hydrolysis is impaired. Protein-protein interactions can also be inhibited by molecules that directly bind key residues or secondary structures that mediate the interface between both partners (Mullard, 2012). In all cases, the ribosome assembly process could be compromised. Hence, it can be valuable to evaluate the potential of different ligands in drug design.

It may also be interesting to crosscheck the effect of each ligand in the interaction of EngA with the 70S, 50S and 30S.1The only structural information available for such complexes regards the *E. coli* 50S subunit. The binding-mode and site of EngA on the 30S is still to be determined. The same applies for binding of EngA to the entire ribosome. The fact that EngA interacts with the 70S and individual subunits may suggest it binds near the interface of 50S and 30S. Indeed, the complex EngA:50S from *E. coli* was recently determined, showing the PTC as the binding-site. However, observation of the structure suggests that the 50S binding-site would not be accessible in the entire 70S ribosome. Hence, another possibility for interactions with the 50S or 30S may be that EngA binds each subunit through distinct

mechanisms. Therefore, ligands may interfere differently with the interactions of EngA and distinct ribosomal partners. The optimisation of the ELISA assay with individual subunits should complement the data obtained with the 70S assay. We have recently started cryo-EM experiments, carried out by Emmanuelle Neumann (IBS, Grenoble), in order to obtain the structure of *B. subtilis* EngA:50S complex. So far, a structure of 6.4 Å has been obtained where EngA seems to be bound to the large ribosomal subunit. The data are being processed to understand if similar mechanisms as the ones observed for *E. coli* are present in *B. subtilis* and possibly unravel some additional aspects of this interaction. Following the same strategy to know the structure of the complex EngA:30S would help us in understanding EngA mechanisms in ribosome assembly and potentially propose a rational for the action of inhibitors of this pathway.

CONCLUSIONS AND PERSPECTIVES

In our work we have studied *B. subtilis* EngA as a target for new antibacterial drugs. In particular, we have addressed the interaction of EngA with the bacterial ribosome in order to identify inhibitors of the ribosome assembly pathway. However, there are still several issues that need to be resolved. When probing interactions of EngA with the 70S ribosomes, individual subunits can be present in the bioassay medium – whose amount is not possible to control – and can possibly affect the result, producing some variability. When using individual ribosomal subunits, the purification protocol is crucial and has to be controlled carefully to assure the presence of a homogeneous sample with intact particles. Nevertheless, targeting these 50S and 30S interactions may be a promising strategy to identify specific inhibitors for the role of EngA in ribosome assembly. A big step forward would be to unveil the exact mechanisms and functions behind the EngA-ribosome interactions. Cryo-EM investigations, as the one we have recently started in collaboration with Emmanuelle Neumann (IBS, Grenoble) in order to unravel the structural bases of these interactions, can potentially provide clues about the role of EngA in the ribosome biogenesis and suggest a rational for the action of inhibitors of this pathway.

Even if the ELISA assay appeared to be a precise and reproducible tool to measure interactions between EngA and the ribosome, problems associated with the nature of the chemical libraries affected the method. The limited solubility of many compounds is a major issue and during these three years, we have tested many solubility strategies in order to improve the behaviour of these inhibitors, with more or less success. In addition, before implementing these strategies in the screening bioassay, such solvents/additives have to be tested on EngA and the ribosome to avoid possible effects on protein stability. These parameters have to be taken into account for future developments.

Structural studies were employed to understand the binding mechanisms of EngA inhibitors, but similarly to the screening assay, solubility issues affected our co-crystallisation trials, limiting the compound concentration used in crystallisation drops. In addition to solubility problems, we are facing the well-known bottleneck of either crystal growth or crystal stability for which trials and errors remain the main strategy.

Due to these challenges in crystallisation, obtaining the structure of the GTP-bound EngA was also not possible during these three years. The combination of different techniques revealed to be of great value to elucidate structural aspects of EngA. Studies in solution, by SAXS and limited proteolysis, allowed us to probe conformational changes that we were so far not able to stabilise in the crystalline state. Crystallography, on the other hand, has provided details of the atomic resolution structure of EngA that revealed some hints on possible mechanistic analyses of nucleotide binding and catalysis. Other valuable

information obtained from our crystal structures provided us clues on how crystallisation may be favouring a specific conformation and how to possibly circumvent this obstacle. While such approach involving crystal contact engineering will certainly require time, it may be a promising solution for obtaining the atomic resolution structure of different conformational states of EngA.

Knowing this conformation would also allow us to propose a mutant to block the GTP-bound form. In the particular scope of developing an inhibitor screening assay, this mutant would be a valuable tool, as this conformation is prone to bind the ribosome. Hopefully, X-ray crystallography or cryo-EM experiments will allow us to obtain the structure of EngA in its ribosome-binding conformation and to proceed with this approach.

BIBLIOGRAPHY

Achila, D., Gulati, M., Jain, N., and Britton, R.A. (2012). Biochemical characterization of ribosome assembly GTPase RbgA in Bacillus subtilis. J. Biol. Chem. *287*, 8417–8423.

Adilakshmi, T., Bellur, D.L., and Woodson, S.A. (2008). Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. Nature *455*, 1268–1272.

Agarwal, N., Pareek, M., Thakur, P., and Pathak, V. (2012). Functional characterization of EngA(MS), a P-loop GTPase of Mycobacterium smegmatis. PloS One *7*, e34571.

Agirrezabala, X., and Frank, J. (2010). From DNA to proteins via the ribosome: structural insights into the workings of the translation machinery. Hum. Genomics *4*, 226–237.

Agrawal, R.K., Sharma, M.R., Yassin, A., Lahiri, I., and Spremulli, L.L. (2011). Chapter 8. Structure and function of organellar ribosomes as revealed by cryo-EM. In Ribosomes: Structure, Function and Dynamics, (Springer Vienna), pp. 83–96.

Alonso, A., Campanario, E., and Martínez, J.L. (1999). Emergence of multidrug-resistant mutants is increased under antibiotic selective pressure in Pseudomonas aeruginosa. Microbiol. Read. Engl. *145 (Pt 10)*, 2857–2862.

Anand, B., Surana, P., Bhogaraju, S., Pahari, S., and Prakash, B. (2009). Circularly permuted GTPase YqeH binds 30S ribosomal subunit: Implications for its role in ribosome assembly. Biochem. Biophys. Res. Commun. *386*, 602–606.

Anand, B., Surana, P., and Prakash, B. (2010). Deciphering the Catalytic Machinery in 30S Ribosome Assembly GTPase YqeH. PLoS ONE *5*.

Audic, S., Lopez, F., Claverie, J.M., Poirot, O., and Abergel, C. (1997). SAmBA: an interactive software for optimizing the design of biological macromolecules crystallization experiments. Proteins *29*, 252–257.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905–920.

Bebbington, C., and Yarranton, G. (2008). Antibodies for the treatment of bacterial infections: current experience and future prospects. Curr. Opin. Biotechnol. *19*, 613–619.

Bharat, A., and Brown, E.D. (2014). Phenotypic investigations of the depletion of EngA in Escherichia coli are consistent with a role in ribosome biogenesis. FEMS Microbiol. Lett. *353*, 26–32.

Bharat, A., Jiang, M., Sullivan, S.M., Maddock, J.R., and Brown, E.D. (2006). Cooperative and critical roles for both G domains in the GTPase activity and cellular function of ribosome-associated Escherichia coli EngA. J. Bacteriol. *188*, 7992–7996.

Bharat, A., Blanchard, J.E., and Brown, E.D. (2013). A high-throughput screen of the GTPase activity of Escherichia coli EngA to find an inhibitor of bacterial ribosome biogenesis. J. Biomol. Screen. *18*, 830–836.

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., and Schäfer, F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. Methods Enzymol. *463*, 439–473.

Blundell T.L., and Johnson Lewis (1994). Protein Crystallography (Academic Press).

Bourne, H.R., Sanders, D.A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. Nature *349*, 117–127.

Britton, R.A. (2009). Role of GTPases in bacterial ribosome assembly. Annu. Rev. Microbiol. *63*, 155–176.

Brown, E.D. (2005). Conserved P-loop GTPases of unknown function in bacteria: an emerging and vital ensemble in bacterial physiology. Biochem. Cell Biol. Biochim. Biol. Cell. *83*, 738–746.

Buckstein, M.H., He, J., and Rubin, H. (2008). Characterization of Nucleotide Pools as a Function of Physiological State in Escherichia coli. J. Bacteriol. *190*, 718–726.

Caldon, C.E., Yoong, P., and March, P.E. (2001). Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function. Mol. Microbiol. *41*, 289–297.

Campbell, T.L., Daigle, D.M., and Brown, E.D. (2005). Characterization of the Bacillus subtilis GTPase YloQ and its role in ribosome function. Biochem. J. *389*, 843–852.

Carroni, M., and Saibil, H.R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. Methods San Diego Calif *95*, 78–85.

CDDEP (2015). The State of the World's Antibiotics 2015 (The Center for Disease Dynamics, Economics & Policy).

Charron, C., Kern, D., and Giegé, R. (2002). Crystal contacts engineering of aspartyl-tRNA synthetase from Thermus thermophilus: effects on crystallizability. Acta Crystallogr. D Biol. Crystallogr. *58*, 1729–1733.

Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. Physiol. Rev. *93*, 269–309.

Corey, R., Naderer, O.J., O'Riordan, W.D., Dumont, E., Jones, L.S., Kurtinecz, M., and Zhu, J.Z. (2014). Safety, Tolerability, and Efficacy of GSK1322322 in the Treatment of Acute Bacterial Skin and Skin Structure Infections. Antimicrob. Agents Chemother. *58*, 6518–6527.

Corrigan, R.M., Bellows, L.E., Wood, A., and Gründling, A. (2016). ppGpp negatively impacts ribosome assembly affecting growth and antimicrobial tolerance in Gram-positive bacteria. Proc. Natl. Acad. Sci. U. S. A. *113*, E1710-1719.

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V.A., Foster, S., Gilmore, B.F., Hancock, R.E.W., Harper, D., et al. (2016). Alternatives to antibiotics-a pipeline portfolio review. Lancet Infect. Dis. *16*, 239–251.

Daigle, D.M., and Brown, E.D. (2004). Studies of the interaction of Escherichia coli YjeQ with the ribosome in vitro. J. Bacteriol. *186*, 1381–1387.

Dale, G.E., Oefner, C., and D'Arcy, A. (2003). The protein as a variable in protein crystallization. J. Struct. Biol. *142*, 88–97.

DeLucas, L.J., Bray, T.L., Nagy, L., McCombs, D., Chernov, N., Hamrick, D., Cosenza, L., Belgovskiy, A., Stoops, B., and Chait, A. (2003). Efficient protein crystallization. J. Struct. Biol. *142*, 188–206.

Di, L., and Kerns, E.H. (2006). Biological assay challenges from compound solubility: strategies for bioassay optimization. Drug Discov. Today *11*, 446–451.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. J. Struct. Biol. *169*, 45–53.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Fabbretti, A., Brandi, L., Petrelli, D., Pon, C.L., Castañedo, N.R., Medina, R., and Gualerzi, C.O. (2012). The antibiotic Furvina® targets the P-site of 30S ribosomal subunits and inhibits translation initiation displaying start codon bias. Nucleic Acids Res. *40*, 10366–10374.

Fabbretti, A., Schedlbauer, A., Brandi, L., Kaminishi, T., Giuliodori, A.M., Garofalo, R., Ochoa-Lizarralde, B., Takemoto, C., Yokoyama, S., Connell, S.R., et al. (2016). Inhibition of translation initiation complex formation by GE81112 unravels a 16S rRNA structural switch involved in P-site decoding. Proc. Natl. Acad. Sci. U. S. A. *113*, E2286-2295.

Feldkamp, M.D., Frank, A.O., Kennedy, J.P., Patrone, J.D., Vangamudi, B., Waterson, A.G., Fesik, S.W., and Chazin, W.J. (2013). Surface reengineering of RPA70N enables cocrystallization with an inhibitor of the replication protein A interaction motif of ATR interacting protein. Biochemistry (Mosc.) *52*, 6515–6524.

Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structure by limited proteolysis. Acta Biochim. Pol. *51*, 299–321.

Foucher, A.-E. (2010). Caractérisation biochimique de YphC, une protéine de. Bacillus subtilis à deux domaines GTPases impliquée dans la biogenèse du ribosome.

Foucher, A.-E., Reiser, J.-B., Ebel, C., Housset, D., and Jault, J.-M. (2012). Potassium acts as a GTPase-activating element on each nucleotide-binding domain of the essential Bacillus subtilis EngA. PloS One *7*, e46795.

GE Healthcare Life Sciences Size Exclusion Chromatography Handbook: Principles and Methods.

GE Healthcare Life Sciences GST Gene Fusion System Handbook.

GE Healthcare Life Sciences Biacore Sensor Surface Handbook.

Gelin, M., Delfosse, V., Allemand, F., Hoh, F., Sallaz-Damaz, Y., Pirocchi, M., Bourguet, W., Ferrer, J.L., Labesse, G., and Guichou, J.F. (2015). Combining “dry” co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. Acta Crystallogr. D Biol. Crystallogr. *71*, 1777–1787.

Gideon, P., John, J., Frech, M., Lautwein, A., Clark, R., Scheffler, J.E., and Wittinghofer, A. (1992). Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity. Mol. Cell. Biol. *12*, 2050–2056.

Goto, S., Muto, A., and Himeno, H. (2013). GTPases involved in bacterial ribosome maturation. J. Biochem. (Tokyo) *153*, 403–414.

Greber, B.J., and Ban, N. (2016). Structure and Function of the Mitochondrial Ribosome. Annu. Rev. Biochem.

Grover, A.K. (2013). Use of allosteric targets in the discovery of safer drugs. Med. Princ. Pract. Int. J. Kuwait Univ. Health Sci. Cent. *22*, 418–426.

Hancock, R.E.W., Nijnik, A., and Philpott, D.J. (2012). Modulating immunity as a therapy for bacterial infections. Nat. Rev. Microbiol. *10*, 243–254.

Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat. Rev. Microbiol. *13*, 298–309.

Henchoz, Y., Bard, B., Guillarme, D., Carrupt, P.-A., Veuthey, J.-L., and Martel, S. (2009). Analytical tools for the physicochemical profiling of drug candidates to predict absorption/distribution. Anal. Bioanal. Chem. *394*, 707–729.

Hörer, S., Reinert, D., Ostmann, K., Hoevels, Y., and Nar, H. (2013). Crystal-contact engineering to obtain a crystal form of the Kelch domain of human Keap1 suitable for ligand-soaking experiments. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. *69*, 592–596.

https://hamptonresearch.com/default.aspx Hampton Research.

http://www.fortebio.com/bli-technology.html ForteBio, BLI Technology.

Hwang, J., and Inouye, M. (2001). An essential GTPase, der, containing double GTP-binding domains from Escherichia coli and Thermotoga maritima. J. Biol. Chem. *276*, 31415–31421.

Hwang, J., and Inouye, M. (2006). The tandem GTPase, Der, is essential for the biogenesis of 50S ribosomal subunits in Escherichia coli. Mol. Microbiol. *61*, 1660–1672.

Hwang, J., and Inouye, M. (2008). RelA functionally suppresses the growth defect caused by a mutation in the G domain of the essential Der protein. J. Bacteriol. *190*, 3236–3243.

Hwang, J., and Inouye, M. (2010). A bacterial GAP-like protein, YihI, regulating the GTPase of Der, an essential GTP-binding protein in Escherichia coli. J. Mol. Biol. *399*, 759–772.

Jacques, D.A., and Trewhella, J. (2010). Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. Protein Sci. Publ. Protein Soc. *19*, 642–657.

Jeon, Y., Ahn, C.S., Jung, H.J., Kang, H., Park, G.T., Choi, Y., Hwang, J., and Pai, H.-S. (2014). DER containing two consecutive GTP-binding domains plays an essential role in chloroplast ribosomal RNA processing and ribosome biogenesis in higher plants. J. Exp. Bot. *65*, 117–130.

Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P.C., and Maddock, J.R. (2006). The Escherichia coli GTPase CgtAE is involved in late steps of large ribosome assembly. J. Bacteriol. *188*, 6757–6770.

Jordana-Lluch, E., Martró Català, E., and Ausina Ruiz, V. (2012). La espectrometría de masas en el laboratorio de microbiología clínica. Enfermedades Infecc. Microbiol. Clínica *30*, 635–644.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kaczanowska, M., and Rydén-Aulin, M. (2007). Ribosome biogenesis and the translation process in Escherichia coli. Microbiol. Mol. Biol. Rev. MMBR *71*, 477–494.

Kalepu, S., and Nekkanti, V. (2015). Insoluble drug delivery strategies: review of recent advances and business prospects. Acta Pharm. Sin. B *5*, 442–453.

Kanjee, U., Ogata, K., and Houry, W.A. (2012). Direct binding targets of the stringent response alarmone (p)ppGpp. Mol. Microbiol. *85*, 1029–1043.

Karbstein, K. (2007). Role of GTPases in ribosome assembly. Biopolymers *87*, 1–11.

Katzung (2006). Basic and Clinical Pharmacology (McGraw Hill Lange).

Kerns, E.H., Di, L., and Carter, G.T. (2008). In vitro solubility assays in drug discovery. Curr. Drug Metab. *9*, 879–885.

Kikhney, A.G., and Svergun, D.I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett. *589*, 2570–2577.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Kriel, A., Bittner, A.N., Kim, S.H., Liu, K., Tehranchi, A.K., Zou, W.Y., Rendon, S., Chen, R., Tu, B.P., and Wang, J.D. (2012). Direct regulation of GTP homeostasis by (p)ppGpp: a critical component of viability and stress resistance. Mol. Cell *48*, 231–241.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Kufareva, I., and Abagyan, R. (2012). Methods of protein structure comparison. Methods Mol. Biol. Clifton NJ *857*, 231–257.

Lafontaine, D.L., and Tollervey, D. (2001). The function and synthesis of ribosomes. Nat. Rev. Mol. Cell Biol. *2*, 514–520.

Lake, J.A. (1976). Ribosome structure determined by electron microscopy of Escherichia coli small subunits, large subunits and monomeric ribosomes. J. Mol. Biol. *105*, 131–139.

Lamb, H.K., Thompson, P., Elliott, C., Charles, I.G., Richards, J., Lockyer, M., Watkins, N., Nichols, C., Stammers, D.K., Bagshaw, C.R., et al. (2007). Functional analysis of the GTPases EngA and YhbZ encoded by Salmonella typhimurium. Protein Sci. Publ. Protein Soc. *16*, 2391–2402.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lee, R., Aung-Htut, M.T., Kwik, C., and March, P.E. (2011). Expression phenotypes suggest that Der participates in a specific, high affinity interaction with membranes. Protein Expr. Purif. *78*, 102–112.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. J. Mol. Biol. *317*, 41–72.

Lin, B., Thayer, D.A., and Maddock, J.R. (2004). The Caulobacter crescentus CgtAC protein cosediments with the free 50S ribosomal subunit. J. Bacteriol. *186*, 481–489.

Loh, P.C., Morimoto, T., Matsuo, Y., Oshima, T., and Ogasawara, N. (2007). The GTP-binding protein YqeH participates in biogenesis of the 30S ribosome subunit in Bacillus subtilis. Genes Genet. Syst. *82*, 281–289.

Long, H., Miller, S.F., Strauss, C., Zhao, C., Cheng, L., Ye, Z., Griffin, K., Te, R., Lee, H., Chen, C.-C., et al. (2016). Antibiotic treatment enhances the genome-wide mutation rate of target cells. Proc. Natl. Acad. Sci. U. S. A. *113*, E2498-2505.

Martínez, J.L. (2012). Natural Antibiotic Resistance and Contamination by Antibiotic Resistance Determinants: The Two Ages in the Evolution of Resistance to Antimicrobials. Front. Microbiol. *3*.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McPherson, A. (2004). Introduction to protein crystallization. Methods San Diego Calif *34*, 254–265.

Mehr, I.J., Long, C.D., Serkin, C.D., and Seifert, H.S. (2000). A homologue of the recombination-dependent growth gene, rdgC, is involved in gonococcal pilin antigenic variation. Genetics *154*, 523–532.

Meyer, S., Wittinghofer, A., and Versées, W. (2009a). G-domain dimerization orchestrates the tRNA wobble modification reaction in the MnmE/GidA complex. J. Mol. Biol. *392*, 910–922.

Meyer, S., Böhme, S., Krüger, A., Steinhoff, H.-J., Klare, J.P., and Wittinghofer, A. (2009b). Kissing G domains of MnmE monitored by X-ray crystallography and pulse electron paramagnetic resonance spectroscopy. PLoS Biol. *7*, e1000212.

Milburn, M.V., Tong, L., deVos, A.M., Brünger, A., Yamaizumi, Z., Nishimura, S., and Kim, S.H. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. Science *247*, 939–945.

Mishra, A.K., and Lambright, D.G. (2016). Small GTPases and their GAPs. Biopolymers.

Mishra, R., Gara, S.K., Mishra, S., and Prakash, B. (2005). Analysis of GTPases carrying hydrophobic amino acid substitutions in lieu of the catalytic glutamine: implications for GTP hydrolysis. Proteins *59*, 332–338.

Mittenhuber, G. (2001). Comparative genomics of prokaryotic GTP-binding proteins (the Era, Obg, EngA, ThdF (TrmE), YchF and YihA families) and their relationship to eukaryotic GTP-binding proteins (the DRG, ARF, RAB, RAN, RAS and RHO families). J. Mol. Microbiol. Biotechnol. *3*, 21–35.

Morimoto, T., Loh, P.C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S., and Ogasawara, N. (2002). Six GTP-binding proteins of the Era/Obg family are essential for cell growth in Bacillus subtilis. Microbiol. Read. Engl. *148*, 3539–3552.

Muench, S.P., Xu, L., Sedelnikova, S.E., and Rice, D.W. (2006). The essential GTPase YphC displays a major domain rearrangement associated with nucleotide binding. Proc. Natl. Acad. Sci. U. S. A. *103*, 12359–12364.

Mullard, A. (2012). Protein-protein interaction inhibitors get into the groove. Nat. Rev. Drug Discov. *11*, 173–175.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Naganathan, A., and Moore, S.D. (2013). Crippling the essential GTPase Der causes dependence on ribosomal protein L9. J. Bacteriol. *195*, 3682–3691.

Nelson, D.L., and Cox, M.M. (2005). Lehninger, Principles of Biochemistry.

Ni, X., Davis, J.H., Jain, N., Razi, A., Benlekbir, S., McArthur, A.G., Rubinstein, J.L., Britton, R.A., Williamson, J.R., and Ortega, J. (2016). YphC and YsxC GTPases assist the maturation of the central protuberance, GTPase associated region and functional core of the 50S ribosomal subunit. Nucleic Acids Res.

Nicastro, G., Taylor, I.A., and Ramos, A. (2015). KH-RNA interactions: back in the groove. Curr. Opin. Struct. Biol. *30*, 63–70.

Nierhaus, K.H., and Dohme, F. (1974). Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *71*, 4713–4717.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. *2*, 2212–2221.

Nobelprize.org “Sir Alexander Fleming - Nobel Lecture: Penicillin”. Nobel Media AB 2014. Web. 5 Feb 2016. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1945/fleming-lecture.html>.

Nowotny, V., and Nierhaus, K.H. (1988). Assembly of the 30S subunit from Escherichia coli ribosomes occurs via two assembly domains which are initiated by S4 and S7. Biochemistry (Mosc.) *27*, 7051–7055.

Nussinov, R., and Tsai, C.-J. (2013). Allostery in disease and in drug discovery. Cell *153*, 293–305.

Paduch, M., Jeleń, F., and Otlewski, J. (2001). Structure of small G proteins and their regulators. Acta Biochim. Pol. *48*, 829–850.

Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. EMBO J. *9*, 2351–2359.

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. *20*, 660–664.

Rafay, A., Majumdar, S., and Prakash, B. (2012). Exploring potassium-dependent GTP hydrolysis in TEES family GTPases. FEBS Open Bio *2*, 173–177.

Ramakrishnan, V. (2009). The ribosome: some hard facts about its structure and hot air about its evolution. Cold Spring Harb. Symp. Quant. Biol. *74*, 25–33.

Rambo, R.P., and Tainer, J.A. (2011). Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law. Biopolymers *95*, 559–571.

Robinson, V.L., Hwang, J., Fox, E., Inouye, M., and Stock, A.M. (2002). Domain arrangement of Der, a switch protein containing two GTPase domains. Struct. Lond. Engl. 1993 *10*, 1649–1658.

Satow, Y., Cohen, G.H., Padlan, E.A., and Davies, D.R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 A. J. Mol. Biol. *190*, 593–604.

Saylor, C., Dadachova, E., and Casadevall, A. (2009). Monoclonal antibody-based therapies for microbial diseases. Vaccine *27 Suppl 6*, G38-46.

Schaefer, L., Uicker, W.C., Wicker-Planquart, C., Foucher, A.-E., Jault, J.-M., and Britton, R.A. (2006). Multiple GTPases participate in the assembly of the large ribosomal subunit in Bacillus subtilis. J. Bacteriol. *188*, 8252–8258.

Schmeing, T.M., and Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. Nature *461*, 1234–1242.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H.D. (2005). Structures of the bacterial ribosome at 3.5 A resolution. Science *310*, 827–834.

Scrima, A., and Wittinghofer, A. (2006). Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. EMBO J. *25*, 2940–2951.

Servier image bank Servier image bank.

Shajani, Z., Sykes, M.T., and Williamson, J.R. (2011). Assembly of bacterial ribosomes. Annu. Rev. Biochem. *80*, 501–526.

Sharma, M.R., Barat, C., Wilson, D.N., Booth, T.M., Kawazoe, M., Hori-Takemoto, C., Shirouzu, M., Yokoyama, S., Fucini, P., and Agrawal, R.K. (2005). Interaction of Era with the 30S ribosomal subunit implications for 30S subunit assembly. Mol. Cell *18*, 319–329.

Signor, L., and Boeri Erba, E. (2013). Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometric analysis of intact proteins larger than 100 kDa. J. Vis. Exp. JoVE.

Speicher, K.D., Gorman, N., and Speicher, D.W. (2009). N-terminal sequence analysis of proteins and peptides. Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al *Chapter 11*, Unit11.10.

Spellberg, B., Bartlett, J.G., and Gilbert, D.N. (2013). The Future of Antibiotics and Resistance. N. Engl. J. Med. *368*, 299–302.

Sprang, S.R. (1997). G protein mechanisms: insights from structural analysis. Annu. Rev. Biochem. *66*, 639–678.

Strickley, R.G. (2004). Solubilizing Excipients in Oral and Injectable Formulations. Pharm. Res. *21*, 201–230.

Suwastika, I.N., Denawa, M., Yomogihara, S., Im, C.H., Bang, W.Y., Ohniwa, R.L., Bahk, J.D., Takeyasu, K., and Shiina, T. (2014). Evidence for lateral gene transfer (LGT) in the evolution of eubacteria-derived small GTPases in plant organelles. Front. Plant Sci. *5*, 678.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. *25*, 495–503.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Crystallogr. *28*, 768–773.

Tan, J., Jakob, U., and Bardwell, J.C.A. (2002). Overexpression of two different GTPases rescues a null mutation in a heat-induced rRNA methyltransferase. J. Bacteriol. *184*, 2692–2698.

Tomar, S.K., Dhimole, N., Chatterjee, M., and Prakash, B. (2009). Distinct GDP/GTP bound states of the tandem G-domains of EngA regulate ribosome binding. Nucleic Acids Res. *37*, 2359–2370.

Tomar, S.K., Kumar, P., Majumdar, S., Bhaskar, V., Dutta, P., and Prakash, B. (2012). Extended C-terminus and length of the linker connecting the G-domains are species-specific variations in the EngA family of GTPases. FEBS Open Bio *2*, 191–195.

Traub, P., and Nomura, M. (1968). Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. Proc. Natl. Acad. Sci. U. S. A. *59*, 777–784.

Tu, C., Zhou, X., Tropea, J.E., Austin, B.P., Waugh, D.S., Court, D.L., and Ji, X. (2009). Structure of ERA in complex with the 3' end of 16S rRNA: implications for ribosome biogenesis. Proc. Natl. Acad. Sci. U. S. A. *106*, 14843–14848.

Uicker, W.C., Schaefer, L., and Britton, R.A. (2006). The essential GTPase RbgA (YlqF) is required for 50S ribosome assembly in Bacillus subtilis. Mol. Microbiol. *59*, 528–540.

Uicker, W.C., Schaefer, L., Koenigsknecht, M., and Britton, R.A. (2007). The essential GTPase YqeH is required for proper ribosome assembly in Bacillus subtilis. J. Bacteriol. *189*, 2926–2929.

Ventola, C.L. (2015). The Antibiotic Resistance Crisis. Part 2: Management Strategies and New Agents. Pharm. Ther. *40*, 344–352.

Verstraeten, N., Fauvart, M., Versées, W., and Michiels, J. (2011). The universally conserved prokaryotic GTPases. Microbiol. Mol. Biol. Rev. MMBR *75*, 507–542, second and third pages of table of contents.

Vetter, I.R., and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. Science *294*, 1299–1304.

WHO (2014). Antimicrobial Resistance: Global Report on Surveillance.

Wicker-Planquart, C., and Jault, J.-M. (2015). Interaction between Bacillus subtilis YsxC and ribosomes (or rRNAs). FEBS Lett. *589*, 1026–1032.

Wilson, D.N. (2009). The A-Z of bacterial translation inhibitors. Crit. Rev. Biochem. Mol. Biol. *44*, 393–433.

Wilson, D.N., and Nierhaus, K.H. (2007). The weird and wonderful world of bacterial ribosome regulation. Crit. Rev. Biochem. Mol. Biol. *42*, 187–219.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327–339.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242.

Wittinghofer, A., and Vetter, I.R. (2011). Structure-function relationships of the G domain, a canonical switch motif. Annu. Rev. Biochem. *80*, 943–971.

Wittinghofer, A., Scheffzek, K., and Ahmadian, M.R. (1997). The interaction of Ras with GTPase-activating proteins. FEBS Lett. *410*, 63–67.

Yusupov, M. (2014). Chapter 2. Recent Progress in Ribosome Structure Studies. In Regulatory Nascent Polypeptides, (Springer), pp. 23–43.

Yusupova, G., and Yusupov, M. (2014). High-resolution structure of the eukaryotic 80S ribosome. Annu. Rev. Biochem. *83*, 467–486.

Zalacain, M., Biswas, S., Ingraham, K.A., Ambrad, J., Bryant, A., Chalker, A.F., Iordanescu, S., Fan, J., Fan, F., Lunsford, R.D., et al. (2003). A global approach to identify novel broad-spectrum antibacterial targets among proteins of unknown function. J. Mol. Microbiol. Biotechnol. *6*, 109–126.

Zhang, X., Yan, K., Zhang, Y., Li, N., Ma, C., Li, Z., Zhang, Y., Feng, B., Liu, J., Sun, Y., et al. (2014). Structural insights into the function of a unique tandem GTPase EngA in bacterial ribosome assembly. Nucleic Acids Res. *42*, 13430–13439.

Abstract

The development of new therapeutics against bacterial infections has aroused great interest over the last years in the context of drug resistance. The starting-point in the pursuit of new antibiotics for which bacterial resistance mechanisms do not exist is the identification of novel cellular targets. Genetics studies in the early 2000s have identified *engA* as a conserved bacterial gene whose product is a GTPase that could represent a potential drug target: it is conserved among bacteria, essential for cell survival, and absent in humans.

Since EngA acts as an assembly factor for the bacterial ribosome, one of our aims was to develop an assay to screen inhibitors of the EngA-ribosome interactions. These interactions are modulated by EngA conformational changes that are in turn triggered by the binding of different nucleotides to the catalytic G-domain. As the interplay between all these events in EngA is still not resolved, we have used a multi-technique approach to explore these questions in order to obtain useful information for the setting up of a robust screening assay.

SAXS and limited proteolysis showed a conformational change occurring in solution upon addition of either di- or tri-phosphate nucleotides. While model validation analysis confirmed the GDP-bound conformation, the GTP-bound state does not match any known EngA structure. Binding studies have revealed modulation of interactions by different nucleotide-bound states. Furthermore, response to nucleotides occurs at high concentrations, suggesting that the role of EngA in promoting ribosome assembly could be monitored by the intracellular nucleotide concentration. Efforts on identifying the GTP-bound state 3D structure by crystallography have resulted in EngA structures in different crystal forms. Although all the obtained structures represent the GDP-bound state, packing analysis has revealed conserved crystal contacts that can potentially stabilise this conformation during nucleation. Specific mutations aiming at disrupting these contacts may help to promote crystallisation of alternative conformations. Cryo-EM investigation has been initiated in order to obtain the structure of the *B. subtilis* EngA:50S complex. So far, an electron density map at 6.4 Å resolution has been obtained and its interpretation is underway.

Résumé

Au cours des dernières années, le développement de nouvelles thérapies contre les infections bactériennes a suscité un grand intérêt face à l'émergence des nombreuses souches résistantes aux antibiotiques. Le point de départ de cette recherche de nouveaux antibiotiques, pour lesquels les bactéries n'ont pas encore acquis de mécanismes de résistance, est l'identification de nouvelles cibles cellulaires. En 2000, des études génétiques ont identifié *engA*, un gène bactérien dont le produit est une GTPase, comme une cible pharmacologique pertinente: elle est essentielle à la survie cellulaire, conservée au sein des bactéries et absente chez les eucaryotes.

Puisque EngA agit comme un facteur d'assemblage pour le ribosome bactérien, un de nos objectifs a été de développer un test de criblage pour identifier des inhibiteurs des interactions EngA-ribosome. Ces interactions sont modulées par des changements conformationnels de EngA, qui sont eux-mêmes déclenchés par la fixation de différents nucléotides dans le domaine catalytique. Cependant, les liens entre ces différents changements restent encore méconnus. Nous avons utilisé une approche multi-technique pour étudier ces questions et obtenir des informations utiles pour l'optimisation de notre test de criblage.

Des analyses de SAXS et protéolyse limitée ont démontré un changement conformationnel en solution après adition de nucléotides di- ou tri-phosphate. La comparaison des données avec des modèles cristallographiques de EngA a confirmé la conformation de la protéine liée au GDP. Cependant, la conformation de la protéine liée au GTP ne correspond à aucune structure connue. Des essais d'interaction ont démontré que la fixation de différents nucléotides au niveau des domaines catalytiques régule l'interaction de EngA avec le ribosome. En outre, les effets des nucléotides se produisent en utilisant des fortes concentrations, ce qui suggère que le rôle de EngA dans la biogenèse du ribosome peut être contrôlé par la concentration intracellulaire de nucléotides. Les travaux visant la détermination de la structure de EngA dans sa conformation liée au GTP par cristallographie nous ont permis d'obtenir la structure d'EngA dans différentes formes cristallines. Cependant, ces structures représentent la conformation liée au GDP. L'analyse de l'empilement des cristaux a montré des contacts intermoléculaires conservés qui peuvent stabiliser cette conformation pendant la nucléation. Des mutations spécifiques permettant la rupture de ces contacts peuvent éventuellement aider à promouvoir la cristallisation de conformations alternatives. Des analyses de cryo-microscopie électronique ont débuté afin d'obtenir la structure du complexe EngA:50S de chez *B. subtilis*. Des résultats préliminaires montrent une carte de densité électronique à 6.4 Å de résolution. L'interprétation de ces résultats est en cours.