Élucidation structurale des facteurs de transcription végétaux à domaines MADS

Sriharsha Puranik

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 7 août 2006

Présentée par

Sriharsha PURANIK

Thèse dirigée par **Chloe ZUBIETA**
codirigée par **Gordon LEONARD**

préparée au sein du **Structural Biology Group, European Synchrotron Radiation Facility**
dans **l'École Doctorale de Chimie et Sciences du Vivant**

Elucidation structurale des facteurs de transcription végétaux à domaines MADS

Structural elucidation of plant MADS domain transcription factors

Thèse soutenue publiquement le **30th May 2016**,
devant le jury composé de :

Dr. Pierre Emmanuel MILHIET	Rapporteur
Dr. Pradeep DAS	Rapporteur
Dr. Darren HART	Président
Dr. Pau BERNADO	Examinateur

Table of Contents

Acknowledgements **4**

Abstract **6**

Résumé en Français **8**

Popularized summary of thesis **10**

I INTRODUCTION **12**

I.1 Summary *12*

I.2 Plant evolution and classification *13*

I.2.1 Role of MADS TFs in plants 14

I.2.2 Molecular mechanisms for flower development 20

I.2.3 MADS TF classification and structure: 22

I.3 Objectives of the thesis *26*

II SEPALLATA3 ..... **28**

II.1 Summary *28*

II.2 Introduction *30*

II.3 Results and Discussion *33*

II.3.1 Construct design and protein purification 33

II.3.2 Biophysical characterisation and DNA-binding studies of $SEP3^{(75\text{-}178)}$ 51

II.4 Conclusions *69*

II.5 Materials and methods *71*

II.5.1 Construct design and purification 71

II.5.2 Biophysical characterisation and DNA binding studies 76

III AGAMOUS **81**

III.1 Summary *81*

III.2 Introduction *82*

III.3 Results and Discussion *84*

III.3.1 Construct design and protein purification 84
III.3.2 Biophysical characterisation of AG$^{(74-173)}$ 95

III.4 Conclusions *104*

III.5 Materials and methods *105*
III.5.1 Construct design and purification 105
III.5.2 Biophysical characterisation 106

IV SHORT VEGETATIVE PHASE 109

IV.1 Summary *109*

IV.2 Introduction *111*

IV.3 Results and Discussion *114*
IV.3.1 Construct design and protein purification 114
IV.3.2 Biophysical characterisation and DNA binding studies of SVP$^{(1-240)}$ 131

IV.4 Conclusions *161*

IV.5 Materials and methods *162*
IV.5.1 Construct design and purification 162
IV.5.2 Biophysical characterization and DNA binding studies 166

V Conclusion 171

Bibliography 173

VI Appendix I 193

VI.1 Library generation and purification techniques *193*
VI.1.1 Multi-vector expression screen 193
VI.1.2 Insect cell expression system at EMBL-Grenoble 203

VI.2 Biophysical techniques *205*
VI.2.1 Atomic force microscopy 205
VI.2.2 Small Angle X-ray Scattering 208

VII Appendix II – Collection of articles 215

Acknowledgements

"Feeling gratitude and not expressing it, is like wrapping a gift and not giving it"

-William Arthur Ward

It is impossible for any work to come to a conclusion without adding a warm note of thanks to all those special people who have been kind enough to encourage me throughout and helped me in completing this dissertation.

Firstly, I would like to express my sincere gratitude to my advisor Dr. Chloe Zubieta. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own and at the same time the guidance to recover when my steps faltered. I would like to thank her for her continuous support during the thesis, for her patience, motivation, immense personal attention and encouragement. I could not have imagined having a better advisor and mentor for my Ph.D. I am also really thankful to my co-supervisor Dr. Gordon Leonard for his stimulating suggestions, optimism, valuable inputs and constructive criticisms during the project and especially during writing of this dissertation. I am also grateful to Dr. Montse Soler-Lopez, for taking out time from her busy schedules and providing insightful comments on my dissertation. Her critical questioning, remarks and suggestions proved to be a helpful in shaping the dissertation. I am grateful to the members of thesis committee, Dr. Francois Parcy, Dr. Renaud Dumas, Dr. Veronique Hugouvieux and Dr. Jean-Luc Pellequer for their motivation and suggestions every year which helped to progress in the right direction.

I wish to express my gratefulness to all the members of the Structural Biology group in ESRF for their unconditional support and help, especially, Samira Acajjaoui for her enormous contributions to the project. I also take this opportunity to express my heartfelt thanks to Dr. Martha Brennich, Dr. Adam Round and Dr. Luca Costa for being resourceful tutors for the subjects which were mystery to me such as SAXS and AFM. Their doors were always open for attending and answering all my naïve queries.

This acknowledgment is incomplete without thanking my friends in Grenoble, especially Lahari and Vipin for the number of drink and dinner meetings which helped me to forget the failed experiments in laboratory; and also Elise and Benedicte for the coffee discussions, for being my French translators and for the memories which I will cherish forever.

Most importantly, none of this would have been possible without the love and patience of my family. I am thankful to my parents Mrs. Medha and Mr. Vivek Puranik for their unconditional love, encouragement and support throughout my PhD thesis work. I admire

them for all of their accomplishments, for their efforts and most importantly for teaching me the value of dedication and hard work. I would like to express my heartfelt appreciation to my wife, Dr. Sakshi Sood for having faith in me and accompanying me on this adventure, I look forward to our next one. Finally, I would like to dedicate this thesis to my late grandparents, Mrs. Pramila and Mr. Gopal Gokhale who played an important role in the development of my identity and shaping the individual that I am today.

I would like to add a note of thanks to all the different laboratories I worked with during this project, especially, the ESRF structural biology laboratory and beamlines (BioSAXS-BM29, ID14-4, ID23-1) , surface science laboratory (Dr. Fabio Comin, ESRF), high throughput protein technologies-ESPRIT (Dr. Darren Hart, IBS/ISBG, Grenoble), eukaryotic expression facility (Dr. Imre Berger, EMBL, Grenoble), cell free expression facility (IBS) and other platforms used as a part of 'Partners in Structural Biology' (PSB).

Abstract

Virtually all terrestrial habitats are dominated by angiosperms, or flowering plants. Their success in colonizing new habitats and supplanting other species is due to the advent of a complex reproductive structure – the flower. The flower unites the male and female organs into one compact structure and encloses the seed. Flowering plants are not only the dominant type of land plants, but also are the primary source of food and habitat for all animals, including humans. In evolutionary terms, flowers are considered a recent development and have been a subject of speculation from the time of Charles Darwin who termed the dominant rise and diversification of flowering plants as "an abominable mystery"* due to the lack of a smooth transition from non-flowering to flowering plants in the fossil record. With the sequencing of multiple genomes from gymnosperms (non-flowering seed plants), basal angiosperms and higher flowering plants, certain gene families have been identified which play a central role in the development and evolution of the flower. My research focuses on one such family of high-level regulators, the MADS transcription factor (TF) family. This TF family helps to orchestrate flower development among other functions. As such, there is great interest in understanding the molecular mechanisms of the MADS family and how these proteins are able to control complex reproductive pathways.

This project integrates different biophysical techniques including x-ray crystallography, small angle x-ray scattering (SAXS) and atomic force microscopy (AFM) to investigate protein-protein and protein-DNA interactions of MADS TFs. No studies to date have investigated the molecular mechanisms of MADS TFs using this integrated structural approach.

One important hurdle in the study of the MADS TFs has been recombinant protein expression and purification. In this project, recombinant purification protocols for several full length MADS TFs were established, allowing the structural and biochemical characterisation of the proteins. The crystal structure of the oligomerisation domain of the MADS family protein SEPALLATA3 (SEP3) is presented and used as a template for understanding the oligomerisation patterns of the larger family and the molecular basis for protein-protein interactions. Investigation of solution structures, derived from SAXS studies, of AGAMOUS (AG) and SHORT VEGETATIVE PHASE (SVP) along with biochemical characterisation of their oligomerisation states are also presented.

**Letter from Charles Darwin to Joseph Dalton Hooker, written 22 July 1879 (Source: Cambridge University Library DAR 95: 485 – 488) (Friedman, 2009b).*

In order to study protein-DNA interactions, complementary methods were used. An important putative property of the MADS TFs is their ability to change the structure of DNA through the formation of DNA loops. MADS TFs are hypothesized to oligomerise and bind DNA at two different sites, potentiating looping of DNA. Using AFM, the first direct evidence of DNA looping by SEP3 is described. The DNA binding characteristics of SVP were studied using electrophoretic mobility shift assay (EMSA), microscale thermophoresis (MST) and AFM. Unlike SEP3, SVP is dimeric and thus exhibits different DNA-binding patterns.

The data presented here provide an atomic and structural basis for MADS TF function. Based on this work, we now are beginning to understand some of the oligomerisation and DNA-binding specificity determinants. These studies demonstrate how the MADS TFs oligomerise and the results show that we can disrupt oligomerisation and potentially DNA-binding very specifically through the introduction of point mutations. Future work will investigate the *in vivo* consequences of altered oligomerisation and how this affects different developmental programs in plant reproduction and floral organ morphogenesis.

Résumé en Français

Virtuellement tous les habitats terrestres sont dominés par les angiospermes, ou plantes à fleurs. Leur capacité à coloniser de nouveaux habitats et supplanter une autre espèce est dûe à l'avènement d'une nouvelle structure reproductrice – la fleur. La fleur uni les organes mâles et femelles dans une structure compacte et contient la graine. Les plantes à fleurs ne sont pas seulement le type dominant des plantes terrestres, mais sont également la principale source de nourriture et l'habitat de tous les animaux, y compris les humains. En termes d'évolution, les fleurs sont considérées comme un développement récent. Elles ont fait l'objet de spéculations depuis l'époque de Charles Darwin qui à nommé l'évolution dominante et la diversification des plantes à fleurs comme «un abominable mystère» en raison de l'absence d'une transition en douceur de la non-floraison vers la floraison des plantes dans le registre fossile. Avec le séquençage de plusieurs génomes de gymnospermes (semences de plantes non-florales), d'angiospermes basals et de plantes à fleurs supérieures, certaines familles de gènes jouant un rôle central dans le développement et l'évolution de la fleur ont été identifiées. Notre recherche se concentre sur une de ces familles de régulateurs de niveau supérieur appelée « famille de facteur de transcription MADS » (TF). Cette famille de TF permet d'orchestrer le développement des fleurs. Nous nous sommes intéressés à la compréhension des mécanismes moléculaires de la famille des MADS et à la façon dont ces protéines sont capables de contrôler les fonctions de reproduction complexes.

Ce projet intègre différentes techniques biophysiques comme la cristallographie aux rayons X, la diffusion des rayons X aux petits angles (SAXS) et la microscopie à force atomique (AFM) afin d'étudier les interactions protéine-protéine et protéine-ADN des FT MADS. Aucune étude n'a, à ce jour, porté sur les mécanismes moléculaires des FT MADS en utilisant cette approche structurale intégrée.

Un obstacle important dans l'étude des FT MADS a été l'expression des protéines recombinantes et leur purification. Dans ce projet, les protocoles de purification de plusieurs recombinants FT MADS entières ont été établis, permettant la caractérisation structurale et biochimique des protéines dans leurs intégralités. La structure aux rayons X du domaine d'oligomérisation de la protéine de la famille MADS, SEPALLATA3 (SEP3) est présenté et utilisé comme modèle pour comprendre les motifs d'oligomérisation de la famille élargie et les bases moléculaires des interactions protéine-protéine. Des solutions de structures

provenant d'études SAXS de AGAMOUS (AG) et de la phase végétative courte (SVP) sont présentées et complétés par la caractérisation biochimique de leur état d'oligomérisation.

Afin d'étudier les interactions protéine-ADN, des procédés complémentaires ont été utilisés. Une propriété importante des FT MADS est leur capacité à modifier la structure de l'ADN grâce à la formation de boucles d'ADN. De manière hypothétique, les FT MADS oligomérisent et fixent l'ADN sur deux sites différents, bouclant potentiellement l'ADN. En utilisant l'AFM, la première preuve directe de la formation de boucle d'ADN par SEP3 est obtenue. Les caractéristiques de liaison d'ADN de SVP ont été étudiées par analyse de décalage de mobilité électrophorétique (EMSA), par thermophorèseà échelle microscopique (MST) et par AFM. Contrairement au cas de SEP3, l'EMSA et l'AFM ont montrés que SVP est un dimère et présente différents modes de liaison à l'ADN.

Ces données fournissent une base atomique et structurale de la fonction des FT MADS. Sur la base de ce travail, nous commençons à comprendre l'oligomérisation et certaines spécificités déterminantes de liaison à l'ADN. Ces études montrent comment les FT MADS s'oligomérisent. De plus, les résultats montrent que nous pouvons très précisément perturber cette oligomérisation et potentiellement la liaison à l'ADN grâce à l'introduction de mutations ponctuelles. Les travaux à venir porteront sur les conséquences de l'altération de l'oligomérisation *in vivo* et son effet sur les différents programmes de développement dans la reproduction des plantes et la morphogenèse des organes floraux.

Popularized summary of thesis

The proliferation of flowering plants (angiosperms) over a short evolutionary period has been credited to the development of a compact reproductive structure - the flower. The significance of flowering plants as the primary source of food and habitat for all animals, including humans, makes the study of their development and evolution an important endeavor. Genomic studies have identified several factors that played significant roles in the evolutionary development of flowers. Some of these factors belong to a class of regulatory proteins called transcription factors that diversified and expanded from early plants to present day higher angiosperms. This project focuses on study of these proteins, called MADS transcription factors, using biophysical and biochemical methods and aims at understanding their molecular mechanism and using this information to understand their control of complex developmental functions *in vivo*.

Résumé vulgarisé de la thèse

La propagation des plantes à fleurs ou angiospermes sur une courte période d'évolution a été attribuée au développement d'une structure reproductive compacte – la fleur. L'importance des plantes à fleurs comme la principale source de nourriture et d'habitat pour tous les animaux, y compris les humains rend l'étude de leur développement et de leur évolution nécessaire. Des études génomiques ont identifié plusieurs facteurs qui jouent un rôle important dans le développement et l'évolution de la fleur. Certains de ces facteurs appartiennent à une classe de protéines régulatrices et sont connus pour avoir évolué et s'être amplifié considérablement depuis les premières plantes jusqu'aux angiospermes supérieurs présents actuellement. Notre projet se concentre sur l'étude de ces protéines appelées facteurs de transcription MADS en utilisant des méthodes biophysiques et biochimiques. Le projet vise à comprendre leur mécanisme moléculaire en rapport au contrôle du développement de leurs fonctions complexes.

Chapter I

INTRODUCTION

I INTRODUCTION

I.1 Summary

The first chapter provides a background and overview to the main topics of this work. The chapter initially introduces the evolution of plants focusing on gymnosperms and angiosperms. The flower, an important factor in the spread of angiosperms is thus discussed with a brief introduction to various regulatory proteins involved in floral development. This chapter elaborates on one of the important groups of regulatory proteins called the MADS transcription factors, which are the subject of the thesis. These are introduced and their importance in plant development is explained with emphasis on their role in floral organ development.

Résumé en Français

Ce premier chapitre introduit de manière approfondie les principaux sujets de ce travail. Le chapitre présente intialement l'évolution des plantes, se concentrant sur les gymnospermes et les angiospermes. La fleur, un facteur important dans la propagation des angiospermes est discuté avec une brève introduction des diverses protéines régulatrices impliquées dans son développement.

En particuler, un groupe de facteurs de transcription appelés facteurs de transcription MADS est introduit et son importance dans le développement de la plante est expliqué en mettant l'accent sur son rôle dans le développement des organes floraux.

I.2 Plant evolution and classification

Plants have evolved over millions of years, with the first land plants being very simple vascular plants such as ferns and clubmosses which developed a vascular tissue to transport sugars, nutrients and water to adapt to life on land. More complex seed-bearing plants, called spermatophytes, quickly displaced these simple plants. One reason for the success of seed-bearing plants is that their reproductive adaptation such as the presence of heterospory, the storage of a fertilized ovule in a protective cover, was in alignment with changing environmental factors. This led to their rapid terrestrial colonisation. The extant seed plants are divided into two sister groups, the gymnosperms, with naked seeds, and the angiosperms in which the seeds are enclosed in a fruit. Gymnosperms include plants such as conifers (pine, fir, spruce and cedar), cycads, gnetophytes and gingko with approximately 1000 extant species. The conifers, which dominate the high altitudes, form a major group of gymnosperm species (Willis and McElwain, 2013). Gymnosperms are mostly evergreen plants with naked seeds and no flowers. They are unisexual and rely primarily on wind for pollination. Angiosperms, on the other hand, are seasonal flowering plants with enclosed seeds. They can be unisexual or bisexual and take advantage of various agents like insects, animals and wind for pollination. Gymnosperms are used widely for lumber and gum while angiosperms form the main source of food for humans and livestock alike.

Fossil records show the presence of early forms of angiosperms even in the mid Cretaceous (120-90 Ma) period and a steadily increasing diversity allowed them to quickly become ecologically significant over a short period of 20-30 Ma (Friis et al., 2010). There have been several hypotheses which consider the role of higher relative growth, improved synthetic capacity, better pollination systems and higher diversification as the prime factors for the spread of angiosperms (Augusto et al., 2014). Among them, the highly developed compact reproductive organ, the flower, is considered one of the most important factors. Although recent paleontological evidence from the early Cretaceous period (145-120 Ma) has tried to pinpoint flower origins, it still fails to explain the sudden diversification of angiosperms (Friedman, 2009a). In order to investigate flower origins, it is important to examine the role of different regulatory pathways and the proteins that direct them.

Most of the plant life processes such as flowering, seed germination, root and fruit development are controlled by a set of regulatory proteins called transcription factors (TFs). One of the key families involved in all these processes is the MADS TF family. While the

MADS TFs are present in all eukaryotes, they are a small and limited family in animals and fungi. However, in plants the family has undergone a significant expansion.

I.2.1 Role of MADS TFs in plants

MADS TFs play important roles in almost all fundamental plant life processes including seed germination, root development, fruit development, and, perhaps most notably, flowering and floral organ development (Rounsley et al., 1995; Alvarez-Buylla et al., 2000a; Shore and Sharrocks, 1995; Favaro et al., 2003; Angenent and Colombo, 1996; Bowman et al., 1989) (Figure I-1). Our current understanding of MADS TF function is based largely on extensive genetic studies of loss-of-function mutants of floral homeotic genes. In addition, the identification of targets for MADS has developed empirically from early days of transcriptome analysis to present day techniques like chromatin immunoprecipitation (ChIP) coupled with high throughput sequencing (ChIP-seq) and yeast n-hybrid assays resulting in a dramatic increase in the knowledge of MADS TF targets as well as their interacting partners (Becker and Ehlers, 2015; Mara and Irish, 2008; Kaufmann et al., 2009; ÓMaoiléidigh et al., 2013; de Folter et al., 2005).

In *Arabidopsis thaliana* at least 107 regulatory MADS-box genes have been identified (Parenicová et al., 2003). These MADS genes can be divided in to two main types- type I and type II which can be further sub-classified. MADS type I genes encode a DNA binding domain and a C-terminal domain, whereas type II encode proteins that have a DNA binding domain, oligomerisation domains and a C-terminal domain.

Several genetic studies have elucidated the role of MADS type I genes in female gametogenesis and in seed development. Some of them include *AGL80 (AGAMOUS LIKE 80)* (Portereiko et al., 2006; Steffen et al., 2008) and *DIANA* (Bemer et al., 2008) involved in central cell and endosperm development, *AGL23 (AGAMOUS LIKE 23*) (Colombo et al., 2008a, 2008b) involved in embryo sac development and *PHERES1* (Köhler et al., 2003, 2005) in seed development (Masiero et al., 2011).

The post-embryonic period in plants forms the major developmental and growth phase marking the transition into vegetative and, later, the reproductive phase. These developmental transitions are controlled by mainly type II encoded MADS TFs whose expression depends on different external and internal cues such as plant age, light, temperature, and autonomous and hormonal pathways (Blazquez, 2000; Poethig, 2003). A few examples of the regulatory functions controlled by the type II transcription factors in *Arabidopsis thaliana* are provided

in Table I-1. Due to the numerous studies related to the role of type II MADS TFs in plant and flower development, only a few which are relevant to the dissertation are mentioned in the table.

MADS TFs have a wide range of functions throughout the plant lifecycle, but their role in reproductive development, specifically flowering and floral organ development, is one of the best characterized (Figure 1.1). The activity of transcription factors is often dependent on the formation of multi-protein complexes resulting in either expression or repression of target genes. MADS TFs form homodimeric and heteromeric complexes with other MADS TFs as well as with proteins of other classes to regulate floral organ morphogenesis (Theissen and Saedler, 2001; Honma and Goto, 2001). By organizing themselves in different levels of interactions, the floral homeotic MADS TFs directly or indirectly regulate the expression or repression of a variety of genes that are significant for the specification of floral organ identity during initial stages as well as differentiation and development of floral organs at later stages (reviewed by (Dornelas et al., 2011; Ito, 2011).

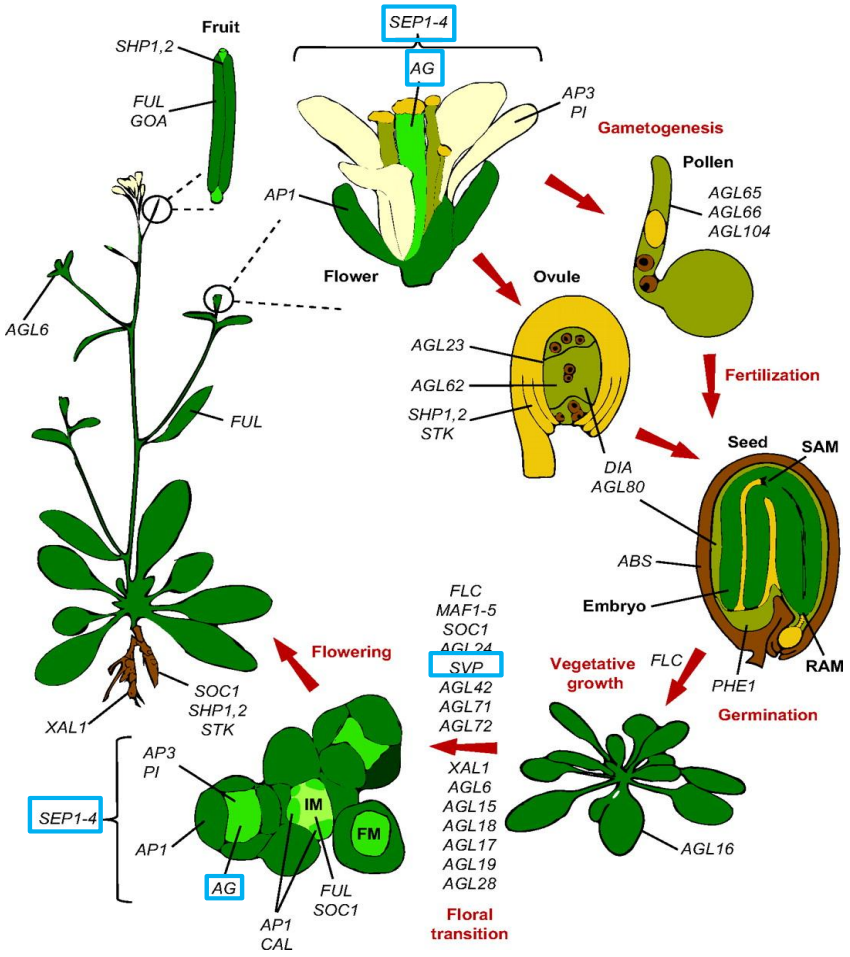


Figure I-1: ***MADS-box genes functions in different stages of Arabidopsis thaliana life cycle****. The proteins encoded by SEP(3), AG, SVP genes marked with blue box are characterized in this dissertation.* (Smaczniak et al., 2012a)

Protein	Function	Reference
FLOWERING LOCUS C (FLC)	Repressor of flowering	(Michaels, 1999)
APETALA1 (AP1)	Sepal and petal development, floral meristem identity	(Mandel et al., 1992; Gustafson-Brown et al., 1994)
AGAMOUS (AG)	Carpel and stamen development	(Yanofsky et al., 1990)
APETALA3 (AP3)	Petal and stamen specification	(Jack et al., 1992)
PISTILLATA (PI)	Petal and stamen specification	(Goto and Meyerowitz, 1994)
SHORT VEGETATIVE PHASE (SVP)	Flowering repressor	(Hartmann et al., 2000)
SEPALLATA 1, SEPALLATA 2, SEPALLATA 3, SEPALLATA 4; (SEP 1-4)	Sepal, stamen and carpel specification	(Pelaz et al., 2000; Ditta et al., 2004; Smaczniak et al., 2012a)
SOC1	Regulator of floral transition	(Samach et al., 2000)

Table I-1: ***Functions of MADS type II transcription factors.*** *MADS TFs are involved in variety of plant developmental processes mainly flower organ formation.*

I.2.1.1 Flower as a reproductive structure; an overview

A typical angiosperm flower (Figure I-2) consists of four concentric whorls. From outer to inner, they are composed of green leaf like protective sepals, followed by the showy organs called petals, which attract pollinators, followed by stamens, which produce pollen that contains the male gametophyte and the innermost carpel, which contains the ovules or female gametophyte. There is a great deal of variation in structure, colour and scent of flowers specialized for promoting pollination. Some plants even show a slight modification in this basic structure with respect to the number and arrangement of floral organs and in some cases fused organs, for example, the presence of tepals, a combination of sepals and petals.

One of the central questions in plant evolutionary developmental biology is how the compact bisexual reproductive flower evolved and what are the molecular mechanisms controlling its formation.

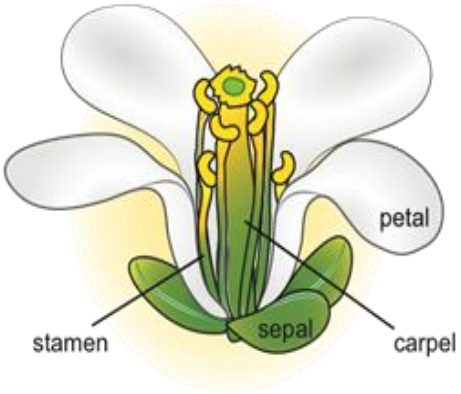


Figure I-2: ***Parts of a typical flower.*** *A typical angiosperm flower consists of four whorls namely sepal, petal, carpel and stamen.*

I.2.1.2 Flower development

As described above in Figure I-2 above, the flower consists of four main parts, variations in which strategically affect the features such as pollination and seed dispersion. The development of flowers is a complex process, and is influenced by external factors and the interaction of numerous regulatory proteins including transcription factors. Figure I-3 gives an overview of the genes involved in flower development.

In plants, leaves, stems and flowers are all products of stem cells in the shoot apical meristem (SAM) (Tucker and Laux, 2007; Steeves and Sussex, 1989). The SAM gives rise to the inflorescence meristem (IM), and the floral meristem (FM) on its flanks, during reproductive development. The main steps in flower formation are floral induction, floral meristem formation and floral organ development. This process has been divided in to 12 stages on the basis of specific morphological parameters (Smyth et al., 1990).

Floral induction depends on plant age, hormone levels such as gibberellic acid and auxin concentration, vernalisation, photoperiod and temperature (Srikanth and Schmid, 2011). These factors trigger the initial flowering time genes as shown in the Figure I-3. Activation of the flowering time genes acts as a switch between the vegetative phase and reproductive phase. During the vegetative phase, SHORT VEGETATIVE PHASE (SVP) acts as a repressor of flowering, by regulating the expression of *FLOWERING LOCUS T (FT), TWIN SISTER OF FT (TSF)* and *SUPPRESOR OF OVEREXPRESSION OF CONSTANS 1(SOC1)* (Lee et al., 2007; Li et al., 2008; Jang et al., 2009). The floral transition marks the conversion of the SAM to the inflorescence meristem (IM). Meristem identity genes function at the interface between the floral transition and floral organogenesis. Floral organ initiation begins with the transition of the IM to the FM. All floral organs develop from the floral meristem. In Arabidopsis, the floral meristem identity genes such as *LEAFY (LFY), APETALA 1 (AP1)* and *CAULIFLOWER (CAL*) act as master regulators (Chandler, 2012). The floral meristem maintains a tight regulation of the number and position of stem cells. The regulatory proteins WUSCHEL (WUS), SHOOT MERISTEMLESS (STM) and CLAVATA (CLV) play a role in maintenance of stem cells (Endrizzi et al., 1996; Sun and Ito, 2010). WUS directly activates *AGAMOUS (AG)* thus also playing a role in floral organ initiation while AG in a negative feedback loop represses *WUS* (Lenhard et al., 2001). The expression of *WUS* is also repressed by CLV. During the early floral stages (1-6), the CLV-WUS pathways regulate and maintain stem cell numbers while at floral stage 6, the AG-WUS pathway acts to terminate stem cell activity (Sun and Ito, 2010). Cells on the uppermost layer of the floral meristem develop into the floral primordial, starting from the outer whorl of sepals and ending with the innermost carpel primordia (Bossinger and Smyth, 1996; Smyth et al., 1990). The formation of floral primordia marks the activation of floral organ identity genes. They form complex networks for activation of specification and differentiation of floral organs (Ó'Maoiléidigh et al., 2014). The mechanisms proposed for formation of different floral organs are described later in this chapter.

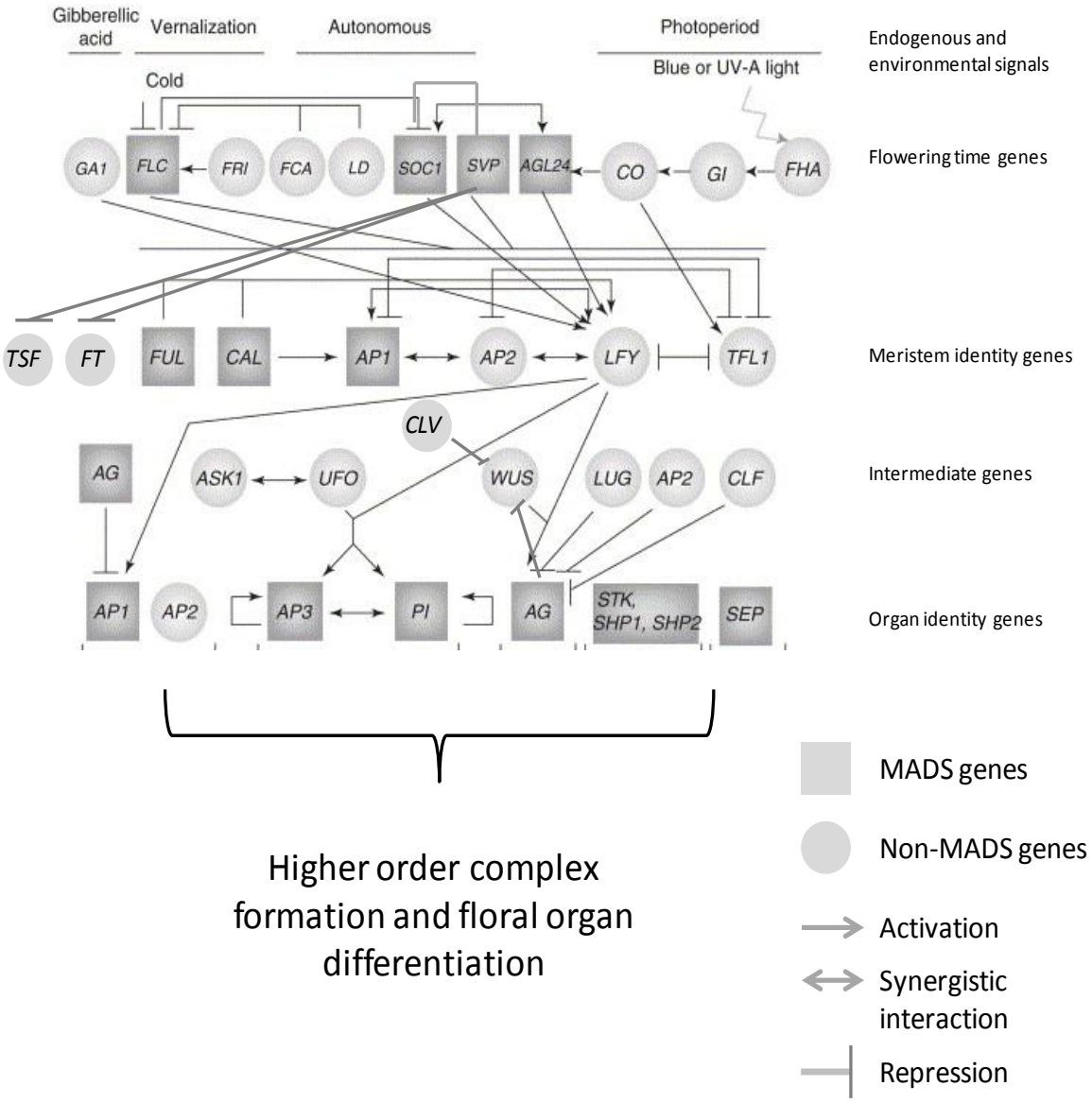


Figure I-3: ***Regulatory cascades in flower development.*** *Flowering time is influenced by endogenous and environmental factors. Interaction between different genes happens at various stages of flower development leading to a cascade which ultimately leads to formation of floral organs. Only a few representative genes are shown here. Adapted from* (Kaufmann et al., 2005) *and* (Soltis et al., 2002).

I.2.2 Molecular mechanisms for flower development

I.2.2.1 ABC(D)E model

The floral organ development genes focused in this work have been extensively studied. Several models have been proposed for floral organ development controlled by MADS of which the ABC(D)E model is the most widely accepted (Theissen and Saedler, 2001; Angenent and Colombo, 1996; Pinyopich et al., 2003; Pelaz et al., 2000; Bowman et al., 1991; Coen and Meyerowitz, 1991). As shown in the Figure I-4, this model categorises

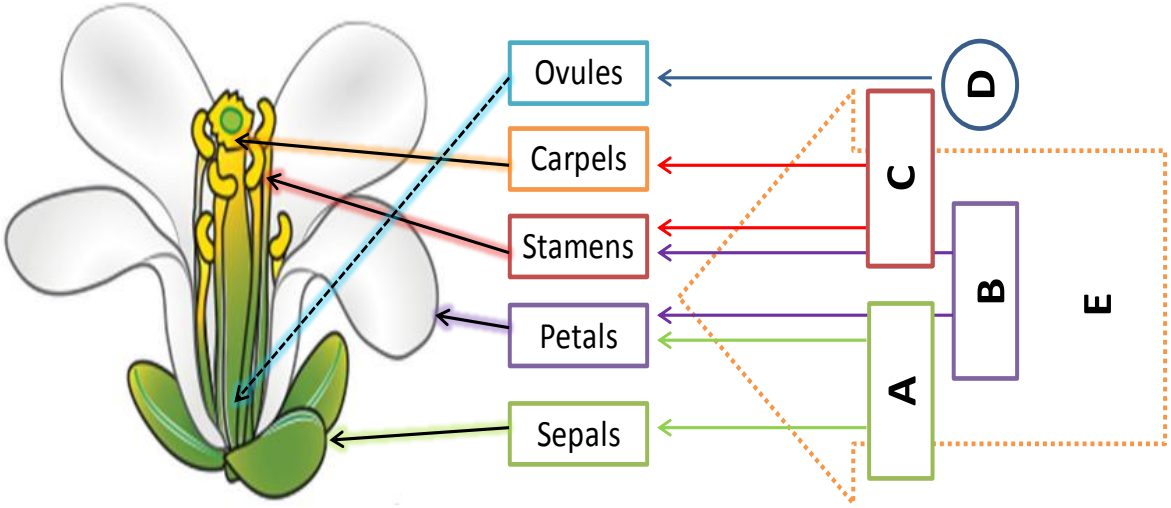


Figure I-4: ***The ABC(D)E model for floral organ development.*** *The floral regulator genes are divided into 5 classes, A, B, C, D and E. The arrows point to the floral organ specified by each class of genes. Some floral organs are specified by a combination of different class of genes. Class E genes are involved in formation of all floral organs.*

MADS genes into five types of functional genes types, A, B, C, D and E- whose overlapping expression patterns are necessary to produce distinct floral organs. The A class genes code for sepal organ identity; A and B together encode petal identity; B and C continued determine stamen organ identity; C alone establishes carpel identity; D class genes are ovule specific; E class genes are required for formation of all floral organs. In *A. Thaliana, APETALA1 (AP1)* and *APETALA2 (AP2)* are A class genes, B class genes are represented by *APETALA3 (AP3)* and *PISTILLATA (PI)*, the C class by *AGAMOUS (AG)* while the *SEPALLATA (SEP1-4)* group are E class genes. The D class function is performed by the combinatorial action of *SHATTERPROOF 1, 2 (SHP1, SHP2)* and *SEEDSTICK (STK)*. All these genes encode MADS TFs with the exception of *AP2*.

I.2.2.2 The floral quartet model

The floral quartet model was postulated to explain the molecular mechanism of action underlying ABCDE protein function in floral organ specification (Theissen and Saedler, 2001; Theissen, 2001).

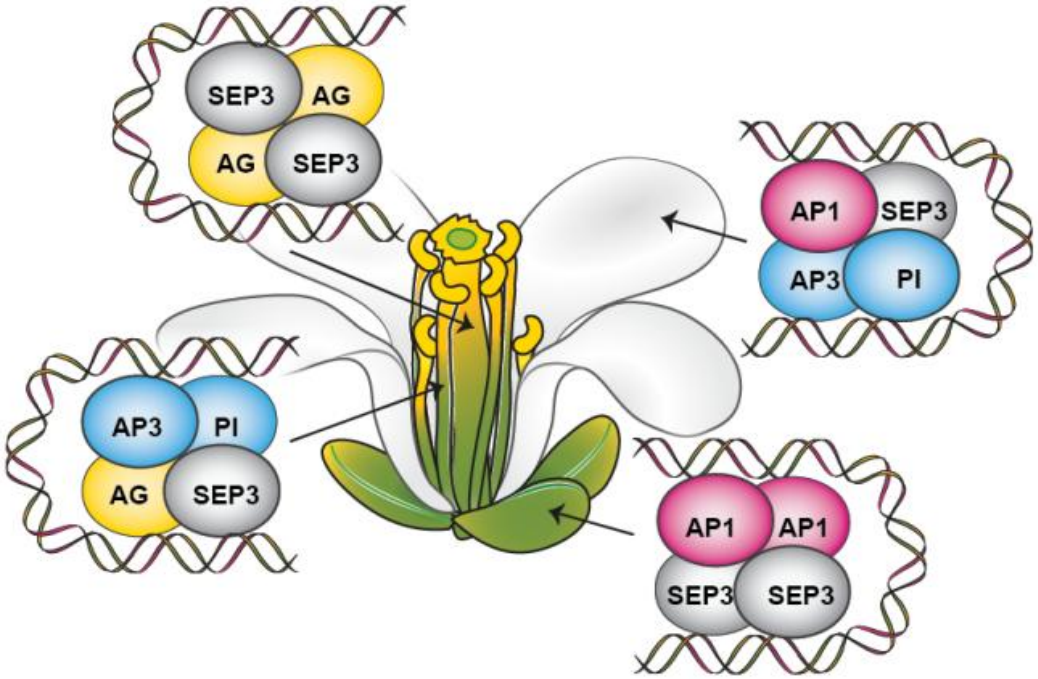


*Figure I-5: The **floral quartet model.** Schematic representation of floral MADS TF tetramers binding to DNA. Different MADS TF tetramers trigger the development of the different floral organs. SEP3, AG, AP1, AP3 and PI are depicted as ovals with SEP3 in gray, AG in yellow, AP1 in pink and AP3/PI in blue. A DNA loop is depicted for each tetramer. Arrows indicate the tetrameric MADS complex that triggers the development of each floral organ (sepals, petals, stamen and carpels).*

The quartet model postulates that two MADS dimers form a tetrameric complex and co-operatively bind specific sites (called CArG boxes) on DNA to regulate the expression of their target genes. As shown in Figure I-5 for Arabidopsis, AP1 and SEP3 homodimer, tetramerise to code for sepals, petals are result of a heterotetramers involving, AP1, AP3, PI and a SEP, stamens are product of heterotetramers of AP3, PI, AG and a SEP protein while AG and SEP homodimers combine to generate carpels.

However, despite the wealth of information about the biological functions of plant MADS function from genetic studies, the molecular mode of action at the protein level is still largely unknown.

I.2.3 MADS TF classification and structure:

Evolution based studies, genome sequencing and computational methods link the origin of the MADS domain to the coding region of DNA- topoisomerase II (Gramzow and Theissen, 2010). The name MADS comes from four proteins in four different species namely MINICHROMOSOME MAINTENANCE 1 (MCM1) from yeast, AGAMOUS (AG) from Arabidopsis, DEFICIENS (DEF) from Antirrhinum and SERUM RESPONSE FACTOR (SRF) from humans which possess the same DNA binding properties of "MADS" domain TFs (Norman et al., 1988; Yanofsky et al., 1990; Passmore et al., 1988; Schwarz-Sommer et al., 1990).

In plants, the MADS family is divided in to two distinct lineages, as mentioned previously- type I and type II based on protein domain structure (Alvarez-Buylla et al., 2000b) (Figure I-6). The type I and type II MADS-box genes of plants have been hypothesized to be orthologous to the SRF-like and MEF2-like MADS-box genes from animals and fungi respectively. The type I genes can be further classified in to three sub classes Mα, Mβ and Mγ on the basis of their phylogeny and the presence or absence of conserved motifs in the carboxy-terminal region of the encoded proteins (De Bodt et al., 2003; Parenicová et al., 2003). The proteins encoded by type I genes contain a ~60 residue MADS domain and a highly variable C-terminal domain and have recently been shown to have roles in female gametophyte, embryo and endosperm development and are less well studied than type II MADS TFs (Masiero et al., 2002; Gramzow et al., 2010).

The type II lineage contains the floral homeotic genes. These genes encode for proteins with a modular domain structure. Type II MADS TFs comprise four domains with well-studied functions; an N-terminal DNA binding MADS domain, an I (Intervening) domain responsible for dimerisation specificity, a K (Keratin-like coiled-coil) domain important for both dimerisation and tetramerisation and a C-terminal domain essential for transactivation and higher order complex formation (Yang et al., 2003; Kaufmann et al., 2005; van Dijk et al., 2010; Smaczniak et al., 2012b).

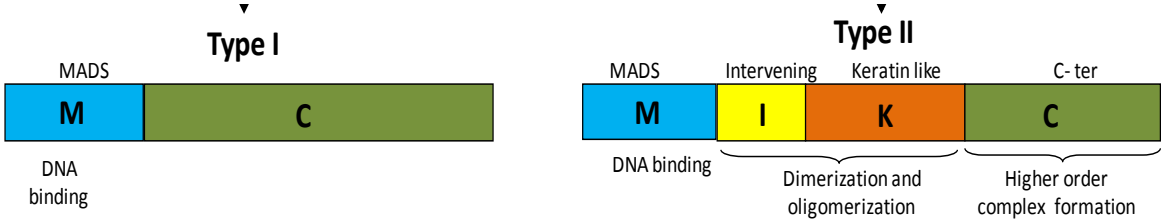


*Figure I-6: C**lassification of MADS TFs.** The MADS TFs are divided in to two main types depending on their domain structure. The different domains are colour coded.*

The type II genes can be further classified in to MIKC^C and MIKC* types depending on the structure of the K domain (Kwantes et al., 2012; Henschel et al., 2002). The MIKC^C subclass can be further divided in to numerous sub-families on the basis of their phylogeny. These sub-families display distinct C-terminal motifs which can influence their functional characteristics (Piwarzyk et al., 2007). In this dissertation, I will mainly focus on a few of the MIKC and sub-family MIKC^C transcription factors.

The majority of plants MADS-domain TFs are part of the MIKC group. Among the four domains, the MADS domain is the most conserved and is required for DNA binding as well as protein dimerisation (Theissen et al., 1996). The currently available structural data is limited to the isolated MADS DNA-binding domains from human proteins MEF2A (Wu et al., 2010; Santelli and Richmond, 2000) and SRF (Pellegrini et al., 1995; Mo et al., 2001) and the yeast protein MCM1 (Tan and Richmond, 1998) (Figure I-7). Plants have elaborated on the eukaryotic MADS domain with plant specific domains, enlarging the repertoire of TF function for this family.

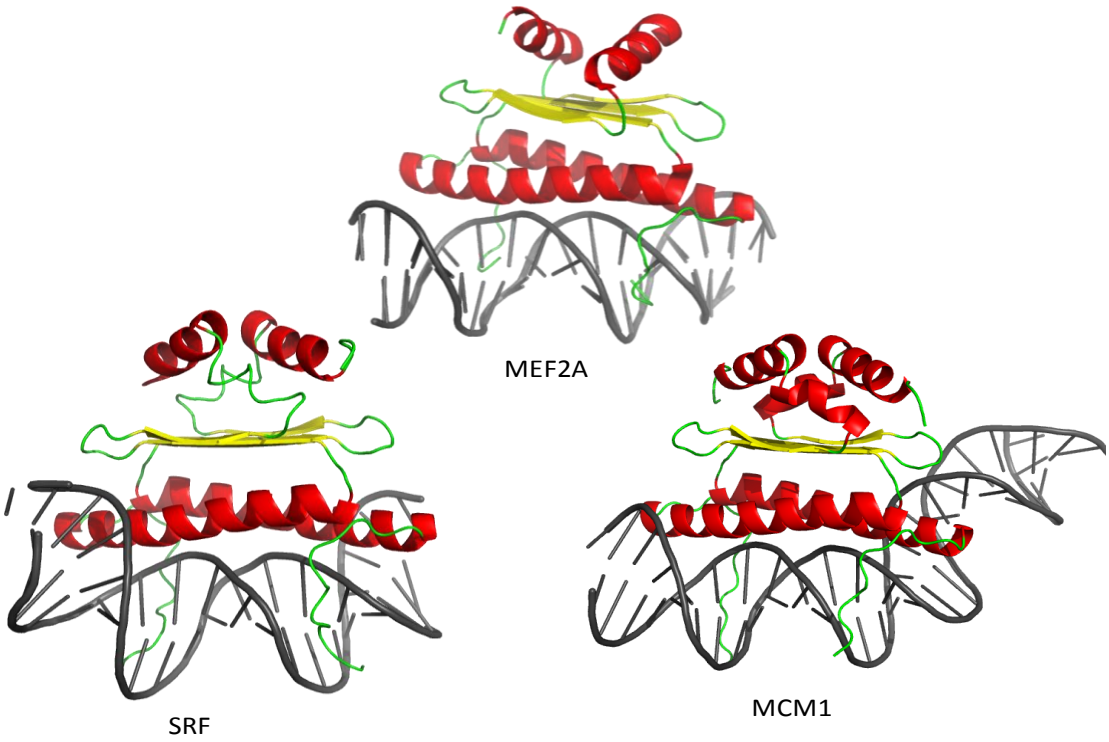


Figure I-7: ***Crystal structures of DNA bound MADS domain of mammalian and yeast MADS TFs.*** *The MADS domains for SRF (PDB 1SRS), MEF2A (PDB 1EGW) and MCM1 (PDB 1MNM1) show very similar structure. The three layer organization consists of a long α helix a middle two β strands followed by a small α helix at the top. The DNA bending is observed to be different with SRF and MCM1 bending the DNA to almost 72° while MEF2A only induces a 17° bend.* (Wu et al., 2010; Santelli and Richmond, 2000; Pellegrini et al., 1995; Mo et al., 2001; Tan and Richmond, 1998)

The N-terminal region of the MADS domain is predominantly hydrophilic with a high proportion of basic residues whereas the C-terminal region is predominantly hydrophobic. Study of different MADS domains has revealed a structural pattern in which the MADS domain sequence folds into a DNA binding α-helix and two antiparallel β strands which are involved in dimerisation. During dimerisation, the α-helices fold into an antiparallel coiled-coil and expose the basic amino acids for DNA binding (Huang et al., 2000; Pellegrini et al., 1995; Santelli and Richmond, 2000). Using this fundamental DNA-recognition domain, plants have added new domains to expand the functional diversity of the MADS TFs.

The I domain plays a role in dimer formation and specificity (Riechmann et al., 1996; Kaufmann et al., 2005). This short domain of ~ 30 residues is predicted to have some helical structure and acts primarily as a linker between the M and K domains. The K domain is the second most conserved domain among MADS TFs after the M domain and is involved in the formation of higher-order complexes with other MADS TFs (Egea-Cortines et al., 1999; Honma and Goto, 2001; Melzer et al., 2009) and the dimerisation and tetramerization of MADS TFs (Yang et al., 2003). The K domain has been postulated to be divided in to three subdomains K_1, K_2 and K_3 (Yang and Jack, 2004; Yang et al., 2003) comprised of amphipathic α-helices which form a hydrophobic interaction surface. Some differences between the role of subdomains during protein-protein interactions within MADS TFs have been noted (Riechmann et al., 1996; Yang et al., 2003; Yang and Jack, 2004). The K domain is predicted to interact with partner proteins via hydrophobic residues (Yang et al., 2003) which are predominantly located at the a and d positions of its characteristic leucine-zipper like heptad repeats $(abcdefg)_n$ (Fan et al., 1997; Yang et al., 2003). The C-terminal domain of MADS TFs is the least conserved domain and is predicted to be mostly unstructured. It is postulated to play a role in enhancing and stabilizing the interactions mediated by the K-domain or to be involved in the formation of higher order complexes with non-MADS proteins (Fan et al., 1997; Pelaz et al., 2001). Thus, each domain of the MIKC organisation has a specific function.

Transcription factors in general direct development by binding specific DNA sequences in the promoter or enhancer regions of target genes thus regulating their expression or repression. Both type I and type II MADS TFs recognize similar sites on DNA called as CArG-boxes ($CC(A/T)_6GG$ or $CTA(A/T)_4TAG$) and they bind these sequences as dimers (Pollock and Treisman, 1990, 1991; Schwarz-Sommer et al., 1990). An additional specificity in the flanking regions of the consensus sequence has also been predicted (Muiño et al., 2013). CArG box sequences are strikingly similar for various MADS domain proteins in

different organisms (de Folter and Angenent, 2006; Smaczniak et al., 2012b). However, several variants with different lengths in the A/T rich region have been identified (Nurrish and Treisman, 1995). As observed in (Figure I-7), human and yeast MADS proteins bind DNA via dimers of their M domains. Although the plant MADS TFs have conserved this property, they have evolved more complex structures through their oligomerisation and higher order complex formation domains which likely play a role in DNA binding through avidity.

As is evident from the regulatory cascade described in Figure I-3 as well as floral development models (Figure I-4 & Figure I-5), several MADS TFs play a significant role both in the early stages of flower development as well as in the later stages of floral organ differentiation. This project focuses on proteins encoded by floral organ differentiation genes *SEP3* and *AG* as well as the protein encoded by flowering time gene *SVP*.

SEP3 is one of the most promiscuous member of the MADS family and has been shown to form more than 50 different complexes, including interactions with SVP and AG (de Folter et al., 2005; Immink et al., 2009; Kaufmann et al., 2009). It also plays a role on formation of all floral organs (Theissen and Saedler, 2001; Theissen, 2001; Bowman et al., 2012, 1991). AGAMOUS is an important MADS TF that also plays two distinct roles. As previously described it acts as a repressor for *WUSCHEL* (Section I.2.1.2) to terminate floral meristem cell maintenance (Lenhard et al., 2001) and also participates in the formation of two floral whorls containing stamens and carpels (Bowman et al., 1989; Yanofsky et al., 1990). SVP is a multi-functional protein which acts as a repressor of the floral transition during the vegetative phase and contributes to the specification of floral meristems during the reproductive phase (Hartmann et al., 2000; Gregis et al., 2008; Li et al., 2008; Lee et al., 2007). It also has been shown to play a role in maintaining temperature sensitivity in plants (Lee et al., 2007).

The following chapters elaborate the functional significance of the three MADS TFs and attempts to characterize them biophysically and biochemically in order to understand the underlying mechanisms governing their action.

I.3 Objectives of the thesis

How TFs are able to selectively form different complexes and direct the transcriptional machinery has been the subject of extensive study for decades. Dynamic interactions occurring at the molecular level result in changes at the cellular and organismal level, but no one technique is able to probe all these scales instantaneously. This work aims at merging different biophysical techniques to unveil the atomic and molecular structure of a subset of the MADS TFs and to investigate their DNA binding and oligomerisation patterns. Such data will bridge the gap between molecular and genetic models of MADS TF function, provide invaluable information about DNA binding and oligomerisation specificity at the amino acid level and will help define set of rules which might predict the activity of all MADS TFs.

The availability of sufficient quantities of soluble pure protein is a prerequisite for pursuing structural or biochemical studies. However, the lack of any structural information of plant MADS TFs and the absence of protocols for recombinant purification of MADS TFs creates a bottleneck in studying these proteins. The primary goal was thus to resolve this rate-limiting step and to follow this up by detailed analysis of the structural and biochemical properties of the protein(s).

Focusing on two floral homeotic TFs, SEPALLATA3 and AGAMOUS, and the flowering time regulator SHORT VEGETATIVE PHASE of the model organism *Arabidopsis thaliana*, this work combines structural and imaging techniques to understand both the atomic level determinants and the molecular mechanisms of MADS TF protein-protein and protein-DNA interactions**.** The strategy developed involved identification and purification of proteins, biophysical characterisation using x-ray crystallography and SAXS, and analysis of the DNA-protein interactions with AFM and EMSA. The results obtained over the course of this PhD work have also been the subject of two published manuscripts, included in Appendix-II.

Chapter II

SEPALLATA3

II SEPALLATA3

II.1 Summary

This chapter focuses on one of the most important plant MADS transcription factors, SEPALLATA3 (SEP3) which is known to form a variety of different complexes. It aims to elucidate the structural and DNA binding properties of SEP3 *in vitro* using a variety of biophysical techniques.

The first part of this chapter describes the different strategies used for obtaining a high yield of soluble protein. Emphasis is given to the high throughput library techniques utilized for this purpose and protocol for purification of the recombinant protein construct spanning the oligomerisation domain is established. As no published protocol for recombinant purification of MADS transcription factors was available, these experiments constitute an advance in the field.

A high-resolution crystal structure of the oligomerisation domain is presented and the effect of different point mutations is discussed. AFM scans are presented which allowed the visualisation of DNA-protein interactions, in particular on DNA looping. Finally, the importance of SEP3, in particular its K domain, in the formation of multimeric protein complexes is discussed.

Résumé en Français

Ce chapitre se concentre sur l'un des plus importants facteurs de transcription végétaux MADS, SEPALLATA 3 (SEP3). Il vise à élucider les propriétés structurales et de liaison à l'ADN *in vitro* de SEP3 en utilisant une variété de techniques biophysiques.

La première partie du chapitre présente les différentes stratégies utilisées pour obtenir une protéine soluble à haut rendement. L'accent est mis sur les techniques de « bibliothèque à haut débit » utilisées à cet effet. Un protocole pour la purification de la protéine recombinante d'assemblage couvrant le domaine d'oligomérisation est établi. En l'absence de protocole publié pour la purification recombinante de facteurs de transcritpion MADS, cette mise au point constitue une partie importante de ce chapitre.

Une structure haute résolution pour le domaine d'oligomérisation est présentée et l'effet de mutations sur les résidus d’interaction est discuté. Les données de Microscopie à Force

Atomique (AFM) sont présentées et permettent de visualiser les interactions ADN-protéines, tout particulièrement l'ADN en boucle. Enfin, l'importance du domaine K de SEP3, en particulier dans la formation de complexes de protéines multimèriques, est discuté.

II.2 Introduction

As described in section I.2.2, floral organ development depends on the coordinated activity of the class A-E MADS box genes. The ABC(D)E model proposes that the overlapping expression of these genes is necessary for floral organ differentiation and development (Theissen, 2001; Coen and Meyerowitz, 1991; Pinyopich et al., 2003; Pelaz et al., 2000; Honma and Goto, 2001). Especially important are the class E genes, which play a role in the formation of all floral organs. In *Arabidopsis thaliana*, the class E MADS genes belong to the *SEPALLATA (SEP)* subfamily and encode different transcription factors (TFs). The *SEPALLATA* genes (*SEP1-4)* exhibit partial functional redundancy as only the *sep1sep2sep3* triple mutant demonstrates a clear phenotype in which all floral organs develop as sepals (Pelaz et al., 2000). The quadruple mutant, *sep1sep2sep3sep4*, results in the formation of leaves suggesting that *SEP4* has a role in sepal development (Ditta et al., 2004). However, it has been shown that the encoded SEP proteins vary in terms of both number of interaction partners and their cooperative DNA binding ability, suggesting only partial redundancy in their *in vivo* function (Jetha et al., 2014).

As established by extensive yeast n-hybrid and immunoprecipitation experiments, SEP3 is one of the most promiscuous members of the MADS TF family and is able to form many different complexes with partner proteins (de Folter et al., 2005; Kaufmann et al., 2009; Immink et al., 2009). SEP3 interacts with most florally expressed MADS-domain TFs implying that many target genes of the A, B and C floral MADS TFs overlap with those of SEP3 (de Folter et al., 2005). In particular, SEP3 is thought to act as a multimerisation mediator, and modulates various fundamental pathways including the hormonal signalling pathways responsible for plant growth and development (Kaufmann et al., 2009). Due to its role in formation of almost 50 different complexes and its influence in the assembly of other MADS type II TF complexes, SEP3 has been termed as a ‘glue’ for binding the MADS family together (Immink et al., 2009). Figure II-1 shows a compilation of all the different partners of SEP3 during various stages of floral initiation and development (Immink et al., 2009). Genetic experiments have also shown a dose-dependent role of *SEP* genes in ovule development (Favaro et al., 2003).

SEP3 can act as both repressor and activator of gene expression depending on its protein partners. SEP3 and LEAFY (LFY) are known to be involved in the activation of *APETALA 3 (AP3), PISTILLATA (PI)* and *AGAMOUS (AG)* (Liu et al., 2009; Kaufmann et al., 2009;

Parcy et al., 1998; Weigel et al., 1992). On the other hand, genes that regulate the floral transition such *SVP*, *AGL24* and *SOC1* are downregulated by SEP3 (Kaufmann et al., 2009). In addition to its role in directing the downstream expression of different floral MADS genes, SEP3 has many other putative targets as suggested by ChIP-seq experiments where over 3000 enriched binding sites were determined (Kaufmann et al., 2009).

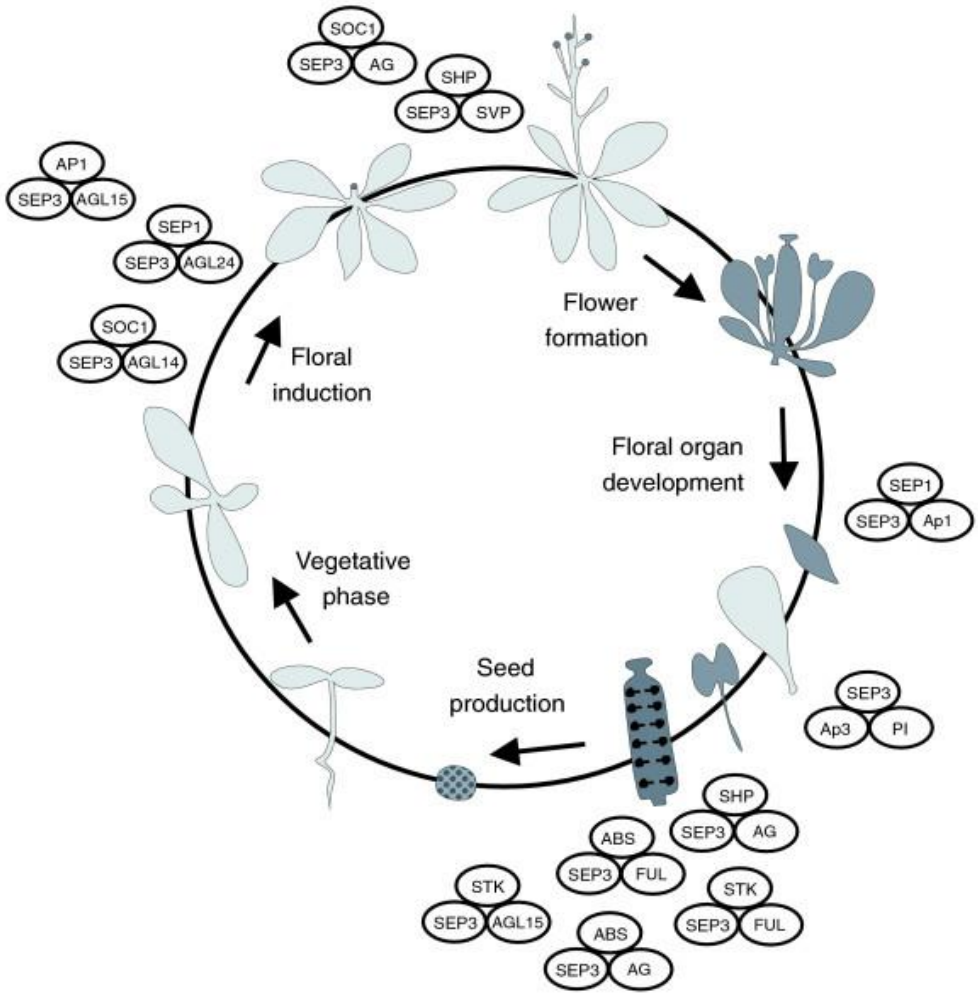


Figure II-1: ***Role of SEP3 complexes in plant development****. Different functions of SEP3 complexes during various stages of Arabidopsis plant development. The stoichiometry of complexes may be different than shown* (Immink et al., 2009).

In this chapter, using different biophysical and biochemical tools, I discuss the structural aspects of SEP3 at the atomic and molecular level. The factors affecting SEP3 oligomerisation and DNA binding are examined. In this chapter, I also elaborate on the different techniques used for gene construct design, an optimisation process for protein

expression and purification and biophysical techniques used for structure determination. The chapter highlights the establishment of a successful recombinant purification for the SEP3 oligomerisation domain, a high resolution crystal structure of the SEP3 oligomerization domain and a visualisation of SEP3 –DNA binding using AFM.

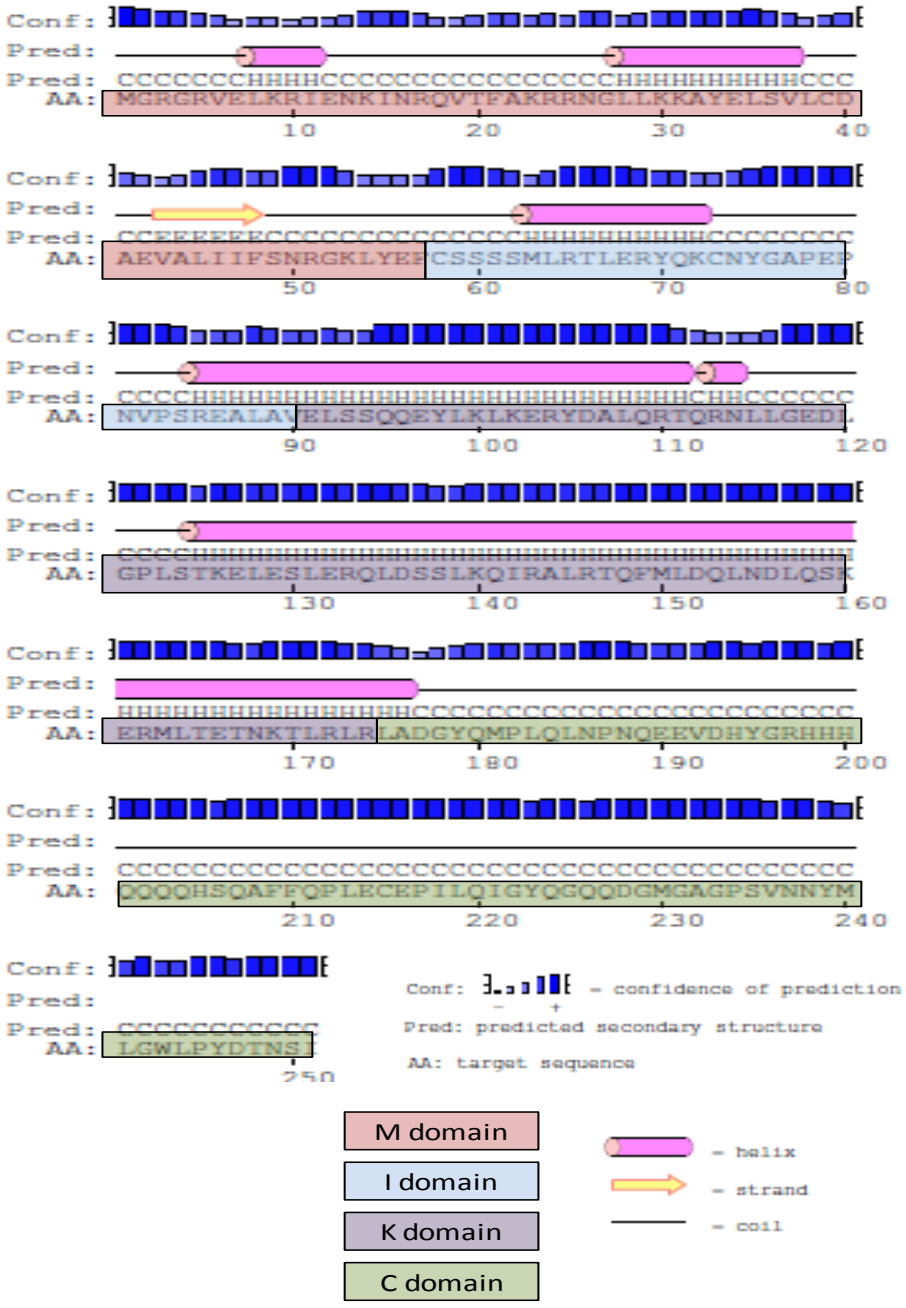


Figure II-2: ***The secondary structure of SEP3 as predicted using PSIPRED*** *(Jones, 1999). The different domains annotated from Uniprot database (The UniProt Consortium, 2014) are marked with different colours. Pink cylinders represent α helices, yellow arrows β strand. The confidence of the prediction is given by blue bars.*

II.3 Results and Discussion

As shown in Figure II-2, SEP3 is a 251 amino acid protein divided into 4 domains in the classic MIKC pattern followed by type II MADS TFs. Secondary structure prediction (PSIPRED) (Jones, 1999) suggests the I and K domains to be mostly α- helical and predicts the C-terminal domain to be largely unstructured. While the DNA binding M domain is conserved across all eukaryotes, the plant-specific oligomerisation domains- I, K and C show a varied degree of sequence variation. An understanding of the MADS TF protein binding and specificity thus requires structural information of the plant specific oligomerisation domains. Indeed, a thorough literature analysis points out the large gap between the genetic and biochemical study of MADS TFs. Although these proteins have been very well studied genetically over almost two decades, the amount of biochemical data available is scarce mostly due to the difficulty in obtaining soluble recombinant protein.

II.3.1 Construct design and protein purification

A fundamental prerequisite for many biochemical and structural studies of biological macromolecules is the availability of pure soluble recombinant protein at high yields. The absence of an established purification protocol for any plant MADS TFs represented an initial bottleneck in the characterisation of SEP3. In order to overcome this, conventional construct design strategies and innovative high-throughput protein expression screening techniques were used.

II.3.1.1 Expression of soluble protein by domain boundary analysis

The standard procedure to obtain soluble protein is to design gene constructs based on protein secondary structure prediction. This technique is extensively reviewed in various structural genomics studies (Gileadi et al., 2008; Yumerefendi et al., 2010; Makrides, 1996; Smith et al., 1983; Stevens, 2000; Burley, 2000). This typically involves removing the unstructured parts of the protein with the aim of optimizing protein solubility, purity and crystallisability. Since a large part of the C terminal domain of SEP3 is predicted to be unstructured (Figure II-2), constructs were designed using conventional domain boundary analysis that in particular excluded the C-terminal region. The different constructs included a fragment spanning the SEP3 MIK domains ($SEP3^{(1-178)}$) and one spanning the SEP3 MI domains ($SEP3^{(1-90)}$), which was a gift from G. Theissen, (Friedrich-Schiller-University, Jena). Furthermore, two full-length protein constructs were designed: 6xHis-$SEP3^{(1-251)}$ and 6xHis-GST-$SEP3^{(1-251)}$. For each construct a thorough expression and purification

optimisation process was undertaken, which involved expression in different bacterial *E.coli* cell strains (BL21-Star, BL21-AI (Arabinose Inducible), BL21-pLysS, Rosetta), different cell growth media (Luria Bertani, Terrific broth, Minimal media), variable expression temperatures (16°C, 20°C, 25°C, 37°C) and different concentrations of the recombinant expression inducer (0.2 -1mM) IPTG (Isopropyl β-D-1-thiogalactopyranoside). We also tested an auto inducible expression system in order to maximize the protein solubility by tuning expression levels (Studier, 2005).

The results presented here (Figure II-3) show the preliminary purification of 6xHis-SEP3$^{(1-251)}$ using an auto-inducible expression system. SEP3$^{(1-251)}$ was cloned in pESPRIT002 (Yumerefendi et al., 2010; Hart and Tarendeau, 2006; Guilligay et al., 2008) with an N-terminal 6xHis tag and expressed in auto inducible media in BL21 RIL (DE3) (Agilent) cells and purified using Ni-NTA affinity and size exclusion chromatography. The elution profile of the gel filtration on a Superdex S200 10/300 GL (GE Healthcare) column shows a sharp peak in the void volume followed by several broad peaks (Figure II-3). The elution peak at the void volume suggests formation of soluble aggregates. In addition, the high absorbance at 254 nm suggests a high amount of DNA contamination. SDS-PAGE analysis of the selected samples shows a band at lower molecular weight as compared to the theoretical molecular weight (~32kDa) suggesting partial degradation of the construct.

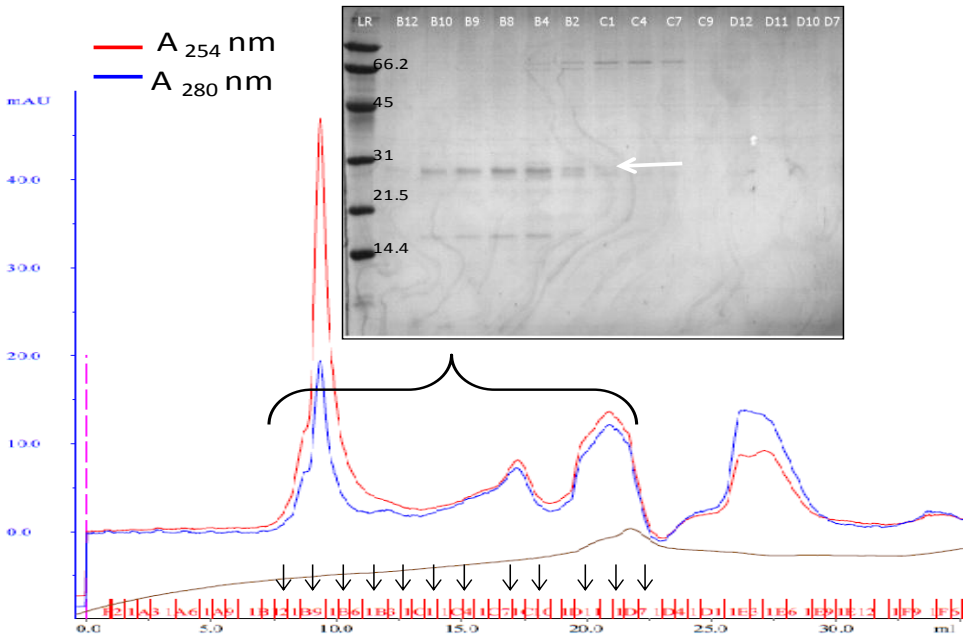


Figure II-3: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-SEP3$^{(1-251)}$.*** *The construct was expressed in BL21-RIL (DE3) and purified with Superdex S200 10/ 300 GL column. (GE Healthcare). The samples from gel filtration marked by black arrows were analyzed by SDS-PAGE(inset). The white arrow marks probable position of truncated protein.*

Better results were obtained for 6xHis-GST-SEP3$^{(1-251)}$ and SEP3$^{(1-90)}$ after a thorough optimisation process. 6xHis-GST-SEP3$^{(1-251)}$ was expressed in *E. coli* Rosetta 2 cells and purified using a three-step purification process comprising a Ni-NTA affinity column, followed by a heparin affinity column and finally a size exclusion chromatography step. The chromatogram from the size exclusion chromatography and the resultant SDS-PAGE analysis (Figure II-4) show the presence of protein at the correct molecular weight (56 kDa) and the A_{254} nm signal indicates a reduction in DNA contamination. However, some lower molecular weight bands representing degradation are also observed. This was one of the first purifications that yielded pure stable protein and was subsequently used in AFM studies described later in this chapter. Unfortunately, repeated attempts for scale up did not prove to be successful.

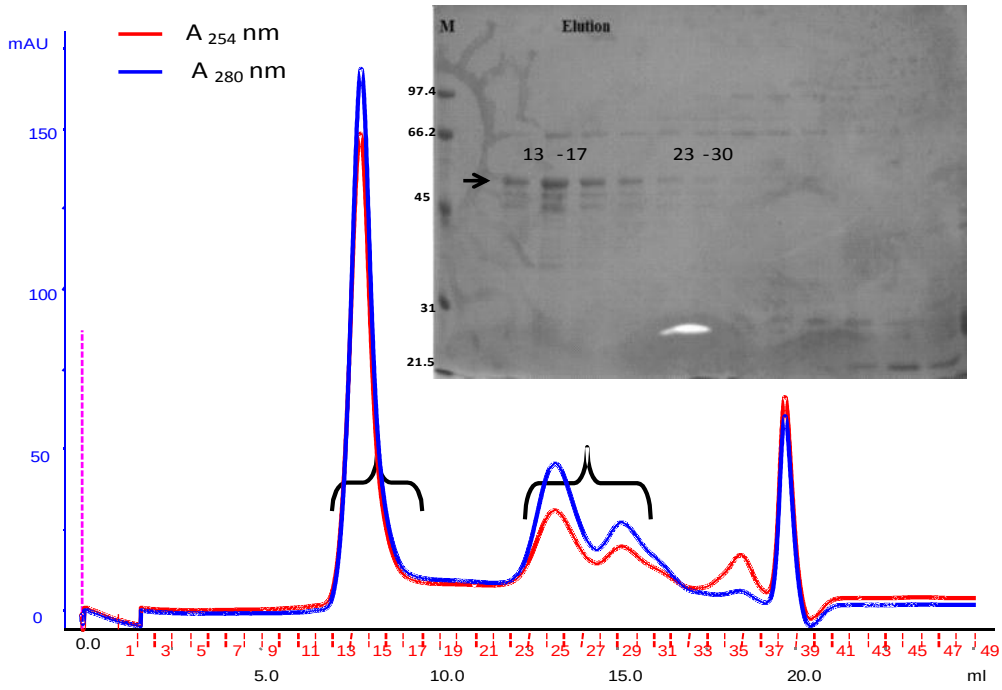


Figure II-4: ***Size exclusion chromatography and SDS-PAGE analysis for 6xHis-GST-SEP3$^{(1-251)}$.*** *The construct was expressed in Rosetta 2 cells and purified with Superdex S200 10/ 300 GL column. The samples from gel filtration marked by black brackets were analyzed by SDS-PAGE (inset). The black arrow marks the position of full length protein.*

The low amount of full-length protein obtained and the degradation bands visible might be attributed to the presence of a long unstructured and proteolytically sensitive C-terminal domain present in SEP3. Therefore, in order to minimize the effect of unstructured regions on degradation, I selected the construct SEP3$^{(1-90)}$ comprising the MI domains and

cloned it in the plasmid pET15b (Figure II-39), with an N-terminal 6xHis tag followed by a thrombin cleavage site. The construct was expressed in BL21 (DE3) RIL (Agilent) cells and purified by Ni-NTA affinity and heparin column chromatography. The 6xHis tag was then cleaved using thrombin protease and the protein was re-purified with size exclusion chromatography. Figure II-5 shows a SDS-PAGE analysis with presence of a substantial amount of protein. The SEP3$^{(1\text{-}90)}$ purification thus yielded a good amount of protein suitable for subsequent crystallisation and AFM studies.

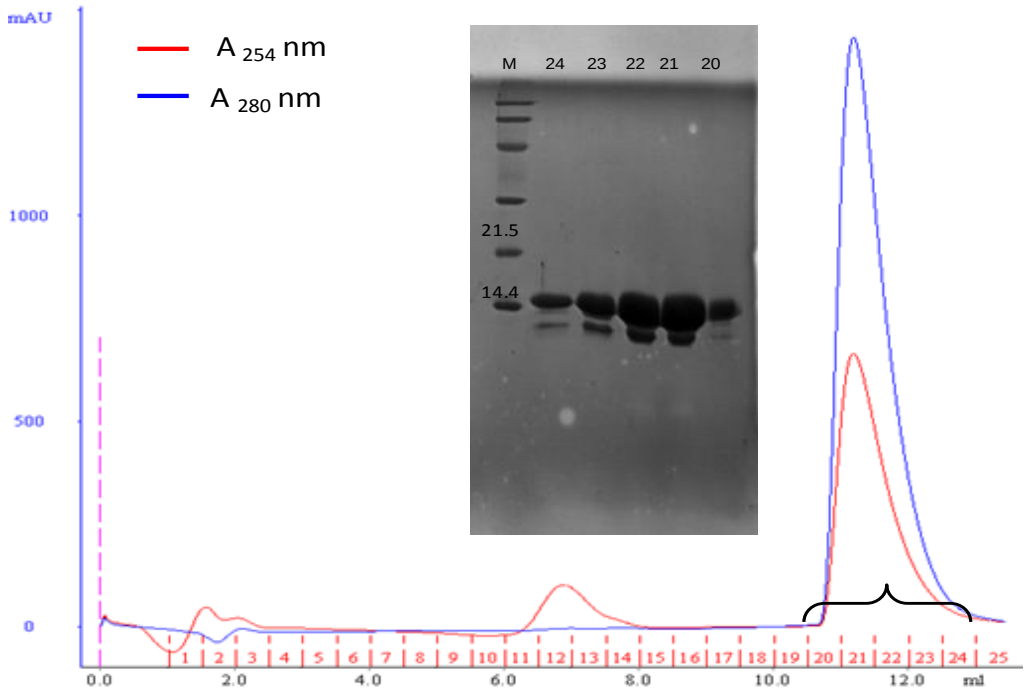
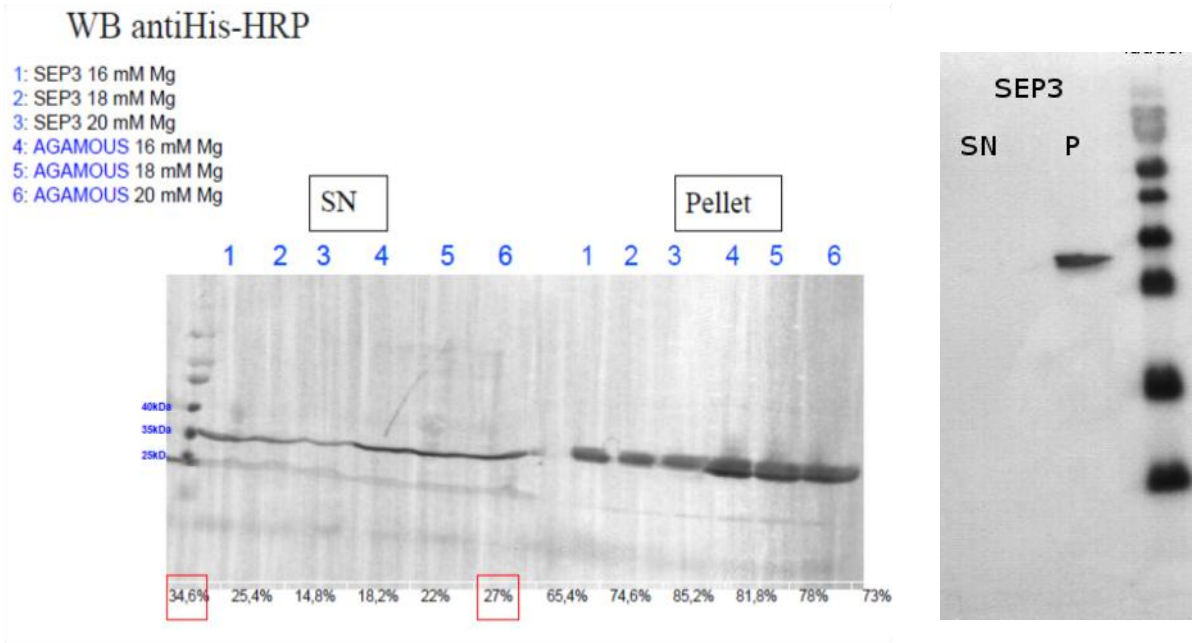


Figure II-5: ***Size exclusion chromatography and SDS-PAGE analysis for SEP3$^{(1\text{-}90)}$ purification****. The construct was expressed in BL21 (DE3) cells and purified with Superdex S200 10/ 300 GL column. The marked samples under the peak were analysed using SDS-PAGE (inset).*

Despite the high quality of the SEP3$^{(1\text{-}90)}$ sample, this protein construct lacks the K and C domains. On the other hand, the yield of 6xHis-GST-SEP3$^{(1\text{-}251)}$ was not enough for structural characterisation using most biophysical techniques. Hence, it was decided to utilize alternative techniques to produce recombinant protein including the use of different expression systems and high-throughput library generation methods.

II.3.1.2 Cell free expression

The *in vitro* full-length SEP3 expression was tested using the IBS cell free expression platform (http://www.isbg.fr/samples-preparation/cell-free-666/). Figure II-6 shows a western blot for the cell free expression trials for 6xHis-SEP3$^{(1-251)}$ and another important MADS TF, AGAMOUS (AG). For both the proteins, a very low percentage of protein is observed in the



soluble fractions. Magnesium concentrations do not seem to have a significant effect on the solubility of proteins. Unfortunately, a large-scale expression (2 ml) also did not yield any soluble protein.

II.3.1.3 Multi vector expression screen

Many factors including post-translational modifications and the presence of chaperones might affect the activity and stability of recombinantly expressed eukaryotic proteins (Dyson et al., 2004; Rosano and Ceccarelli, 2014; Sahdev et al., 2008; Khow and Suntrarachun, 2012; Ellis, 2006; Khoury et al., 2011; van den Berg et al., 2000). Sometimes an absence of these factors in bacterial expression systems can lead to improper folding and thus aggregation of these proteins. Hence, the expression of SEP3 was tested in both insect and mammalian cell expression systems. The screens were performed at the Oxford Protein Production Facility (OPPF), Oxford, UK as part of the Instruct fellowship program (http://www.oppf.rc-harwell.ac.uk/OPPF/). Different solubility tags were also screened in parallel. The details of the vectors used and the tags tested can be found in Appendix I.

Figure II-6: ***Cell free expression for SEP3.*** *Western blot analysis for (left) test expression (50µl) SEP3(1-251) (lanes 1-3) and AG (lanes 4-6) and (right) large scale expression (2ml) using cell free expression system. An increasing concentration of Mg2+ was used for test expression. Anti-his HRP antibody was used to detect the 6xHis tag. A small amount of soluble protein is observed in test expression. The large scale expression showed complete absence of protein in soluble fraction.*

Furthermore, considering the possibility of stabilizing the protein using its binding partner, a co-expression of SEP3 and AG was also performed. AG is an important MADS TF and has been shown to interact with SEP3 during carpel formation (Theissen, 2001; Theissen and Saedler, 2001). The SDS-PAGE gel for Ni-NTA purification shows a positive band for 6xHis-MBP-SEP3$^{(1-251)}$ (lane A3) tagged with an N-terminal 6xHis and MBP tags (Figure II-7 & Table II-1). Several positive constructs are also observed in the co-expression screen, although the weaker bands may represent lower expression levels.

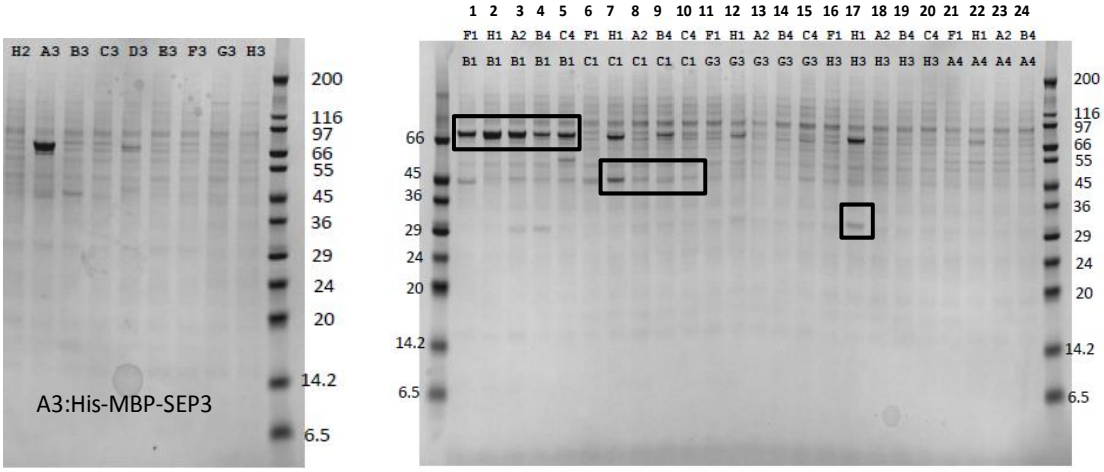


Figure II-7: ***Insect cell expression test for SEP3$^{(1-251)}$ performed at OPPF, Oxford. (left)*** *SDS-PAGE analysis for Ni-NTA purification for SEP3$^{(1-251)}$ expressed in insect cells. The cells were infected with 3ul of first generation virus and grown for 72 hours post day of proliferation arrest. The lane involving SEP3 is annotated in the legend. The results remained consistent even with addition of 30ul of virus (data not shown).* ***(right)*** *A co-expression screen was performed with AG. Bands representing SEP3 are marked with boxes and annotated in the Table II-1.*

Lane	Proteins	Expression?
	SEP 3	**SEP3**
1-5	B1 (His-MBP-SEP3)	+++
7-9	C1 (His-SUMO-SEP3)	++
17	H3 (SEP3-His)	+

Table II-1: ***Annotation for co-expression test shown in Figure II-7(right).*** *The lanes 1-5 show high expression of 6xHis-MBP- SEP3$^{(1-251)}$ while 7-9 show mild expression for 6xHis-SUMO- SEP3$^{(1-251)}$. Lane 17 shows small expression for SEP3$^{(1-251)}$-6xHis.*

The 6xHis-MBP-SEP3 (A3) construct from insect cell expression tests (Figure II-7) was subsequently scaled up at the EMBL-Grenoble eukaryotic expression facility (http://www.embl.fr/multibac/multiexpression_technologies/eef/) and purified using Ni-NTA affinity and size exclusion chromatography. Although full-length protein was obtained, the

gel filtration elution profile shows protein elution in the void volume of a Superdex S 200 10/300 GL column suggesting a possible formation of soluble aggregates (Figure II-8).

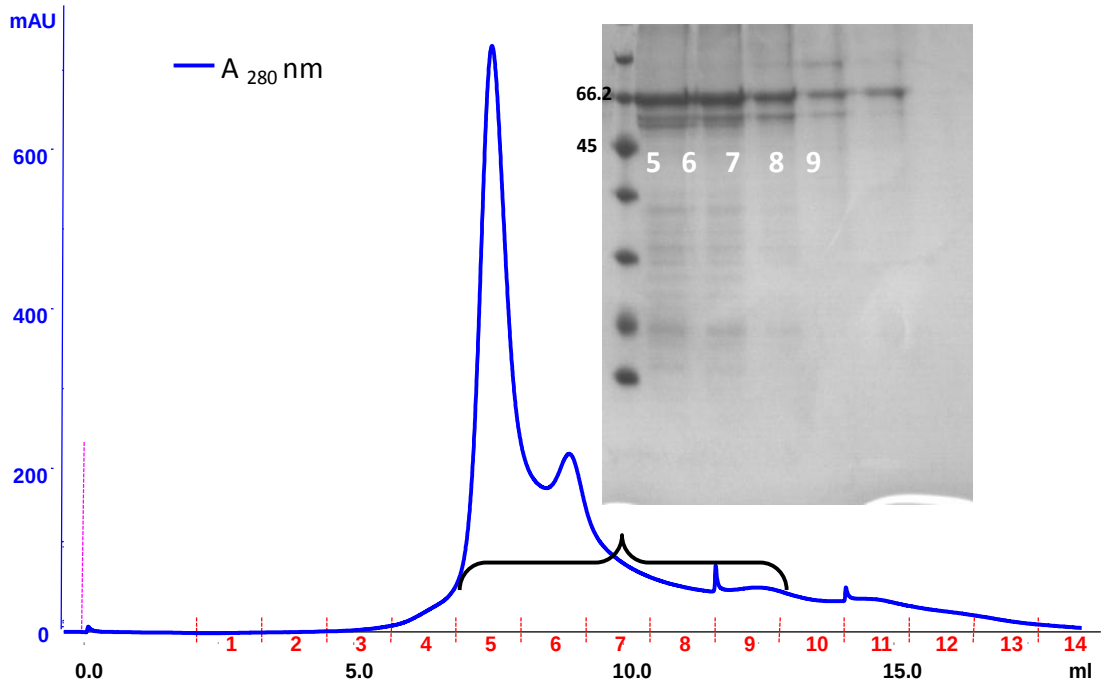


Figure II-8: ***Size exclusion chromatography and SDS-PAGE analysis for 6xHis-MBP-SEP3$^{(1-251)}$.*** *The construct was expressed in insect cells and purified with Superdex S200 10/300 GL (GE healthcare). A_{280} nm shows a sharp peak followed by a small peak. The selected samples were analysed using SDS-PAGE. The SDS-PAGE shows presence of protein in both the lager sharp peak and smaller broad peak. A lower band is also observed probably representing partial degradation.*

Expression tests in HEK-293 mammalian cells were also carried out and showed positive bands for 6xHis-MBP-SEP3$^{(1-251)}$ (A3) and 6xHis-SUMO-SEP3$^{(1-251)}$ (B3) constructs (Figure II-9). However, the mammalian expression system was not pursued further due to absence of a mammalian cell facility in our laboratory and the complications involved in producing large quantities of protein for biophysical studies.

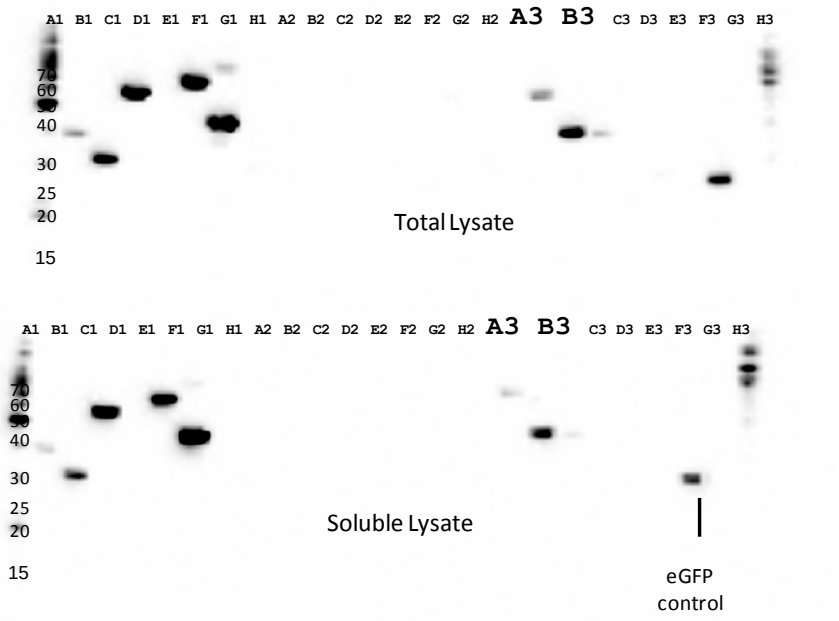


Figure II-9: ***Western blot for HEK 293 mammalian expression test of SEP3 probed with anti-His antibody****. The targets of interest A3 (6xHis-MBP-SEP3$^{(1-251)}$) and B3 (6xHis-SUMO-SEP3$^{(1-251}$) are printed in bold and are observed both in total and soluble fractions.*

Taken together, all the above expression and purification tests for full-length protein were of limited success, with the exception of 6xHis-GST-SEP3$^{(1-251)}$ and SEP3$^{(1-90)}$. Nevertheless, while we could obtain reasonable expression levels, the purified samples were not homogeneous and tended to aggregate, regardless of the use of different expression systems and solubility tags.

The low amount of protein obtained for SEP3 and aggregation observed in the expression trials made it necessary to attempt alternate methodologies. Conventional construct design strategies involve a process of prediction of secondary structure followed by cloning, test expression and optimisation. This process is time consuming, expensive and more challenging for proteins with no structural homology to known high-resolution structures. It also overlooks a host of underlying factors like folding efficiency, need for chaperones, induction methods and temperature. (Dyson et al., 2004; Khow and Suntrarachun, 2012; Sahdev et al., 2008; Yumerefendi et al., 2010; Rosano and Ceccarelli, 2014) In order to avoid repeated iterations of cloning and test expression, we decided to use an approach that aims at generating thousands of constructs in parallel and screening them for solubility and 'purifiability'. The following section focuses on a unique high throughput technique that allows the generation of a library of thousands of clones followed by a thorough screening

process to short list the highly soluble constructs within a short span of 3-4 weeks in ideal cases.

II.3.1.4 Expression of Soluble Proteins by Random Incremental Truncation (ESPRIT)

Expression of Soluble Proteins by Random Incremental Truncation (ESPRIT) is a high throughput technique that is used to obtain soluble protein constructs from a large construct library (Hart and Tarendeau, 2006; Yumerefendi et al., 2010). ESPRIT circumvents the limitations of standard approaches and focuses on generating a large library of thousands of constructs and screening them for soluble expression in a single experimental setup (Hart and Tarendeau, 2006; Yumerefendi et al., 2010). In ESPRIT, Exonuclease III is used to truncate the ends of target genes sequentially to generate a large pool of constructs with variable termini. As shown in the Figure II-10, the truncation can be performed either from the 5´ end or from the 3´end. A scanning library with bidirectional truncation can also be designed.

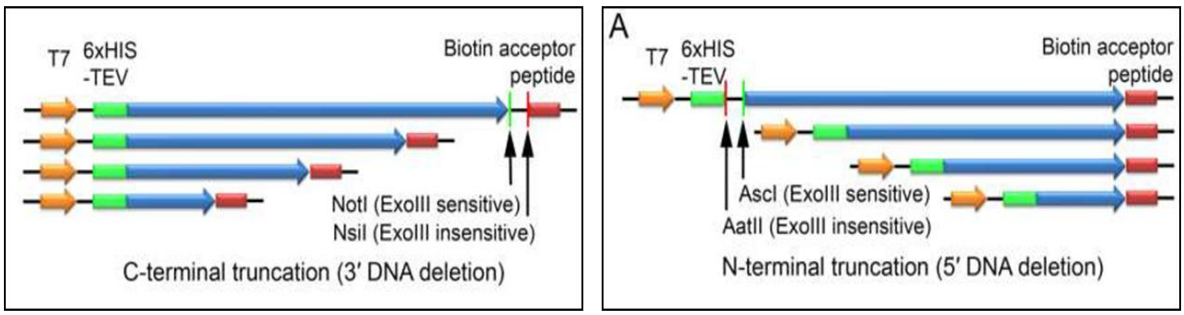


Figure II-10: ***ESPRIT library generation.*** *A C-terminal (left) or N-terminal (right) truncation library can be generated using ESPRIT technique. Digestion with two restriction enzymes generates an ExoIII sensitive and ExoIII resistant ends. An ExoIII treatment is used and samples are aliquoted every minute in quenching buffer. The time may differ depending on the size of gene. Salt concentration is adjusted to tune the rate of reaction.(Image courtesy: Dr. Darren Hart, IBS/ISBG)*

In the ESPRIT protocol, the target gene is cloned into pESPRIT002 (Hart and Tarendeau, 2006; Guilligay et al., 2008) (Figure II-11), replacing the MBP fragment between AatII-NotI and AscI-NsiI depending on the direction of the library to be generated. A restriction digestion using NotI-NsiI or AatII-AscI leads to formation of one Exonuclease III resistant and one susceptible end. Exonuclease III is used to excise the gene and samples are collected at fixed time intervals to generate a pool of different sized constructs. An agarose gel electrophoresis is used for verification of truncation efficiency and acts as a method to divide the library into a specified construct size range. Colony PCRs ensure the linearity and even distribution of the construct library. The number of clones generated depends on the size of

the gene construct with an addition of 2-3 fold oversampling to cover for the out-of-frame constructs with a maxima being 56,000 (28,000 in duplicate) due to technical limitations. Colony picking robots are used to select clones, inoculate small scale expression and to generate nitrocellulose blots. Protein expression is induced on colony arrays in duplicates. The blots are probed with a monoclonal antibody against the N-terminal 6xHis tag and fluorescent streptavidin against C-terminal BAP tag. The presence of the 6xHis tag helps to purify the construct while the presence of the C-terminal BAP confirms solubility of the construct (Figure II.11). The constructs expressing both the tags are statistically ranked, short listed and verified by 4 ml small-scale purifications. Figure II-13 shows the workflow followed in ESPRIT construct screening method.

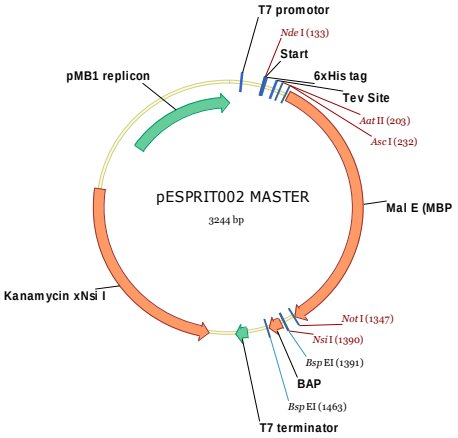


Figure II-12: ***Plasmid map for vector pESPRIT002.*** *The plasmid contains a N-terminal 6xHis tag and a C-terminal BAP tag. The sites BspEI can be used to remove the BAP tag from the final construct. The other restriction sites marked in red are used for gene cloning and library generation* (Hart and Tarendeau, 2006; Yumerefendi et al., 2010)

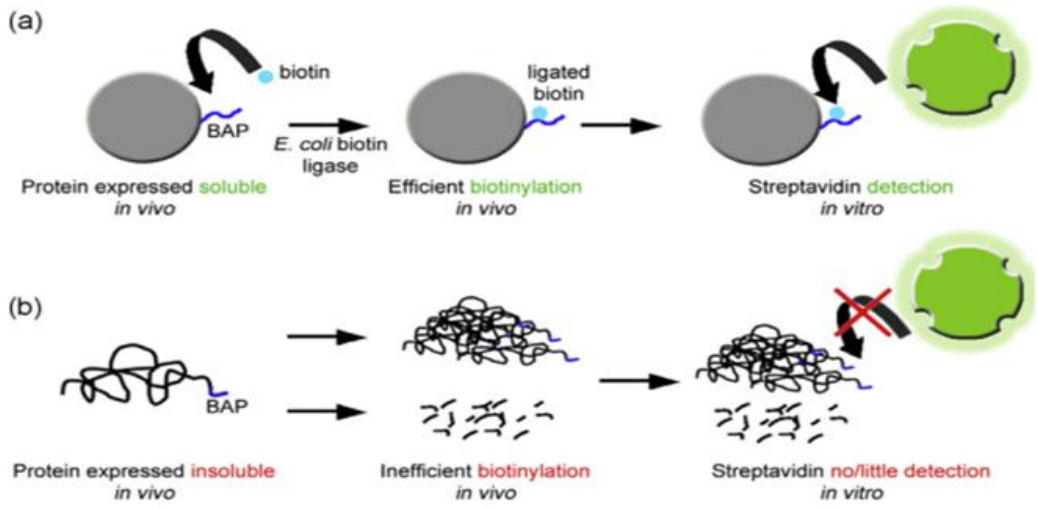


Figure II-11: ***Soluble construct screening using biotin acceptor peptide (BAP) tag.*** *The target gene is cloned with C-terminal BAP tag and a short linker.* ***a)*** *In case of soluble proteins, the BAP is efficiently biotinylated and thus easily detected on addition of fluorescently labelled streptavidin.* ***b)*** *Insoluble or degraded proteins are inefficiently biotinylated and show less or no signal on addition of fluorescent streptavidin.* (Yumerefendi et al., 2010).

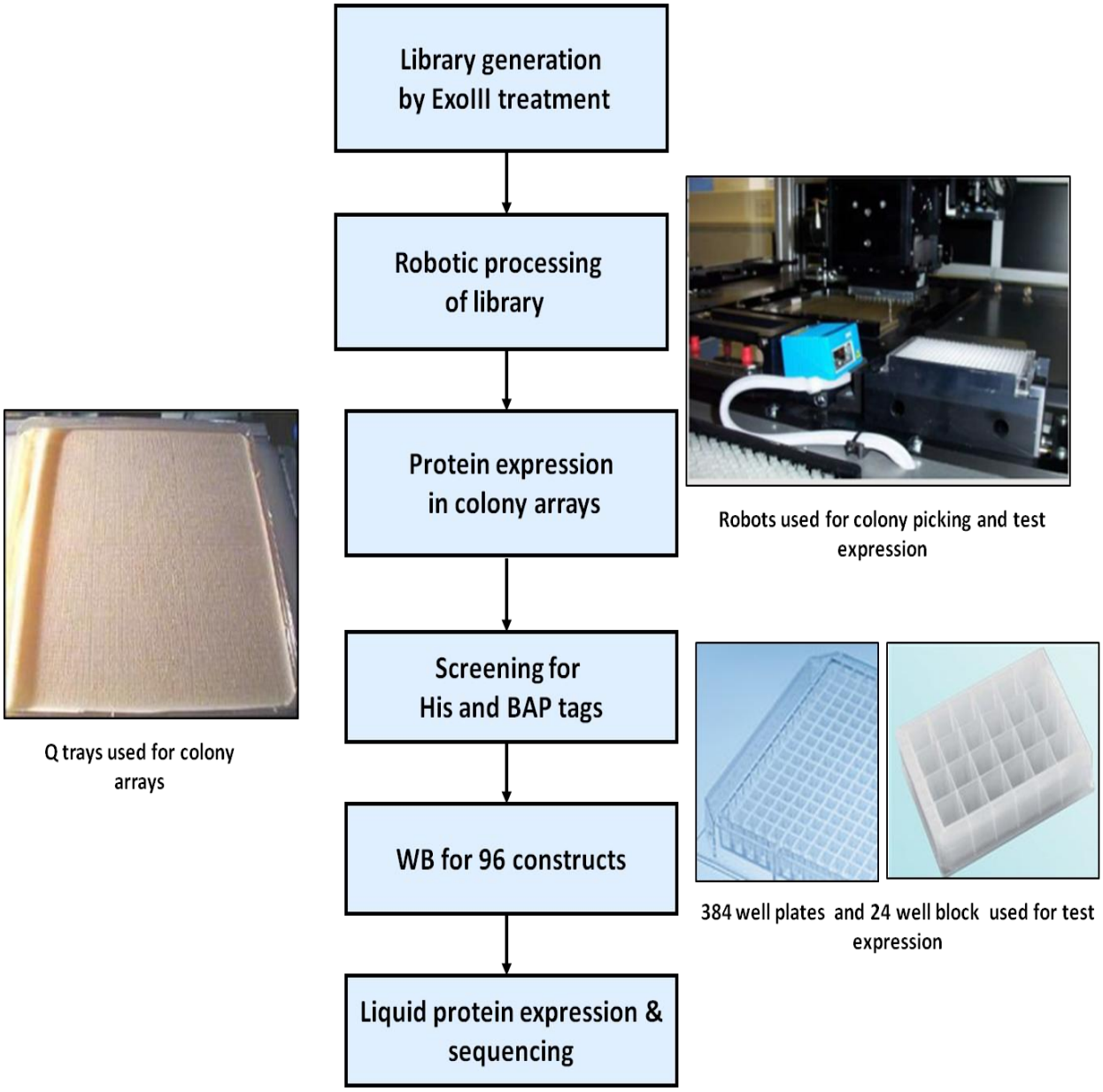


Figure II-13: ***Workflow of ESPRIT library generation.*** *The ESPRIT library generation procedure involves cloning of gene in pESPRIT002 followed by ExoIII truncation. The library is analyzed for its size distribution and colony arrays are set up using robots in Q trays. The 6xHis and BAP tags are screened and positive clones are shortlisted for 4ml test purification. The final constructs are tested with western blot and sequence verified.*

In the case of *SEP3*, a unidirectional truncation method was used with a three fixed N-terminal start sites (Figure II-14) and the exonuclease reactions were initiated at the C-terminal end of the gene. The salt concentration was tuned to control the rate of truncation.

Aliquots were removed at fixed time intervals to generate a pool of different size constructs and analysed using agarose gel electrophoresis.

The required size range was excised and gel purified. In order to avoid bias in library size distribution, construct 1 (spanning a longer *SEP3* gene region) (Figure II-14) was treated separately than constructs 2 and 3. Figure II-15 shows a smear in the agarose gel electrophoresis for SEP3 (constructs 2 & 3) indicating linear digestion. The gel-purified and religated clones were transformed and analysed using colony PCR as shown in the Figure II-16. The colony PCR shows a linear size distribution for SEP3 clones thus confirming absence of bias towards a particular size range.

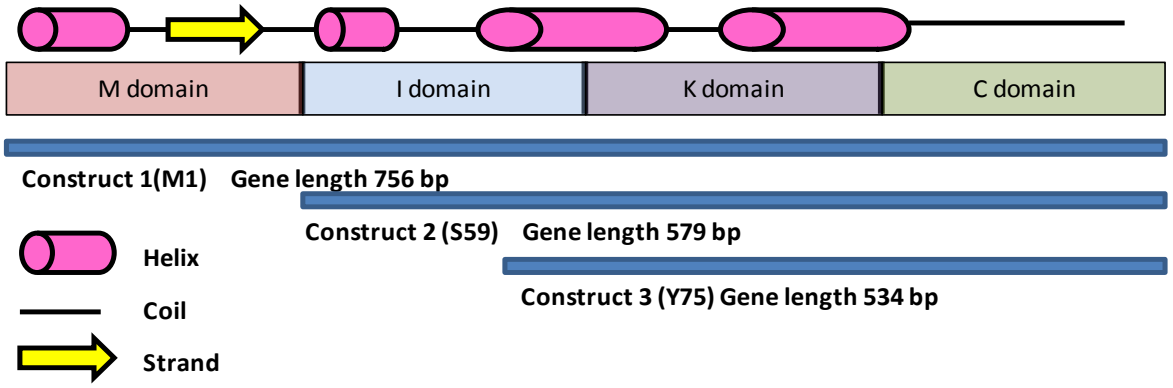


*Figure II-14**: Domain boundaries for the initial constructs of SEP3 used for ESPRIT library generation.** The secondary structure prediction from PSIPRED* (Jones, 1999) *and the domain boundaries for SEP3 are shown at the top. The three initial genes have different N-terminal fixed ends.*

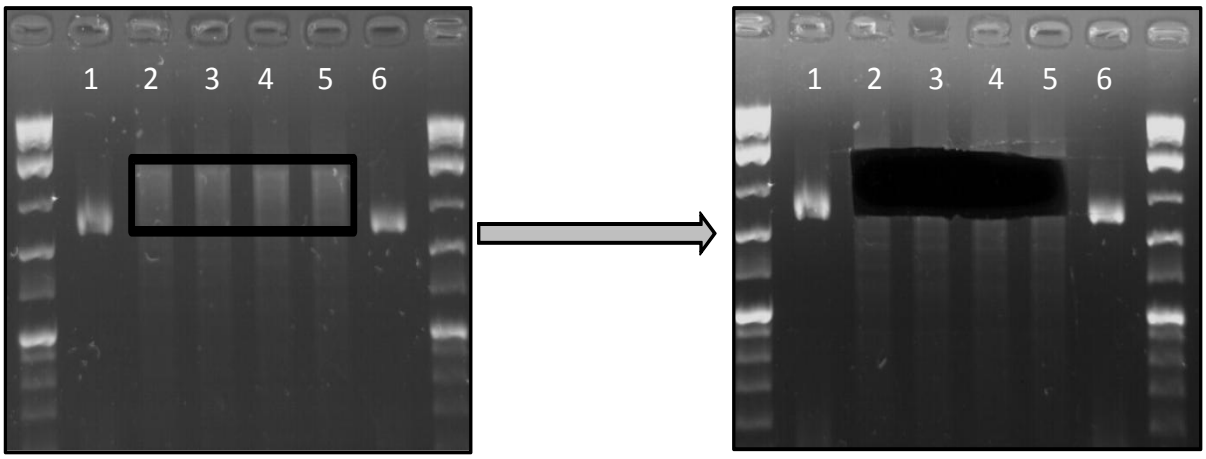


*Figure II-15: **Agarose gel electrophoresis for ExoIII truncation.** SEP3 constructs 2 and 3 were treated with ExoIII and analyzed with agarose gel electrophoresis. A smear indicates proper truncation of the gene. The lanes 1 and 6 represent empty vector used as a lower marker. Part of smear between the whole plasmid and empty vector is excised A similar library was generated for construct 1.*

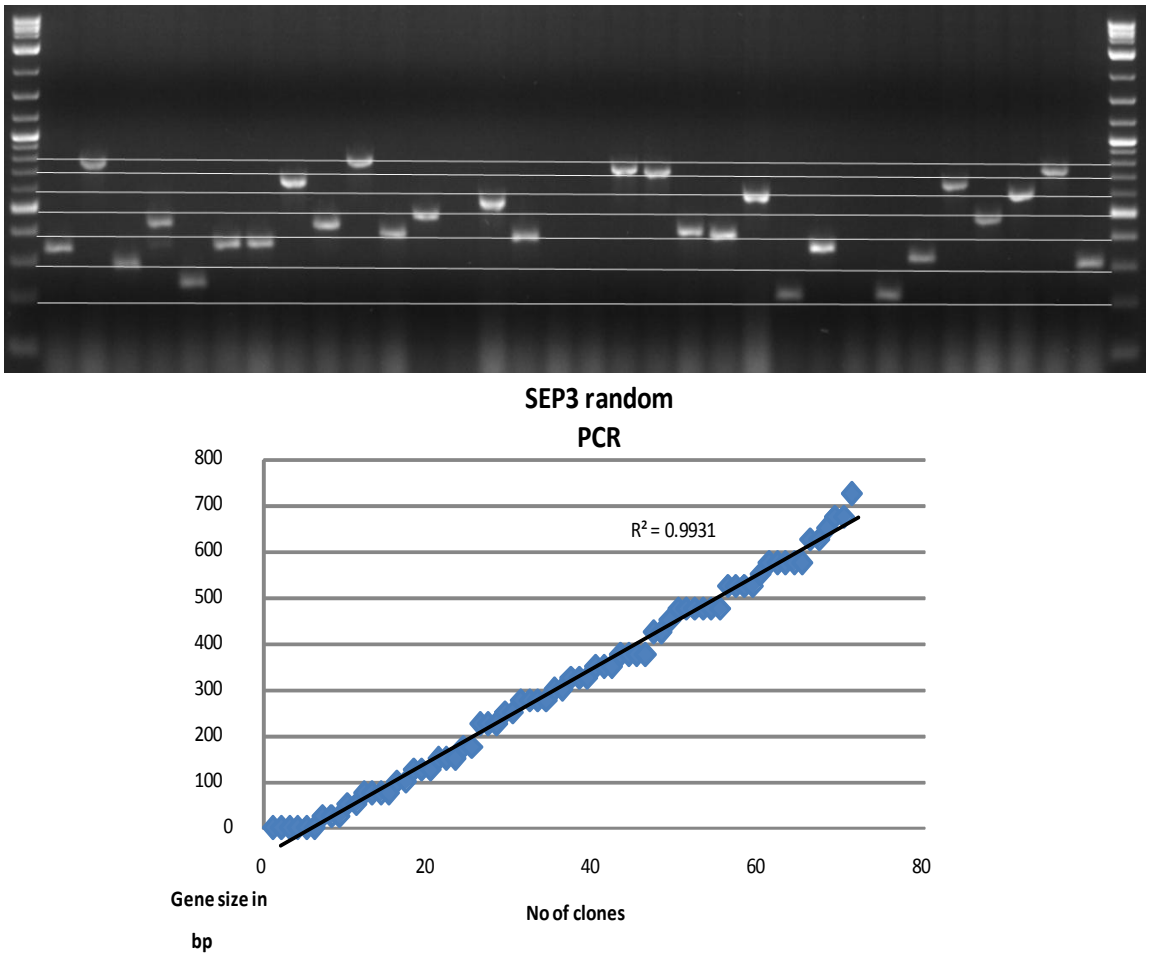


Figure II-16: ***Size distribution analysis of SEP3 library with colony PCR library. (top)*** *Representative colony PCR for SEP3 library. The size of bands was measured and plotted to analyze size distribution.* ***(bottom)*** *The linear graph and the R^2 value close to 1 confirm a linear truncation and unbiased library. The data from 3 other colony PCR agarose gels was pooled for the plot.*

Once the size distribution was validated, a colony array expression was performed for thousands of colonies. The number of colonies to be screened depends on the gene size and ensures 2-3 times oversampling to account for the possibility of constructs which are out of the reading frame. After a robotic expression and colony array test expression, the final constructs were shortlisted by screening for the presence of the 6xHis and BAP tags. The presence of both tags indicated soluble constructs that had an intact 6xHis tag for purification. Overall, the ESPRIT library screening allowed us to generate a pool of ~ 5000 constructs and scan through the poorly annotated region of the protein, resulting in a range of constructs spanning almost the entire protein.

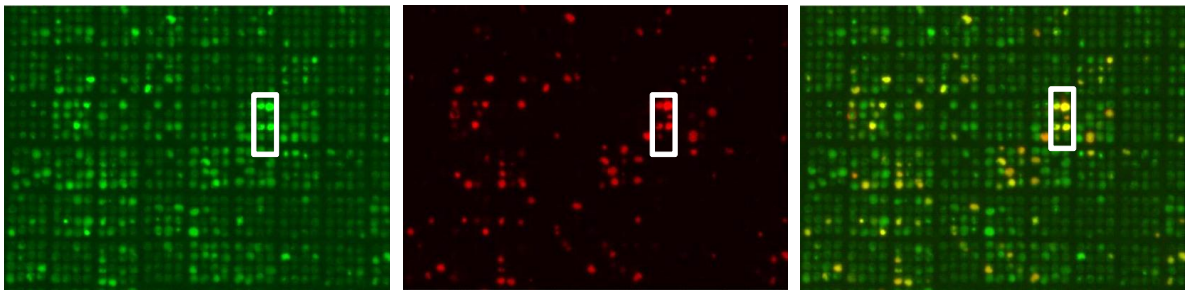


Figure II-17: ***SEP3 colony array analysis.*** *Nitrocellulose membranes were blotted with streptavidin (green) and anti-His antibody (red) and the intensity was scanned with Typhoon Biomolecular Imager (GE Healthcare). Several positive spots are observed in both the scans. The white box marks an example of positive clones for comparison. The clones positive for both 6xHis and BAP tags are represented by yellow spots in the composite image.*

The protein constructs have an N-terminal 6xHis tag and a C-terminal BAP tag. The 6xHis tag is screened with an anti-His antibody while the BAP tag is *in vivo* biotinylated with the endogenous *E.coli* enzyme, BirA, and detected by addition of streptavidin. The nitrocellulose blots in (Figure II-17) show a number of positive colonies represented by green, red and yellow spots. The green spots show constructs which were successfully biotinylated, and thus have high probability of being soluble. The red spots indicate constructs that have an accessible 6xHis tag for purification while the composite image shows yellow spots, which mark constructs positive for both BAP and 6xHis. The presence of signal for both tags confirms the solubility and hence the "purifiability" of a particular protein construct (Yumerefendi et al., 2010). The positive constructs were ranked according to intensity for both the signals and shortlisted. A small-scale (4ml) expression test was then performed for the selected constructs using liquid handling robots and their solubility verified by western blot analysis, similar to that performed for nitrocellulose blots. The western blot for SEP3 as shown in the Figure II-18, shows several positive constructs for both N-terminal 6xHis tag and C-terminal BAP tag with a variation in construct sizes. The positive constructs were sequenced verified and scaled up, followed by an optimisation of the purification protocol to analyse their stability, yield and ease of purification.

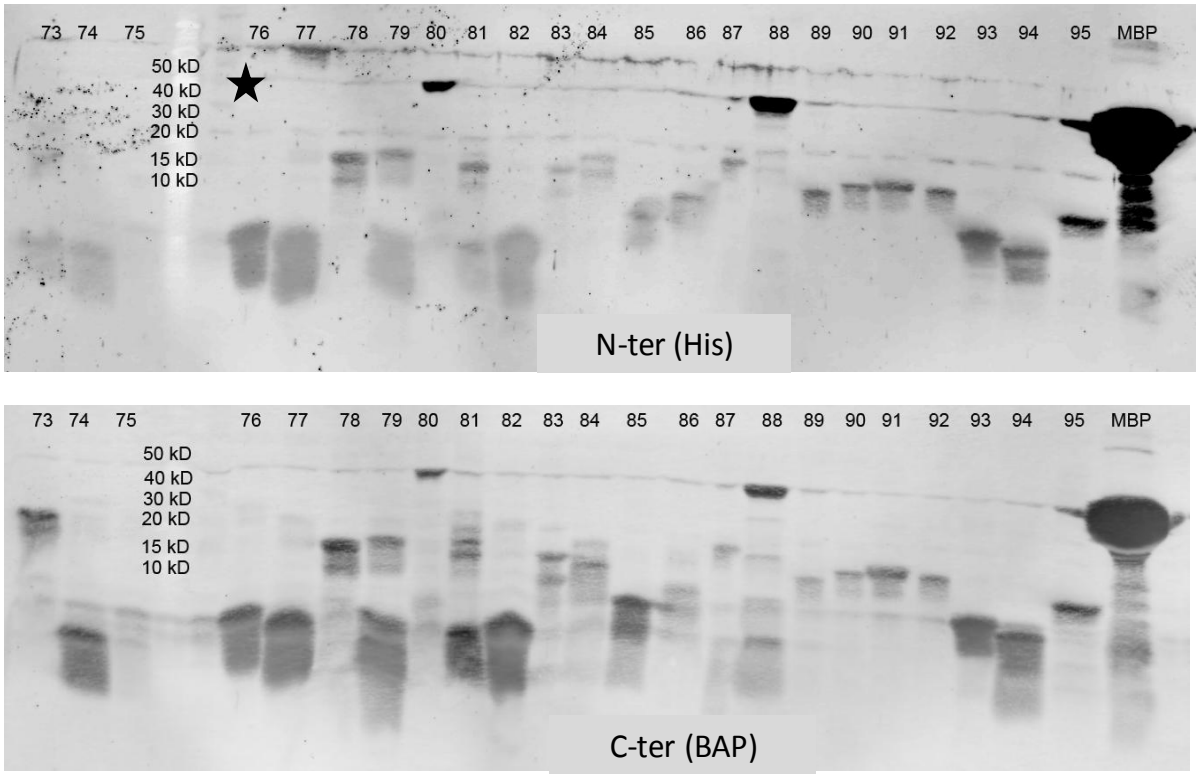


Figure II-18: ***western blot analysis for SEP3 small scale purification tests.*** *The anti-His antibody (top) and streptavidin (bottom) were used as probes. Several positive constructs with variable size range are observed in both the blots. The positive control MBP is used for validation. The protein construct used later for characterization is marked with a star.*

After a thorough analysis, $SEP3^{(75\text{-}178)}$ was selected for large scale expression and biophysical characterisation (Figure II-18, construct 76) due to its stable nature and high yield in large scale expression trials. $SEP3^{(75\text{-}178)}$ was expressed in *E.coli* BL21 (DE3) CodonPlus RIL cells (Agilent Technologies) in LB media and purified using a two step purification process comprising a Ni-NTA affinity column and gel filtration chromatography step. The 6xHis tag was cleaved with TEV protease and separated from the cleaved protein using Ni-NTA affinity column. The cleaved protein was applied to a size exclusion column for final purification. Gel filtration elution analysis shows the protein to be a mixture of dimer and tetramer with a higher percentage of a tetramer (Figure II-19).

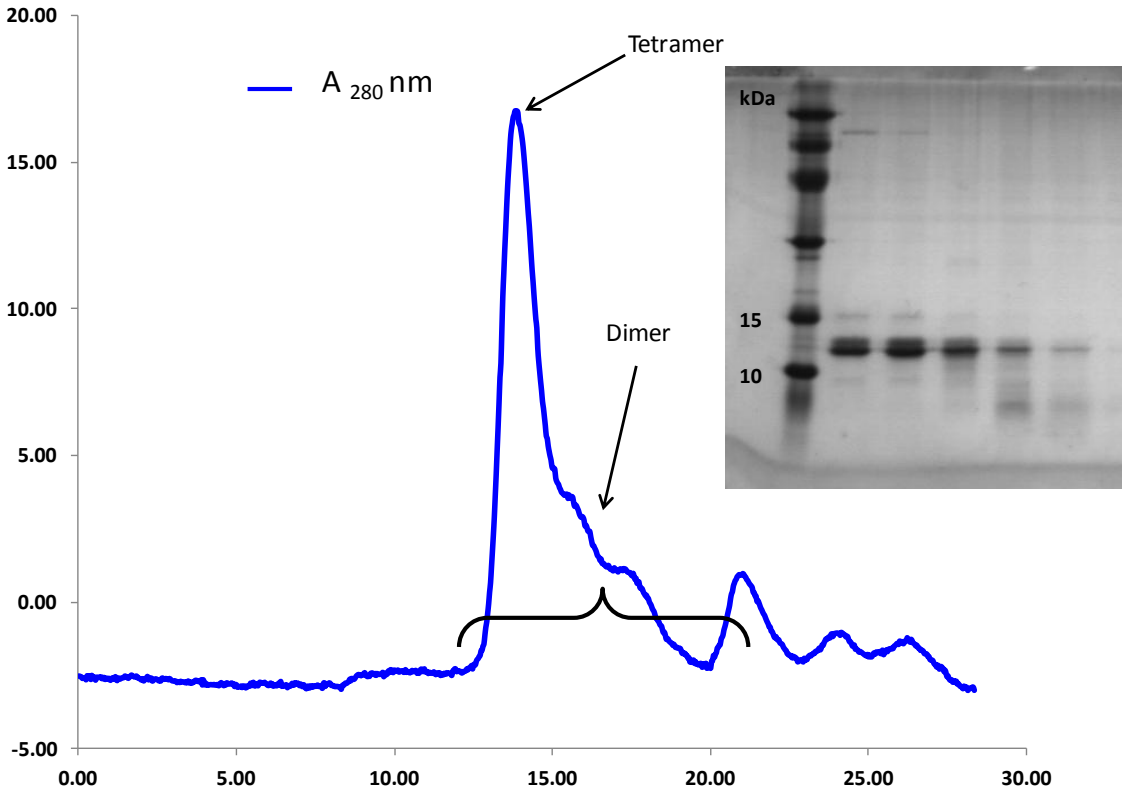


Figure II-19: ***Gel filtration chromatography and SDS-PAGE analysis for SEP3$^{(75-178)}$.*** *The protein shows a sharp peak and a shoulder representing presence of two oligomeric states a tetramer and dimer. The samples marked by black bracket under the peak were analyzed with SDS-PAGE and show presence of protein (~12.5kDa).*

As summarized in the Table II-2, a combination of traditional construct designing and high throughput library generation techniques produced soluble and well expressing constructs for SEP3 for downstream experiments. The constructs SEP3$^{(1-90)}$, 6xHis-GST-SEP3$^{(1-251)}$ and SEP3$^{(75-178)}$ were shortlisted for further structural studies. 6xHis-GST-SEP3$^{(1-251)}$ is a full length protein with an N-terminal 6xHis-GST double tag; SEP3$^{(1-90)}$ contains the complete M and I domains; while SEP3$^{(75-178)}$ comprises a small part of the I domain (residues 75-90), the entire K domain (residues 91-173) and a small part of the C domain (174-178). The details of the domain boundaries for the selected constructs are shown in Figure II-20.

No.	Strategy	Construct	Comments
1	Conventional	SEP3$^{(1-251)}$	Soluble aggregates, truncated, very low yield
2	**Conventional**	**His-GST-SEP3$^{(1-251)}$**	**Low yield but used for AFM**
3	**Conventional**	**SEP3$^{(1-90)}$**	**High yield, Crystallization trials performed, used for AFM**
4	Conventional	SEP3$^{(1-178)}$	Soluble aggregates
5	Insect cell expression, OPPF, Oxford	His-MBP-SEP3$^{(1-251)}$	Soluble aggregates
6	Insect cell expression, OPPF, Oxford	His-SUMO-SEP3$^{(1-251)}$ SEP3$^{(1-251)}$-His	Test expression successful for co-expression screen; not scaled up
7	Mammalian expression system, OPPF, Oxford	His-MBP-SEP3$^{(1-251)}$ His-SUMO-SEP3$^{(1-251)}$	Test expression successful but not pursued further
8	Cell free expression	His-SEP3$^{(1-251)}$	failed
9	ESPRIT	SEP3$^{(75-148)}$,SEP3$^{(75-165)}$ SEP3$^{(75-184)}$,SEP3$^{(61-221)}$ SEP3$^{(61-220)}$)	Test expression successful, some scaled up but not pursued for characterization
10	**ESPRIT**	**SEP3$^{(75-178)}$**	**Crystal structure resolved**

Table II-2: ***Soluble construct screening for SEP3.*** *Various strategies were used to obtain soluble high yielding protein for structural studies. The constructs that were pursued further are printed in bold. The term 'conventional' implies protein constructs designed using domain boundary analysis and expressed with bacterial expression system.*

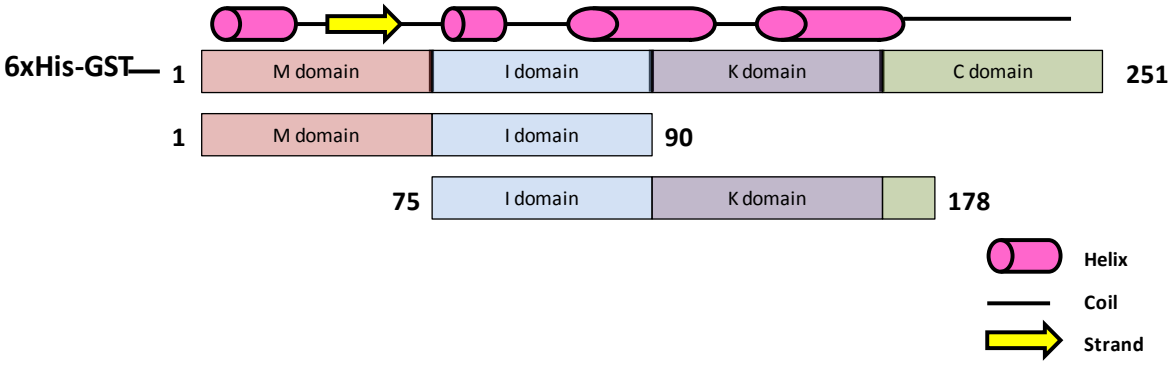


Figure II-20: ***Domain boundary analysis for the soluble SEP3 constructs obtained.*** *The three protein constructs 6xHis-GST-SEP3$^{(1-251)}$, SEP3$^{(75-178)}$ and SEP3$^{(1-90)}$ were selected for further analysis. The domain boundary analysis shows that barring the full length construct, the other two constructs are composed of mainly predicted structured domains.*

As one of the most promiscuous members of the MADS family, SEP3 is involved in a host of different interactions with various other MADS and non-MADS proteins. In order to study the interactions of SEP3, it was necessary to understand the interactions and binding characteristics of the protein at the atomic level. A high-resolution crystal structure can provide us this valuable information, which can be applied to other proteins of the same family based on sequence and structural homology. Due to the high yield and stability observed during purification, $SEP3^{(75\text{-}178)}$ was targeted for biochemical and structural studies. This construct spans the oligomerisation domain of the protein and as such allows the determination of the amino acids critical for dimer and tetramer formation. In order to identify these residues, crystallisation trials and site-directed mutagenesis were performed as detailed below.

II.3.2 Biophysical characterisation and DNA-binding studies of SEP3$^{(75-178)}$

II.3.2.1 Crystallisation and crystal structure determination

SEP3$^{(75-178)}$ was purified via affinity and size exclusion chromatography. The higher molecular weight peak corresponding to a tetramer (Figure II-19) was concentrated and used for crystallisation trials. High-throughput crystal screening was performed using the HTX crystallisation facility, EMBL-Grenoble outstation (https://embl.fr/htxlab/). The screening identified multiple crystallisation conditions, which resulted in well diffracting crystals (Figure II-21). Additive screening and macroseeding was performed to improve the crystal quality. The details of the crystal screens and the optimisation steps are provided in the in section II.5.2.1.

The optimised crystals (Figure II-21) grew in space group $P2_12_12$ and diffracted to d_{min}= 2.5 Å. Due to the absence of similar crystal structures for the oligomerisation domain, molecular replacement could not be used to phase the data. After several trials with heavy atom soaking, seleno-methionine derivatisation was used for phasing. The crystals of seleno-methionine derivatised protein diffracted to d_{min} = 3.2 Å. Based on the obtained experimental phases, a partial structure was built and used as a molecular replacement model for the higher resolution native data. The structure solution and refinement was performed by Dr. Chloe Zubieta. I will briefly summarise the structural data and then detail my studies on the structure of SEP3 using site directed mutagenesis.

Figure II-21: ***Crystallization trials of SEP3$^{(75-178)}$.*** *The crystallization conditions were screened using the high throughput facility at HTX lab in EMBL, Grenoble. Several initial conditions were obtained which were optimized using additive screening and macroseeding. The images above show images of crystals obtained in different conditions. The crystals shown in far right image grown in 0.1 MTris pH 8.0, 25% ethylene glycol were used for native data collection.*

II.3.2.2 Structural characterisation of the SEP3 oligomerisation domain

A native diffraction data set from a crystal grown in 0.1M Tris pH 8.0, 25% ethylene glycol was collected at 100 K on the ESRF ID14-4 beamline, Grenoble, France (Gabadinho et al., 2010; McCarthy et al., 2009). The crystal belonged to space group $P2_12_12$ and data were collected at a wavelength of 0.9393 Å. The crystals had a Mathews coefficient of 4.49 $Å^3$ Da^{-1} with a solvent content of 72% with four molecules per asymmetric unit (Kantardjieff and Rupp, 2003; Matthews, 1968). Table II-3 summarizes the data collection and refinement statistics. The protein crystallized as a tetramer with a distinct dimerisation and tetramerisation interface (Figure II-22). Each monomer consists of two α-helices almost perpendicularly joined by a kink region. The dimer is formed by formation of a coiled coil between the first α-helices, the hydrophobic interactions in the kink region, and the hydrophobic interactions from the N-terminal part of second α-helices. The biological tetramer is a dimer of dimers formed via a crystallographic 2-fold rotation. An alignment of oligomerisation domain of different MADS TFs shows a number of conserved residues in the dimerization and tetramerization interface (Figure II-23). Electron density is visible for residues 83 to 175 (monomer A), 83 to 177 (monomer B), 88 to 178 (monomer C), and 93 to 175 (monomer D) (Figure II-22). Disordered N and C-terminal residues were not modelled.

Data collection	**SEP3 (75-178)**
Beamline	ID 14-4 (ESRF)
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	123.1, 143.2, 48.77
α, β, γ (°)	90, 90, 90
Resolution (Å)	60.-2.49 (2.55-2.49) *
R_{merge}	6.1 (40.1)
$<I / \sigma(I)>$	17.6 (4.3)
Completeness (%)	77.2 (20.2)
Redundancy	5.9 (6.1)

Refinement	
Resolution (Å)	34.4-2.49
No. reflections	23,723
R_{work} / R_{free} (%)	23.0 / 27.4
No. atoms	3,238
Protein	3,183
Ligand/ion	0
Water	55
B-factors (Å^2)	
Wilson B factor	43.7
Mean B value	68
Protein	69.4
Ligand/ion	Na
Water	62
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.2

*Table II-3: **Data collection and refinement statistics for SEP3$^{(75-178)}$.** The values in '()' refer to highest resolution shell.*

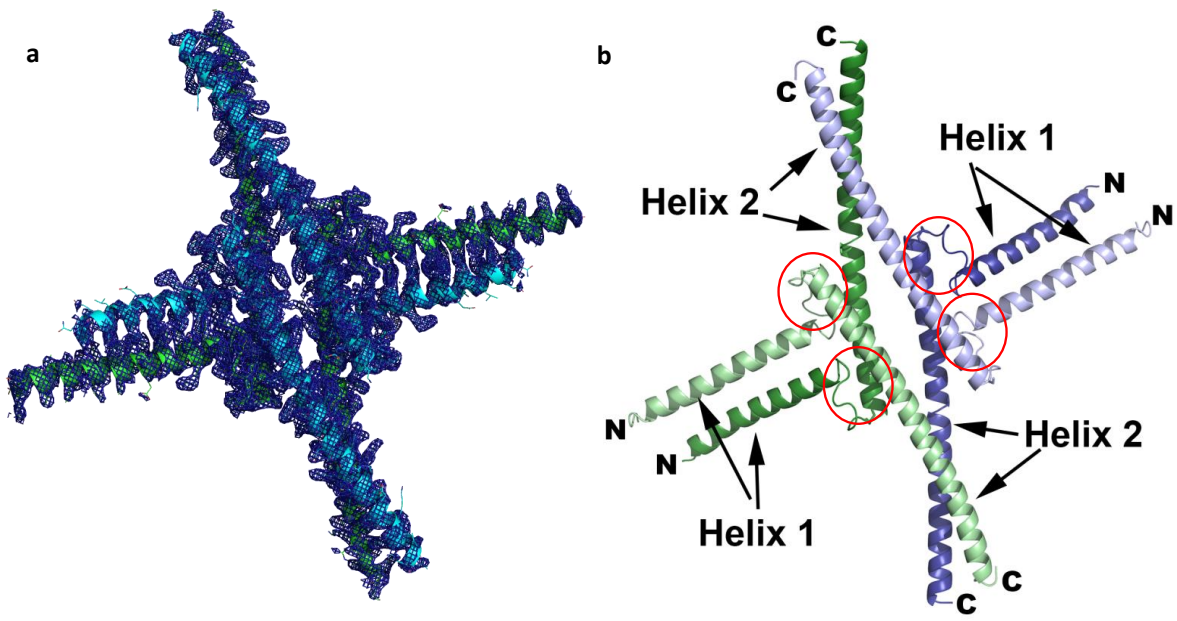


Figure II-22: ***Structure of SEP3$^{(75-178)}$. a)*** *The electron density map for SEP3$^{(75-178)}$ is shown in blue is displayed with 2 Fo-Fc contoured at 1.5 sigma.* ***b)*** *SEP3$^{(75-178)}$ tetramer depicted as cartoon with each monomer coloured uniquely in light green, dark green, light purple, and dark purple, respectively, with the N and C termini labelled. Helix 1 and helix 2 are labelled and indicated by arrow. The kink region is marked by red circles. Adapted from (Puranik et al., 2014).*

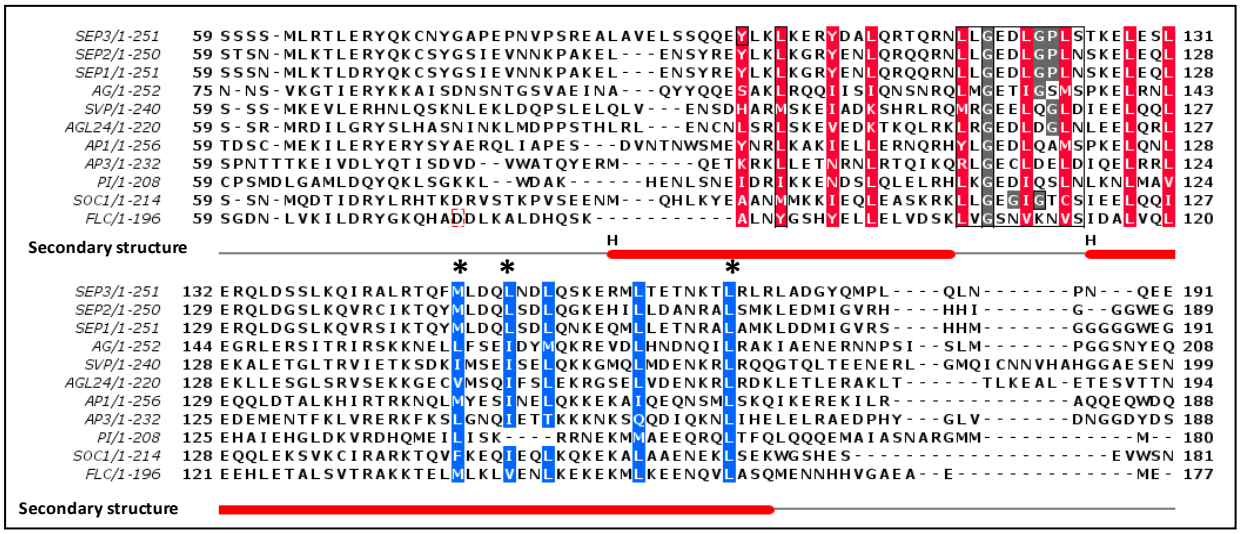


Figure II-23: ***Analysis of conserved residues in dimerization and tetramerization interface.*** *Sequence alignment of representative MADS TFs from Arabidopsis spanning the oligomerisation domain. The structure obtained from crystallized construct SEP3$^{(75-178)}$ is represented below by H followed by red bar representing alpha helix. The black box marks the kink region. Residues involved in dimerization and tetramerization are highlighted in red and blue respectively. The conserved Glycine and Proline residues in the kink region are coloured in grey. Point mutants are marked with a star.*

II.3.2.2.1 Monomer

The SEP3 construct spanning residues 75-178 is largely α-helical. The protein construct has two α-helices, helix 1 and helix 2 with a break in between due to the presence of Gly-Pro residues (Figure II-24a). This results in formation of a kink between helices 1 and 2. Glycine offers conformational flexibility to the structure while the proline residue, due to its side chain, is unable to form hydrogen-bonding interactions between the carbonyl backbone and amide proton. This results in the formation of a tight turn of almost 90° between the two helices. Extensive hydrogen bonding interactions between Glu-127 and Ser-124, a salt bridge between residues Arg-113 and Glu-118 and intramolecular hydrophobic interactions of multiple leucine residues (Leu-115, -120, -123, -128, -131, and -135) provide further stability to the kink region and impede self-association in to a single dimeric coiled-coil (Figure II-24b). An analysis of the kink region in representative MADS TFs shows that the region in the SEP (1-3) proteins is strikingly different and is characterized by presence of a glycine-proline motif which varies within the family, but is conserved for SEP1 and SEP2 (Figure II-23).

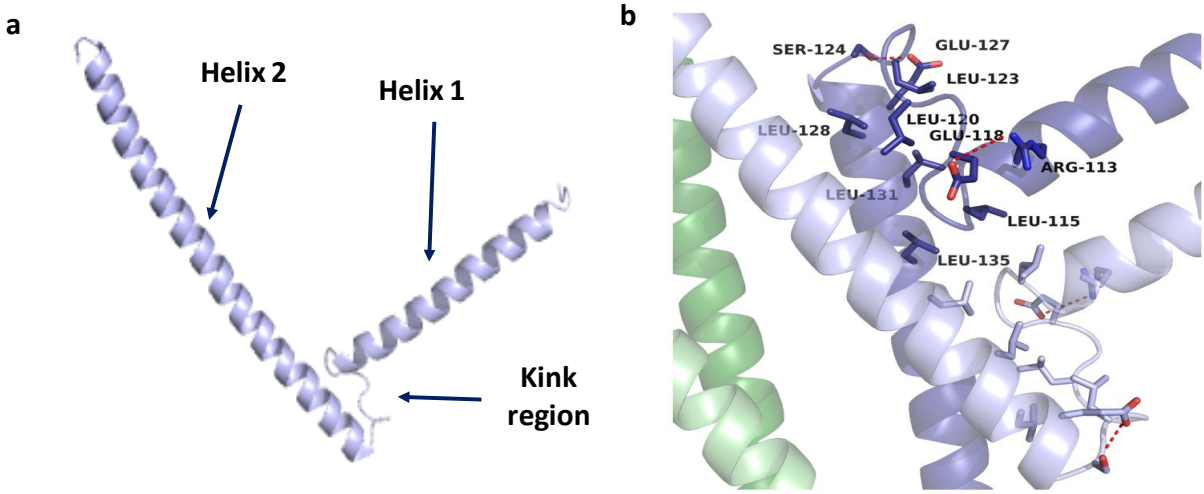


Figure II-24: ***Structure of SEP3(75-178) monomer*** *a) The monomer consists of two α-helices joined by kink region. b) The hydrophobic kink region is shown for one monomer with the view as per (Figure II-22b). Residues are labelled and drawn as sticks coloured by atom. Hydrogen bonds are shown as dashed red lines. Adapted from (Puranik et al., 2014).*

II.3.2.2.2 Dimer interface

Leucine zipper like heptad repeats $(abcdefg)_n$ (Fan et al., 1997; Yang et al., 2003) are observed between the helices 1 and 2. The hydrophobic residues are present at the 'a' and 'd' positions while the charged residues are located at the 'e' and 'g' positions. The heptad repeat is interrupted by the glycine and proline-rich kink region (residues 117 to 127; Gly-117, Gly-121, and Pro-122) thus preventing the formation of a single leucine zipper-like structure. The N-terminal region of helix 1 of SEP3 comprises two heptad repeats of hydrophobic residues tyrosine and leucine, 98-YxxLxxxYxxLxxx-111, forming a large interaction surface. Each monomer interacts via the N-terminal region of a partner forming a left-handed coiled coil and pushing the C-terminal helix (helix 2) 90° apart from the N-terminal helix 1. This prevents intramolecular association of the C-terminal amphipathic alpha helix. Three pairs of salt bridges formed between partner monomers of helix 2 comprising residues Glu-129/Arg-146, Glu-132/Arg-146, and Asp-136/Arg-143 (Figure II-25), further stabilize the dimer interface, which buries over 3000 $Å^2$ (17% of the total accessible surface area of the dimer), as calculated with AREAiMOL (Lee and Richards, 1971; Winn et al., 2011).

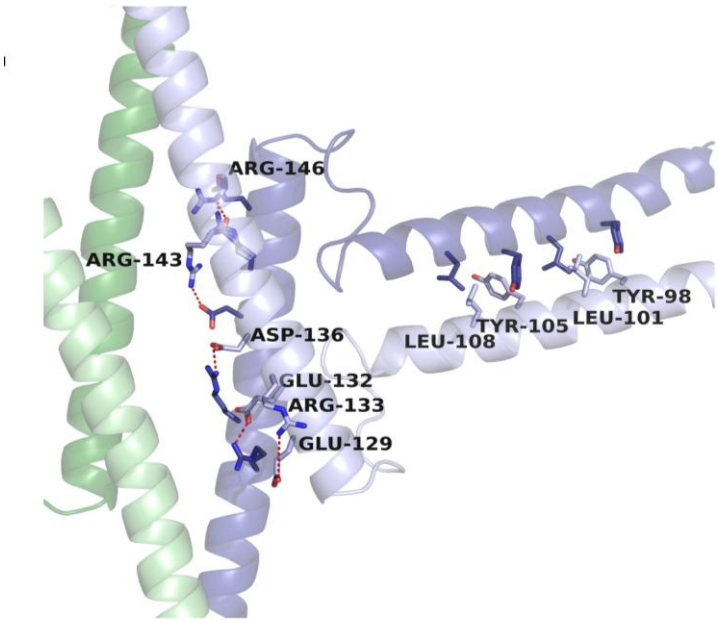


Figure II-25: ***Dimerization interface of SEP3$^{(75-178)}$****. The hydrophobic interaction surface formed by Leu and Tyr of helix1 and salt bridges of Glu, Arg and Asp form the dimerization interface. Residues are labelled and depicted as sticks coloured by atom. Hydrogen bonds are drawn as dashed red lines. For clarity, residues are labelled for one monomer (Puranik et al., 2014).*

II.3.2.2.3 Tetramer interface

The 'bent' arrangement of SEP3$^{(75-178)}$ contributes towards the formation of tetramers due to exposure of the hydrophobic face of the second alpha helix in the K-domain. As described above, the formation of dimer through the N-terminal region of helix 1 and the glycine-proline kink region between the two helices pushes the C-terminal helices apart and orients them at an angle of 90°, thus exposing a number of hydrophobic residues. As the K-domain cannot form a single elongated leucine-zipper due to the presence of glycine and/or proline residues, tetramerisation is favoured to bury the exposed hydrophobic residues. Through these hydrophobic residues (150-MxxxLxxLxxxxxxLxxxxxxL-171), helix 2 interacts with a partner dimer, thereby forming a tetramer. The interacting hydrophobic surface buries a total of 2700 Å^2 (9% of the total surface area of the tetramer). The tetramerisation interface is further stabilised by salt bridges between Lys-160 and Glu-161 and a hydrogen bond between Thr-167 and Asn-168 (Figure II-26).

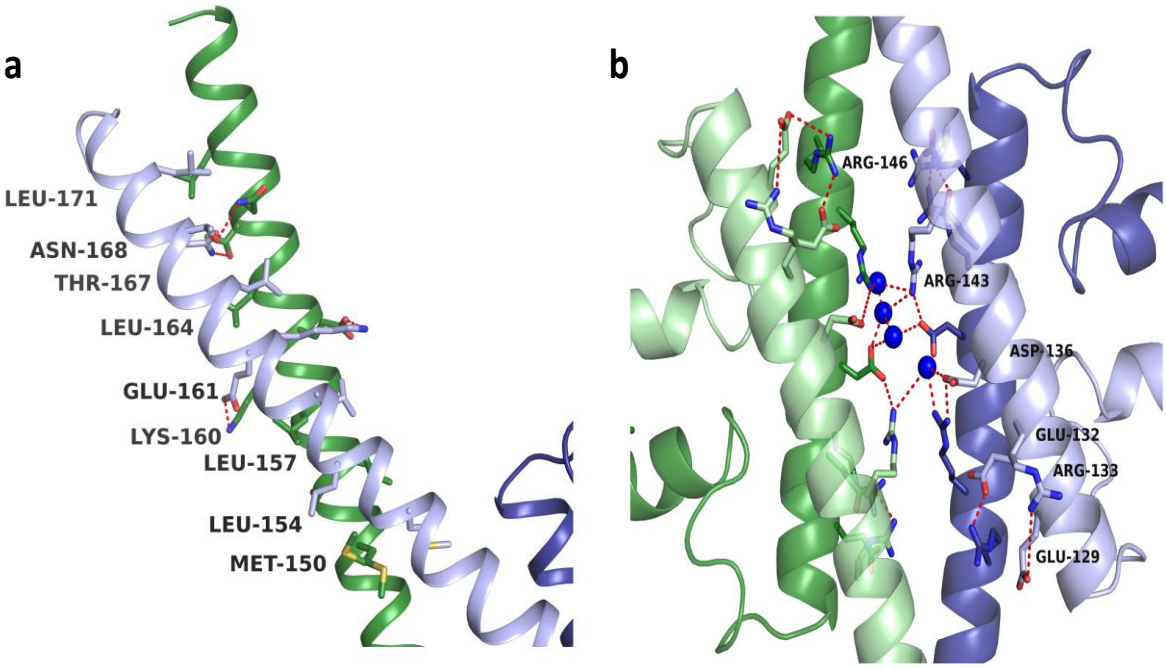


*Figure II-26: T**etramerization interface of SEP3$^{(75-178)}$. a)** The tetramerization interface of SEP3$^{(75-178)}$ showing different residues involved in hydrophobic and hydrogen bonding interactions. Interacting residues are depicted as sticks and coloured by atom. Hydrogen bonds are drawn as dashed red lines. For clarity, residues are labelled for one monomer. **b)** View down the 2-fold crystallographic axis that forms the tetramerization interface shows the intermolecular water mediated hydrogen bonding network. Residues are depicted as sticks and coloured by atom, water molecules are in dark blue, and residues labelled for a single monomer (green) for clarity. Hydrogen bonds are show as dashed red lines (Puranik et al., 2014).*

II.3.2.3 Site-directed mutagenesis

The sequence alignment of the K domain of different plant MADS TFs is shown in Figure Figure II-23 and highlights the conserved residues involved in the tetramerisation interface (Figure II-26). In order to investigate the importance of these conserved hydrophobic residues in helix 2, site-directed mutagenesis was performed. Three alanine mutants M150A, L154A, and L171A were generated and analysed by size exclusion chromatography. Very interestingly, even a single mutation was sufficient to disrupt the tetramerisation even at high protein concentrations of ~ 10-12 mg /ml (Figure II-27). In addition to the mutants, the two naturally occurring splice variants obtained from the TAIR database (www.arabidopsis.org) were also analysed. The SEP3$^{(\Delta V90)\ (75\text{-}178)}$ fragment with a valine deletion at position 90 in helix 1, shows the presence of both tetrameric and dimeric species, similar to the wild type, while SEP3$^{(\Delta 161\text{-}174)\ (75\text{-}178)}$, which lacks a part of the K domain, shows a complete transition into a dimer. The crystal structure does not show any interaction of Val-90 in the dimerisation or tetramerisation interface, thus explaining the predominance of tetrameric species and its resemblance to the wild-type construct. However, for the deletion mutant SEP3$^{(\Delta 161\text{-}174)}$, a part of the K domain involved in tetramerisation is deleted, thus leading to complete loss of the tetramer species, as would be predicted from the crystal structure of the wild type.

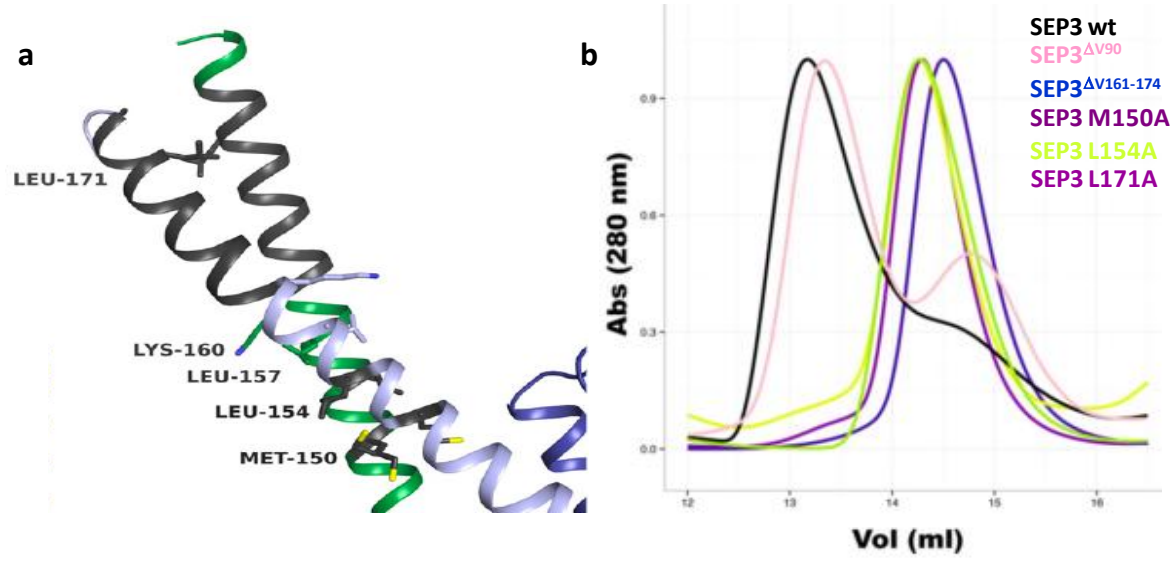


Figure II-27: ***Mutagenesis of SEP3$^{(75\text{-}178)}$ oligomerisation. a)*** *Tetramerization interface labelled as per Figure II-26a. Residues that are deleted in SEP3$^{(\Delta 161\text{-}174)}$ are shown coloured in grey. Three alanine point mutants M150A, L154A, and L171A generated are depicted as sticks.* ***b)*** *The size exclusion chromatograms for SEP3$^{(75\text{-}178)}$ wild type were compared with three alanine point mutants and two splice variants. The wild type and SEP3$^{(\Delta V90)}$ show a mixture of dimer and tetramer species. The different constructs are colour coded. All three point mutants and splice variant SEP3$^{(\Delta 161\text{-}174)}$ show formation of dimers (Puranik et al., 2014).*

Taken together, the crystal structure of SEP3$^{(75-178)}$ presented here and the accompanying mutagenesis studies provide important information about the tetramerisation of SEP3 and the residues involved in maintaining the quaternary structure of SEP3. However, in order to study whether the tetramer observed in the crystal structure also plays a role in orienting the DNA-binding domains for protein-DNA binding, it is important to study the full-length protein. The floral quartet model postulates a relation between control of transcription by MADS TFs and looping of the target DNA (Theissen and Saedler, 2001). Tetramer formation of MADS TFs has been previously linked to looping of DNA using indirect methods such as gel shift assays and tethered particle motion (Melzer et al., 2009; Mendes et al., 2013). However, with the availability of recombinantly purified SEP3 protein, we decided to utilize a more direct method, Atomic Force Microscopy (AFM). The main advantage of AFM is its ability to visualize macromolecular interactions such as DNA-protein complexes in solution using nanomolar quantities of samples and at the single molecule level. AFM can be used with a variety of buffers, thus more closely mimicking *in vivo* conditions. Although the purification of 6xHis-GST-SEP3$^{(1-251)}$ provided low yields of protein, it was sufficient for AFM studies. AFM studies were also performed with SEP3$^{(1-90)}$ in order to study the effect of the lack of an oligomerisation domain on DNA binding and as a control which abrogates putative DNA looping.

II.3.2.4 Atomic Force Microscopy (AFM)

AFM is a tool that allows visualising biological interactions at a single molecule level using nanomolar quantities of samples. Over the years, the use of AFM for studying biological samples has steadily increased (Müller and Dufrêne, 2008; Lyubchenko et al., 2011; Casuso et al., 2010, 2011; Parot et al., 2010). In addition to mapping of protein position along DNA molecules, the local DNA curvature, flexibility, changes in DNA structure on protein binding such as bending and wrapping of DNA can also be measured (Scipioni et al., 2002; Larquet et al., 1996; Chen et al., 2002; Beloin et al., 2003; Rivetti et al., 1999; Heddle et al., 2004).

AFM relies on the principle of touch rather than that of sight. As applied in these experiments, it consists of cantilevers where a sharp nanostylus (tip) with a radius of ~ 2 to 10 nm scans the surface of the sample. As shown in the Figure II-28, a laser is reflected from the back of cantilever on to a photodiode and is used to detect topographical information. A piezoelectric element controls the movement of the sample with respect to the tip and is

connected with a feedback loop. By controlling the deflection, the feedback loop is used to control displacement of the piezoelectric elements. AFM can be performed in air as well as in liquids and in different modes, of which contact and tapping are the most commonly used (Figure II-29).

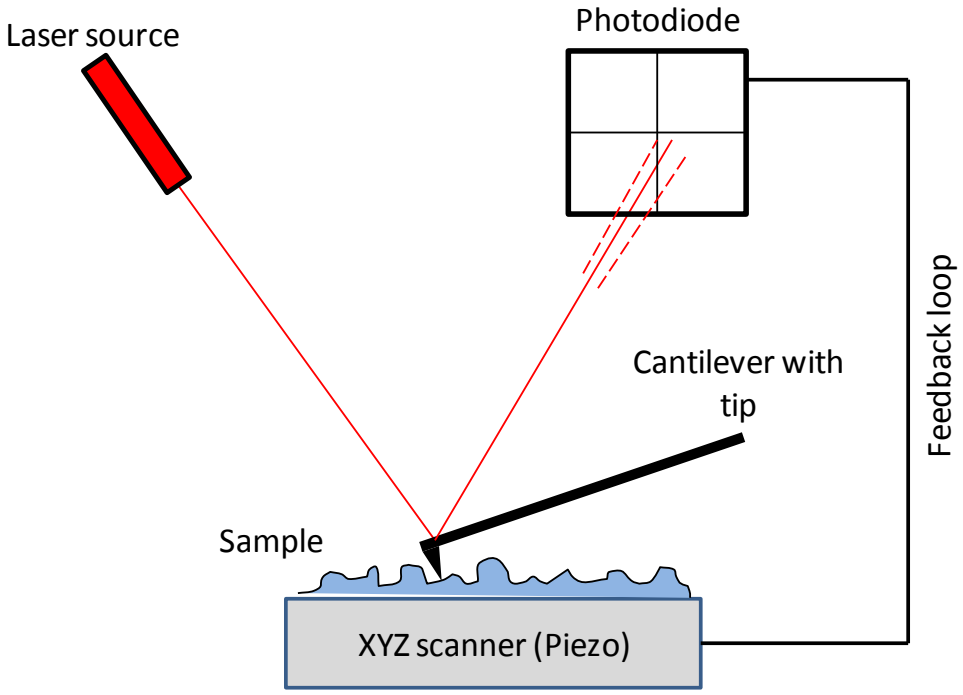


Figure II-28: ***Mechanism of AFM.*** *A cantilever with sharp tip scans the surface of sample. The laser is reflected from back of cantilever on a photodiode. The piezo electric element connected to the feedback loop controls the tip-sample interaction.*

In contact mode, the tip is in constant mechanical contact with the sample during the scan. The tip-surface interaction leads to a change in the deflection of the cantilever. A feedback loop maintains the deflection of the cantilever at a fixed set-point value by applying a varying voltage to the piezo-electric actuator. Consequently, the z-direction piezo-electric actuator moves up and down which, in turn, translate these displacements into topographical information of the sample. This mode is feasible in air and in liquids and yields significant information for hard samples. However, the high lateral force exerted during the scan can lead to displacement or deformation of the soft biological samples (Moreno-Herrero et al., 2004; Hansma et al., 1994; Putman et al., 1994).

In dynamic mode or tapping mode, the cantilever is oscillated near or at its resonance frequency. As the cantilever approaches the surface, the tip-sample interaction results in reduction of the amplitude. The scan is performed when the amplitude is equal to the fixed set point value, which is ~ 60% of the free amplitude. The feedback mechanism maintains the amplitude at a constant value by tuning the Z-scanner surface, thus representing the sample height. The tapping mode exerts significantly less force on the sample as it touches the sample intermittently and leads to lower tip and sample degradation and thus is the method of choice for biological samples (Zhong et al., 1993). Working in liquid maintains DNA and proteins in their native state and allows the possibility of varying external parameters such as buffer conditions over the course of an experiment. In this project, tapping mode was used and samples were scanned in buffered solutions in order to maintain the stability of the macromolecules. The instruments and probes used during the experiments are discussed in the Appendix I.

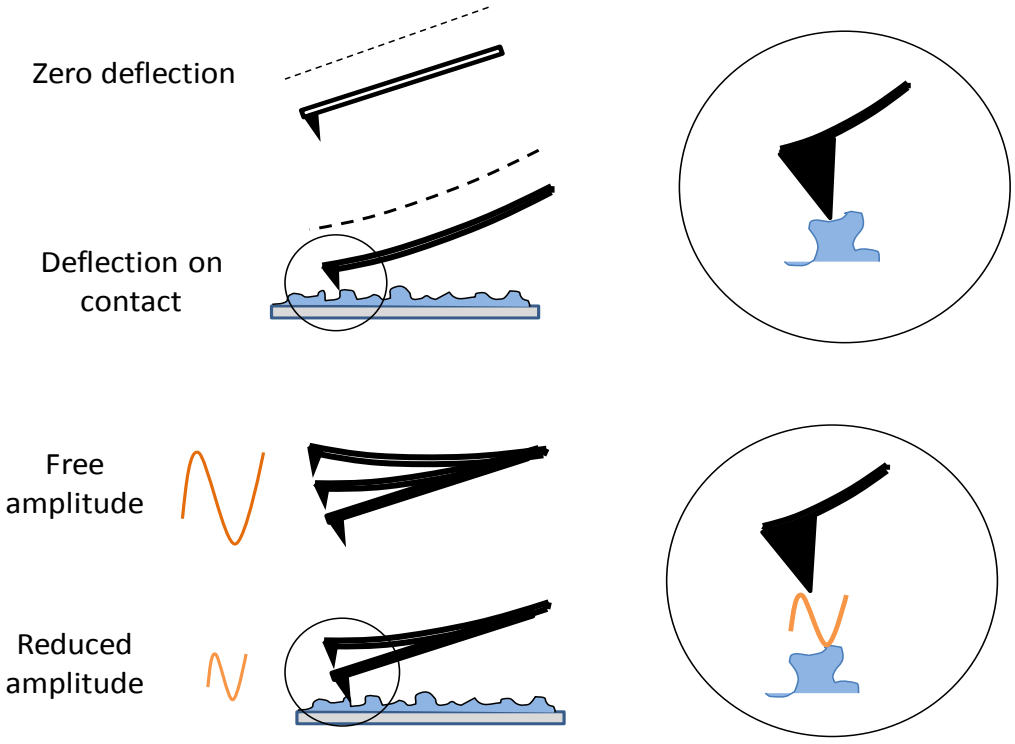


Figure II-29: ***Modes of imaging in AFM.*** *(top) In contact mode the tip is in constant contact with the sample while the voltage applied differs while in (bottom) tapping mode, the tip is oscillated at fixed frequency and is only intermittently in contact with sample.*

For our studies, although the purification of 6xHis-GST-SEP3$^{(1-251)}$ provided low yields of protein, it was sufficient for AFM analyses. AFM analyses were also performed with truncated SEP3$^{(1-90)}$ in order to study the effect of lack of oligomerisation domain on DNA binding. The selection of the DNA for AFM studies was based on ChIP-seq analysis and logos from the JASPAR database (Figure II-30) (Mathelier et al., 2015) along with the genetic data available on SEP3 binding targets. Two DNA sequences from the *AP3* (999 bp) and *SOC1* (1024 bp) promoters, both of which are shown to be regulated by SEP3 (Kaufmann et al., 2009) were used and PCR amplified from *Arabidopsis thaliana* genomic DNA. The details of the primers and the PCR reactions are described in section II.5.1.2 and Table II-4. The *AP3* and *SOC1* promoter DNA sequences with wild type CArG box elements highlighted in yellow are shown (Figure II-31). In case of *AP3*, two mutants, M1 and M2 were generated by replacing the underlined region in the sequence with the CArG box sequence (CCTTTCATGG) printed in bold in Figure II-31. The *SOC1* promoter DNA used for AFM has two CArG box binding sites (Figure II-31) (Mathelier et al., 2015). The smaller 557 bp DNA strand was also amplified using the longer DNA strand as a template. The different DNA targets with the locations of CArG box sequences are shown in the Figure II-31.

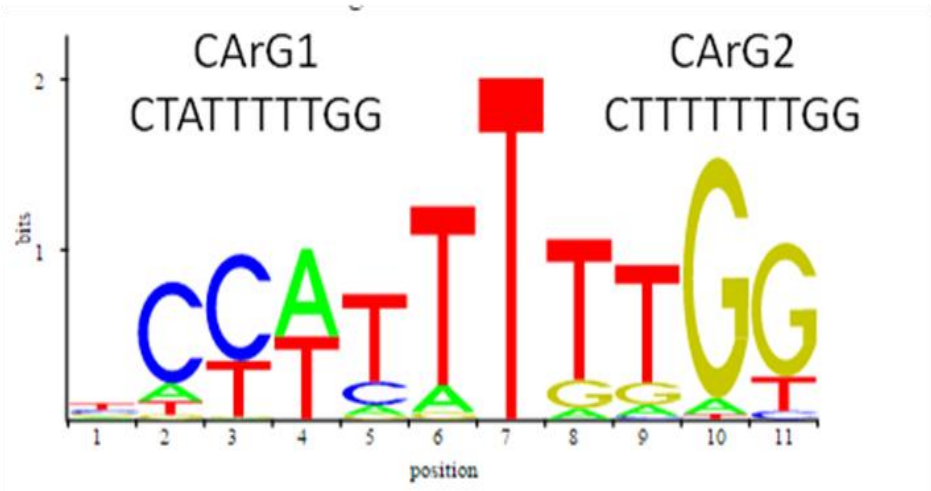


Figure II-30: ***CArG box regions for SEP3 binding.*** *The sequence logo represents the SEP3 binding profile obtained from JASPAR database* (Mathelier et al., 2015). *The height of nucleotides indicates its relative conservation at that position. The CArG box sequences from SOC1 promoter used in this study are provided for comparison.*

AP3 promoter DNA

CCATTTTCGTGACTCACTCACTGATTTCCATTGCTTGAAAATTGATGATGAACTAAGATCAATCCATGTTAGTTTCAAAA
CAACAGTAACTGTGGCCAACTTAGTTTTGAAACAACACTAACTGGTCGAAGCAAAAAGAAAAAAGAGTTTCATCATATATC
TGATTTGATGGACTGTTTGGAGTTAGGACCAAACATTATCTACAAACAAAGACTTTTCTCCTAACTTGT (M1-
CCTTTCATGG)CATTCCTTCTTAAACCCTAGGGGTAATATTCTATTTTCCAAGGATCTTTAGTTAAAGGCAAATCCGGGA
AATTATTGTAATCATTTGGGGAAACATATAAAAGATTTGAGTTAGATGGAAGTGACGATTAATCCAAACATAT (M2-
CCTTTCATGG)ATATCTCTTTCTTCTTATTTCCCAAATTAACAGACAAAAGTAGAATATTGGCTTTTAACACCAATATAAA
AACTTGTTCACACCTAAACACTTTTGTTTACTTTAGGGTAAGTGTAAAAAGCCAACCAAATCCACCTGCACTGATTTGACGT
TTACAAACGCCGTTAAGTTTGTCACCGTCTAAACAAAAACAAAGTAGAAGCTAACGGAGCTCCGTTAATAAATTGACGAA
AAGCAAACCAAGTTTTTAGCTTTGGTCCCCCTCTTTTACCAAGTGACAATTGATTTAAGCAGTGTCTTGTAATTATACAACC
ATCGATGTCCGTTGATTTAAACAGTGTCTTGTAATTAAAAAAATCAGTTTACATAAATGGAAAATTTATCACTTAGTTTTCA
TCAACTTCTGAACTTACCTTTCATGGATTAGGCAATACTTTCCATTTTTAGTAACTCAAGTGGACCCTTTACTTCTTCAACTC
CATCTCTCTCTTTCTATTTCACTTCTTTCTTCTCATTATATCTCTTGTCCTCTCCACCAAATCTCTTCAACAAAAAGATTAAACA
AAGAGAGAAGAATATGGCGAGAGGGAAGATCCAGATCAAGAGGATAGAGAACCAGACAAACAG

SOC1 promoter region

TACTTATTTTGGAAAAAAGCCTTAAGAAAGACCAAAAATAGCATATTTTGATACATATGGACATTTTTACATACACATC
TTTTTCCCTCCCATAAATATCTAGTTGGATGGAAATGCCTGTCAAATTATTTTTATTAAATTTCTTCTAACAAAAATCTTTTTT
TGTTTTTAATGTGGACACAAACACTCTCTCGTACCTATATGCCCCCACTTGTAATTGATATGAAAAATAAGAATATCGTTTTT
TCTTACCTGTGAGTAATACAACTATATTGGTTTTAATCAATTTTTATAATATCTTTCCATCCCAACAGATAAAATTCATTTGA
TTTGTATATTTTCTATGCTTGTTTGTCCATTGTTTCTTATTGTTGTTCTTAAATTGTGTTTTTAGTTATATGACATAATCTTACA
TCTCCATGTGAATAATTAATTACTTGCAATTGTAAGTTTTTATAAGTTGGACAAAACCACATTACAAAATACTACAATTTCCT
CATTACCTAAGTGACCATTAATCCAAATATTTGGAGGCTAGTACAGAGACAATGGGGCAATGGCGTTAAGTAGCGACGTG
TCTAAAGAGGCATTTGCTATTTTTGGTCCCTCGGATTACTAAAGAAAACGTAACTTAGAAATCCAATAATAATTCAGCTTAT
CGAACGTCTTGTCTAGCTAGTGGCACCAAAAAAATATCCTTTTTTGGAGATAAAATTGTTAATCGATATCTAATGATCTTTT
ATCTATATCGGGAGGAGGACCACACAAACTAATTATTTCGAAAAATACACATCTTCTCTGGTGAAATATGTTAACGTATTT
ATATATATTAAATATCAAACAGATAGGTCCGAAAATCTGTATGGATAATTTTTATATAAATACACATCTTCTCTGGTGAATT
GTCTCTATCTTTATATTTTCGTGTTTGTATAAAGTATAAACCCGGTTTGAAGTTTAGTTATGTATTTTGTAGACCCAATTCAG
ATTTAAAGAAACATCATCGTATACCACTAGTAAATACGATCACTTCTTAACTGGTTGAGACAAATCCAAAACAAAAAGTAT
CTAACGTAACTGAACCAACTATACATTACTTAAAGGATTCGTCCTATAGATTAAAACAATCGTGTTAGTAATGTGTCACATC
GAAAGTAAGATAGAACCAGGCTATGCTTCAGGTTTTTGTTCACTGCGATTAATTTTAGATATGATTGACCACTACAATTATT
TTTAAGTGGTAGTAATATCCGATTTTTTTTGTACATAATCATATAAACTAATCTAATTTTCGCTTCGAGATCCCAAAGTATAA
TTTTTTTAAAAATAT

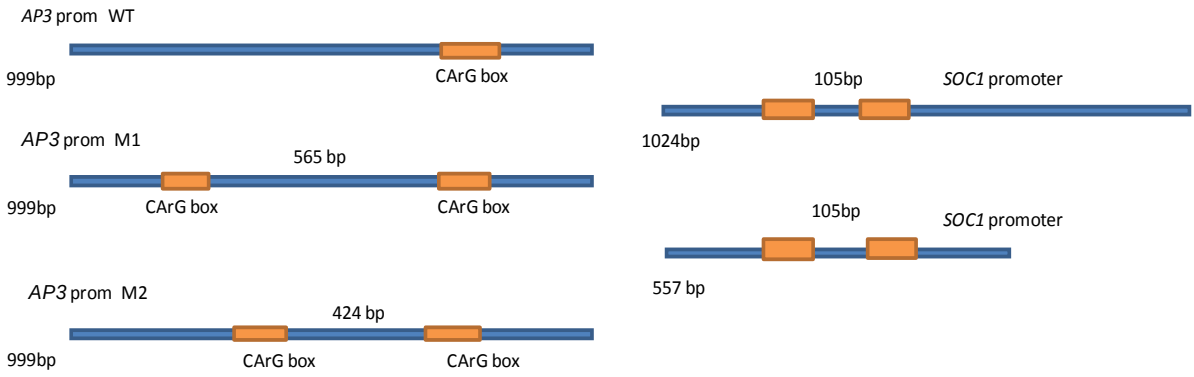


*Figure II-31: D**esign of target DNA sequences for AFM experiments. (top)**DNA sequences for AP3 and SOC1 promoters used for analysis. **(bottom)** The target DNA is represented by long blue bars while the CArG box sequences are represented by orange boxes. The two mutants for AP3 promoter differ in the position of CArG box elements.*

Being a single molecule technique, AFM is highly sensitive to the sample and scanning mode. All the experiments were performed in liquid using tapping mode to ensure minimum force exerted on the sample. This helps to maintain the sample integrity in the case of sensitive biological macromolecules. Several parameters like sample concentration, composition and pH of buffers, the percentage and type of divalent cations and DNA-protein immobilisation methods were optimized in the AFM experiments.

The series of AFM experiments have helped us to understand the DNA binding characteristics for SEP3. Binding of different constructs of SEP3 with DNA from the promoter region of *AP3* and *SOC1* was performed. Figure II-32 shows a DNA-protein interaction for 2-5 nM SEP3$^{(1\text{-}90)}$ and *AP3* promoter DNA. Using image processing software, Gwyddion (Nečas and Klapetek, 2011) and ImageJ (Schneider et al., 2012), an analysis of the protein-DNA binding interaction was performed.

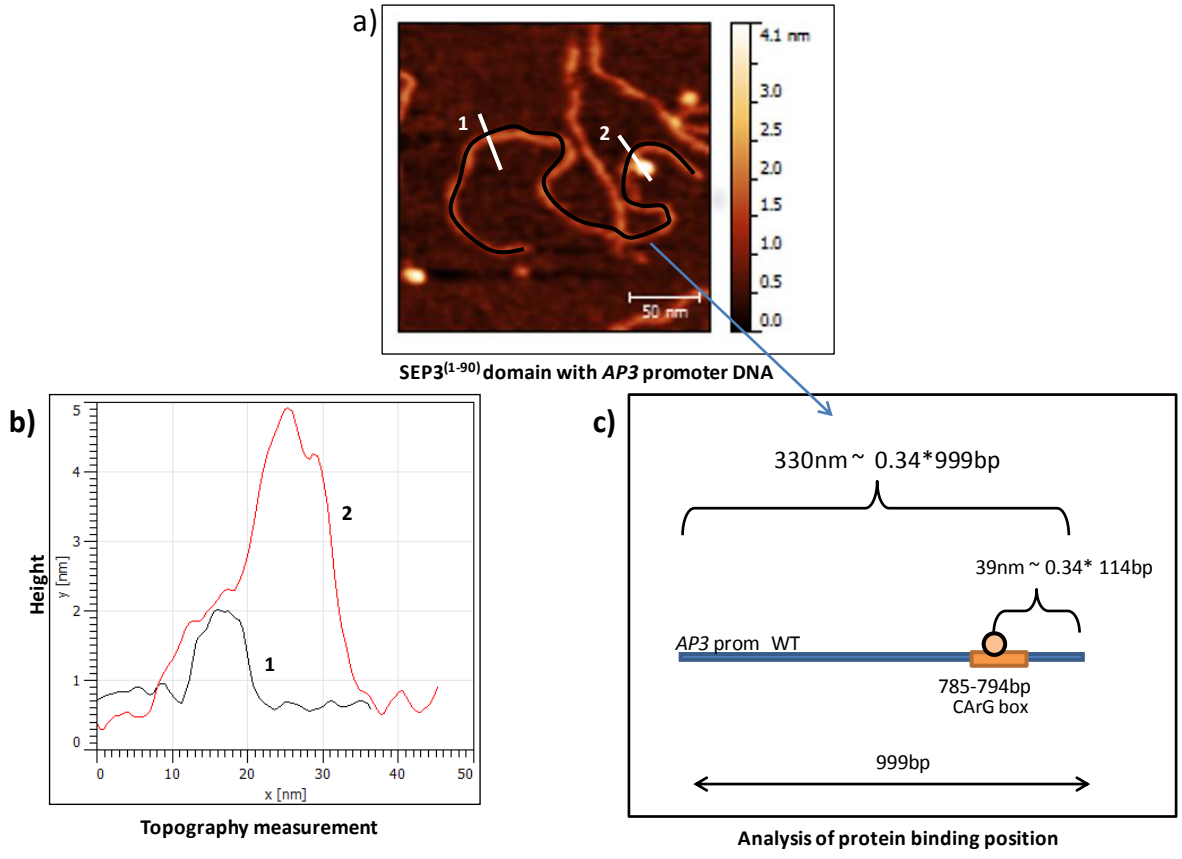


Figure II-32: ***Preliminary AFM analysis. (a)*** *AFM scan of SEP3$^{(1\text{-}90)}$ with wtAP3 promoter DNA.* ***(b)*** *Topography measurements performed using Gwyddion. The numbers 1 and 2 on graphs correspond to DNA alone and DNA-protein complex as shown in (a).* ***(c)*** *Protein binding position calculated corresponding to DNA length.*

Topography measurement for the complex as observed in the Figure II-32b, show that the free DNA is approximately 1.5- 2 nm in height while the complex measures around 5 nm. These results are compatible with the diameter of B type DNA and topography measurements in liquids previously performed (Moreno-Herrero et al., 2003; Cerreta et al., 2013). The length of the DNA strand of 330 nm (Figure II-32c) is also compatible with the B type DNA base-base distance of 0.34 nm (999*0.34=339 nm) while the position of protein at 39 nm approximately corresponds to the location of CArG box sequence (39/0.34= 114 bp).

According to the floral quartet model explained in the Introduction section I.2.2.2, the MADS domain TFs form a tetramers and bind 2 CArG boxes simultaneously, resulting in looping of the DNA (Theissen and Saedler, 2001). This phenomenon has been further supported by gel shift assays and tethered particle motion, (Melzer and Theissen, 2009; Mendes et al., 2013) but never directly visualised. In order to investigate DNA binding events and potential DNA looping, we generated two mutants of *AP3* promoter DNA, M1 and M2, each with two CArG box sequences (Figure II-31). Similar AFM experiments were then performed with wild-type and mutated *AP3* DNA strands and 6xHis-GST-SEP3$^{(1-251)}$ protein. The use of the GST tag helps to stabilize the protein during purification and reduces the propensity for aggregation (Hammarström et al., 2002; Dyson et al., 2004; Harper and Speicher, 2011).

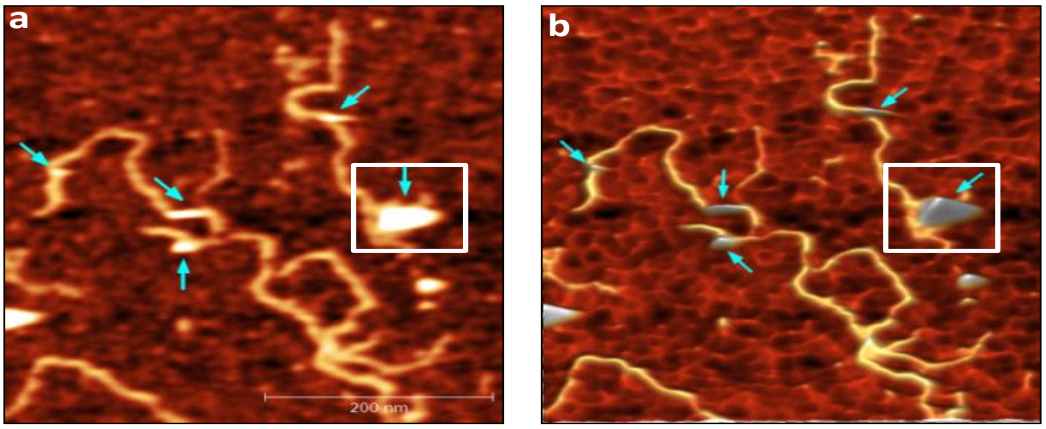


Figure II-33: ***AFM scan of AP3 promoter (M2) DNA interaction with 6xHis- GST-SEP3$^{(1-251)}$.*** ***a)*** *2-5 nM protein was treated with 1-2 nM DNA and bound on freshly cleaved mica with 10 mM HEPES, pH 7.0, 10 mM $NiSO_4$ adhesion buffer. The proteins to DNA are marked by blue arrows. The white box shows formation of higher order complexes or aggregates.* ***b)*** *3D interpolation of figure (a) for better visualization generated with Gwyddion* (Nečas and Klapetek, 2011).

Figure II-33 shows binding of 6xHis-GST-SEP3$^{(1-251)}$ protein to *AP3* promoter DNA mutant (M2) at both of its CArG box sequences. However DNA looping is not observed. The larger size of one protein complex observed may either be due to development of higher order oligomer through the K and C domains or formation of soluble aggregates with well-folded and DNA-binding competent M domains.

According to (Mendes et al., 2013), for loop formation in a protein-DNA complex, the distance between the two CArG boxes should be preferably less than 300 bp. In the current experiment, the distance of 424 bp between the two CArG box elements may be one of the reasons for the absence of loop formation. In order to confirm whether this was the case, the AFM experiment was repeated using *SOC1* promoter DNA. This DNA strand contains two CArG box sequences (CTATTTTTGG and CTTTTTTGG) spaced at 105 bp. The binding of SEP3 to *SOC1* promoter as a tetramer has been predicted previously (Muiño et al., 2013).

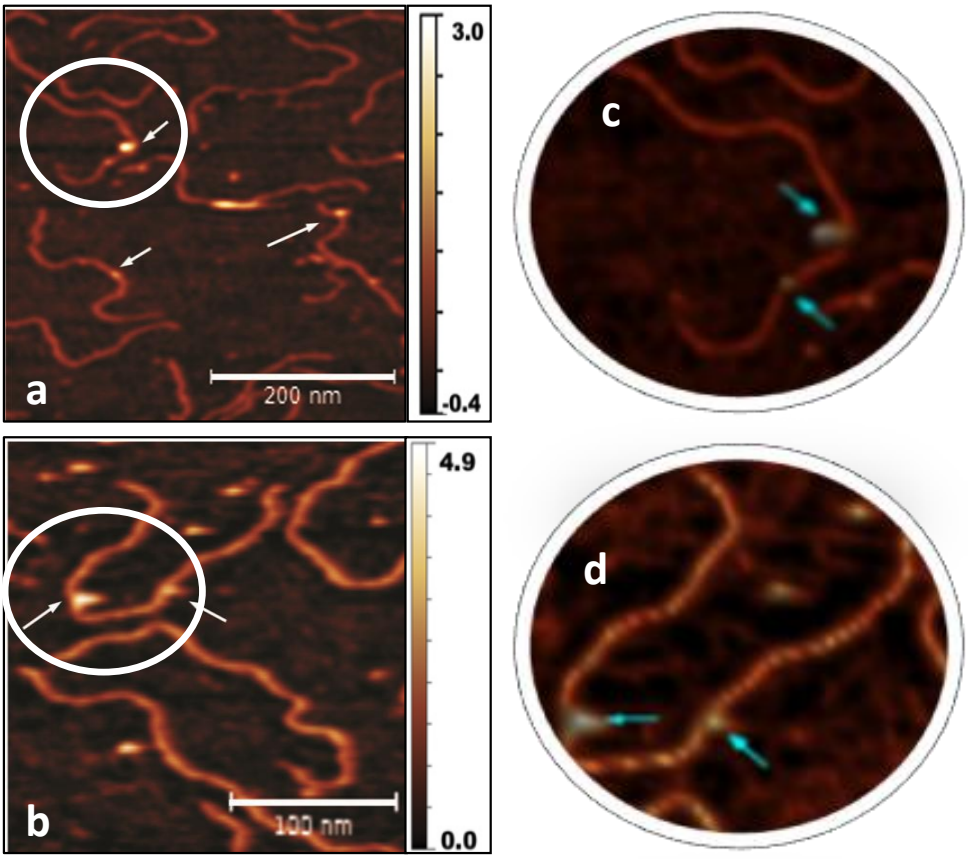


Figure II-34: ***AFM scan of SOC1 promoter DNA interaction with 6xHis-GST-SEP3$^{(1-251)}$. a & b****) 2-5nM of DNA was treated with ~ 5nM of protein and bound on freshly cleaved mica with 10mM HEPES, pH 7.0, 10mM $NiSO_4$ adhesion buffer. The sample was further diluted for imaging. The bound protein is marked by arrows.* ***c & d)*** *Three dimensional interpolation of the encircled white region focusing on protein-DNA complex for better visualization. Adapted from (Puranik et al., 2014).*

Figure II-34 shows the interaction of approximately 5 nM 6xHis-GST-SEP3$^{(1-251)}$ protein with 2-5 nM *SOC1* promoter DNA. Even though we can observe the protein binding on DNA represented by the bright spots, no DNA loop formation is observed. In order to enhance the probability of tetramer formation the protein concentration was increased 3-5 fold while keeping the DNA concentration constant.

Figure II-35 shows the interaction for 10-15 nM 6xHis-GST-SEP3$^{(1-251)}$ protein with *SOC1* promoter DNA. The image clearly shows formation of a loops and the binding of two DNA strands around a protein complex. In cases where looping is not observed, intermolecular protein-DNA interactions are observed in which different DNA strands are bound at their protein binding site. The position of the site-specific binding was confirmed by calculating the contour length of DNA and comparing it with the location of CArG boxes, as previously performed in Figure II-32.

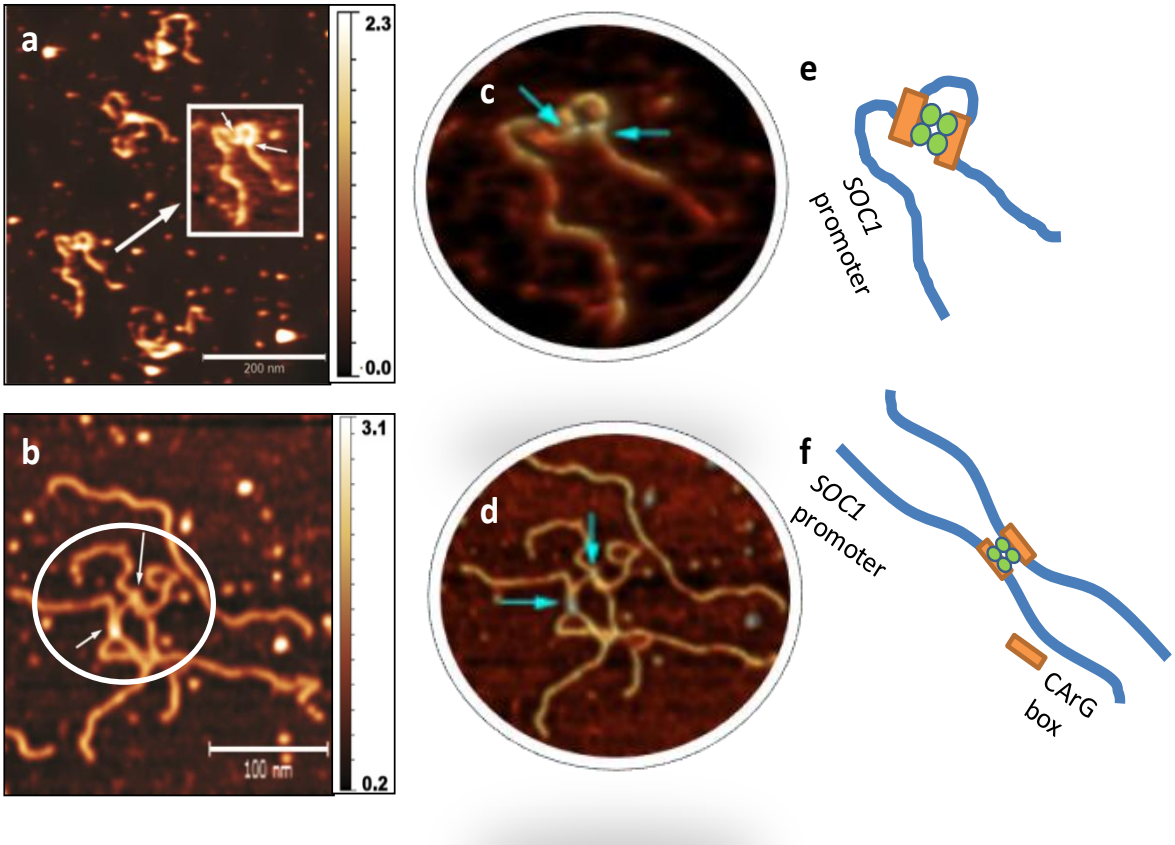


Figure II-35: ***AFM scan of SOC1promoter DNA interaction with 6xHis-GST-SEP3$^{(1-251)}$.*** ***a & b)*** *2-5 nM of DNA was treated with ~10-15 nM of protein and bound on freshly cleaved mica with 10 mM HEPES, pH 7.0, 10 mM $NiSO_4$ adhesion buffer. The sample was further diluted for imaging. The bound protein is marked by arrows. In Figure (a), the inset shows magnified DNA loop.* ***c & d)*** *Three dimensional interpolation of the marked regions from (a) & (b) for better visualization.* ***e & f)*** *DNA-protein interaction depicted in cartoons. Adapted from (Puranik et al., 2014).*

The distance between two CArG boxes is around 105 bp, indicating they are separated by ~10 helical turns. Thus the CArG boxes are on same side of DNA, increasing the probability of DNA binding and looping (Melzer et al., 2009). Formation of a loop for *SOC1* promoter DNA with a distance of 10 turns compared to the absence of observed looping for the *AP3* DNA with a contour length of around 450 bp, or 45 turns is likely due to decreasing probability of loop formation due to longer distance (Melzer et al., 2009).

The AFM scans were successful in not only confirming the protein-DNA binding but also for visualising the DNA looping mediated by a protein binding. However it was necessary to verify that the formation of loop during protein-DNA binding is a characteristic imparted by oligomerisation and tetramerisation domains and not an experimental artefact. Hence, we decided to repeat the AFM binding experiments with a truncated construct comprising only the 'M' and the 'I' domains. The AFM scans in Figure II-36 show an interaction of approximately 5 nM $SEP3^{(1-90)}$ protein with *SOC1* promoter DNA. Although protein-DNA complex formation through the M domain is observed, no loop formation is seen.

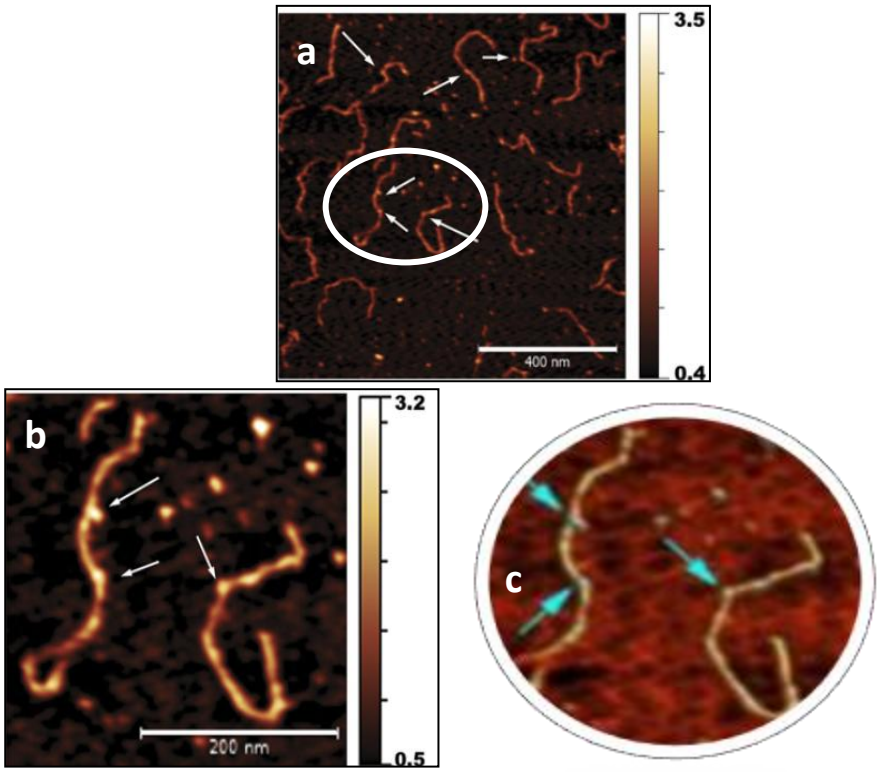


Figure II-36: ***AFM scan of $SEP3^{(1-90)}$ protein with SOC1 promoter DNA.*** ***a)*** *2-5 nM of DNA was treated with ~ 5 nM of protein and bound on freshly cleaved mica with 10mM HEPES, pH 7.0, 10 mM $NiSO_4$ adhesion buffer. The sample was further diluted for imaging. The bound protein is marked by arrows.* ***b)*** *Zoomed image focusing on protein-DNA complex from the encircled region of (a).* ***c)*** *Three dimensional interpolation of the marked region from (b) for better visualization. Adapted from (Puranik et al., 2014).*

To check the concentration effect on DNA binding, the experiments were repeated with aproximately 20-25 nM protein. However, even though non-specific binding on the DNA was observed, no active DNA looping was seen (Figure II-37).

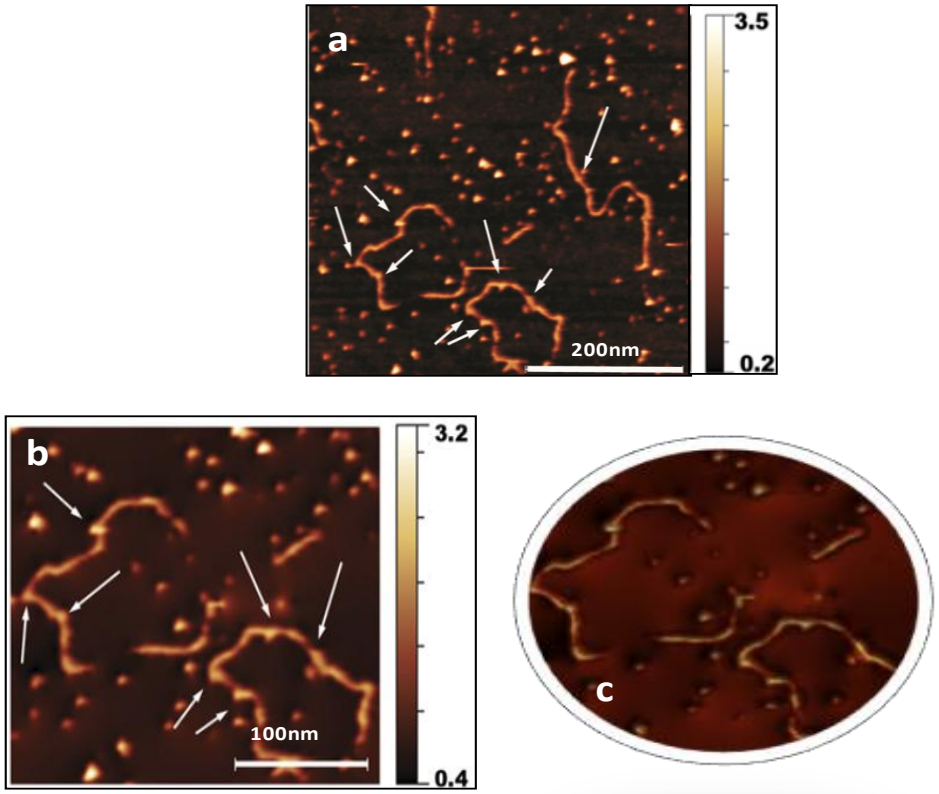


Figure II-37: ***AFM scan SEP3$^{(1-90)}$protein interacting with SOC1 promoter DNA. a)*** *2-5 nM of DNA was treated with ~25 nM of protein and bound on freshly cleaved mica with 10mM HEPES, pH 7.0, 10 mM $NiSO_4$ adhesion buffer. The sample was further diluted for imaging. The bound protein is marked by arrows. Non-specific DNA-protein interactions are observed due to high concentrations.* ***b)*** *Zoomed image focusing on protein-DNA complex.* ***c)*** *Three dimensional interpolation of the marked region from (b) for better visualization. Adapted from (Puranik et al., 2014).*

The ability to visualize protein-DNA complexes at the single molecule level, the use of nanomolar quantities of samples and the capability to work in liquids with almost no limitations on buffer conditions are some of the greatest attributes of this technique (Seong et al., 2002). The AFM results show the formation of looping for DNA-protein binding in the case of the full-length SEP3 protein. It is also observed that the formation of DNA loops is concentration dependent probability due to change in availability and proximity of interacting molecules. Attempts to form DNA loops with longer spacing between CArG boxes were unsuccessful, suggesting an optimum spacing of binding sites is necessary for looping to occur.

II.4 Conclusions

Using a combination of conventional and high throughput approaches I was able to establish a protocol for the expression and purification of various constructs of SEP3. This enabled biochemical and biophysical studies for a plant MADS TFs, SEP3. A thorough crystallisation screening and optimisation process led to the first high-resolution crystal structure for the plant-specific oligomerisation domain of a MADS TF. The crystal structure presented here shows that the SEP3 K domain is composed of two alpha helices aligned at an angle of 90° from each other and separated by a hydrophobic kink region and. This configuration provides a versatile oligomerisation interface to the protein allowing it to form homo and heterodimers and tetramers and provides each alpha helix with an independent interaction interface. The ability of SEP3 to form multimeric complexes with other MADS TFs as described previously (Immink et al., 2009; Kaufmann et al., 2009) can be attributed to this characteristic structural arrangement. Considering the homology of K domain within the MADS TFs, it can be postulated that tetramerisation through the K domain plays an important role in the activity of the MADS TFs. The hydrophobic residues in helices 1 and 2 in the K domain are highly conserved in the SEP class of proteins while more variable in other MADS TFs. This may provide a versatile and plastic interaction interface and contribute to the promiscuous activity of SEP proteins. The conserved nature of the gly-pro motif only in the SEP proteins might suggest their role in formation of almost all currently known MADS TF tetrameric complexes.

The mutagenesis studies have shown that the SEP3 is highly sensitive to mutations in its hydrophobic residues. These mutations tend to disrupt tetramerisation leading to formation of a dimeric species. This might suggest that the homotetrameric form obtained from the crystallisation studies might not be the most stable form and SEP3 might prefer to form heterotetramers over homotetramers. It can also suggest a role of additional cofactors in stabilisation of the tetrameric species. Although the exact physiological significance of SEP3 homotetramers is not yet clear, the co-operative binding leading to DNA looping has been postulated to play a role in floral quartet formation and recruiting other proteins to the quartet (Melzer et al., 2009). Also, it has been previously suggested that in absence of appropriate heterotetramerisation partners, SEP proteins tend to homotetramerise especially in the early stages of flower development (Jetha et al., 2014; Melzer et al., 2009).

Apart from hetero-oligomerisation, the functional diversity of SEP3 is further increased by the presence of naturally occurring splice variants. The splice variant $SEP3^{(\Delta 161\text{-}174)}$ lacking a

portion of K domain shows impaired homotetramerisation and a prevalence of a dimeric species in the gel filtration studies (Figure II-27). This may have significant *in planta* effects. Previous studies have shown that the majority of characterized splice variants in plant MADS TFs cluster to I and K domains, suggesting a general method for increasing functional diversity via alterations in oligomerisation state and/or oligomerisation partners (Severing et al., 2012).

The AFM studies presented here show the first visualisation of DNA-SEP3 binding that demonstrates SEP3 is able to tetramerise and loop DNA. This may indicate that even weak tetramers can occur when the dimeric species are brought into close spatial proximity due to binding of DNA. This would enable the MADS TFs to act as a dynamic interaction network, exploiting chromatin events that temporarily bring distal bound MADS dimers together to enable tetramer formation, as well as forming tetramers when bound on adjacent sites on the DNA. This dynamic oligomerisation may be a key factor in the ability of the MADS TFs to regulate diverse target genes.

With the data obtained from the crystal structure, homology modelling of the missing M domains, mutagenesis experiments and AFM scans, a model for SEP3-DNA binding and tetramerisation can be postulated. As shown in Figure II-38, SEP3 forms a dimer through its M domain and binds two binding sites on DNA. The DNA binding of SEP3 dimers would lead to changes in the DNA structure and ultimately looping of DNA thus promoting tetramerisation. The same model could occur *in vivo*, with multiple different MADS dimers binding proximal sites, resulting in the formation of different DNA looped structures depending on site spacing and affinities of the resulting heteromeric complexes. In addition, as SEP3 has been crystallized as a tetramer, we cannot exclude the possibility of tetramer formation occurring prior to DNA binding for the MADS TFs.

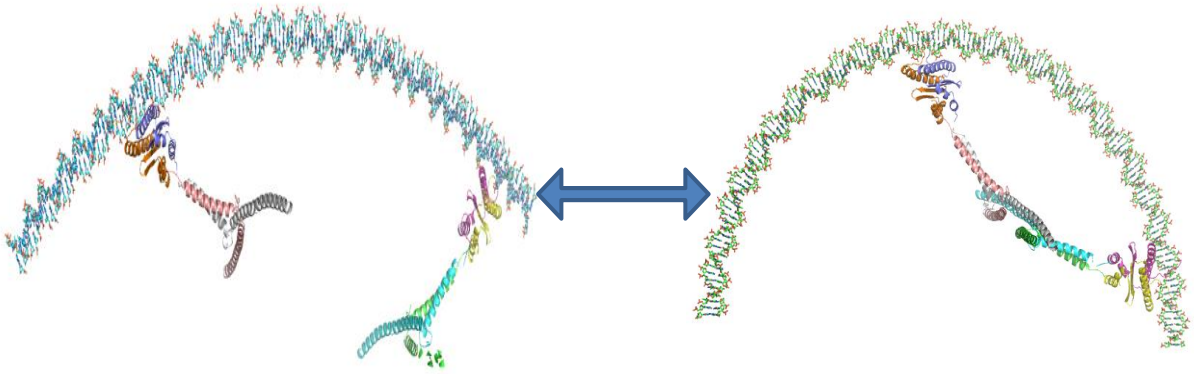


Figure II-38: ***Model for SEP3 mediated DNA looping.*** *SEP3 binds to DNA as a dimer and leads to bending of DNA. This reduces the spatial distance between two distinct SEP3 dimers which interact through K domain to form a tetramer. Only the MIK domains of SEP3 model are shown here.*

II.5 Materials and methods

II.5.1 Construct design and purification

II.5.1.1 Strains and plasmids

SEP3$^{(75-178)}$ was cloned in to the expression vector pESPRIT 002 (Hart and Tarendeau, 2006; Guilligay et al., 2008; Yumerefendi et al., 2010) using the AatII and NotI sites. The plasmid contains an N-terminal 6xHis tag followed by a TEV protease cleavage site. SEP3$^{(1-251)}$ was cloned into the expression vector pETM-30 using the NcoI and XhoI sites. The plasmid has an N-terminal 6xHis/GST double tag followed by a TEV protease cleavage site. The SEP3$^{(1-90)}$ construct was cloned into the expression vector pET15b (Merck Millipore, Darmstadt, Germany) using the XhoI and BamHI sites. The plasmid contains an N-terminal 6xHis tag followed by a thrombin cleavage site. All mutants were generated from the SEP3$^{(75-178)}$ construct in the pESPRIT002 vector, and oligonucleotides for the mutants used for PCR are given in Table II-4. Mutants based on SEP3$^{(75-178)}$ were generated according to the manufacture's protocol using Phusion polymerase (Thermo Fisher Scientific Inc.). All SEP3 proteins were overproduced in *Escherichia coli* BL21 (DE3) CodonPlus RIL (Agilent Technologies) except the 6xHis-GST-SEP3$^{(1-251)}$ construct, which was overproduced in Rosetta 2 (Novagen). Three mutants M150A, L154A, L171A were generated from SEP3$^{(75-178)}$ in pESPRIT002 with Q5 high fidelity polymerase (New England Biolabs Inc.) using the primers mentioned in Table II-4. A standard mutagenesis reaction was set up in a 50 µl reaction volume, 10 ul of 5x polymerase buffer, 10mM dNTPs, 10 µM forward and reverse primers, 0.5 µl Q5 polymerase, 2 mM DMSO and 10 ng plasmid template was amplified with a 98°C, 30 s denaturation, 25 cycles of 98°C, 10 s, 63°C, 30 s, 72°C, 2 min and final extension at 72°C for 2 min. The PCR product was digested with DpnI for 1 h at 37°C and transformed in E. *cloni* chemical competent cells (Lucigen, Middleton, WI, USA) using heat shock treatment (30 min ice; 42°C for 30 s, 2 min ice), revived with SOC medium (500 µl, 1 hr. 37°C) and plated on LB-Agar with 50 mg/ml kanamycin and incubated overnight at 37°C.

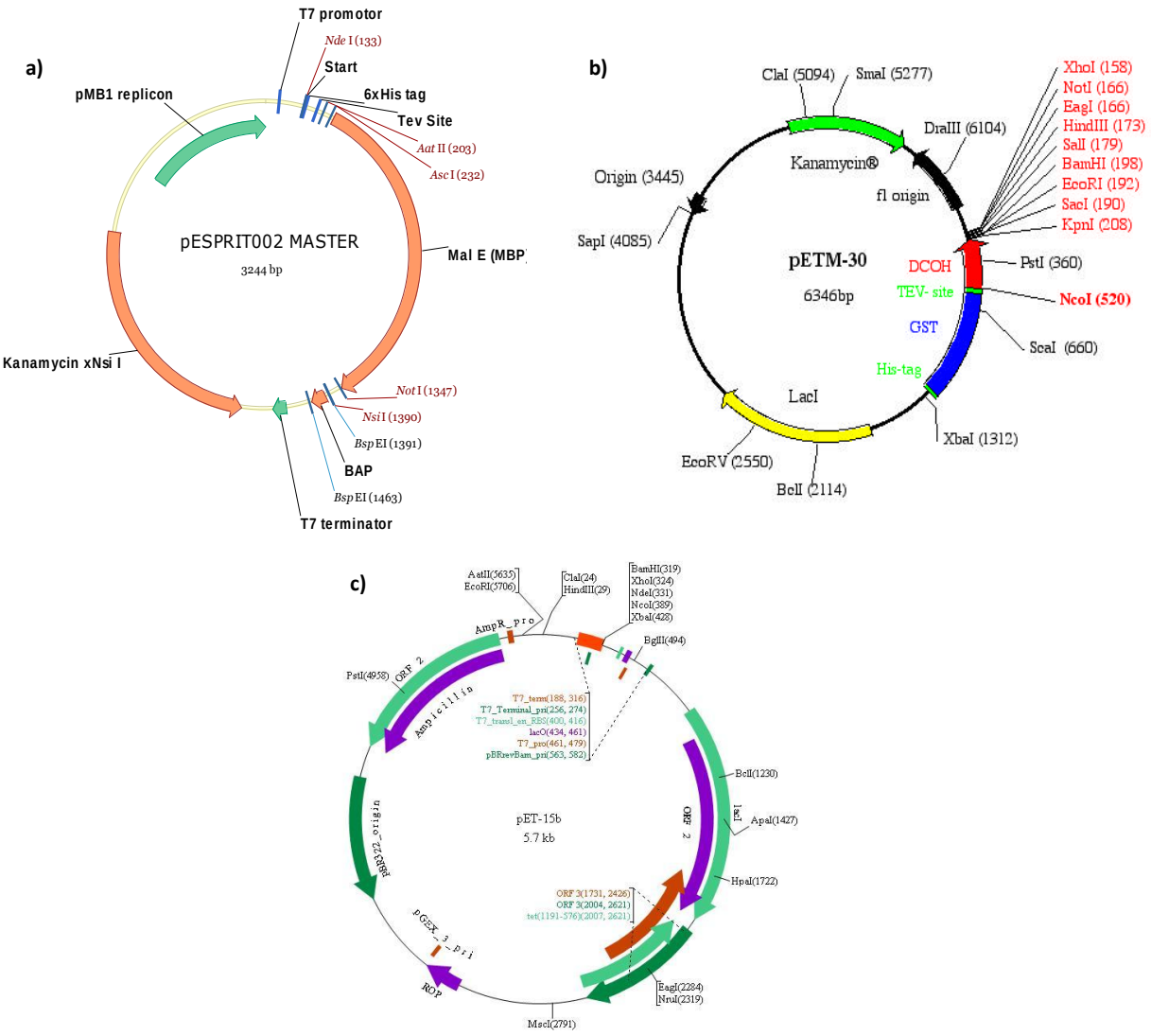


Figure II-39: ***Plasmids used for protein expression. a)*** *pESPRIT002 was used as an expression vector for ESPRIT screen and purification of SEP3*$^{(75\text{-}178)}$*.* ***b)*** *pETM30 encoding an N-terminal 6xHis/ GST tag was used for expression and purification of 6xHis.GST- SEP3*$^{(1\text{-}251)}$*.* ***c)*** *SEP3*$^{(1\text{-}90)}$ *was cloned and expressed in pET15b. Image Source: Gentle DNA editing software* (Manske, 2006), *(http://www.biovisualtech.com/bvplasmid/pET-15b.htm), (https://www.embl.de/pepcore/pepcore_services/strains_vectors/vectors/pdf/pETM-30_map.pdf)*

II.5.1.2 ESPRIT

In ESPRIT, a unidirectional truncation method was performed with a fixed 5´ end and an exonuclease susceptible 3´end. Three initial constructs were cloned in a pESPRIT002 (Hart and Tarendeau, 2006; Yumerefendi et al., 2010; Guilligay et al., 2008) vector using AatII and NotI restriction sites flanked by N-terminal 6xHis tag and a C-terminal biotin acceptor peptide.

12 µg of plasmid was digested with NotI - NsiI at 37° C for 2 h and verified by agarose gel electrophoresis. The double digest leads to an exonuclease sensitive 5´ and resistant 3´ overhangs. Exonuclease III treatment was initiated at 22° C and 2 µl aliquots were removed at 1 min intervals in 250 µl quenching solution, NT buffer (Macherey Nagel Gmbh and Co., Germany) on ice. The low temperature and high salt ensures complete arrest of the exonuclease reaction. Mung bean nuclease treatment (30°C for 30 min) was used to remove the single stranded overhangs followed by a *Pfu* DNA polymerase treatment at 72°C for 10 min. The linear DNA library was extracted and purified from the smear observed in agarose gel electrophoresis performed at 4°C.

The vector was recircularized using Rapid ligation kit (Roche), transformed in Mach1 electrocompetent cells (Life technologies, Carlsbad, CA, USA) and plated on 24 cm*24 cm Q trays (Genetix-Molecular Devices, CA, USA). After overnight incubation at 37°C, around 15-20 k colonies were pooled in and DNA extracted using Plasmid Midi kit (Qiagen, Germany). A colony PCR was performed using *Taq* polymerase to verify the size distribution of the obtained constructs. The three initial constructs were of different sizes and hence were treated separately until the colony PCR stage to avoid bias in size distribution of constructs. The DNA was transformed in BL21-AI, an arabinose inducible strain with pACYC, encoding rare t-RNAs and plated on 24 cm^2 Q trays (Genetix- Molecular Devices, CA, USA) in different dilutions. After overnight incubation at 20°C, single colonies were picked and inoculated in 384 well plates (Genetix-Molecular Devices, CA, USA) using standard colony picking robots. The robots are programmed to pick colonies with an optimal size and spacing between them. Protein expression was induced in high-density colony array format blotted on a nitrocellulose membrane and probed using fluorescent streptavidin and a monoclonal anti-6xHis antibody. The intensity of these spots was measured and the values obtained were statistically analysed. 96 high scoring candidates with signals for both tags were selected and subjected to 4 ml test purification. Each candidate was grown in Terrific broth at 37°C until O.D. of 0.8 and induced with 0.2% arabinose at 20°C. After overnight incubation, the cells were lysed, subjected to Ni-NTA affinity purification and proteins verified using western blot analysis. The probes used were similar as for the nitrocellulose membrane blotting. The clones with appropriate size and signals for both N and C terminal tags were sequence verified.

DNA target	Primer sequence
AP3 mutant1	5 '-CAAACAAAGACTTTTCTCCTTTCATGGGATTCCTTCTTAAACCCTAG-3' 5'-CTAGGGTTTAAGAAGGAATCCCATGAAAGGAGAAAAGTCTTTGTTTG-3'
AP3 mutant2	5'-GATGGAAGTGACGATTAATCCTTTCATGGATATCTCTTTCTTCTTATTTC-3' 5'-GAAATAAGAAGAAAGAGATATCCATGAAAGGATTAATCGTCACTTCCATC-3'
SOC1 DNA1 (557bp)	5'-CTTACCTGTGAGTAATACAACTATATTGG 3' 5'-CACCAGAGAAGATGTGTATTTTCG -3'
SOC1 DNA2 (1024bp)	5'-CTTACCTGTGAGTAATACAACTATATTGG 3' 5'-GCGAAAATTAGATTAGTTTATATGATTATGTAC-3'
M150A	5'-CTCTCAGGACACAGTTTGCGCTTGACCAGCTCAAC-3' 5'-GTTGAGCTGGTCAAGCGCAAACTGTGTCCTGAGAG-3'
L154A	5' –CAGTTTATGCTTGACCAGGCGAACGATCTTCAGAGTAAGG- 3' 5' –CCTTACTCTGAAGATCGTTCGCCTGGTCAAGCATAAACTG- 3'
L171A	5' -CTGACTGAGACAAATAAAACTGCAAGACTAAGGTTAGCTGATGG- 3' 5' -CCATCAGCTAACCTTAGTCTTGCAGTTTTATTTGTCTCAGTCAG- 3'
SEP3$^{(\Delta V90)}$	5'-GAGAGGCCTTAGCAGAACTTAGTAGCCAGC-3' 5'-GCTGGCTACTAAGTTCTGCTAAGGCCTCTC-3'
SEP3$^{(\Delta 161-174)}$	5'-CCAGCTCAACGATCTTCAGAGTAAGCTAGCTGATGGATGAGAGACAAAT AAAACTCTAAGACTAAGG-3' 5'-CCTTAGTCTTAGAGTTTTATTTGTCTCTCATCCATCAGCTAGCTTACTCTG AAGATCGTTGAGCTGG-3'
SEP3$^{(1-251)}$ In pIVEX 2.3d	5'-GATATACCATGGTGGGAAGAGGGAGAGTAGAATTGAAGAG-3' 5'-ACCCCCCCCGGGAATAGAGTTGGTGTCATAAGGTAACCAAC-3'

Table II-4: ***Primers used for DNA amplification.*** *The different primers used for amplification of DNA target sequences are provided here.*

II.5.1.3 Cell free expression and multi-vector expression screen

SEP3$^{(1-251)}$ was cloned in pIVEX 2.3d (Roche Diagnostics) with a C-terminal 6xHis tag and test expression was performed in batch mode in 50 µl for 2 h at 28°C using the *E.coli* BL21 (DE3) S30 extract at the IBS Cell free expression platform (http://www.isbg.fr/samples-preparation/cell-free-666/). A titration of increasing Mg^{2+}

concentration was also tested for its effect on expression and solubility. After 2 h, the sample was centrifuged at 11,000 g for 10 min and 12 µl of supernatant was loaded on 12.5% SDS-PAGE gel. A western blot was performed using anti-His tag–HRP antibody. The details of the protocols used for multi-vector expression screening are provided later in section IV.5.1.1.

II.5.1.4 Large scale expression and purification

Several promising constructs obtained from ESPRIT and other conventional construct design strategies were scaled up as described below.

II.5.1.4.1 SEP3$^{(75\text{-}178)}$

The SEP3$^{(75\text{-}178)}$ construct in the pESPRIT002 (Hart and Tarendeau, 2006; Yumerefendi et al., 2010; Guilligay et al., 2008) plasmid obtained from the ESPRIT screen was transformed in *Escherichia coli* BL21 (DE3) CodonPlus RIL (Agilent Technologies) cells using the heat shock method. Cells were grown in LB medium in the presence of 50 mg/ml kanamycin and 35 mg/ml chloramphenicol at 37°C and 180 rpm and until the OD600 nm was about 0.8 followed by induction with 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20°C. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min and resuspended in lysis buffer (30 mM Tris pH 8.0, 300 mM NaCl, 5 mM β-mercapto-ethanol (BME), 5%(v/v) glycerol, 20%(w/v) sucrose) protease inhibitors (Roche- EDTA free), lysozyme (1 mg/ml), and benzonase (1 mg/ml). After cell lysis by sonication, the culture was centrifuged at 25,000 rpm at 4°C for 30 min and the supernatant was applied to 5 ml Ni–NTA column pre-equilibrated with binding buffer (30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol), washed with binding buffer supplemented with 1 M NaCl and eluted with elution buffer (30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol 250 mM Imidazole). The 6xHis tag was removed by overnight dialysis with TEV protease in binding buffer followed by a re-purification on Ni-NTA column to remove TEV and uncleaved protein. The cleaved protein eluted in the flow through and was concentrated to approximately 5 mg/ml and purified using size exclusion chromatography with a Superdex S 200 10/300 GL (GE Healthcare) on an AKTA purifier (GE Healthcare) pre-equilibrated with binding buffer. Seleno-methionine incorporated protein was expressed following the protocol described in (Doublié, 1997) and purified as for the native protein.

All mutants were expressed and purified under the same conditions as wild type protein SEP3$^{(75\text{-}178)}$. Two naturally occurring splice variants SEP3$^{(\Delta V90)}$ and SEP3$^{(\Delta V161\text{-}174)}$ were also recombinantly purified using a similar protocol.

II.5.1.4.2 SEP3$^{(1-90)}$

The construct SEP3$^{(1-90)}$ was cloned into pET15b with an N-terminal 6xHis tag followed by a thrombin cleavage site. The over-expression and Ni-NTA affinity purification protocol was similar to that for the SEP3$^{(75-178)}$ except for the overnight induction at 15°C and lower concentration of salt (300 mM NaCl) in the wash buffer. After Ni-NTA purification, the sample was further purified with a heparin column pre-equilibrated with binding buffer (30 mM Tris pH8.0, 300 mM NaCl, 5% glycerol, 5 mM BME) and eluted with a high salt elution buffer (30 mM Tris pH 8.0, 1.5 M NaCl, 5% glycerol, 5 mM BME). The fractions of interest were pooled in and dialysed with 30 mM MES pH6.5, 300 mM NaCl, 5 mM BME, 5% glycerol and 2mM $CaCl_2$ with thrombin in order to cleave the 6xHis tag. The protein was applied to a benzamidine column to separate the thrombin from the cleaved pure protein. The protein was concentrated to approximately 10 mg/ml and applied to a size-exclusion column (Superdex S 200 10/300 GL) pre-equilibrated with 30 mM Tris, pH 8.0, 8.0, 300 mM NaCl, 2 mM TCEP (tris(2-carboxyethyl)phosphine), and 5% glycerol.

II.5.1.4.3 6xHis-GST-SEP3$^{(1-251)}$

The 6xHis-GST-SEP3$^{(1-251)}$ protein construct was overexpressed in Rosetta 2 cells using a similar protocol as for SEP3$^{(75-178)}$. The purification protocol was as used for SEP3$^{(1-90)}$.The fractions of interest from heparin column elution were pooled in concentrated to 0.3 mg/ml, and applied to a size- exclusion column (Superdex 200 10/300 GL) pre-equilibrated with 50 mM Tris, pH 8.0, 1.2 M NaCl, 5% glycerol, and 2 mM TCEP.

II.5.2 Biophysical characterisation and DNA binding studies

II.5.2.1 Crystallisation and structure determination

SEP3$^{(75-178)}$ at a concentration of 10 mg/ml in 30 mM Tris pH 8.0, 300 mM NaCl, 5 mM BME, 5% glycerol was subjected to high-throughput crystallisation trials using the EMBL HTX facility (European Molecular Biology Laboratory, Grenoble, France).200 nl sitting drops were set up with the Cartesian PixSys 4200 (Genomic Solutions, UK) crystallisation robot using the Greiner CrystalQuick plates (flat bottom, untreated) and imaged with a Formulatrix Rock Imager (Formulatrix Inc.,USA) at 277 K (Dimasi et al., 2007). Several conditions such as (a) 0.1 M Tris pH 8.0, 20% ethanol, (b) 0.1 M Tris pH 8.0, 20% (±)-2-

methyl 2,4- pentanediol (MPD), (c) 0.1 M Tris pH 8.0, 20% ethylene glycol and (d) 0.1 M sodium/potassium phosphate pH 6.2, 25% 1,2-propanediol,10% glycerol yielded crystals during the robotic trials in commercial screens (Hampton Research, CA, USA). Crystals grew to dimensions of ~200*200*100 µm clusters over one month. They were harvested and cryocooled in liquid nitrogen without prior cryoprotection for initial screening.

An additive screen for pH and precipitation concentration of best diffracting crystals from the robotic screen was performed using commercial kits (Additive Screen and Detergent Screen, Hampton Research). Best diffracting single crystals were obtained in 0.1 M Tris pH 8.0, 25% ethylene glycol after successive rounds of macroseeding. Large crystal clusters were broken manually and the pieces placed in 2 µl hanging drops consisting of a 1:1 ratio of protein to well solution (25% ethylene glycol, 0.1 M Tris pH 8.0. The final large crystals were obtained when the above crystals were reharvested and placed in fresh 2µl hanging drops within 24-48 h. Crystals were harvested and cryocooled in liquid nitrogen for data collection. Crystals of the seleno-methionine derivatized protein were grown under the same conditions as the native protein with micro seeding of the crystallisation drops using crystals of the native protein. The details of the crystal optimisation, screening and data collection are published in (Acajjaoui and Zubieta, 2013).

A native diffraction dataset was collected from crystal grown at 0.1 M Tris pH 8.0, 25% ethylene glycol at d_{min} = 2.53 Å on beamline ID14.4 at ESRF (Gabadinho et al., 2010; McCarthy et al., 2009). The crystals belonged to space group $P2_12_12$. Due to absence of suitable molecular replacement structures, it was difficult to obtain phase information. Hence, heavy atom soaking was performed in two different concentrations (10 mM and 100 mM) and soaking time (5 min, 30 min, 2 h and 24 h) with potassium tetrabromo platinate and diammino platinum dinitrite at different concentrations of ethylene glycol. However, the soaking procedure resulted in poorly diffracting crystals necessitating the use of seleno-methionine derivatives. Based on the obtained phases, a partial structure was built and used as a molecular replacement model for the higher resolution native data. Data for the seleno-methionine derivative were collected at ID23-1, ESRF (Gabadinho et al., 2010) to d_{min} = 3.18 Å, integrated and scaled using XDS and XSCALE (Kabsch, 2010) and phasing was performed using SHELX (Sheldrick, 2008). A fluorescence scan employed gave peak absorbance of 1.07 Å as calculated by *CHOOCH* (Evans and Pettifer, 2001). Based on the obtained phases a partial structure was built and used as a molecular replacement model for the higher resolution native data. The structure refinement was performed by Dr. Chloe Zubieta.

II.5.2.2 Atomic Force Microscopy

DNA substrate synthesis: AFM was performed on protein-DNA complexes with 999 bp *AP3* and 1024 bp *SOC1* promoter DNA strands amplified from *Arabidopsis thaliana* genomic DNA. A 50 µl reaction volume comprising 5x Phusion HF buffer, 10mM dNTPs, 0.5 µl Phusion polymerase (Thermo Fisher Scientific Inc.), 10 µM forward and reverse primers (Table II-4) and 50 ng genomic DNA template were used to amplify *AP3* and *SOC1* promoter DNA strands. Two mutant *AP3* DNA strands with additional CArG box sequences (CCTTTCATGG) at different locations were synthesized using Quick-change site directed mutagenesis kit (Agilent Technologies CA, USA). A 50 µl reaction mix comprising of 10x buffer, 10 ng template, 10 µM forward and reverse primers, 10 mM dNTPs and 1 µl *Pfu* Ultra polymerase was amplified using standard Quick change protocol; 95°C 30 s initial denaturation; 18 cycles of 95°C 30 s, 55°C, 1 min and 68°C 4 min followed by a final extension at 68°C for 5 min.

Imaging: 6xHis-GST-SEP3$^{(1-251)}$ and *SOC1* promoter DNA were mixed at two different protein concentrations (2 to 5 nM and 10 to 15 nM) to 2-5 nM DNA in a buffer comprising 10 mM HEPES and incubated on ice for 15 min. For SEP3$^{(1-90)}$ - *SOC1* protein-DNA complexes, protein concentrations were 5 and 25 nM to 2-5 nM DNA prior to dilution for imaging. For AFM imaging, the preformed complexes were diluted in 10µl of adsorption buffer (10 mM $NiSO_4$ and 10 mM HEPES, pH 7.0) to obtain a final DNA concentration of 0.5-2 nM and deposited on freshly cleaved mica (Agar Scientific, Essex, UK). The mica sheet was rinsed 2-3 times with imaging buffer (10mM HEPES, pH 7.0) to remove unbound molecules and scanned under 200 µl of imaging buffer. MFP 3D and Cypher S atomic force microscopes (Asylum Research, CA, USA) were used with MSNL-10 (Bruker Co. ,CA, USA) and Biolever mini BL-AC40TS-C2 (Olympus Co. Tokyo, Japan) silicon nitride cantilevers respectively in tapping mode for imaging. The free amplitude was set to 0.4V and set point around 250 mV with an integral gain at 10. However a constant monitoring was necessary to obtain high quality uninterrupted scans. Images were obtained both at 256 * 256 pixels and 512 * 512 pixels with a scan size between 0.2- 2 µm and processed using Gwyddion (Nečas and Klapetek, 2011). Image processing comprised of filtering, flattening and masking in order to remove the background noise as per requirement. Similar procedure was followed for *AP3* promoter DNA and 6xHis-GST-SEP3$^{(1-251)}$ and SEP3$^{(1-90)}$ domain construct. A titration of protein, DNA and NaCl concentration was performed to study the

specificity and the concentration effect on protein-DNA binding. Different variables in AFM experiment include, choice of divalent cations (Mg^{2+}, Ni^{2+}, Co^{2+}), difference in concentration of protein and DNA, choice of pH and buffer conditions, scanning mode and system (tapping mode, constant mode; liquid or air). A considerable amount of experiments were performed with AFM to optimize the reaction and scanning parameters.

Chapter III

AGAMOUS

III AGAMOUS

III.1 Summary

This chapter focuses on the structural characterisation of a second important floral organ identity TF, AGAMOUS (AG), using a structure-based construct design based on homology to the oligomerisation domain of SEP3 (see Chapter II). The chapter describes attempts to use a high-throughput library technique to obtain soluble constructs of AG, the development of a protocol for the purification of the recombinant protein using denaturation and refolding and the characterisation of the protein using SAXS. From the latter experiments two models of the solution structure are proposed. The effect of conserved residues on the oligomerisation state of protein is discussed.

Résumé en Français

Ce chapitre se concentre sur un autre facteur de transcription régulateur d'organes floral important, AGAMOUS (AG), grâce à l'utilisation d'une construction basée sur l'homologie d'AG au domaine d'oligomérisation de SEP3 et la structure de cette dernière. Le chapitre décrit l'utilisation de « bibliothèque à haut débit » pour obtenir des constructions solubles d'AG, le développement d'un protocole de purification de la protéine recombinante par dénaturation et renaturation, ainsi que la caractérisation de la protéine par SAXS. Ces dernières expériences ont permis de proposer deux modèles structuraux d'AG. L'effet des résidus conservés sur l'état d'oligomérisation de la protéine est discuté.

III.2 Introduction

The ABC(D)E model of floral organ development (Coen and Meyerowitz, 1991; Bowman et al., 1991; Angenent and Colombo, 1996; Favaro et al., 2003; Pelaz et al., 2000) describes the role of class A-E MADS box genes in determining floral organ identity. A host of genetic experiments and bioinformatic studies have been performed to study the floral organ regulator genes; however, little structural knowledge of the encoded proteins is available. In the previous chapter, the structure of the oligomerisation domain of the MADS TF SEP3 was described and its DNA binding properties illustrated using AFM. In the current chapter, I will focus on structural characterisation of another MADS TF, AGAMOUS (AG).

In *Arabidopsis thaliana*, the class C gene of the ABC model is *AG*. The gene product of *AG* is an MIKC MADS TF involved in stamen and carpel identity, microsporogenesis, organ maturation and floral meristem identity (Yanofsky et al., 1990; Bowman et al., 1989; Ito et al., 2004, 2007). AGAMOUS has also been shown to indirectly regulate genes associated with cell cycle and DNA repair (Ó'Maoiléidigh et al., 2013). ChIP-seq studies, performed using flowers at early developmental stages have identified about 2000 high confidence targets for AG in Arabidopsis (Ó'Maoiléidigh et al., 2013). However, while much is known about the downstream targets of AG and its role in determinacy and floral organ morphogenesis, (Ito et al., 2004, 2007; Gómez-Mena et al., 2005; Sun et al., 2009; Liu et al., 2011), the molecular basis, specifically its oligomerisation state, underlying these activities is still not clear.

The sequence of AG follows the classic MIKC structure common to most MADS TFs although it possess an additional region of 17 residues, N-terminal to the MADS domain (Figure III-1). However, previous data has showed that AG retains its DNA binding and oligomerisation properties *in vitro* in the absence of this N terminal extension (Mizukami et al., 1996) thus all future reference to AG in the chapter refers to $AG^{(18-252)}$. The domain structure (Figure III-1) of $AG^{(18-252)}$ comprises a DNA binding MADS (M) domain, a dimerisation Intervening (I) domain, an oligomerisation Keratin like (K) domain and a C-terminal domain responsible for higher order complex formation. Secondary structure prediction using PSIPRED (Jones, 1999) suggests a mostly α-helical structure in the I and K domains with the M domain structure forming the conserved DNA binding domain characteristic of all MADS TFs (Figure III-1). The C-terminal region is predicted to be

unstructured. In the following chapter, I will describe construct design- screening, expression, purification and structural studies of oligomerisation domain from AG.

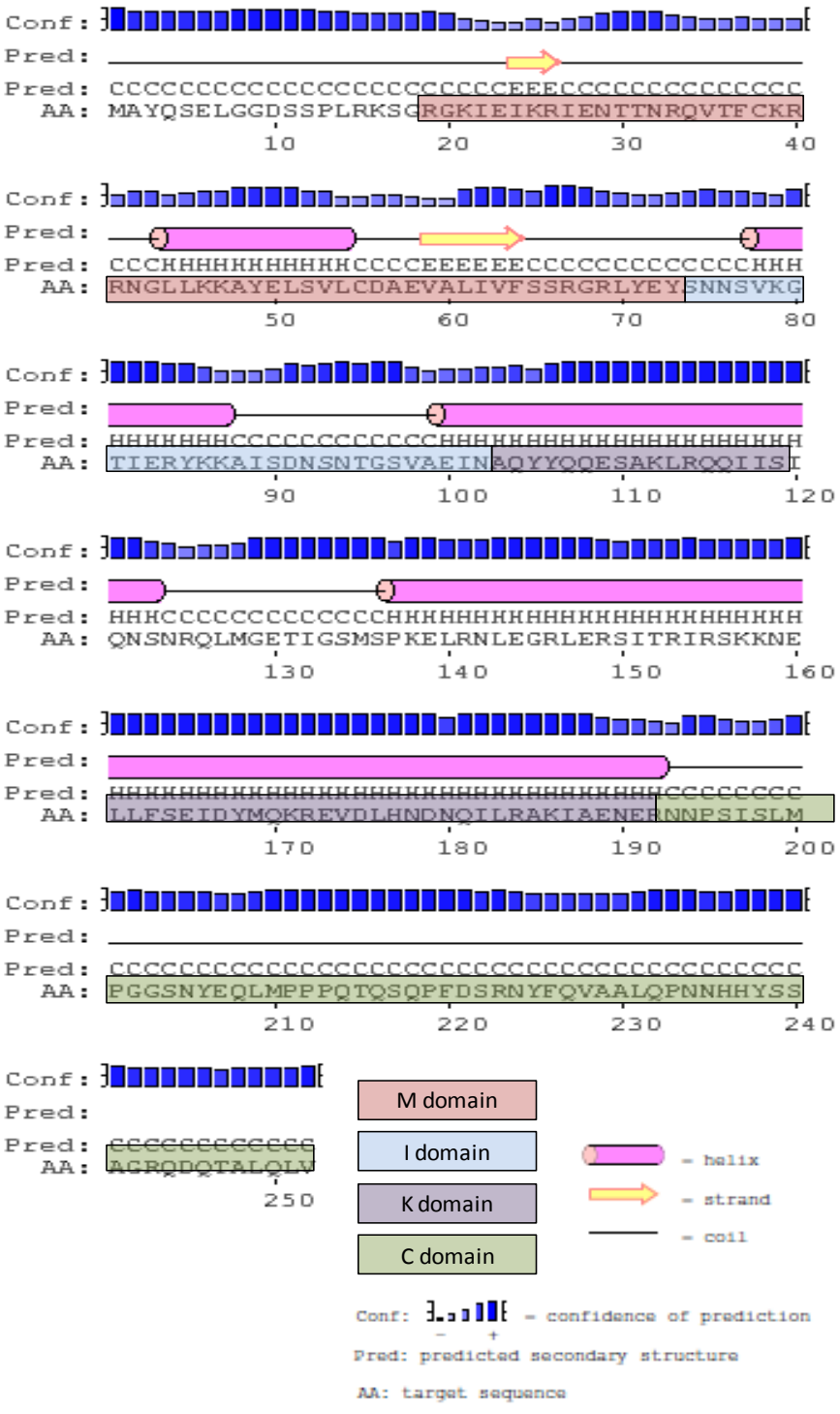


Figure III-1: ***The secondary structure of AG as predicted using PSIPRED*** **(Jones, 1999).** *The different domains annotated from Uniprot database* (The UniProt Consortium, 2014) *are marked with different colours. The pink cylinders represent α helices, yellow arrows β strands. The confidence of prediction is given by blue bars.*

III.3 Results and Discussion

III.3.1 Construct design and protein purification

In the previous chapters, the obstacles to the characterisation of plant MADS TFs were discussed. The absence of protein production and purification protocols is the most prominent barrier. Also, the presence of a large unstructured C-terminal domain makes protein crystallisation difficult. Different strategies were used to try to obtain pure, soluble recombinant constructs of AG. These included the use of alternate expression systems (cell free, mammalian and insect cells), high throughput expression screens and library generation techniques ultimately followed by denaturation-based purification and subsequent refolding of the protein.

III.3.1.1 Cell free expression trials

An initial attempt to produce AG$^{(18\text{-}252)}$ exploited an *in vitro* expression system using the IBS cell free expression platform (http://www.isbg.fr/samples-preparation/cell-free-666/). Here, however, a western blot (Figure III-2) showed a low percentage of protein in the soluble fraction in test expression trials with no significant effect of magnesium concentration on solubility being observed. Moreover, a large-scale expression (2 ml) trial also did not yield any soluble protein.

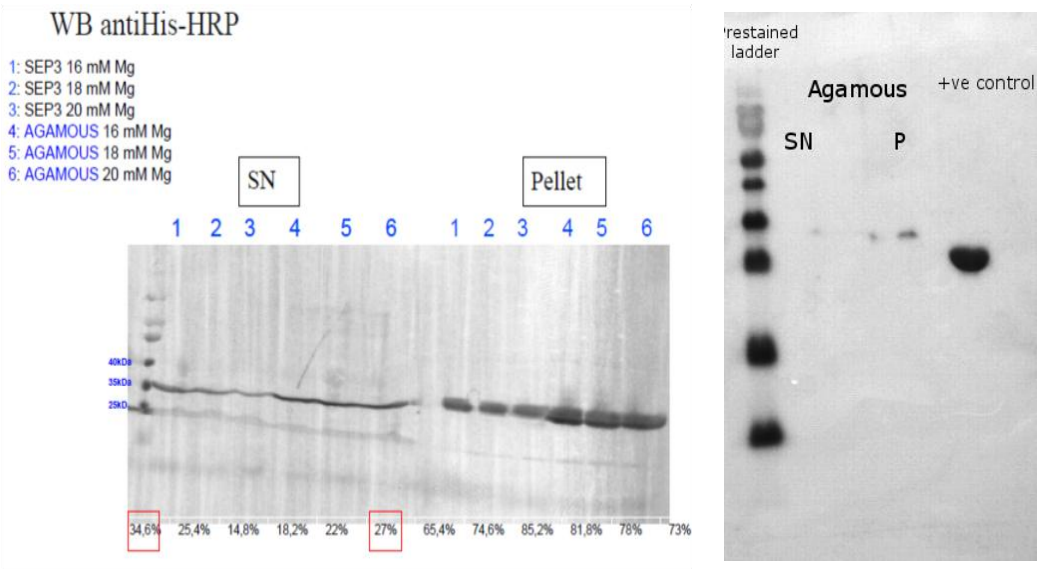


Figure III-2: ***Cell free expression tests for AG. (left)*** *Western blot analysis for test expressions (50 µl) of SEP3 (lanes 1,2,3) and AG (lanes 4,5,6) showing only a small amount of protein in soluble fractions. As can be seen, increasing Mg^{2+} concentration did not improve solubility.* ***(right)*** *No soluble protein was observed in large scale (2 ml) expression tests for AG. Anti His-HRP antibody was used to detect the 6xHis tag.*

III.3.1.2 Multi vector expression screening

Using the Oxford Protein Production Facility (OPPF), Oxford (https://www.oppf.rc-harwell.ac.uk/OPPF/), AG protein expression was performed exploiting various solubility tags (6xHis, MBP, GST) and alternate expression systems (*E.coli*, insect and mammalian cells). A large number of conditions were tested (see Appendix I) and analysed with SDS-PAGE. However, none of these trials, see Figure III-3 for an example, produced soluble AG construct.

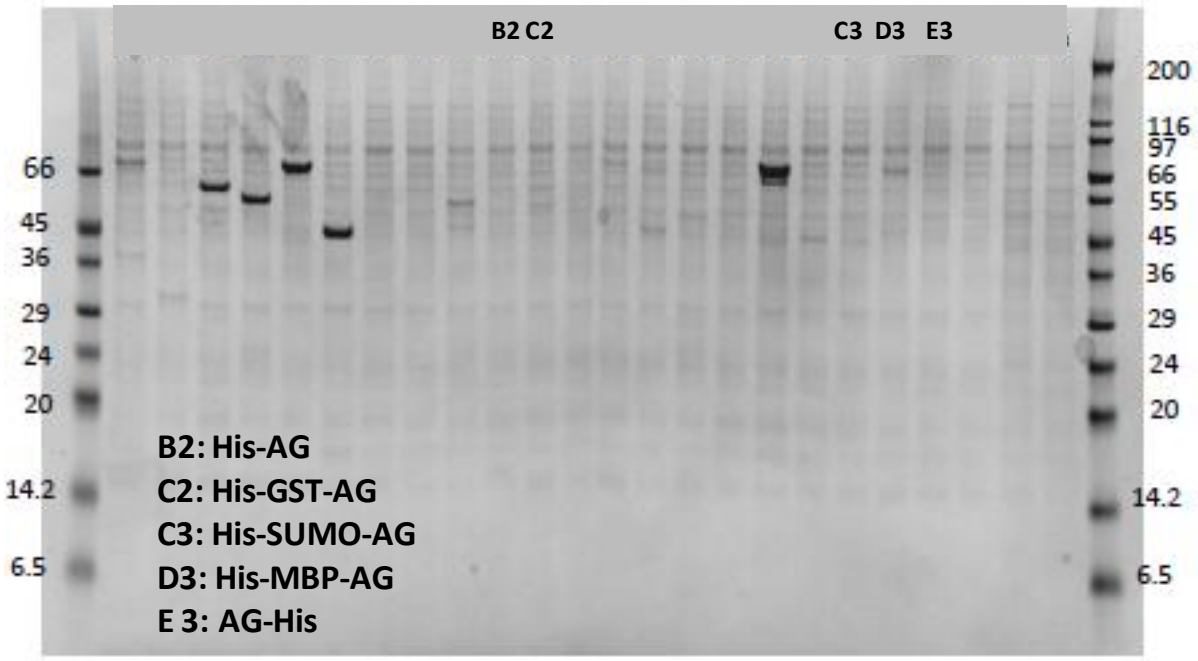


*Figure III-3**: SDS-PAGE analysis for Ni-NTA purification for various AG constructs expressed in insect cells* **at OPPF.** *The cells were infected with 3 ul of first generation virus and grown for 72 h post day of proliferation arrest. Only the lanes involving AG are annotated. No soluble protein was observed in the tests. The bands observed represent other proteins such as SVP and FLM that were screened in parallel. The results remained consistent even with addition of 30 ul of virus (data not shown).*

According to the floral quartet model, SEP3 interacts with AG during carpel formation (Theissen, 2001; Theissen and Saedler, 2001). Thus, a co-expression screen using SEP3 as an AG binding partner was performed in the OPPF insect cell expression system. About 24 co-expression tests were performed for different combinations of constructs of SEP3 and AG by varying their solubility tags (see Appendix I for details). Figure III-4 shows the SDS-PAGE for Ni-NTA purification of co-expression test for AG and SEP3. The 6xHis-

tagged protein constructs (AG-6xHis and 6xHis-AG) yielded small amounts of protein while strong bands were observed for the N-terminal MBP tagged construct (6xHis-MBP-AG).

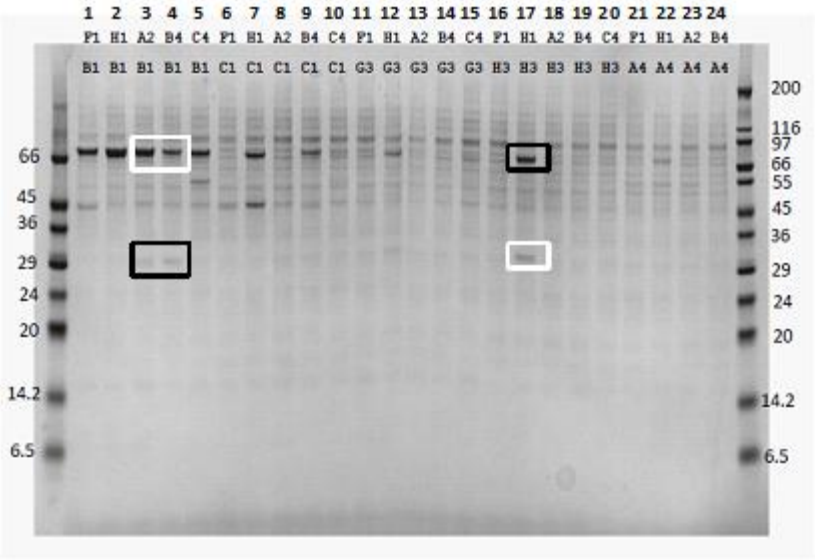


Figure III-4: ***Insect cell SEP3/AG co-expression.*** *SDS-PAGE analysis for Ni-NTA purification for SEP3-AG co-expressed in insect cells. The cells were infected with 3 ul of first generation virus and grown for 72 h post day of proliferation arrest. The lanes showing positive bands for AG are marked by black boxes, those for SEP3 with white boxes. The positive bands are annotated in Table III-1.*

Lane	Proteins		Expression?	
	AG	SEP 3	AG	SEP3
3	A2 (AG-His)	B1 (His-MBP-SEP3)	+	+++
4	B4 (His-AG)	B1 (His-MBP-SEP3)	+	+++
17	H1 (His-MBP-AG)	H3 (SEP3-His)	+++	+

Table III-1: ***Annotation of the co-expression tests shown in Figure III-4.*** *The table describes the construct design and expression levels for the marked bands. Lanes 3 & 4 show a high expression of 6xHis-MBP-SEP3 and low expression for AG-6xHis while lane 17 shows high expression for 6xHis-MBP-AG and low level expression for SEP3-6xHis.*

During the trials of the expression and purification of SEP3 described in II.3.1.3, it was observed that although the N-terminal MBP-tagged constructs showed high expression levels, the proteins were found to be soluble aggregates on gel filtration chromatography. Hence, in the case of 6xHis-MBP-AG, alternate strategies for protein expression and purification were explored before focusing on optimisation of the production of the MBP

construct. A test expression was performed in Human Embryonic Kidney (HEK) 293 cells using the mammalian expression system. Figure III-5 shows a small level of expression for N-terminal SUMO tagged AG (6xHis-SUMO-AG). The low amount of protein observed in the western blot and the absence of a mammalian expression facility in our laboratory, was a deterrent from continuing with this construct.

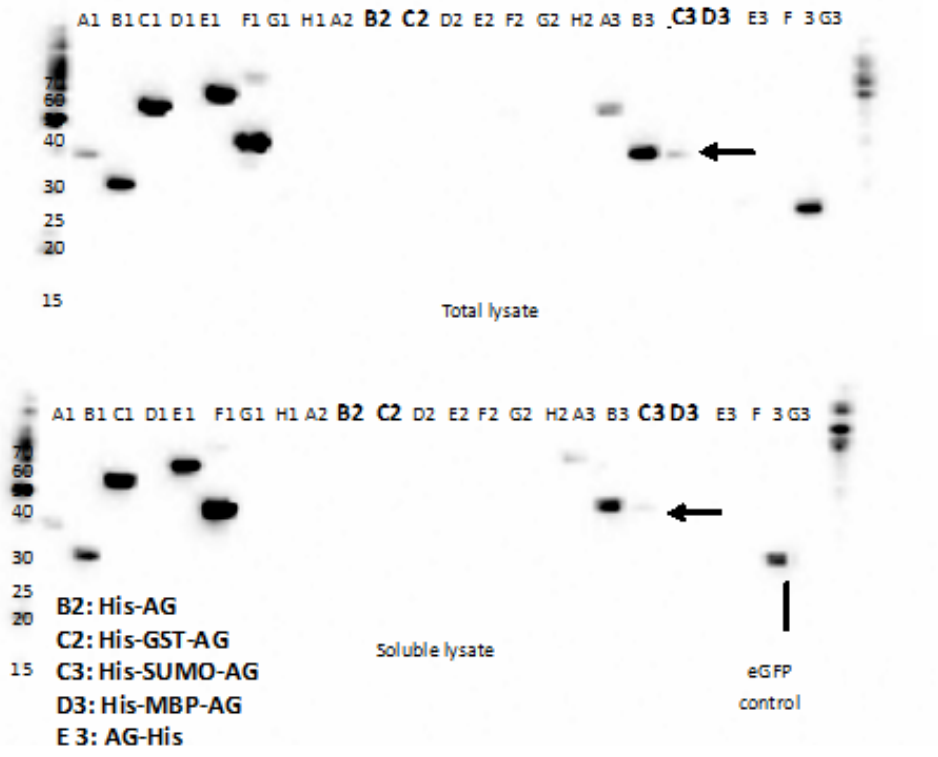


*Figure III-5**: Western blot analysis for HEK 293 mammalian expression test of AG probed with anti-His antibody.** The lanes containing AG protein constructs are printed in bold and annotated. A light band is observed in lane C3 (6xHis-SUMO-AG).*

III.3.1.3 Expression of Soluble Proteins by Random Incremental Truncation (ESPRIT)

In the previous chapter, a high throughput library generation technique ESPRIT (Hart and Tarendeau, 2006; Yumerefendi et al., 2010) yielded highly soluble constructs that eventually lead to the elucidation of the crystal structure of the SEP3 oligomerisation domain. ESPRIT was therefore also used for brute-force soluble construct screening for AG. Three initial constructs for AG, varying in their N-terminal starting points were used to generate a C-terminal truncation library (Figure III-6). Exonuclease III was then used to truncate the C-terminal ends and samples were aliquoted at regular time intervals to generate large pool of constructs with different C-termini.

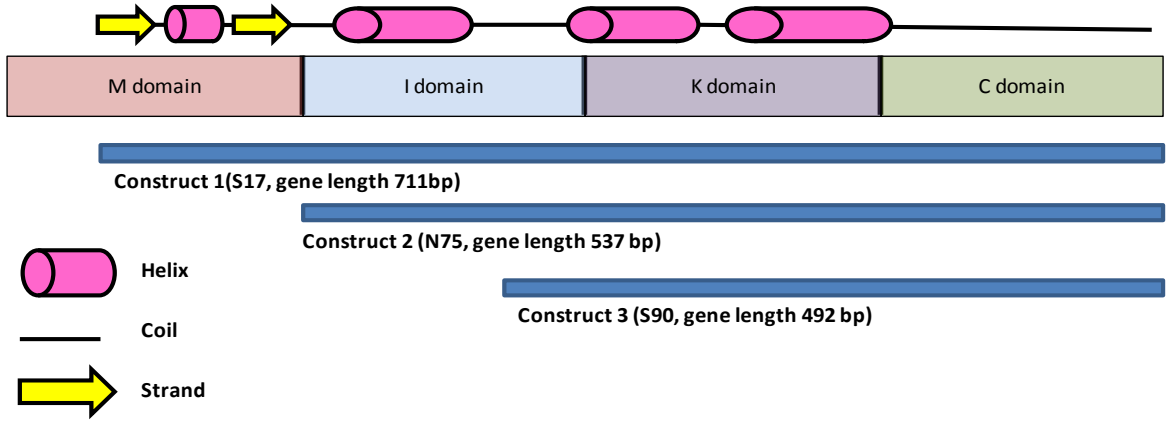


*Figure III-6**: AG construct design for ESPRIT.** Domain boundaries for the initial constructs of AG used for ESPRIT library. The secondary structure prediction from PSIPRED* (Jones, 1999) *and the domain boundaries for AG are shown at the top. The three initial genes have different N-terminal fixed ends.*

An agarose gel electrophoresis was used to isolate the regions of interest. The library was gel purified; religated and size distribution was analysed using colony PCR as shown in Figure III-7. A colony array expression was performed for about 7500 colonies taking in to

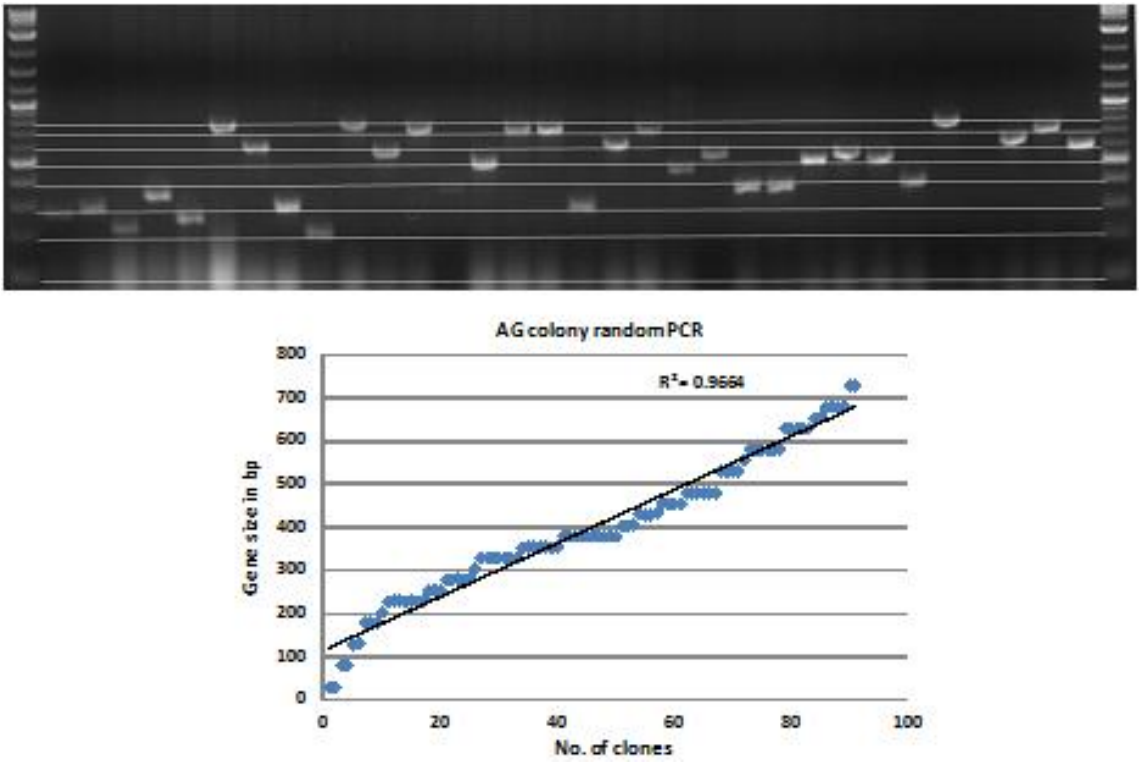


*Figure III-7: **Size distribution analysis of the ESPRIT-generated AG library.** **(top)** Representative colony PCR for AG library. The size of bands was measured and plotted to analyze size distribution **(bottom).** The linear graph and the R^2 value close to 1 suggests a linear truncation and unbiased library. The data from 3 other gels was pooled for the plot.*

account an approximately 3 fold oversampling required to cover for out-of-frame constructs. Colony picking robots were used to generate nitrocellulose blots and protein expression was induced in duplicates by transferring the blots on arabinose containing agar trays. Monoclonal antibodies against the N-terminal 6xHis tag and fluorescent streptavidin against the C-terminal BAP tag were used to probe the blots. The positive clones were statistically analysed and short-listed for 4 ml test expression screening. The protein constructs were purified using Ni-NTA purification and analysed with western blot (Figure III-8) using anti-His antibody and streptavidin similar to that for the nitrocellulose blots. The positive constructs observed in the western blot were sequenced and selected for large-scale expression.

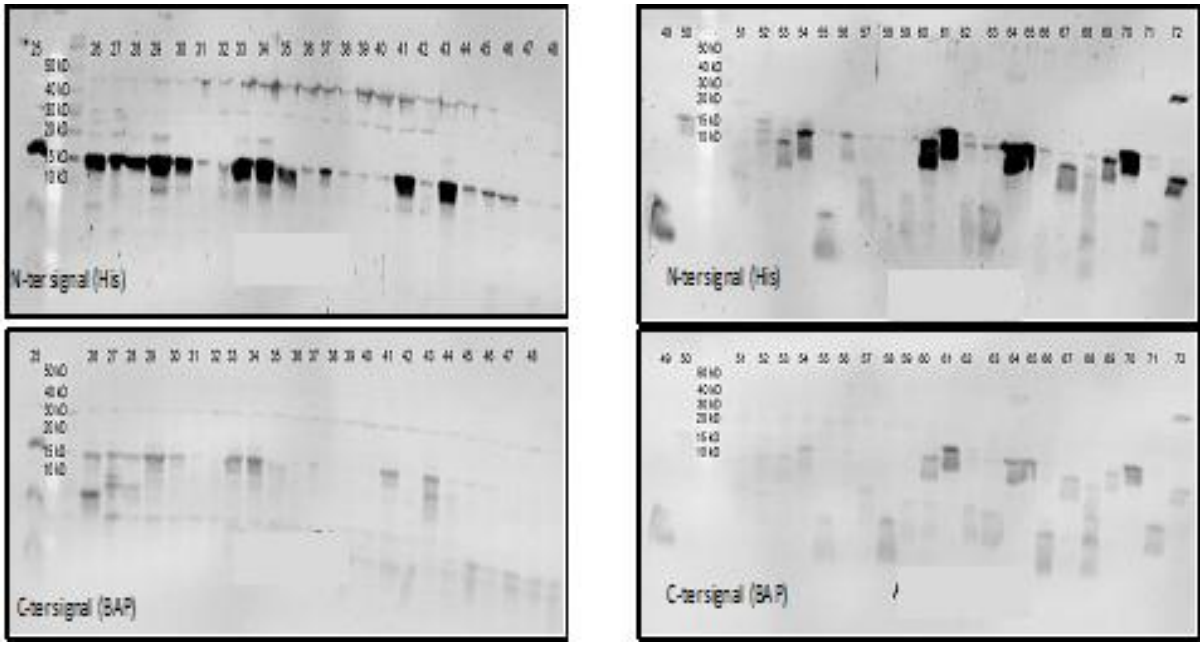


Figure III-8: ***Western blot analysis for small scale purification test of ESPRIT-generated AG constructs.*** *The western blot for AG small scale (4 ml) purification tests probed with anti-His antibody* ***(top panels)*** *and streptavidin* ***(bottom panels).***

All the constructs with positive bands for both 6xHis and BAP tag in the western blot from the 4 ml test expressions were scaled up and expressed in 50 ml of Terrific Broth media in BL21(AI) with the pACYC–RIL (Stratagene) plasmid and purified using Ni-affinity chromatography. The pACYC-RIL plasmid supplies tRNA for four rare codons (AUA, AGG, AGA, CUA). Figure III-9 shows the results of test purification following 50 ml expression trials for five AG constructs. Based on these, constructs AG26 and AG41 (names chosen according to well numbers in ESPRIT test expression) were scaled up to 200 ml and the

purification process was repeated (Figure III-9). Almost all the protein constructs screened in 50ml test expression seemed to have a similar size. This was indeed unexpected considering the linear size distribution observed during the library generation. A high molecular weight band observed for all the protein constructs was likely an *E. coli* contaminant. The 200 ml expression for AG26 and AG41 shows a small amount of protein in the washes (W1 and W2) with increasing amount of imidazole and a sharp band in the elution. The high molecular weight contaminant is still observed.

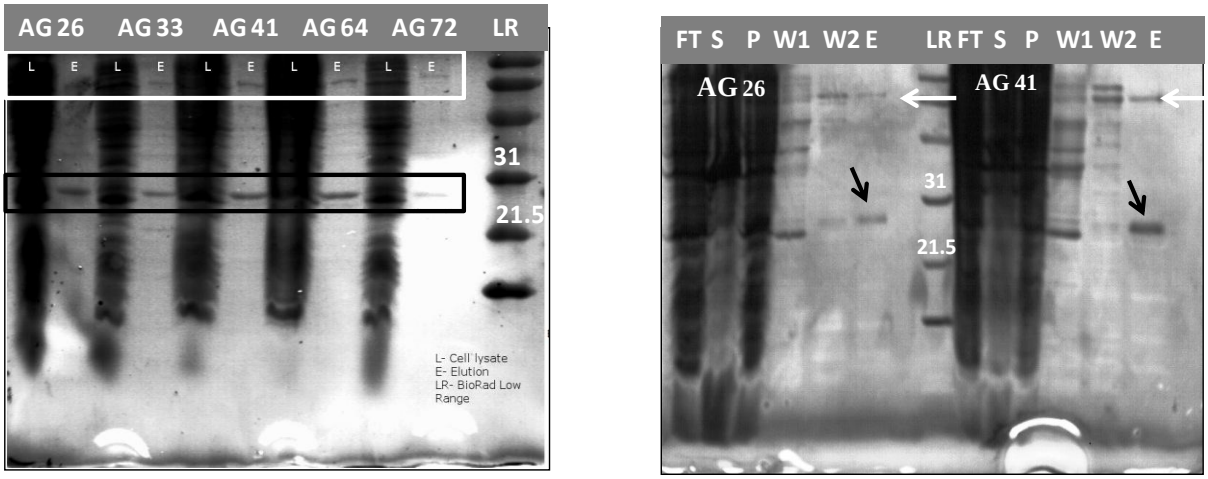


Figure III-9: ***Test expression for 5 ESPRIT-generated AG constructs. (left)*** *50 ml Ni-NTA purification test for 5 AG constructs obtained from ESPRIT showing lysate (L) and elution (E). The black box shows a band observed in elution for all 5 protein constructs at about the same size. Higher molecular weight contaminants are marked by white box.* ***(right)*** *200 ml Ni-NTA purification test for AG26 and AG41. A band is observed in both washes and in elution (marked by black arrows). The higher molecular weight contaminant is also observed (marked by white arrows).*

For both AG26 and AG41 the protein obtained in the eluted fraction was subjected to TEV cleavage to remove the 6xHis-tag and repurified using Ni-NTA affinity chromatography. In the absence of 6xHis tag, the protein should be observed in the flow through. Also a band shift to lower molecular weight should be visible to account for the loss of the 6xHis-tag and linker residues. Figure III-10 shows a SDS-PAGE analysis of the 6xHis-tag cleavage test for AG26 and AG41. The SDS-PAGE shows a band representing the proteins for both AG26 and AG41 in the 'uncleaved' lane. As observed in the gel, the band remains intact even after treatment with TEV protease and is detected in the flow through (FT) and elution (E) samples suggesting still presence of 6xHis-tag. This pattern is similar to the band observed in elution of the previous purification tests. An additional band is obtained which might represent the excess amount of TEV protease. In order to verify the presence of a 6xHis-tag, a western blot

was performed for AG26, which shows the absence of a band in the 'uncleaved' fraction. The excessive 6xHis tagged TEV protease is observed in the 'FT and 'E' fractions. The absence of protein bands in western blot and unsuccessful digestion suggests that the bands obtained for the elution might belong to an endogenous histidine-rich bacterial protein that was enriched during the purification process.

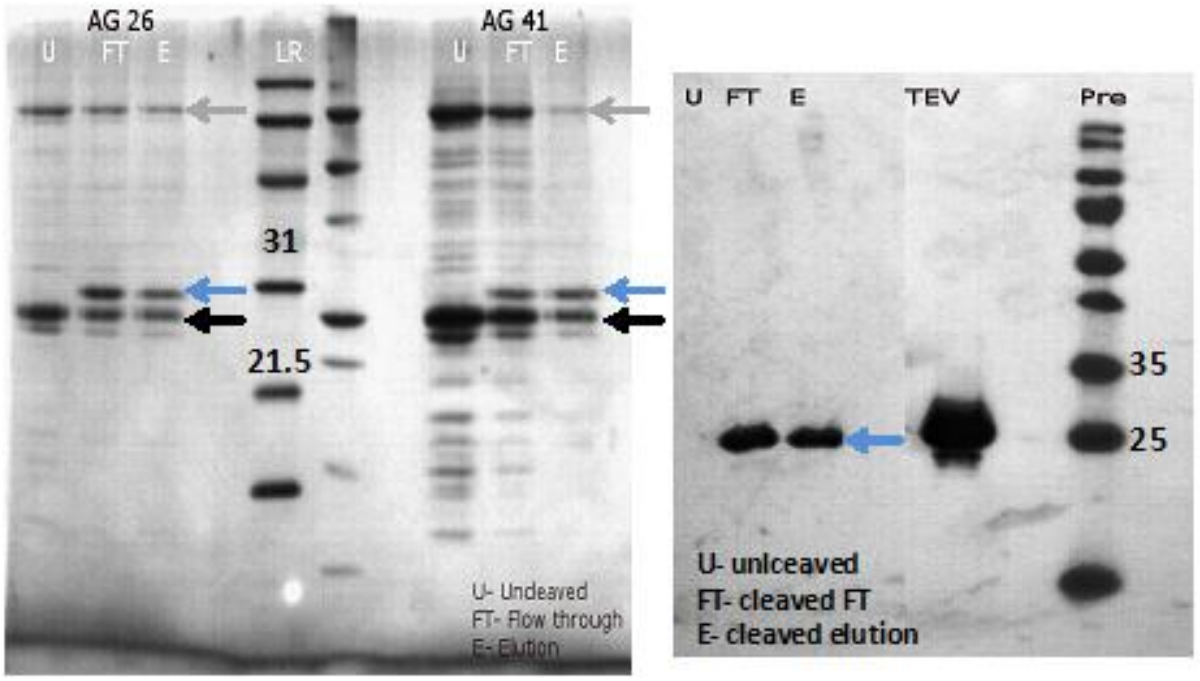


Figure III-10: ***Cleavage test for AG26 and AG41. (left)*** *SDS-PAGE for AG26 and AG41 cleavage by TEV protease shows presence of bands in all the lanes suggesting lack of cleavage (black arrow). An additional band representing excess TEV protease (blue arrow) and a higher molecular weight contaminant is observed (grey arrow).* ***(right)*** *Western blot for the cleavage test shows absence of protein in uncleaved fraction thus confirming absence of 6xHis-tagged protein. An excess TEV protease is observed in flow through and elution fractions (blue arrow).*

III.3.1.4 AG constructs obtained using structure based design

Despite exhaustive trials, soluble AG constructs were not obtained either with the use of high throughput expression screening (Section III.3.1.3) or alternate expression systems (Sections III.3.1.2 & III.3.1.1). Thus, in order to produce soluble AG constructs, a structure-based construct design approach was used exploiting the SEP3 oligomerisation domain as a template. Figure III-11 shows a ClustalW alignment (Larkin et al., 2007) of SEP3 with AG and the construct $AG^{(74-173)}$ was thus designed on basis of the $SEP3^{(75-178)}$ construct. Using the secondary structure prediction from PSIPRED (Jones, 1999), a second construct $AG^{(81-171)}$ was designed as shown in Figure III-12.

Figure III-11: ***Design of AG$^{(74-173)}$ based on sequence alignment with the SEP3 oligomerisation domain.*** *AG was aligned with SEP3 using ClustalW* (Larkin et al., 2007). *The boxed part shows the final construct designed.*

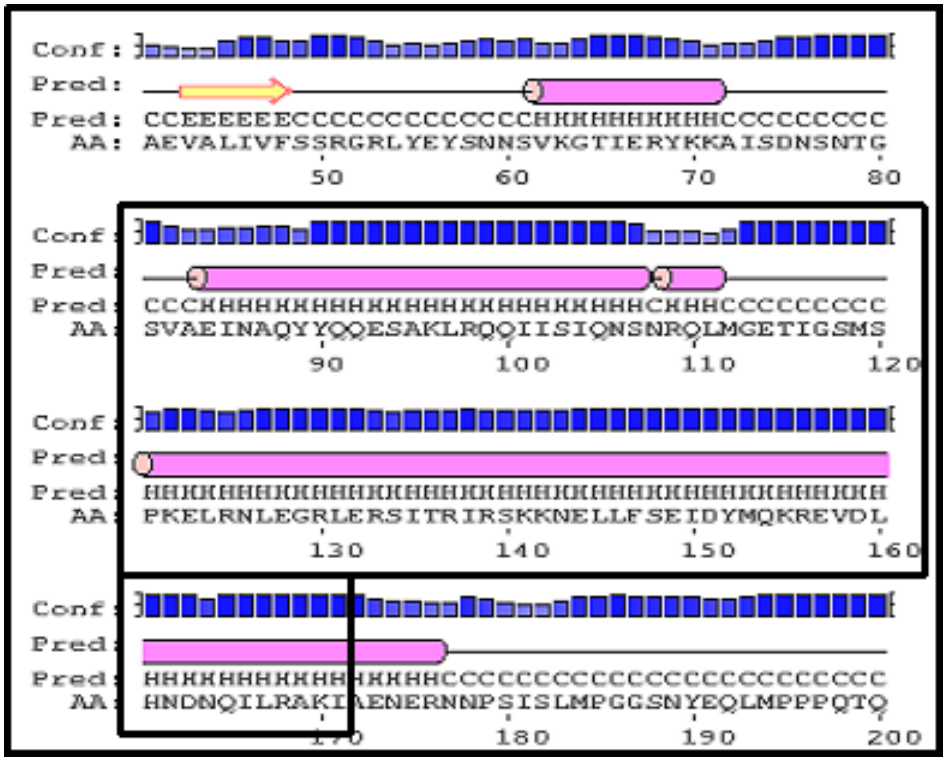


Figure III-12: ***Design of AG using secondary structure prediction.*** *The secondary structure prediction was performed using PSIPRED* (Jones, 1999). *α- helices and β- sheets are represented by pink cylinders and yellow arrows respectively. The confidence of prediction is shown by the blue columns. The boxed shows the final construct.*

The two AG constructs AG$^{(74-173)}$ and AG$^{(81-171)}$ were cloned in pESPRIT002 (Yumerefendi et al., 2010; Hart and Tarendeau, 2006; Guilligay et al., 2008) and expressed in BL21(DE3)pLysS (Life technologies) using Luria Bertani media. These constructs yielded high expression levels, but all protein was insoluble.

III.3.1.5 Large scale purification of AG protein constructs

Purification and refolding of denatured proteins from inclusion bodies using urea and guanidium hydrochloride has been used previously in numerous cases of insoluble protein aggregates (Reviewed in (Singh and Panda, 2005; Tsumoto et al., 2003; Li et al., 2004; Burgess, 2009). Similar protocols were tested on the insoluble $AG^{(74-173)}$ and $AG^{(81-171)}$ constructs obtained during this work. In both cases, the lysate was subjected to 8M urea denaturation and the construct purified using Ni-NTA affinity chromatography. The constructs were then gradually refolded by stepwise dialysis with decreasing urea concentrations followed by final purification by gel filtration chromatography. Figure III-13 shows the SDS-PAGE analysis and gel filtration purification for both $AG^{(74-173)}$ and $AG^{(81-171)}$ following this protocol.

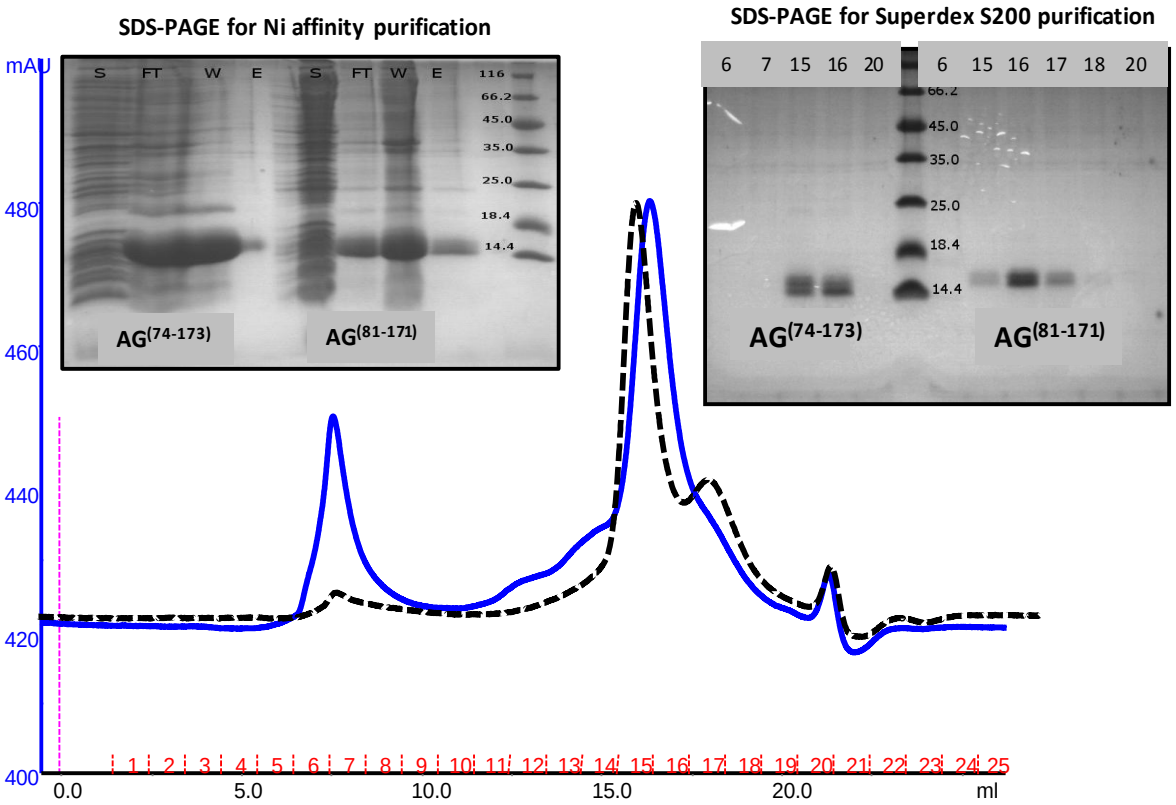


Figure III-13: ***Gel filtration and SDS-PAGE analysis of AG constructs purified by denaturation-reconstruction.*** *The SDS-PAGE for Ni-NTA affinity purification after denaturation shows strong bands for both the constructs in FT, W and E. gel filtration following refolding of the sample in the elute elution sample shows a clear separation of aggregates and protein for both $AG^{(74-173)}$ (blue) and $AG^{(81-171)}$ (dashed black). A small shoulder is observed for both the proteins. The SDS-PAGE analyses of the samples following gel filtration confirm presence of pure protein.*

As can be seen, SDS-PAGE analysis of fractions collected during Ni-NTA affinity purification in 8 M urea shows the presence of protein in the flow through, wash and elution. To avoid contamination, only the sample from the elution fraction was collected and subjected to refolding and gel filtration. The gel filtration performed using a Superdex 200 10/300 GL (GE Healthcare) column shows the presence of an early peak for A_{280} representing aggregated protein and a second peak for pure non-aggregated protein. The overlay for the two chromatograms ($AG^{(74-173)}$ and $AG^{(81-171)}$) shows that both the proteins are eluted at the same volume thus belonging to the same oligomeric state. The SDS-PAGE analysis for the gel filtration confirms the presence of pure protein for both $AG^{(74-173)}$ and $AG^{(81-171)}$ protein constructs.

The approach described above thus allowed the successful purification of $AG^{(74-173)}$ and $AG^{(81-171)}$, both of which contain the K domain of AG, for downstream experiments. In the previous chapter, I have shown the role played by the K domain in dimerisation and tetramerisation of SEP3. Similar to SEP3, AG is also known to bind DNA as a dimer and putatively form regulatory tetrameric complexes, at least with SEPALLATA family proteins (Honma and Goto, 2001; Airoldi et al., 2010; Melzer and Theissen, 2009). Due to the absence of the M domain in the purified constructs, the DNA binding properties could not be evaluated. However, the two constructs obtained did allow a structural characterisation of the oligomerisation properties of AG.

III.3.2 Biophysical characterisation of AG$^{(74-173)}$

III.3.2.1 High throughput crystallisation trials

A high-resolution crystal structure can be used to obtain valuable information about the atomic level determinants of protein complex formation. Of the two purified AG constructs, AG$^{(74-173)}$ was selected for crystallisation trials due to its higher yield from refolding protocols as observed from repeated purification trials. These were performed at EMBL HTX facility (https://embl.fr/services/ht_crystallisation/) and yielded a few crystalline precipitates. Unfortunately, no diffraction quality crystals were obtained.

III.3.2.2 Small Angle X-ray Scattering

The lack of success in obtaining diffraction quality crystals made it impossible to obtain high resolution structural information for AG$^{(74-173)}$. However, the conformation of the protein in solution can be still determined using low-resolution techniques such as small angle x-ray scattering (SAXS), which can provide information about the shape, oligomerisation state, molecular size and flexibility of a macromolecule (Putnam et al., 2007; Doniach, 2001; Svergun and Koch, 2003; Koch et al., 2003).

SAXS scattering curves for AG$^{(74-173)}$ (Figure III-15) were collected at BioSAXS (BM29) ESRF beamline (http://www.esrf.eu/home/UsersAndScience/Experiments/Beamlines/content/content/bm29.html) using an on-line HPLC system (with a Superdex 75 3.2/300 PC column (GE Healthcare) to ensure sample purity and homogeneity (Pernot 2010, Pernot et al., 2013; Round et al., 2013). The A_{280} chromatogram from the HPLC run for AG$^{(74-173)}$ is shown in Figure III-14. The analysis of scattering intensity I(0) shows a single peak and a stable radius of gyration and confirms the presence of single species in solution (Figure III-14).

SAXS data analysis consists of determining preliminary characteristics of the protein such as the radius of gyration (R_g), maximum particle dimension (D_{max}) and molecular weight followed by modelling using reconstruction algorithms with dummy beads or known domains where possible which are representative of the molecule in solution (Brennich et al., 2016). Model-independent data analysis was performed using tools from the ATSAS package (Petoukhov et al., 2012). SAXS measurements are highly sensitive to the presence of interparticle interactions at low angles, which result in skewing of the Guinier region (Guinier, 1938). The linear portion of the Guinier region from which a value for R_g of 2.7 ± 0.02 nm was derived using the program GNOM (Svergun, 1992) is shown in Figure III-15.

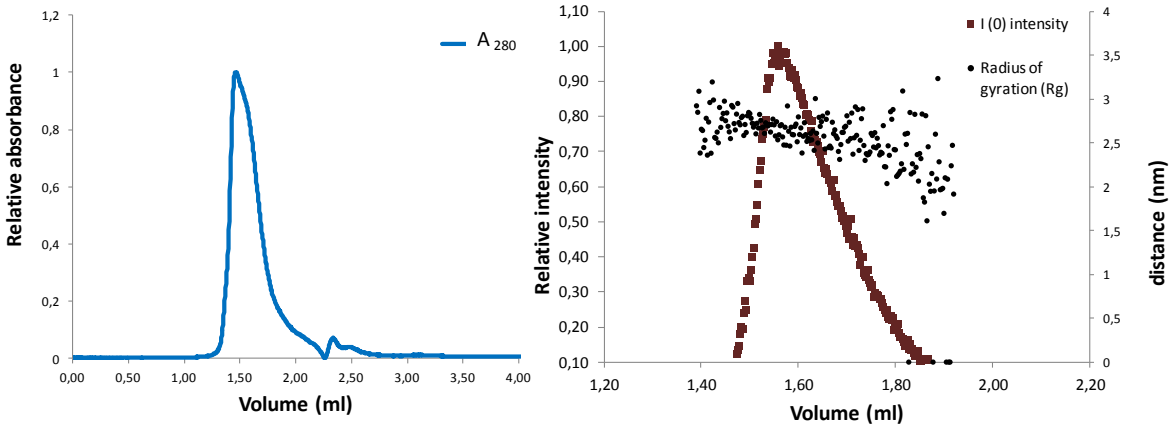


Figure III-14: ***SAXS analysis for AG$^{(74-173)}$.*** ***(left)*** *The protein AG$^{(74-173)}$ was applied to a Superdex S75 3.2/30 PC column on a online HPLC system. The A_{280} shows presence of single peak.* ***(right)*** *Analysis of I(0) and Rg shows presence of single peak for scattering intensity and a constant Rg throughout the region of peak. Frames with constant Rg were pooled for further analysis.*

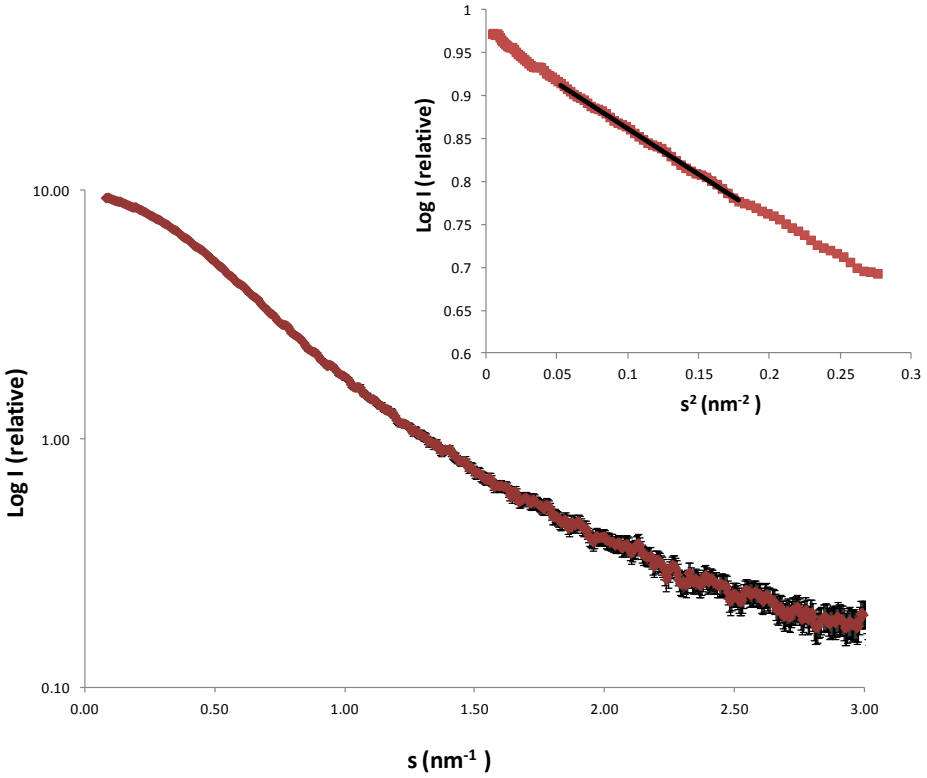


Figure III-15: ***SAXS scattering curves and Guinier analysis for solutions of AG$^{(74-173)}$.*** *The scattering curve is plotted showing Log (I) as a function of s.* ***(inset)*** *The Guinier plot shows Log(I) as a function of s^2. The upper range was limited by s*Rg < 1.3. The linear region used for analysis is indicated by a black line. The linear Guinier plot suggests absence of aggregation and inter-particle interactions.*

Information regarding the shape and volume of the protein can be obtained from the distance distribution function (P(r)) shown in the Figure III-16. The presence of an elongated tail to the distribution suggests an elongated shape of $AG^{(74-173)}$ in solution (Putnam et al., 2007). The volume obtained from the distance distribution plots was 36 nm^3 as calculated from GNOM (Svergun, 1992) while the molecular weight of the construct in solution was calculated to be 21 kDa. Thus, unlike the homologous $SEP3^{(75-178)}$ construct, $AG^{(74-173)}$ is a dimer and not a tetramer in solution.

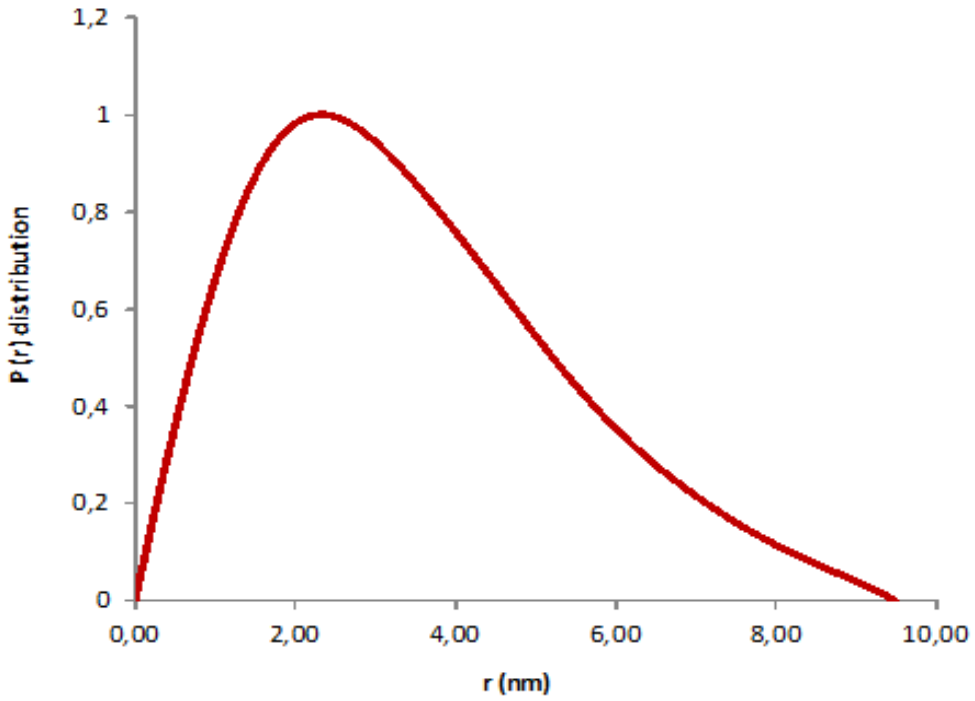


Figure III-16: ***Normalized distance distribution function for $AG^{(74-173)}$.*** *The plot shows P(r) as a function of r. The plot has been normalized for better visualisation. An elongated tail conformation is observed. A D_{max} value of 9.46 nm is obtained from the x intercept of the plot.*

The globularity and flexibility of the construct was assessed with a Kratky plot (Durand et al., 2010; Receveur-Brechot and Durand, 2012) which suggests that in solution $AG^{(74-173)}$ is flexible and non-globular. The evaluation of flexibility by the dimensionless Kratky plot depends on the divergence of the peak from that obtained for globular protein (peak maxima of 1.104 at $q*R_g = \sqrt{3}$) (Figure III-17).

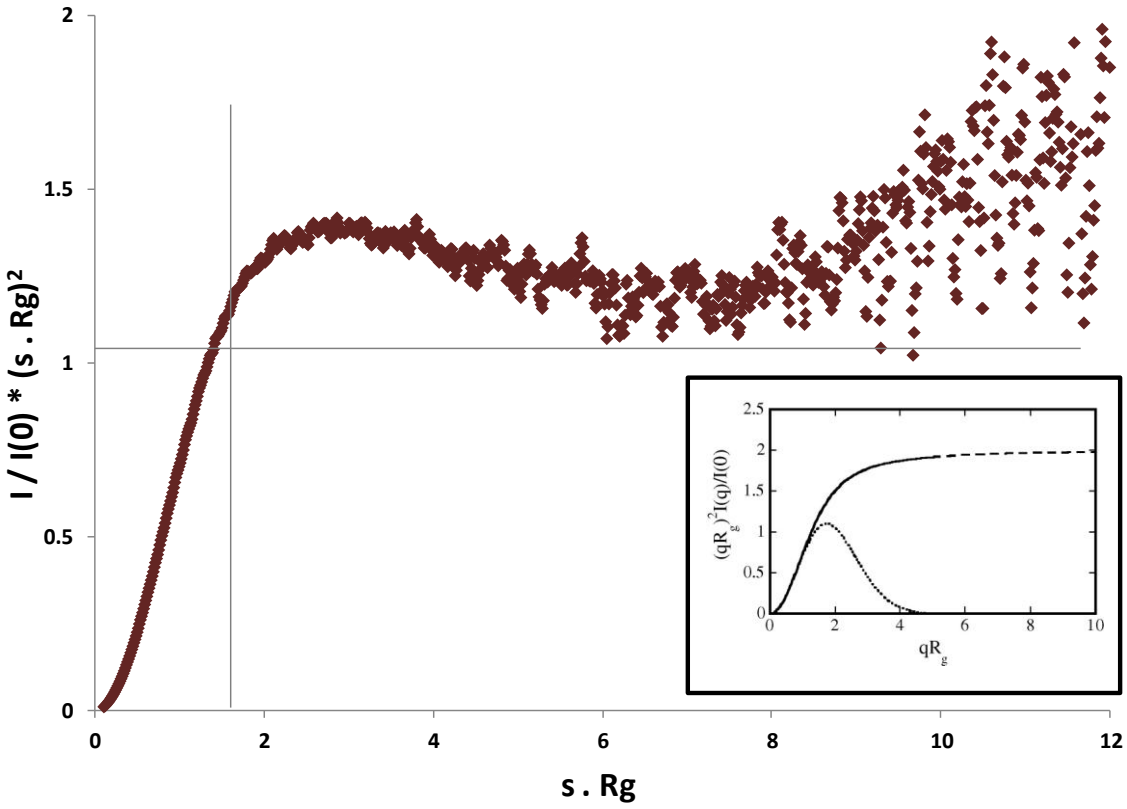


Figure III-17: ***Dimensionless Kratky plot for AG$^{(74-173)}$.*** *The crosslines mark the position for a well folded monodomain protein. The protein shows a peak shift towards right. This is indicative of a high degree of flexibility or a partial unfolding of the protein.* ***(inset)*** *Theoretical dimensionless Kratky plots representing the two extreme cases: globular protein (dotted line with peak) and completely unfolded protein (continuous line followed by dashes in plateau region* (Durand et al., 2010)

The model free SAXS analysis was followed by the construction of three-dimensional models for the solution structure of AG$^{(74-173)}$ using the *ab initio* and rigid body modelling techniques. The homology of AG$^{(74-173)}$ to SEP3$^{(75-178)}$, the availability of high resolution crystal structure for the later and the confirmation of a dimeric form for AG$^{(74-173)}$ all provided important information in the modelling studies.

III.3.2.2.1 Ab-initio modelling

Ab-initio model building generates a large number of molecular envelopes consistent with the one-dimensional scattering curve and any symmetry constraints provided. These are then averaged and filtered to rank them according to the quality of fit with the experimental data. In the case of AG$^{(74-173)}$, DAMMIF (Franke and Svergun, 2009) was used to generate 50 *ab*

initio models which were averaged and filtered using DAMAVER (Volkov and Svergun, 2003). Superimposition of averaged model generated by DAMAVER and the filtered model obtained from DAMFILT is shown in Figure III-18.

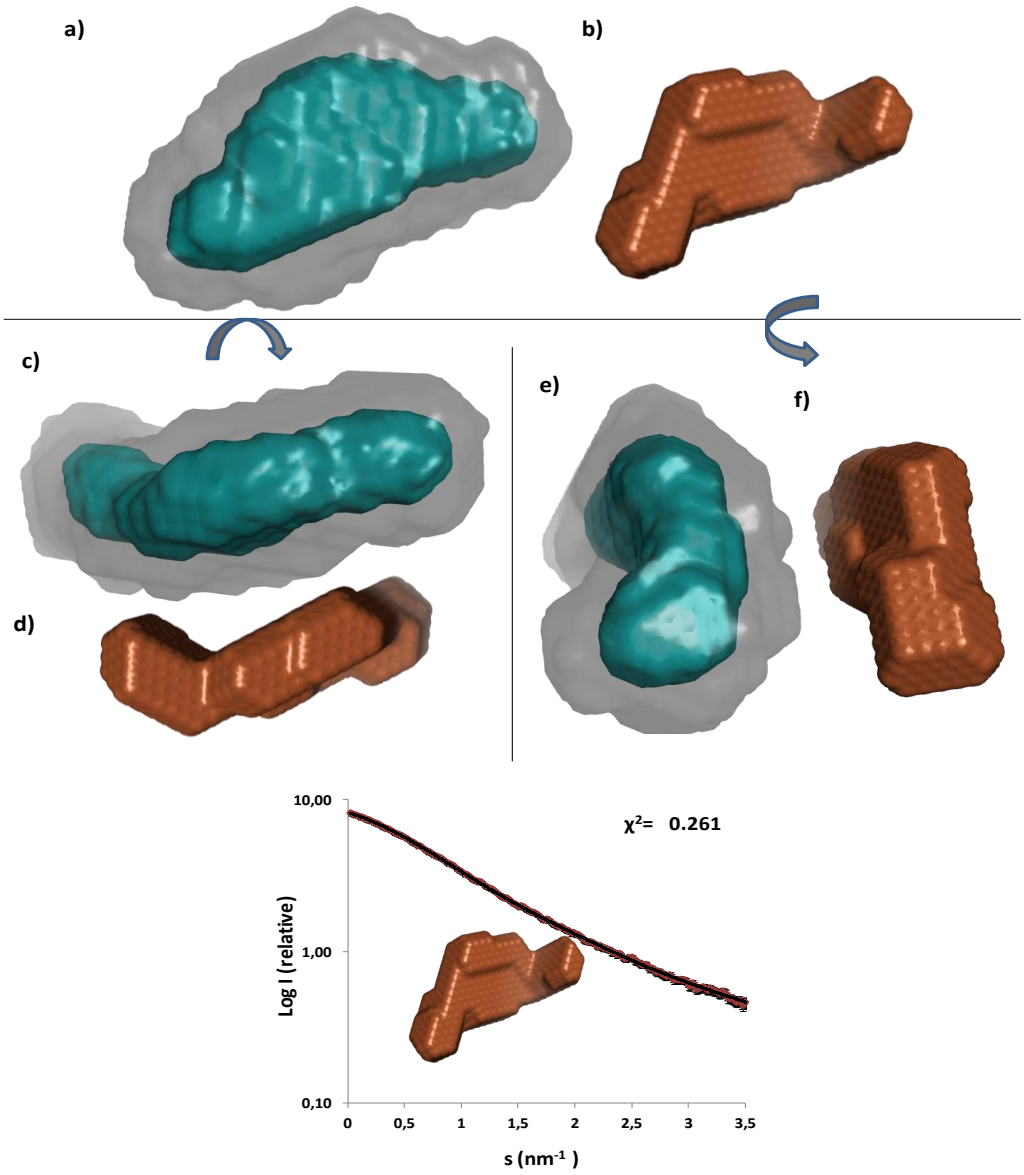


Figure III-18: ***Ab-initio models for the solution structure of AG$^{(74-173)}$. (a)*** *Overlay of average and filtered ab initio model obtained from DAMFILT (green) and that from DAMAVER (grey);* ***c,e)*** *rotation at x = 90° and y =90°* ***. (b)*** *Most representative model obtained from DAMMIF.* ***d,f)*** *rotation at x = 90° and y =90°* ***(bottom)*** *Comparison of the experimental scattering curves(brown) and the theoretical curve (black) calculated using the representative model.*

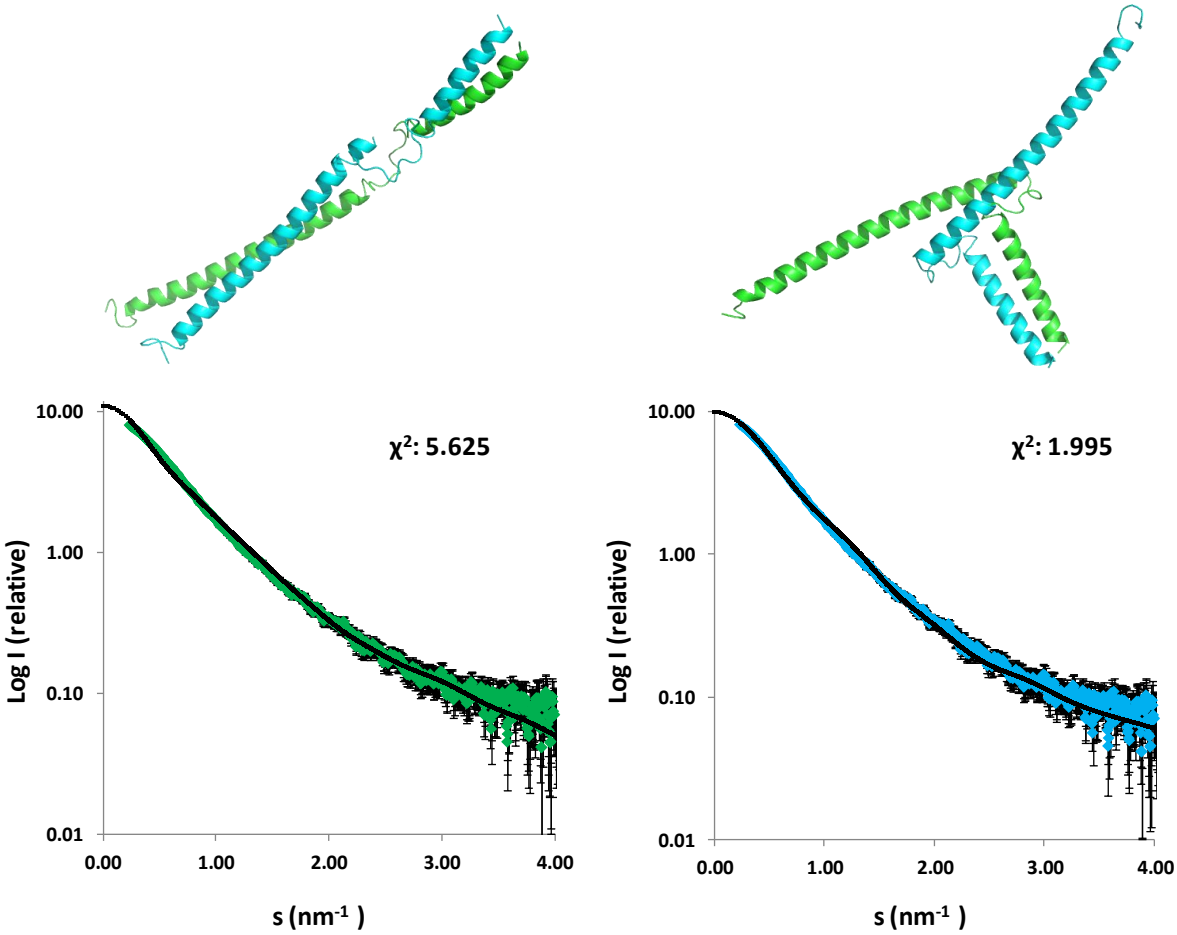


Figure III-19: ***CRYSOL analysis for AG$^{(74-173)}$ dimeric models.*** *Two conformations were postulated using the homologous region of SEP3$^{(75-178)}$. (left) shows an elongated conformation while (right) shows a bent conformation. CRYSOL fitting with the experimental data is shown below the respective models. The coloured (green & blue) lines show the experimental scattering while the black line represents the theoretical scattering of the model. The χ^2 value is also shown for each fit.*

The *ab initio* model for AG$^{(74-173)}$ provides a low resolution "structure" which fits the measured data. However, the availability of the high resolution crystal structure for the homologous domain from SEP3, provides an advantage in modelling a pseudo-atomic resolution solution structure for AG$^{(74-173)}$. The homologous domain of SEP3$^{(75-178)}$ exists as a tetramer (PDB 4OX0, Section II.3.2.2, Figure II-24) formed from a dimer of dimers. Homology models for AG$^{(74-173)}$ were generated by arranging two SEP3 monomers into two distinct dimeric conformations as shown in Figure III-19. These conformations represent the extremes - one with a 90° bend angle, the second completely linear. Theoretical scattering curves for AG$^{(74-173)}$ dimers in these two conformations were calculated and compared to the experimental scattering data using CRYSOL (Svergun et al., 1995) as shown in Figure III-19.

The analysis shown in Figure III-19 clearly shows a better fit of the theoretical curve for the 'bent' dimer (χ^2 = 1.99) than for the elongated conformation (χ^2 = 5.65). The R_g calculated for the 'bent' dimer was 3.1 nm while that for elongated was 3.6 nm, both being larger than that calculated from the experimental scattering curve (2.7 nm ± 0.02). This discrepancy can likely be attributed to the flexibility of the solution structure of $AG^{(74\text{-}173)}$ as observed from the Kratky plot shown previously. The above analysis suggests that the dimeric $AG^{(74\text{-}173)}$ is potentially able to sample a 'bent' dimeric configuration with higher probability than a completely linear conformation. However, the presence of a hydrophobic amphipathic α-helix in the later parts of K domain makes it unfavourable for the protein to be locked in to a bent configuration with exposed hydrophobic residues. An analysis of D_{max} for the bent (11.98 nm) and elongated (12 nm) and its comparison with the experimental data (9.46 nm) suggests that neither model completely fits the data. Thus, although both the possibilities cannot be completely ruled out, the presence of alternate models apart from the two proposed is also very likely as they present extremes of possible configurations.

However, it can be confidently stated that $AG^{(74\text{-}173)}$ does not adopt a tetrameric $SEP3^{(75\text{-}178)}$ – like conformation in solution. To test this point, a CRYSOL comparison of the experimental data to the theoretical scattering obtained from the $SEP3^{(75\text{-}178)}$ tetramer was carried out. This shows an extremely poor fit (Figure III-20) and the SAXS experiments described here clearly indicate that $AG^{(74\text{-}173)}$ exists as a homodimer in solution.

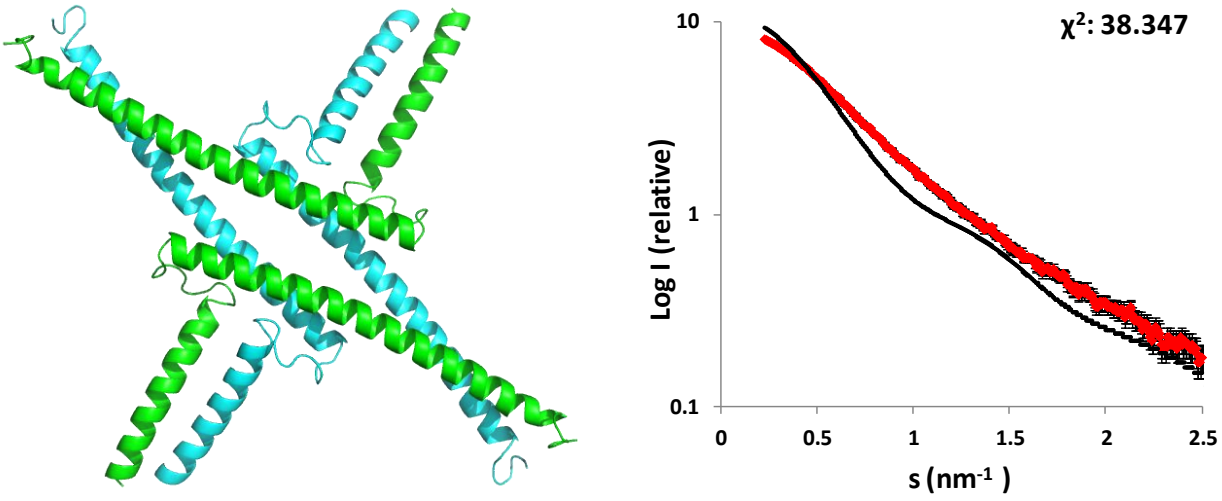


Figure III-20: ***CRYSOL analysis for $AG^{(74\text{-}173)}$ tetramer model.*** *A tetrameric model was generated similar to $SEP3^{(75\text{-}178)}$. CRYSOL fitting with the experimental data is shown on the right. The red curve represents the experimental scattering while the black line represents the theoretical scattering of the model. The chi squared value is also mentioned.*

Alignment of the amino acid sequences of different plant MADS TFs (Figure III-21) reveals a high degree of sequence homology, especially of the hydrophobic residues in the oligomerisation K-domain. In the previous chapter, the role of these conserved residues in oligomerisation of SEP3 has already been emphasized. It was noted that even single point mutations in the interacting resides lead to disruption of the SEP3 tetramer (section II.3.2.3 Figure II-27). This suggests that even small differences between the amino acid sequences can lead to a major shift in the oligomeric state of MADS TFs.

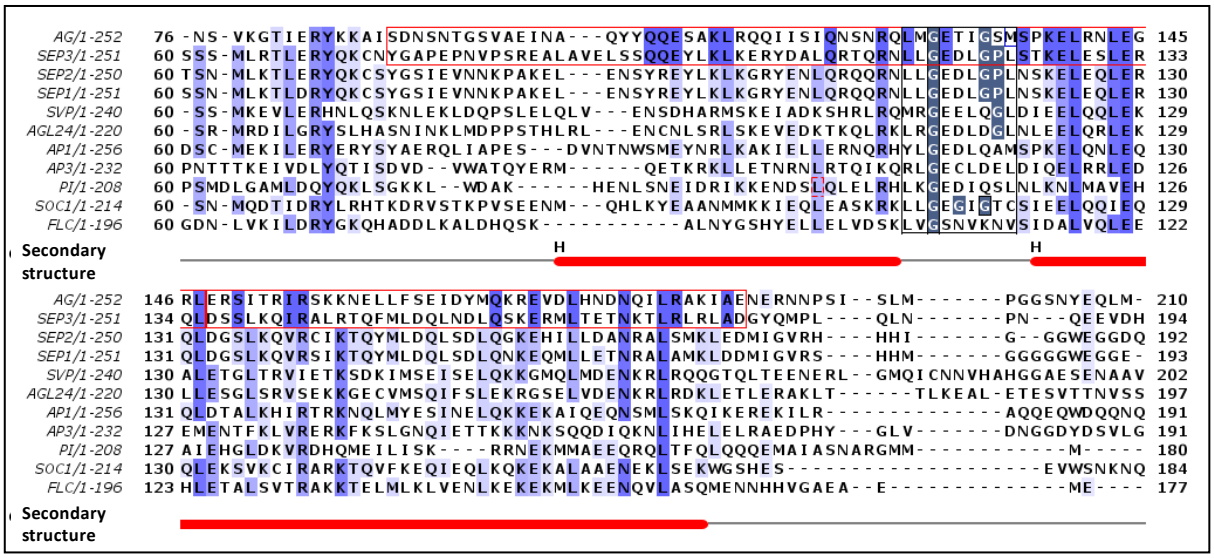


Figure III-21: ***Alignment of MADS TFs oligomerisation domain.*** *Sequence alignment comparing the amino acid sequence of AG$^{(74-173)}$ with oligomerisation domain of important MADS TFs. The red box shows AG$^{(74-173)}$ and its homology with SEP3$^{(75-178)}$. The secondary structure elements of SEP3$^{(75-178)}$is shown below with 'H' followed by red bar representing alpha helix. The black box marks the kink region. The conserved Gly and Pro residues in kink region are highlighted in grey. The conserved residues are highlighted in violet with increasing dark.*

The crystal structure of SEP3$^{(75-178)}$ shows a kink region in SEP3$^{(75-178)}$ monomers containing a Gly-Pro motif. This prevents formation of single elongated helical structure and the formation of leucine-zipper type dimeric structure. Prolines act as "breakers" in α-helices due to their inability to form the appropriate hydrogen bonding interactions between the carbonyl backbone and amide proton due to the presence of the proline side chain. Glycine residues exhibit a high degree of conformational flexibility and have been shown to lead to kinks in α-helices in soluble and membrane proteins (Wilman et al., 2014). This configuration plays a role in tetramerization of SEP3. However, comparison of different

MADS sequences shows that this kink region is highly variable. Although Gly-117 (numbering as per SEP3) is highly conserved and scattered glycine and proline residues are also observed in the kink region, the typical Gly-Pro is not observed in AG. Also, unlike SEP3, AG is dimeric in solution and SAXS studies show that the protein is highly flexible and is probably not locked into an open/bent conformation as shown for SEP3. This flexibility offered by the kink region might lead to several intermediate configurations (Figure III-22) for the AG dimer other than those provided in Figure III-19.

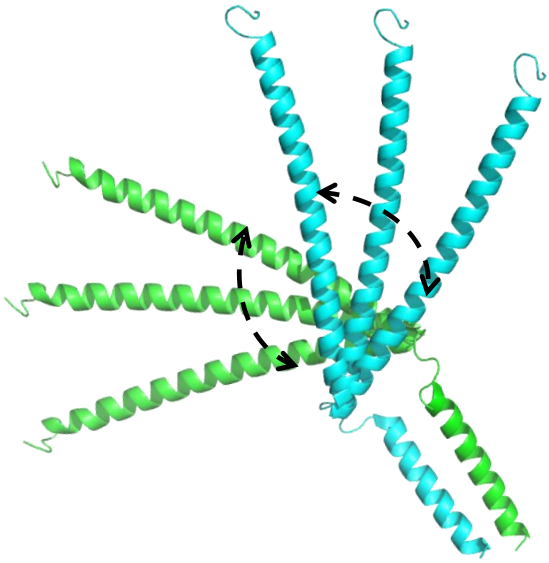


Figure III-22: ***Alternate configuration for AG$^{(74-173)}$.*** *Several intermediate configurations can be postulated for AG$^{(74-173)}$ due to the presence of flexible kink region.*

In the presence of an appropriate binding partner, such as SEP3, to drive tetramerization, the hydrophobic residues in the helical region would be buried in the complex and a tetramerization might occur. Thus, in the absence of a binding partner, AG is likely to preferentially form homodimers *in vivo*. The analysis performed here suggests that the conformation of these homodimers is dynamic and highly flexible with no one representative single conformation. However, it can be suggested that the K domain and its bent or elongated conformation may be important for determining *in vivo* function through the formation of either dimers or tetramers depending of the availability of binding partners.

III.4 Conclusions

AG is the second MADS TF to be characterized as a part of the thesis project. The role of AG in various plant developmental processes, including meristem determinacy and stamen and carpel development, made it an important target for our study. The difficulty in obtaining pure recombinant proteins, makes the MADS TFs challenging targets for biophysical characterisation. This chapter describes a purification protocol for the oligomerisation domain of AG. The structural characterisation and oligomerisation studies were performed using SAXS. An *ab initio* model was built to estimate the shape of protein. The SAXS data revealed that $AG^{(74-173)}$ exists as a dimer, as was expected by SEC, and possess a high degree of flexibility. The R_g and Kratky analysis supports the flexible nature of the dimeric complex. A study of conserved residues in the kink region of AG revealed small differences in the conserved residues when compared to the $SEP3^{(75-178)}$ protein studied previously.

AG has well-defined roles in Arabidopsis- stamen identity, carpel identity and floral meristem determinacy. These functions are likely dependent on the formation of complexes with other MADS TFs and subsequent binding to the specific promoter sequences of target genes. The ability of AG to form stable dimers via its M domain and to bind DNA has been previously demonstrated (Mizukami et al., 1996). Experiments demonstrated that the three main functions of AG are genetically separable by generating two variants lacking 12 and 14 amino acids from the K domain and a R173M mutant. The separation of these functions may suggest that these changes influence the formation of specific complexes, which initiate distinct functions *in vivo* (Sieburth et al., 1995). In addition, previous yeast two hybrid studies, have established that the individual K domain from AG is sufficient to bind the K domain of SEP proteins (Fan et al., 1997). However, these experiments were unable to detect the binding of the AG K domain with itself. Our study demonstrates that the K domain can indeed form homodimers in solution. Thus, homodimers of AG will likely form *in vivo* and may compete for heterodimerisation partners.

Based on the studies presented here, I speculate that the absence of the Gly-Pro motif favours formation of a dimer over the formation of a homotetramer. Thus, a change in the oligomerisation state due to a small change in amino acid residues might have important functional implications for the MADS TF family.

III.5 Materials and methods

III.5.1 Construct design and purification

III.5.1.1 Strains and plasmids

$AG^{(74-173)}$ and $AG^{(81-171)}$ were cloned in to the expression vector pESPRIT 002 (Hart and Tarendeau, 2006; Guilligay et al., 2008) (Figure II-39 Section II.5.1.1) using the AatII and NotI sites. The plasmid contains an N-terminal 6xHis tag followed by a TEV protease cleavage site.

DNA target	Primer sequence
$AG^{(74-173)}$	5'-TTCAGGGACGTCGGTCGGACAATTCTAACACCGGA-3‘ 5'-ATTACGCCGCGGCCGCCTCATTCAGCTATCTTTGCACGAAG-3‘
$AG^{(81-171)}$	5'-TTCAGGGACGTCGGTCGGTGGCAGAAATTAATGCA-3' 5'-ATTACGCCGCGGCCGCCTCATATCTTTGCACGAAGAATCTGG-3'
AG full length In pIVEX 2.3d	5'-GATATACCATGGCTGGGAGAGGAAAGATCGAAATCAAACG-3‘ 5'-AACCCCCCCCGGGTAACTGGAGAGCGGTTTGGTCTTGGCG-3‘

*Table III-2**: Primers used for DNA amplification.** The different primers used for amplification of DNA target sequences are provided here.*

III.5.1.2 ESPRIT and Cell free expression system

The protocols for ESPRIT library generation and cell free expression trials were similar to that used for SEP3 in the previous chapter (Refer section II.5.1.2 and II.5.1.3). The protocol for multi-vector expression screening is described later in section IV.3.1.1.

III.5.1.3 Large scale expression and purification

$AG^{(74-173)}$ was cloned into a pESPRIT002 vector (Hart and Tarendeau, 2006; Guilligay et al., 2008) using NotI and AatII restriction sites. The construct contained an N-terminal TEV protease cleavable 6x-His tag. The protein was overexpressed in *E. coli* BL21 Star (DE3) pLysS cells (Life Technologies). Cells were grown in Luria Bertani medium in the presence of 50 mg/ml kanamycin and 35 mg/ml chloramphenicol at 37 °C and 180 rpm to an optical density A_{600} = 0.8 after which time the temperature was lowered to 20°C and 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for induction. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min and resuspended in lysis

buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP , 5%(v/v) glycerol, 20% (w/v) sucrose and 1x protease inhibitors (Roche EDTA-free)). Cells were lysed by sonication and the insoluble fraction pelleted by centrifugation at 25000 rpm and 4 °C for 30 min. The pellet was resuspended in denaturation buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol, 8 M Urea) and incubated for 1 h at room temperature. The solubilised fraction was applied to a 5 ml Ni-NTA (Hi-trap. HP GE Healthcare) column pre-equilibrated with denaturation buffer, followed by a wash with 10 CV of wash buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 30 mM imidazole) and eluted with 3 CV of elution buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 300 mM Imidazole). The eluted fraction was dialysed step-wise against 6M, 4M, and 2M urea plus 30mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol. After the final dialysis step, the protein was applied to a size exclusion chromatography column (Superdex 75 10/300 GL, GE Healthcare) pre-equilibrated with gel filtration buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol). The purity of the final fractions was assessed using SDS-PAGE. Fractions of interest were pooled and incubated overnight with TEV protease to remove the 6xHis tag. After depletion of TEV and uncleaved protein over a 5 ml Ni-NTA column, the cleaved $AG^{(74-173)}$ was loaded onto a Superdex S75 10/300 GL (GE Healthcare) column as a final purification step and the fractions of interest pooled and concentrated to approximately 4 mg/ml for SAXS studies.

III.5.2 Biophysical characterisation

III.5.2.1 High throughput crystallisation trials

$AG^{(74-173)}$ at a concentration of 3.5 mg/ ml in 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol was subjected to high throughput crystallisation trials using the EMBL-Grenoble HTX facility (https://embl.fr/htxlab/). 200 nl sitting drops were set up using Cartesian PixSys 4200 (Genomic Solutions, UK) crystallisation robot using the Greiner CrystalQuick plates (flat bottom, untreated) and imaged with a Formulatrix Rock Imager (Formulatrix Inc.,USA) at 277 K (Dimasi et al., 2007). Commercial crystal screens from Hampton Research (Aliso Viejo, California, USA), Qiagen (Hilden, Germany) and Molecular Dimensions (Suffolk, UK) were used in robotic screening trials.

III.5.2.2 SAXS data collection

An on-line HPLC system (Viscotek, Malvern Instruments) was attached directly to the sample inlet valve of the BM29 sample changer (European Synchrotron Radiation Facility,

bioSAXS bending magnet beamline 29) (Pernot et al., 2013; Round et al., 2013). The protein sample (50 µl) was injected onto the column (Superdex 75 3.2/300 PC, GE Healthcare) after column equilibration. Buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP) was degassed prior to the run and a flow rate of 0.1 ml/min at room temperature was used. All data from the run was collected using a sample to detector (Pilatus 1M Dectris) distance 350 of 2.86 m corresponding to an s range of 0.04-4.9 nm^{-1}. Approximately 1800 frames (1 frame/s) per HPLC run were collected. Initial data processing was performed automatically using the EDNA pipeline (Incardona et al., 2009), generating radially integrated, calibrated and normalised 1-D profiles for each frame. All frames were compared to the initial frame and matching frames were merged to create the reference buffer. Any subsequent frames which differed from the reference buffer were subtracted and then processed within the EDNA pipeline using tools from the EMBL-HH ATSAS suite (Petoukhov et al., 2012). The invariants calculated by the ATSAS auto R_g tool were used to select a subset of frames from the peak scattering intensity. The 49 frames corresponding to the highest protein concentration were merged manually and used for all further data processing and model fitting. Molecular weight for the protein was estimated based on the correlated volume (Rambo and Tainer, 2013). The volume was calculated using the GNOM interface of the cross platform version of PRIMUS for the 4 ATSAS software suite (Svergun, 1992; Konarev et al., 2003). *ab initio* models (~50) representing the solution structure were generated using the distance distribution function P(r) as input files with program DAMMIF (Franke and Svergun, 2009) without any symmetry constraints. The resultant models were aligned, averaged and the most probable models were selected with the program suite DAMAVER (Volkov and Svergun, 2003). The final bead models generated were visualized using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

Homology models for $AG^{(74-173)}$ were generated based on the SEP3 structure (PDB 4OX0; (Puranik et al., 2014). For the elongated conformation, the kink between helices 1 and 2 was removed, the helices superposed and residues corresponding to the flexible region between the helices built in manually using COOT with idealized geometry and no secondary structure restraints. The model for the bent conformation was generated by threading the sequence of $AG^{(74-173)}$ directly on to the SEP3 dimer (4OX0). Structures corresponding to two different dimer conformations (bent and elongated) were used to calculate theoretical scattering curves. These curves were compared with the experimental data using CRYSOL (Svergun et al., 1995)

Chapter IV

SHORT VEGETATIVE PHASE

IV SHORT VEGETATIVE PHASE

IV.1 Summary

The significance of the oligomerisation domain on complex formation for the MADS TFs responsible for floral organ differentiation has been described in previous chapters. In this chapter, I will focus on a MADS TF that acts as an important flowering time regulator, SHORT VEGETATIVE PHASE (SVP). This chapter will present the biochemical and biophysical characterisation of the protein, its oligomerisation state in solution and its interactions with DNA *in vitro*. Here, the first purification protocol for a full length MADS TF is presented as are the results of SAXS experiments performed in order to study the structure of SVP protein in solution. Several representative models fitting the experimental data are presented and discussed. In addition, the DNA binding characteristics of SVP investigated by microscale thermophoresis (MST), electrophoretic mobility shift assay (EMSA) and atomic force microscopy (AFM) are presented. This chapter also introduces a cutting edge development in AFM, high speed AFM (HS-AFM) that can be used to study DNA-protein interactions in real time.

Résumé en Français

L'importance du domaine d'oligomérisation pour la formation d'un complexe de facteur de transcription MADS, responsable de la différentiation des organes floraux, a été décrite dans les chapitres précédents. Dans ce chapitre, je me concentre sur l'importance d'un régulateur du moment de floraison, Phase Végétative Courte (SVP). Ce chapitre présente la caractérisation biochimique et biophysique de la protéine, son état d'oligomérisation en solution et ses interactions *in vitro* avec l'ADN. Ici, je présente le premier protocole de purification pour un fracteur de transcription MADS entier, SVP. La diffusion des rayons-X aux petits angles (SAXS) a été utilisée afin d'étudier la structure de la protéine en solution. Plusieurs modèles représentatifs correspondant aux données expérimentales sont présentés et discutés. En outre, les caractéristiques d'interaction à l'ADN de SVP sont étudiées par thermophorèse à échelle microscopique (MST), par des expériences de décalage de mobilité électrophorétique (EMSA) et par microscopie à force atomique (AFM). Ce chapitre présente également l'utilisation d'une technique AFM de pointe, l'AFM haute vitesse (HS-AFM), qui peut être utilisée pour étudier les interactions ADN-protéines en temps réel. Enfin, de futures

expériences pouvant être effectuées afin de mieux comprendre le processus d'oligomérisation et de liaison à l'ADN des facteurs de transcriptions MADS sont discutées.

IV.2 Introduction

Plants, being sessile organisms, have to continuously adjust to changes in the surrounding environment for their survival and proliferation. The floral transition, the change from the vegetative phase to the generative phase of growth, is a key event, which is affected by environmental factors such as temperature and length of the day, in addition to endogenous signals. In the model plant *Arabidopsis thaliana,* MADS TFs have been previously shown to regulate various pathways that are involved in the control of the floral transition in response to environmental and hormonal changes (Andrés and Coupland, 2012; Lee et al., 2007; Mouradov et al., 2002). The regulation of flowering time is important for the success of any given plant species as it correlates with availability of sunlight and pollinators.

SHORT VEGETATIVE PHASE (SVP) is a MIKC-type MADS TF that plays distinct roles in both the vegetative and reproductive phases. During the vegetative phase, it acts as a repressor of the floral transition while during the reproductive phase it contributes to the specification of floral meristems (Hartmann et al., 2000; Gregis et al., 2008; Li et al., 2008; Lee et al., 2007). ChIP–seq studies performed on 2 week old seedlings, representing the vegetative phase, and on early stage flowers, representing the reproductive phase, have identified about 3000 genes that are potential targets of SVP (Gregis et al., 2013). Some of these were found to be specific for one stage, while others, such as those involved in meristem development, were regulated during both phases. Indeed, SVP is known to modulate and control the expression levels of floral pathway integrator genes such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and *FLOWERING LOCUS T (FT)* in response to changes in temperature, hormone levels and the plants developmental state in order to control flowering time (Andrés and Coupland, 2012; Jang et al., 2009). SVP also interacts with FLOWERING LOCUS C (FLC) to repress the expression of *FT* and of other floral transition initiator genes (Li et al., 2008; Jang et al., 2009; Smaczniak et al., 2012a).

SVP also regulates flowering through multiple pathways independent of *SOC1* and *FT*. SVP targets are involved in several pathways, including the circadian pathway (through *GIGANTEA* repression), the photoperiodic pathway (through *PSUDO-RESPONSE REGULATOR 7*) and the autonomous pathway (via *FLOWERING LATE KH MOTIF* and *FLOWERING LOCUS D*) (Gregis et al., 2013; Li et al., 2008; Jang et al., 2009). SVP's role in the specification of the floral meristem occurs through its direct interactions with

APETALA1 (AP1). The SVP-AP1 complex helps establish floral meristem identity and restricts the expression of the downstream floral homeotic genes such as *APETALA3 (AP3), PISTILLATA (PI), AGAMOUS (AG)* and *SEPALATA3 (SEP3)* (Gregis et al., 2009). Once the sepal primordial commences differentiation, SVP expression stops, resulting in de-repression of floral homeotic genes.

Temperature is among one of the important factors that can have a wide-spread effects on the reproductive success of plants. Considering the visible effects of climate change and global warming (Kelly and Goulden, 2008), understanding temperature regulation pathways and the proteins involved in these is an important challenge. Two main mechanisms have been proposed for the functioning of SVP with respect to temperature sensing. At colder temperatures, SVP is postulated to form a complex with FLM (FLOWERING LOCUS M) variant β, which directly represses the expression of floral activator genes including *FT*, *TSF (TWIN SISTER OF FT (TSF)* and *SOC1*. However, at warmer temperatures, the activation of flowering occurs, putatively due to increased transcriptional levels of FLM-δ, a second FLM splice variant, leading to reduced levels of the active SVP-FLM- β complex (Lee et al., 2013; Posé et al., 2013; Nilsson, 2013). A second mechanism proposes that the temperature dependent degradation of SVP is responsible for the de-repression of flowering genes (Lee et al., 2013). Whether one or both mechanisms are important for SVP activity is still under investigation. However, it is clear that *svp* mutants show loss of sensitivity to temperature and an early flowering phenotype (Lee et al., 2007; Hartmann et al., 2000).

It has been suggested that interaction of SVP with different binding partners results in the differential targeting of SVP complexes to different genomic regions and might be an important factor conferring the ability of the TF to participate in distinct regulatory pathways. In order to probe the molecular and atomic level determinants of oligomerisation and DNA-binding, structural characterisation of SVP using complementary techniques including SAXS, SEC-MALLS, AFM and MST was performed and presented here.

SVP is a 240 amino acid protein divided into 4 domains in the classic MIKC pattern followed by type II MADS TFs as shown in Figure IV-1. Secondary structure prediction (PSIPRED) (Jones, 1999) shows the presence of mainly alpha helices in the I and K domain and a large unstructured C terminal domain. In this chapter, I will discuss the biochemical and biophysical characterisation of SVP. The parallel techniques used to obtain soluble constructs for SVP, biophysical techniques, including SAXS, used for structure determination and DNA binding studies performed using AFM, EMSA and MST will be presented. The

chapter concludes by introducing high speed AFM, a cutting-edge application of AFM that can track protein-DNA binding in real time. These techniques have opened new vistas for the study of plant MADS TFs and possess the potential to understand the mechanisms regulating flowering.

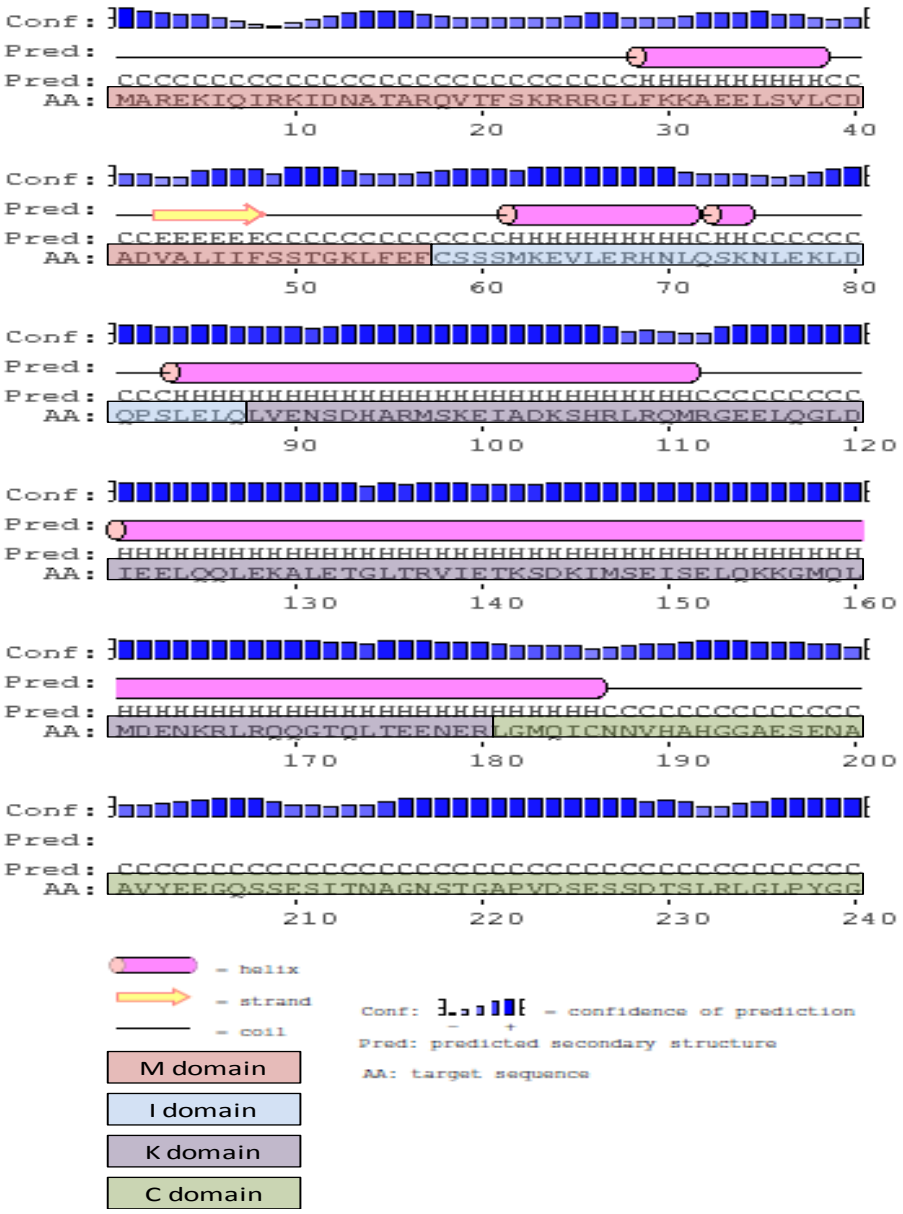


Figure IV-1: ***Secondary structure prediction from PSIPRED*** **(Jones, 1999)**. *The different domains annotated from Uniprot database* (The UniProt Consortium, 2014) *are marked with different colours. The pink cylinders represent β strands while the yellow arrows represent α helices. The confidence of prediction is given by blue bars where the height of bars.*

IV.3 Results and Discussion

One of the main bottlenecks in the study of the MADS TFs, as has been discussed previously in this work, is the production of soluble protein. I have previously presented brute force screening techniques (ESPRIT) (Section II.3.1.4) and construct design based on secondary structure prediction and homology models aimed at production of recombinant proteins. Here, I describe a screening technique, briefly mentioned in the previous chapters, that was successfully used to obtain a high yield of soluble protein for SVP. The technique utilizes a high throughput multi-vector screen and different expression systems aimed at increasing the solubility of the target protein by the addition of different solubility tags.

IV.3.1 Construct design and protein purification

IV.3.1.1 Multi vector expression screen

A multi-vector expression screen consists of parallel expression tests using a variety of plasmid vectors from the pOPIN suite in bacterial, insect and mammalian expression systems (Berrow et al., 2009; Bird, 2011; Bird et al., 2014). Here, high throughout expression screening was performed at the Oxford Protein Production Facility (OPPF), UK (http://www.oppf.rc-harwell.ac.uk/OPPF/) and funded by an Instruct Fellowship. The vectors used incorporate different solubility tags such as 6xHis, 6xHis-GST, 6xHis-SUMO and 6xHis-MBP, which are known to increase the solubility of target proteins (Hammarström et al., 2002; Dyson et al., 2004; Bird, 2011). Along with the primary target protein SVP, the expression of few other important MADS TFs such as SEP3 (Chapter II), AG (Chapter 0), AP3, PI and FLM were also screened using this method. The '6xHis' tag consists of 6 histidine residues and is routinely used for purification of recombinant proteins using Ni-affinity chromatography. The GST tag consists of the 26 kDa protein, Glutathione S-transferase (220 a.a) that can increase the solubility of the target protein and assist in purification as immobilized glutathione beads can be used for purification of GST tagged proteins using affinity chromatography (Smith and Johnson, 1988). The MBP tag is a large, 45 kDa tag comprising the highly soluble Maltose Binding Protein (di Guana et al., 1988; Kapust and Waugh, 1999). The SUMO tag, Small Ubiquitin-like Modifier, is an approximately 12 kDa protein. Almost all the tested constructs possessed an N-terminal 6xHis tag in front of the other solubility tags to facilitate downstream purification steps.

The MADS genes, including *SVP*, were cloned in a range of vectors using the In-Fusion™ (Takara- Clontech, CA, USA) cloning technique. These vectors were selected on the basis of their solubility tag, antibiotic resistance and compatibility for baculovirus and mammalian expression systems (Bird, 2011). 32 constructs for single expression and 36 for co-expression were designed for screening in *E.coli* while 22 constructs were selected for Sf9 insect cell and mammalian HEK 293 cell expression (Appendix I). The *E. coli* screen also tested expression in two different strains (Rosetta2 (DE3) pLacI and Lemo21 (DE3)) as well as two growth media (Power broth (AthenaES, Baltimore, MD USA) and auto inducible media) making the total screen about 200 conditions. The general workflow of the OPPF expression screening system is shown in the Figure IV-2.

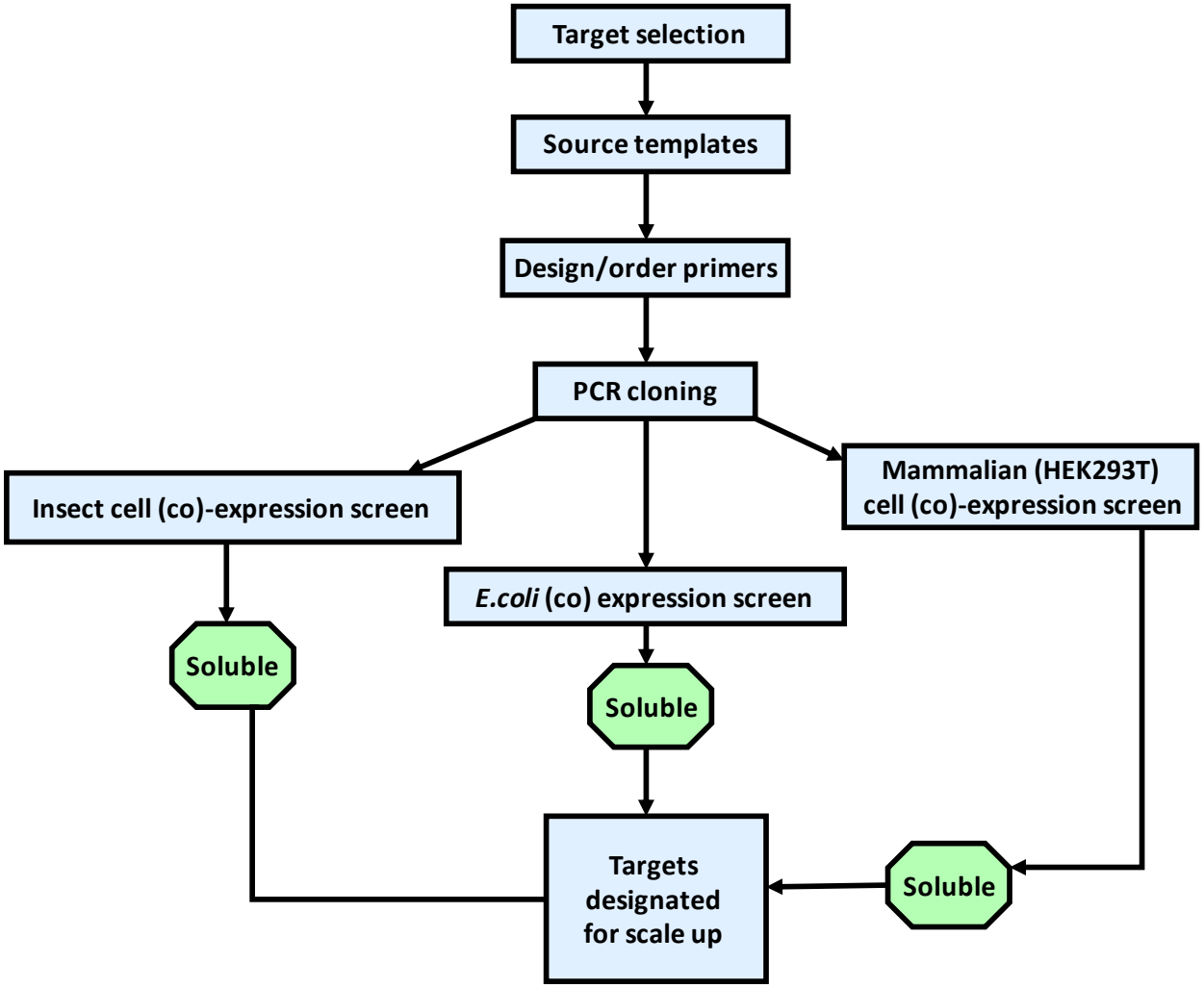


Figure IV-2: ***Workflow of OPPF comprising of three parallel expression systems.*** *The soluble constructs obtained from each are pooled and scaled up.*

Due to the large number of conditions tested in the expression screen, a list of all the constructs tested and the resultant SDS-PAGE gel images of the Ni-NTA affinity purifications are provided in Appendix I. Only a few representative results, which yielded soluble protein, are reported here (Figure IV-3, Figure IV-4 and Table IV-1). A number of positive constructs are observed both in Lemo21 (DE3) and Rosetta2 (DE3) cell lines. Table IV-1 provides an annotation for the SDS-PAGE gels. Positive bands for the target protein SVP tagged with N-terminal GST, SUMO and MBP are observed in all the conditions.

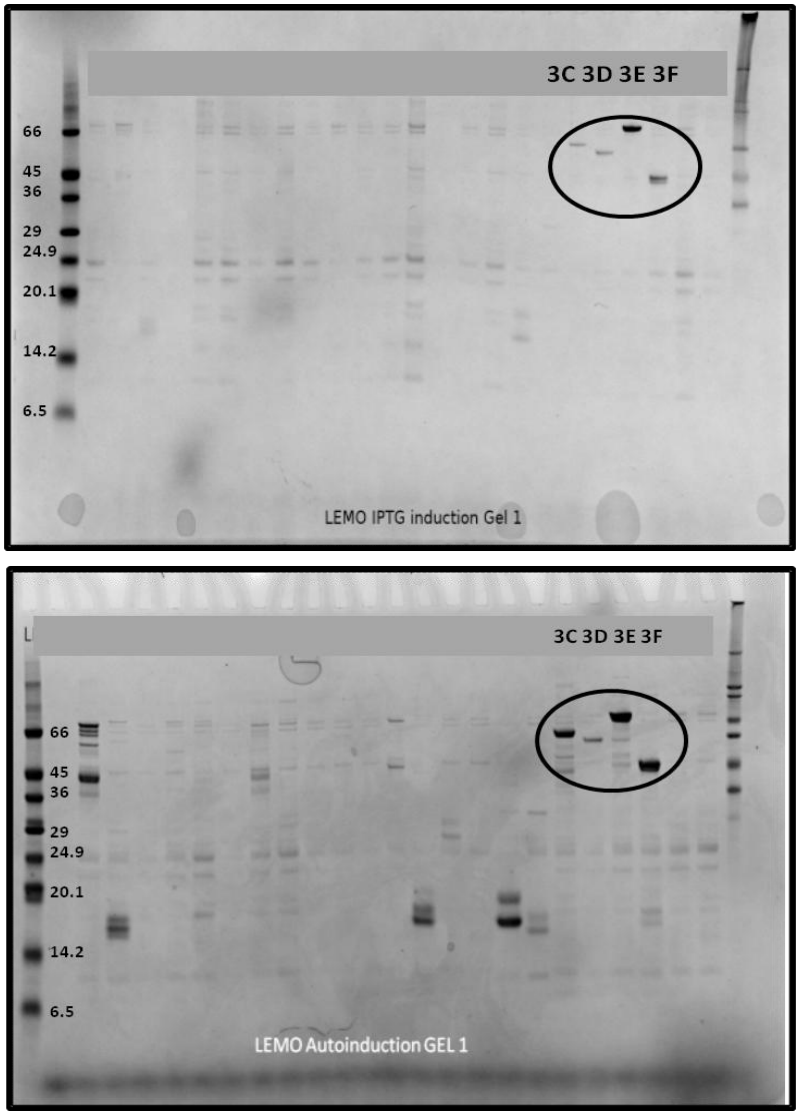


Figure IV-3: ***SDS-PAGE analysis for Ni- affinity purification from E.coli expression screen.*** *The constructs were expressed in LEMO 21 cells and induced with (top) 1mM IPTG induction and (bottom) auto inducible system. The positive constructs are printed in bold and marked by white arrows. Table IV-1 below provides details of the constructs.*

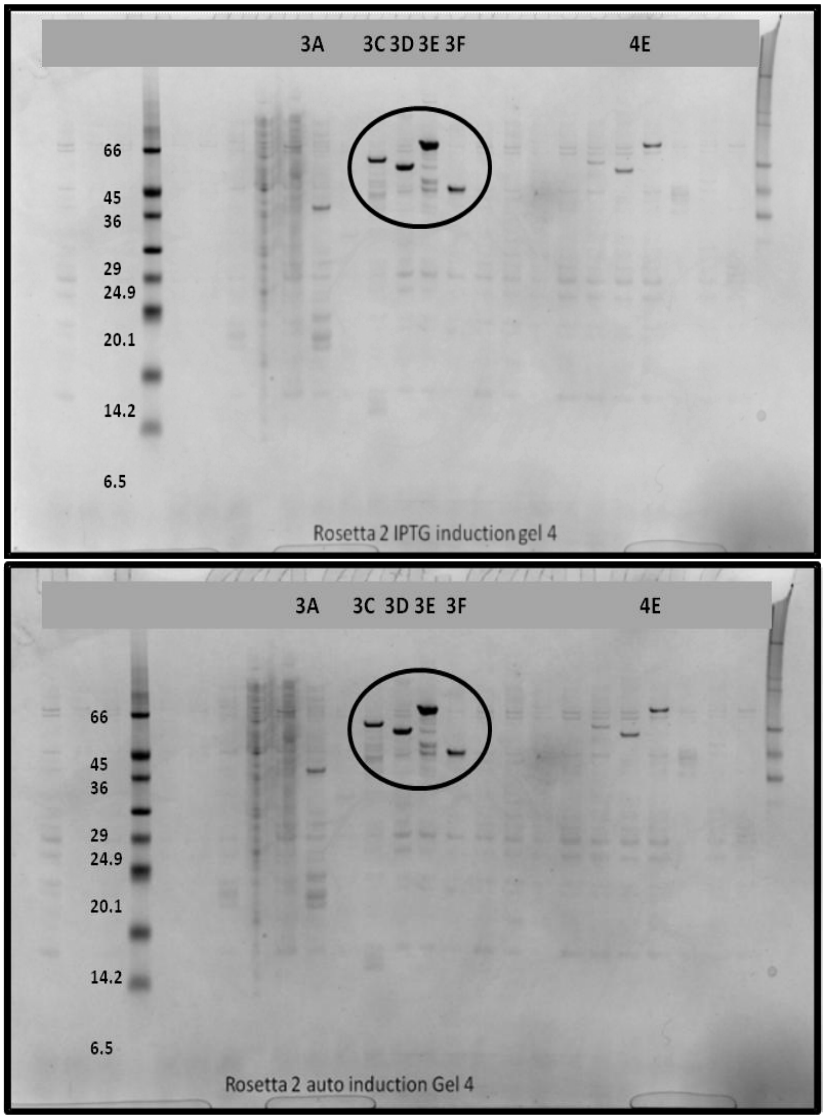


Figure IV-4: ***SDS-PAGE analysis for Ni-affinity purification for E.coli expression screen****. The constructs were expressed in Rosetta 2 cells and induced with (top) 1 mM IPTG and (bottom) auto inducible system. The positive constructs are printed in bold and marked by white arrows. Table IV-1 below provides details of the constructs.*

Cell lines	Well	Construct	Vector	Tag	Mol Wt (kDa)
Rosetta2	3A	PI	pOPINS3C	HIS6-SUMO-3C-POI	35.9
Rosetta2 /Lemo21	3C	FLM	pOPINM	HIS6-MBP-3C-POI	58.1
Rosetta2 /Lemo21	**3D**	**SVP**	**pOPINJ**	**HIS6-GST-3C-POI**	**54.2**
Rosetta2 /Lem21	**3E**	**SVP**	**pOPINM**	**HIS6-MBP-3C-POI**	**69.0**
Rosetta2 /Lemo21	**3F**	**SVP**	**pOPINS3C**	**HIS6-SUMO-3C-POI**	**39.5**
Rosetta2	4E	AP3	pOPINJ	HIS6-GST-3C-POI	53.3

Table IV-1: ***Analysis of SDS-PAGE gels from Figure IV-3 and Figure IV-4*** *The name of construct, the vector used for cloning, the solubility tag use and the theoretical molecular weight of the protein is provided. The target protein discussed in this chapter is highlighted in bold. 3C-protease to cleave tag, POI- Protein of Interest*

As can be seen the positions of the bands observed in the SDS-PAGE gels are consistent with the theoretical molecular weight of the constructs expressed. Along with our priority target SVP, several positive constructs for AP3, PI and FLM are also observed.

Expression of constructs was also tested in an insect cell expression system (Figure IV-5 and Table IV-2). Here, positive results are seen for 6xHis-GST, 6xHis-MBP and 6xHis-SUMO tagged SVP.

Finally, expression screening in mammalian cells was also performed (Figure IV-6 and Table IV-3). These also yielded some positive hits for 6xHis-GST, 6xHis-MBP and 6xHis-SUMO tagged SVP. In this test screen, several positive constructs for SEP3, PI, AG and FLM were also observed.

The expression tests carried out at OPPF thus yielded several positive clones in all three expression systems. However, bacterial expression using *E. coli* has several advantages over insect and mammalian systems notably, the ease of use, speed, simplicity and economy of the expression system. Hence, the bacterial expression system (Figure IV-3 & Figure IV-4) was

decided as method of choice for the scale up of constructs obtained from the multi-vector expression screen.

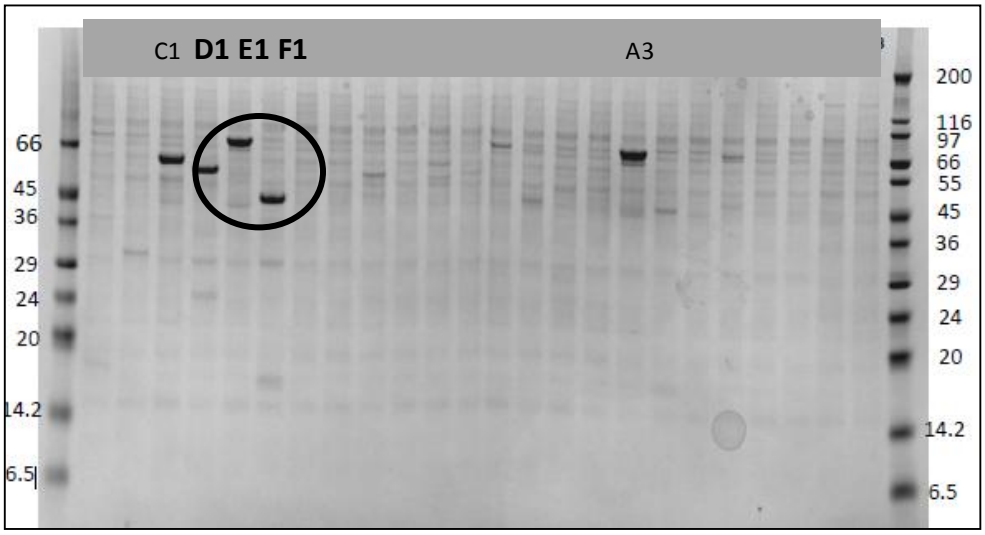


Figure IV-5: ***SDS-PAGE for Ni-affinity purification of insect cell expression screen.*** *The test expression was performed using Sf9 cells in Sf900II medium in monolayer at 27°C. The samples were grown for 72 h using 3 µl of 1st generation virus (P1) for infection. The lanes for SVP are printed in bold and bands are encircled. The annotation of the well numbers is given in Table IV-2. Similar results were obtained with infection by 30 ul of virus P1*

Well	Construct	Vector	Tag	Mol wt (kDa)
C1	FLM	pOPINM	HIS6-MBP-3C-POI	58.1
D1	**SVP**	**pOPINJ**	**HIS6-GST-3C-POI**	**54.2**
E1	**SVP**	**pOPINM**	**HIS6-MBP-3C-POI**	**69.0**
F1	**SVP**	**pOPINS3C**	**HIS6-SUMO-3C-POI**	**39.5**
A3	SEP3	pOPINM	HIS6-MBP-3C-POI	70.2

Table IV-2: ***Analysis for SDS-PAGE gels from*** *Figure IV-5. The name of construct, the vector used for cloning, the solubility tag use and the theoretical molecular weight of the protein is provided. The target protein is highlighted in bold.*

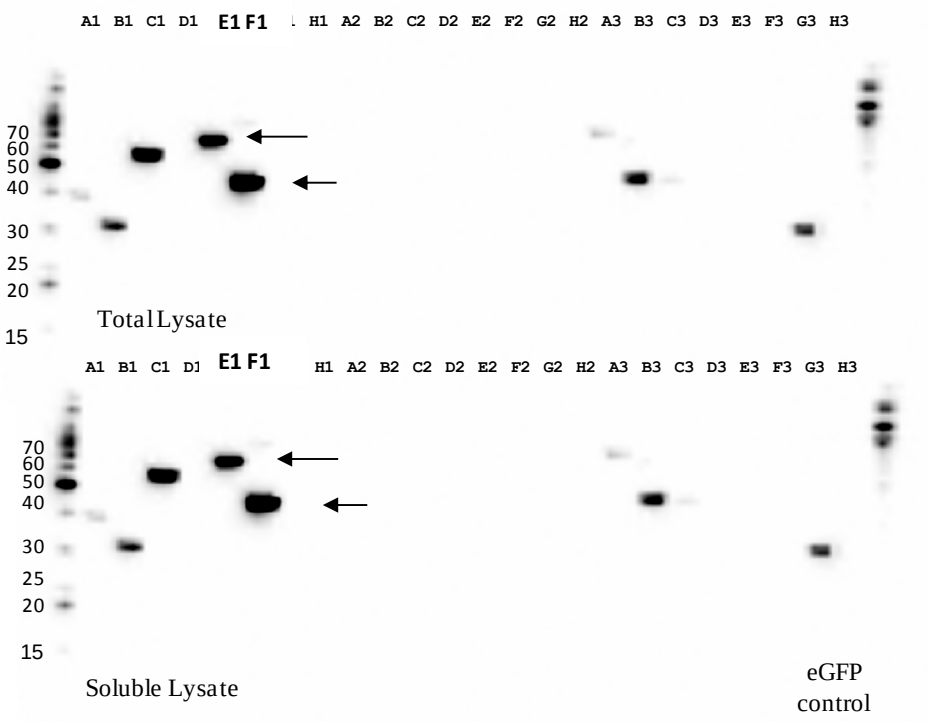


Figure IV-6: ***Western blot for HEK 293 expression tests probed with anti-6xHis antibody.*** *The bands are visible in both total and soluble lysate. The positive lanes for SVP are printed in bold and the bands are marked by arrows. The annotation of lanes is given in Table IV-3.*

Well HEK screen	Construct well	Construct name	Vector	Mol wt (KDa)
A1 (low)	A3	PI	pOPIN S3C	35.9
B1	B3	FLM	pOPINS3C	28.6
C1	C3	FLM	pOPINM	58.1
D1	D3	SVP	pOPINJ	54.2
E1	**E3**	**SVP**	**pOPINM**	**69.0**
F1	**F3**	**SVP**	**pOPINS3C**	**39.5**
A3(low)	B1	SEP3	pOPINM	70.2
B3	C1	SEP3	pOPINS3C	40.7
C3(low)	F1	Agamous	pOPINS3C	38.9

Table IV-3: ***Analysis for western blot from Figure IV-6,*** *The name of construct, the vector used for cloning, the solubility tag use and the theoretical molecular weight of the protein is provided. The target protein is highlighted in bold.*

IV.3.1.2 E.coli expression scale-up

For scale up, the 6xHis-GST-SVP construct cloned in pOPINJ was expressed in Rosetta 2 cells and grown in auto-induction media. The protein construct was then purified using a three step purification comprising of Ni-NTA affinity, GST affinity and size exclusion chromatography.

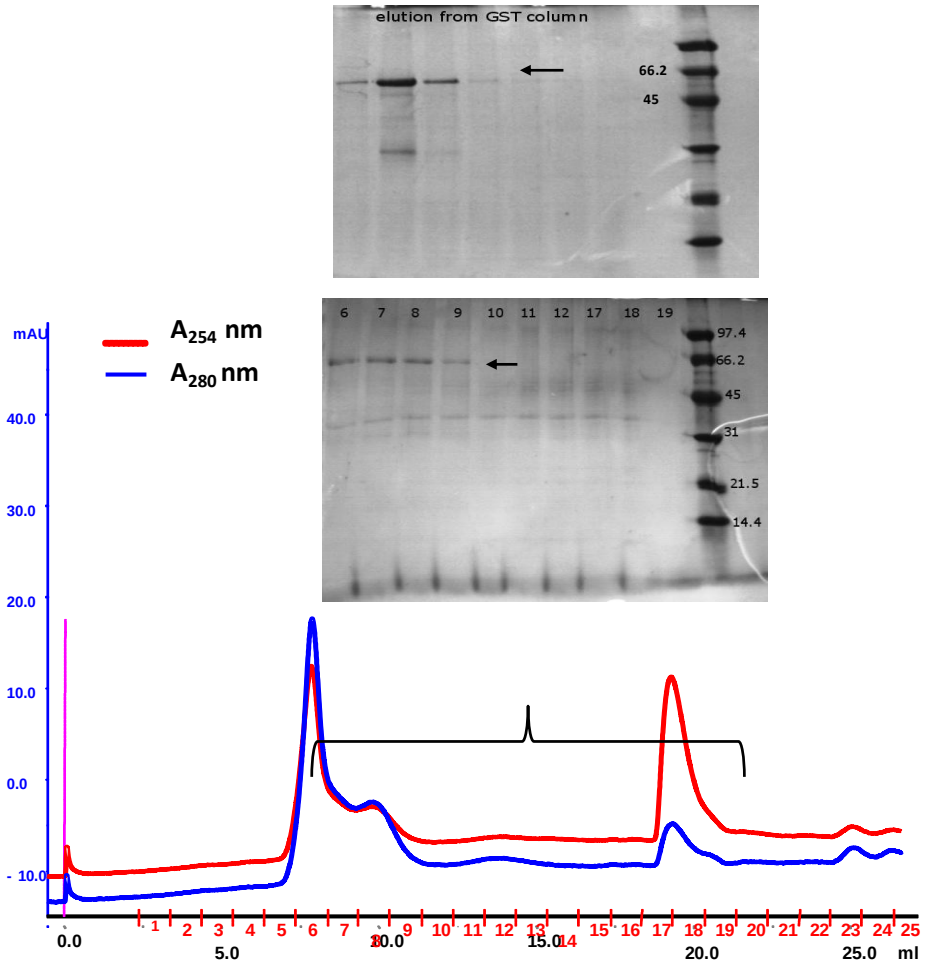


Figure IV-7: ***(top) SDS-PAGE gel showing elution from GST column.*** *The first three lanes were pooled in and applied to Superdex S 200 10/300 GL column.* ***(bottom) Gel filtration chromatography and SDS-PAGE analysis.*** *The A_{280} shows protein elution in void volume of column. The labelled samples were analyzed using SDS-PAGE. The back arrow points to the protein bands.*

As seen in the Figure IV-7, the 6xHis-GST-SVP is eluted approximately at 8 ml, which is around the void volume of the gel filtration column Superdex 200 10/300 (GE healthcare). This suggests the presence of soluble aggregates. Absence of proper folding, post-translational modifications or specific chaperones can often lead to aggregation of eukaryotic

proteins when expressed in bacterial systems. Thus, large-scale expression tests were performed using an insect cell expression system to circumvent these difficulties.

IV.3.1.3 Insect cell expression system

The baculovirus expression system is used to produce eukaryotic proteins by utilizing recombinant baculovirus carrying heterologous genes to infect insect cells. Since its first use in 1983, it has been used extensively for recombinant protein expression (Smith et al., 1983). The system involves cloning of the target gene in a donor vector, shuttling the gene into a bacmid using site-specific recombination and transfection of bacmids in insect cells to produce recombinant viruses. The viruses are amplified and used to infect insect cells at different MOIs (Multiplicity Of Infection) to screen for increased protein yield. An overview of the technique undertaken at OPPF and in the Eukaryotic Expression Facility (EEF) at EMBL, Grenoble (http://www.embl.fr/multibac/multiexpression_technologies/eef/) is provided in the Appendix I. The insect cell expression tests shown in Figure IV-5 identified several SVP constructs that could be potentially scaled up to eventually obtain pure non-aggregated protein.

Following the cloning and primary virus generation for various MADS TFs at OPPF, expression scale-up was performed at the EEF in EMBL, Grenoble. The 1st generation viruses (P1) obtained from OPPF for 6xHis-SVP-GST and 6xHis-SUMO-SVP were used to infect 50-100 ml of Sf21 cells and harvested 48-72 h post proliferation arrest. The constructs were then purified using Ni-NTA affinity and gel filtration chromatography using Superdex S200 (10/300) GL (GE Healthcare). Figure IV-8 shows the gel filtration chromatogram for 6xHis-GST-SVP, which elutes at around 10 ml as observed by the A_{280} trace. The SDS-PAGE confirms the purity and the molecular weight of the construct eluting from the column.

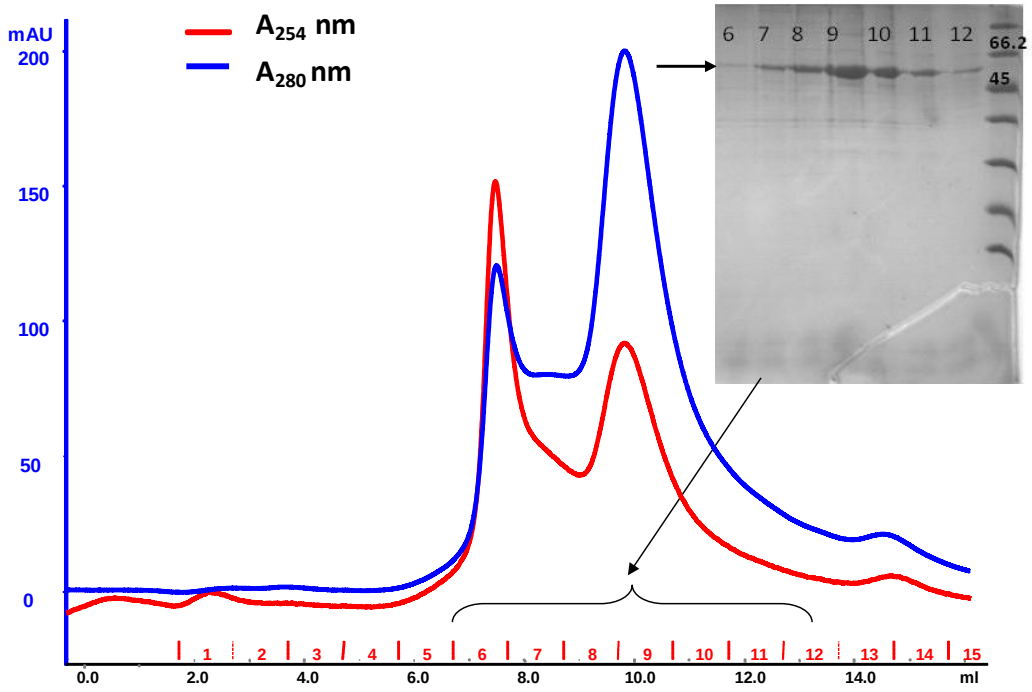


Figure IV-8: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-GST-SVP.*** *The construct was expressed in insect cells and purified with Superdex S200 10/300 GL column. The marked samples were analyzed using SDS-PAGE. The bands representing protein are marked by arrow.*

As the protein elution in the Superdex S200 10/300 GL (GE Healthcare) column was observed near to the void volume of the column (~7 ml), the resulting sample was analyzed using the Superose 6 PC 3.2/30 (GE Healthcare) analytical column (Figure IV-9). The analytical purification confirmed the absence of aggregates. A band in subsequent SDS-PAGE gel at ~27 kDa suggests a possible degradation product likely corresponding to 6xHis-GST which has a theoretical molecular weight of ~ 27.9 kDa. However, as the unstructured C-terminal of SVP protein is exposed, a possible degradation of protein through the C-terminus cannot be ruled out.

Cleavage of 6xHis-GST tag was carried out overnight at 4°C using 3C protease. Subsequent SDS-PAGE analysis (Figure IV-10) suggests successful tag cleavage and this result was taken as confirmatory step for protein ientification. However, the native protein SVP seems to be running as a slightly higher band (~30 kDa) as compared to its theoretical molecular weight of 26.7 kDa.

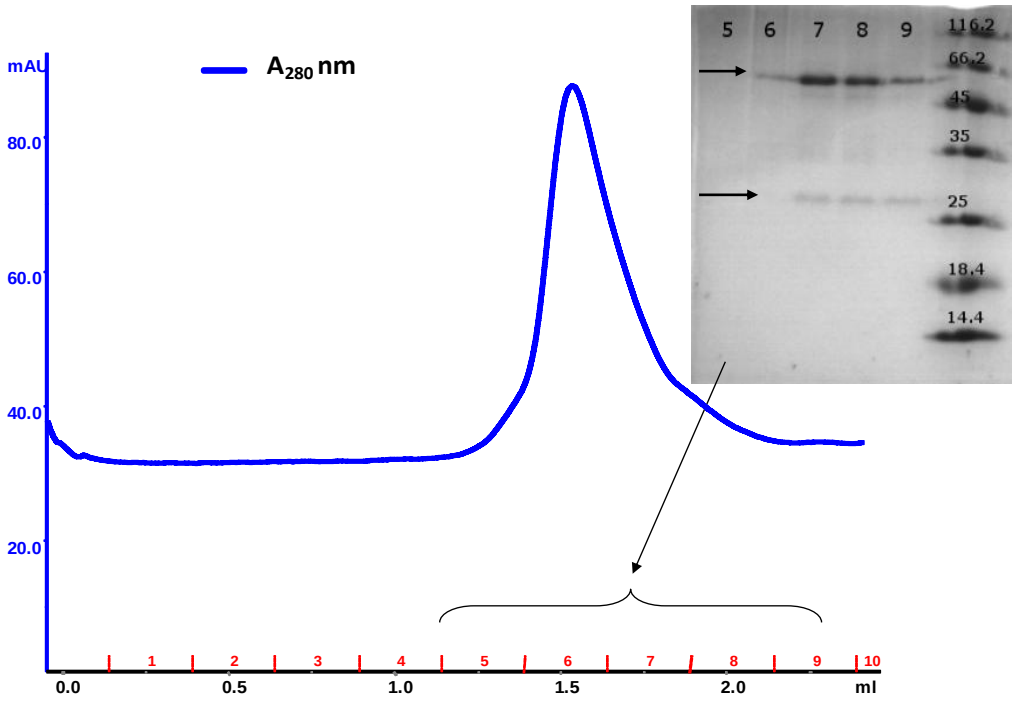


Figure IV-9: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-GST-SVP.*** *The construct was expressed in insect cells and applied to a Superose 6 PC 3.2/ 30 PC column. The marked samples are analyzed using SDS-PAGE. The upper bands represent intact proteins while the lower bands suggest degradation of protein construct.*

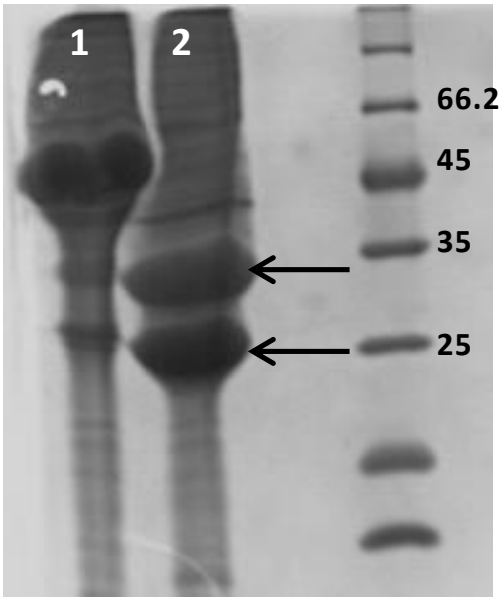


Figure IV-10: ***SDS-PAGE analysis for tag cleavage for 6xHis-GST-SVP.*** *The 6xHis-SVP-GST construct is observed in lane 1 and the cleaved protein in lane 2. The top arrow in lane 2 indicated native SVP while the bottom arrow probably represents 6xHis-GST.*

A similar purification protocol was followed for 6xHis-SUMO-SVP. However, size exclusion purification and western blot analysis confirmed the gradual degradation of the protein as shown in Figure IV-11.

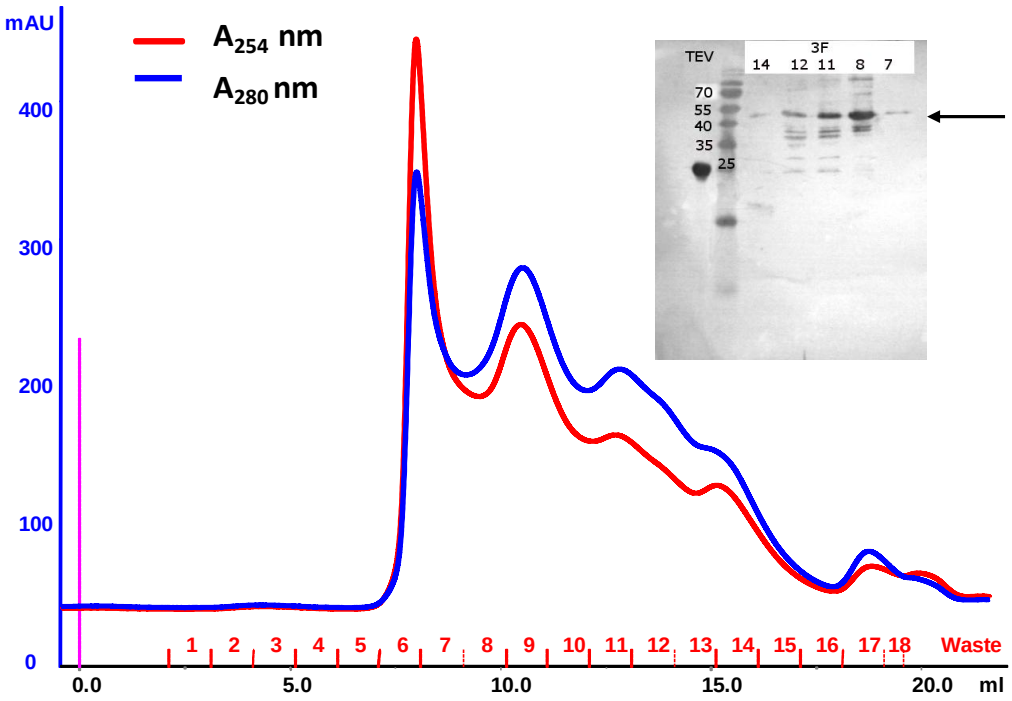


Figure IV-11: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-SUMO-SVP.*** *The construct was expressed in insect cells and purified with Superdex S200 10/300 GL (GE healthcare). A_{280} nm shows several broad peaks The western blot analysis of the selected samples suggests gradual degradation of the protein.*

The activity of the virus generated from the OPPF, Oxford, UK significantly deteriorated over time leading to lack of infection and protein expression (Figure IV-12). A plaque assay performed to measure the virus titre confirmed the loss of infectivity as no plaque formation was observed even for 10^{-1} dilution (data not shown). Due to the absence of a fluorescent marker protein, the inefficiency of virus infection was difficult to detect in the preliminary stages and could be confirmed only after harvesting and purification leading to inefficient use of time and resources. In order to cope up with the frequent inactivation of the OPPF viruses and to ease the monitoring of protein expression, genes were cloned in pAceBac1 (Berger group EMBL, Grenoble) and the virus production was repeated. The virus amplification and infection of Sf21 cells was undertaken at EEF in EMBL, Grenoble. Yellow fluorescent

protein (YFP) was used a marker to detect infection (Bieniossek et al., 2008; Trowitzsch et al., 2010).

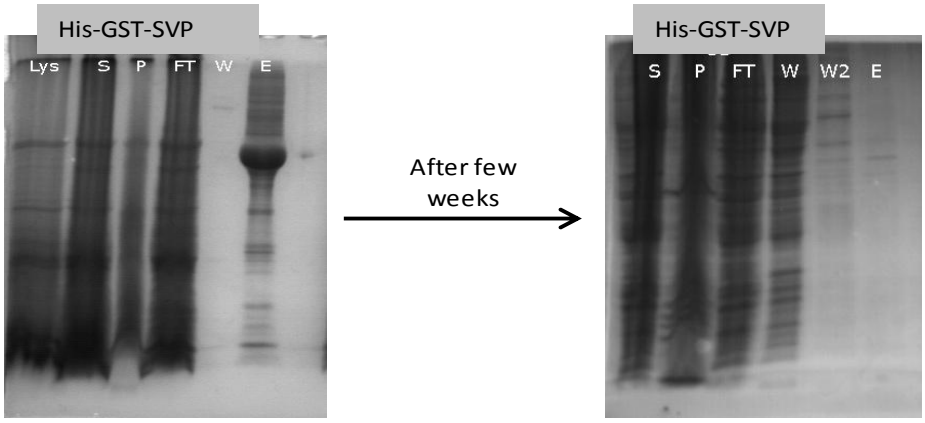


Figure IV-12 : ***SDS-PAGE gels showing variation of expression of 6xHis-SVP-GST construct over few weeks****.* ***(left)*** *The insect cell expression test shows presence of protein in elution while* ***(right) e****xpression levels are diminished with no protein observed in elution after few weeks. Lys- lysate; S-supernatant; P-pellet; FT-Flow through; W, W2- Washes; E-Elution*

Following the generation of new viruses, 6xHis-GST-SVP and 6xHis-SUMO-SVP were both expressed in ~ 200 ml of Sf21 cells using the Sf900 II medium (Gibco-Life technologies). 6xHis-SVP-GST was purified using Ni-NTA affinity and size exclusion chromatography.

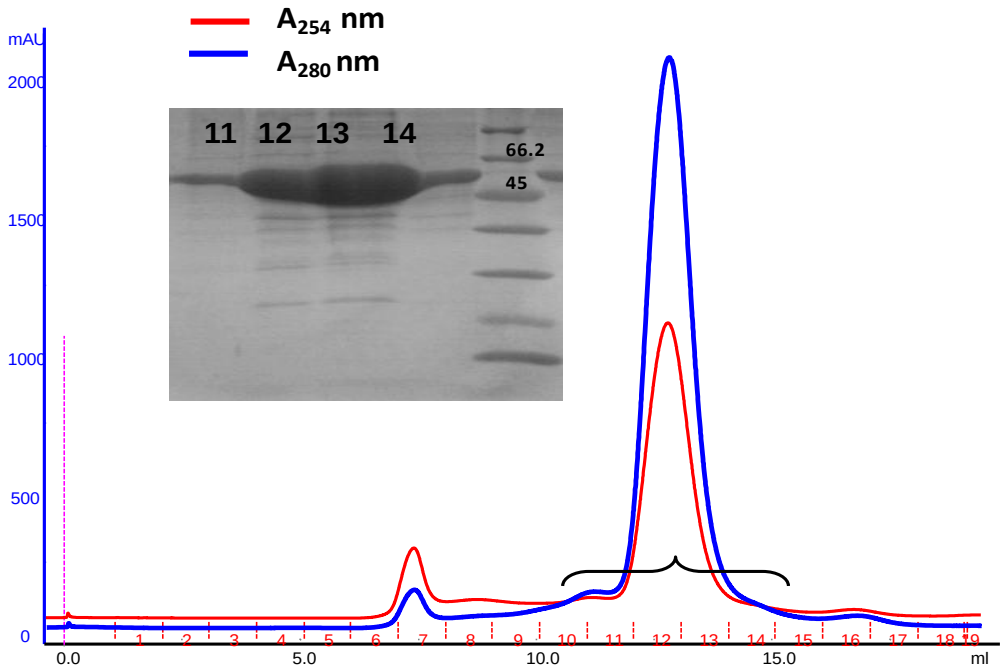


Figure IV-13: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-GST-SVP.*** *The construct was expressed in insect cell and purified with Superose 6 10/ 300 GL column****.*** *The A_{280} shows a small aggregate peak in void volume while the construct is eluted as a single peak. The SDS-PAGE analysis of marked samples shows presence of protein.*

Figure IV-13 shows a single peak obtained in size exclusion chromatography with a Superose 6 10/300 GL column (GE Healthcare). The early elution of protein as compared to its molecular weight might be due to formation of higher order oligomers. The gel electrophoresis confirms the presence of protein.

The SUMO tagged SVP construct (6xHis-SUMO-SVP) was also purified using a three-step purification comprising Ni-NTA affinity chromatography, heparin column chromatography and size exclusion chromatography. The sample from nickel purification was pooled and loaded onto a heparin column in a low salt buffer. A salt gradient was used to separate the pure protein from DNA and other contaminants. The heparin column purification shows a sharp peak and the presence of the protein of interest confirmed by the SDS-PAGE (Figure IV-14).

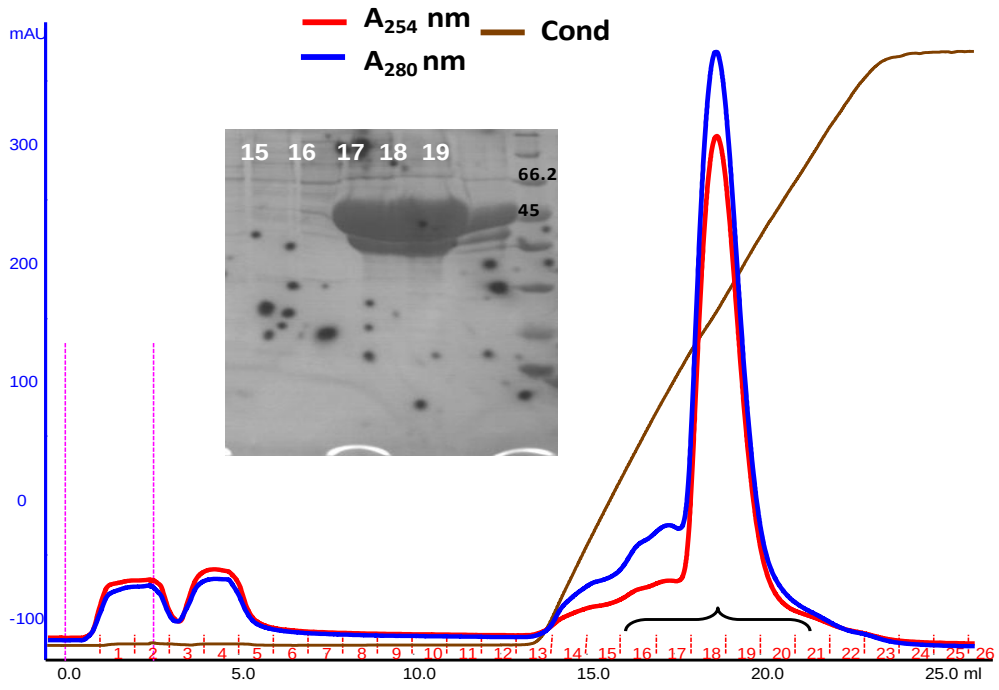


Figure IV-14: ***Heparin column chromatography and SDS-PAGE analysis for 6xHis-SUMO-SVP****. The construct was expressed in insect cells and purified using Hi-Trap Heparin HP (GE Healthcare) column. The* A_{280} *shows initial elution of contaminants followed by a single protein peak. The SDS-PAGE analysis of selected sample shows high yield of protein.*

The protein eluted from the heparin column was pooled, concentrated and purified with size exclusion chromatography. Size exclusion chromatography (Figure IV-15) is used as a final step as it separates the contaminants on the basis of molecular size. It also confirms the

absence of soluble aggregates. A small band observed in the gel electrophoresis might be truncated version of the protein.

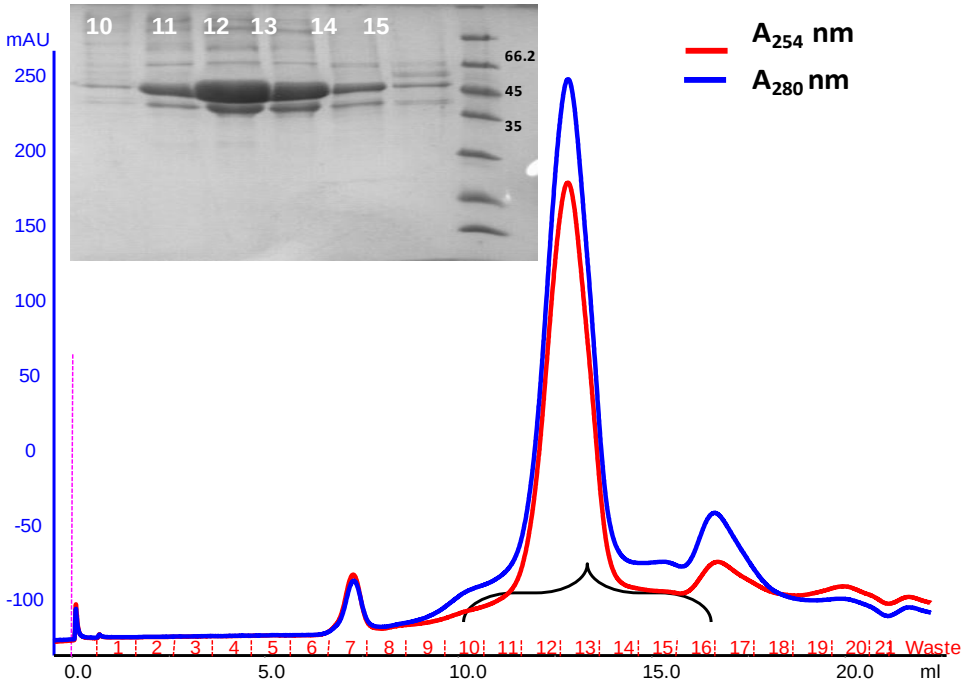


Figure IV-15: ***Size exclusion purification and SDS-PAGE analysis for 6xHis-SUMO-SVP.*** *The construct was expressed in insect cells and purified with Superose 6 10/ 300 GL (GE Healthcare) column. The A_{280} shows a sharp peak for protein elution. SDS-PAGE of the selected sampled shows strong bands representing protein. A lower band suggests presence of partial degradation product.*

The work described here marked the successful purification of two SVP constructs, 6xHis-SUMO-SVP and 6xHis-GST-SVP, establishing a high yielding expression and purification protocols using an insect cell expression system. However, in order to avoid any potential deleterious effect of the solubility tags in structural characterization of protein, they were subsequently removed. Native SVP protein was obtained by cleaving the 6xHis-GST tag with 3C protease and tag removal using GST affinity chromatography. Heparin column (Figure IV-16) and size exclusion chromatography (Figure IV-17) were then performed to remove any remaining contaminants. Figure IV-16 shows small amounts of contaminants being eluted at lower salt concentration while SVP elutes as a single peak.

Figure IV-17 shows a clear separation of soluble aggregates eluted in the void volume, pure protein eluted in fractions 11-14 and contaminants eluted in fractions 15-17. Subsequent SDS-PAGE analysis confirmed the successful separation of native SVP and contaminating 6xHis-GST tag. The fractions corresponding to the pure protein were pooled and concentrated.

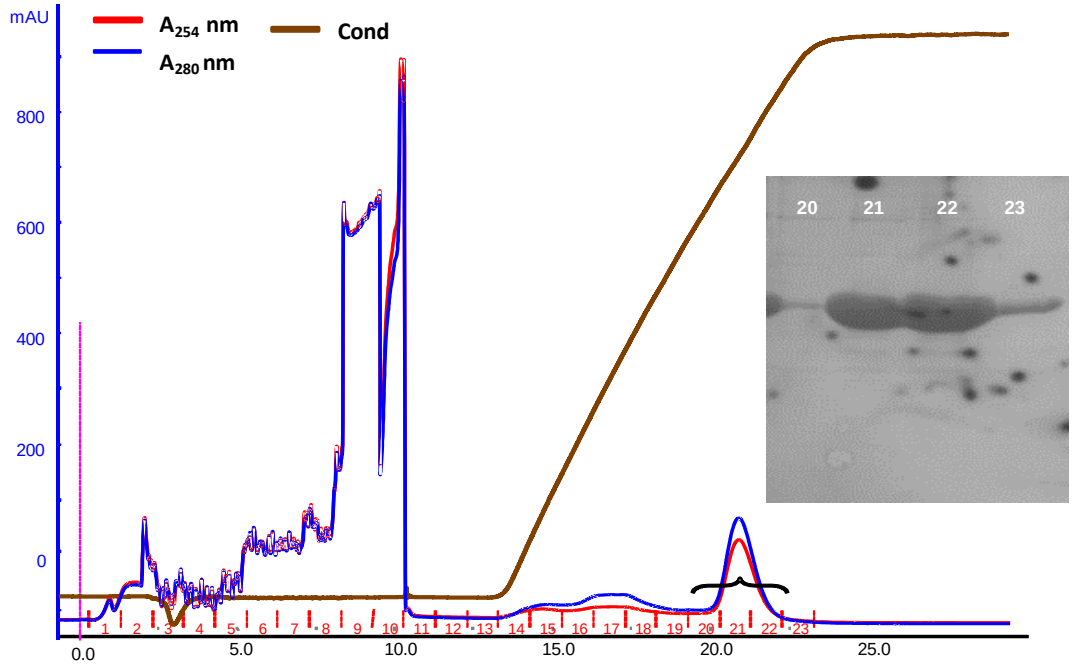


Figure IV-16: ***Heparin column chromatography for native SVP construct.*** *The GST affinity purified construct was applied to Hi-Trap Heparin HP column to separate contaminants from native SVP. The marked fractions were analyzed with SDS-PAGE. A high yield of pure protein is observed on the SDS-PAGE gels.*

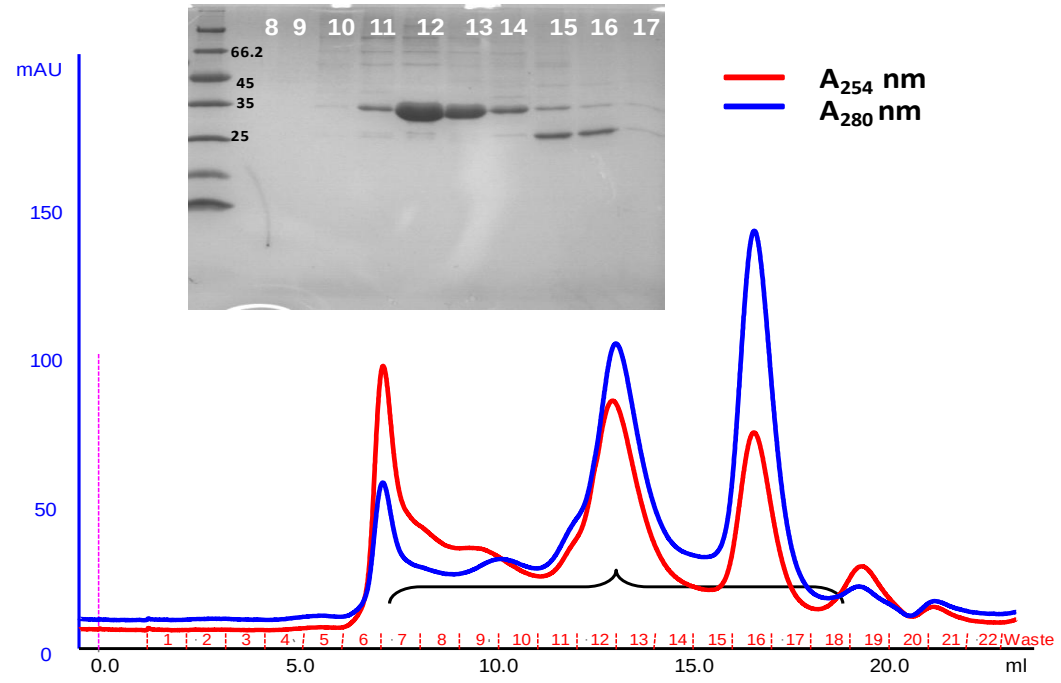


Figure IV-17: ***Size exclusion profile for native SVP construct.*** *The heparin purified protein construct from Figure IV-16 was applied to a Superose 6 10/ 300 GL column. The A_{280} shows three peaks which were analyzed using SDS-PAGE. The SDS-PAGE of the selected samples shows separation of pure protein and contaminating 6xHis-GST tag.*

To summarize, a parallel strategy was used to screen and optimize the recombinant expression and purification of SVP. After screening about 200 expression conditions, a protocol for obtaining a high yield of soluble full-length soluble protein was established. This is the first time that a full length MADS TF was purified in high yield from recombinant sources. SVP is one of the main proteins involved in flowering response to temperature changes. In order to reveal the mechanism of SVP, it is important to understand its structure and oligomerisation properties. The success in purifying full length SVP provided a unique opportunity to study the protein *in vitro*. In the following sections a series of biochemical and biophysical techniques such as SAXS, AFM, MST and EMSA are described that were used to gain an overview of the SVP structure, its oligomerisation state and DNA binding characteristics.

IV.3.2 Biophysical characterisation and DNA binding studies of SVP$^{(1-240)}$

The characterisation of SVP included the use of complementary methods such as SEC-MALLS to determine the oligomerisation state of the protein and low-resolution structural techniques such as SAXS to provide information as to the conformation of the protein in solution. AFM and gel shift assays were also performed to study the interaction of SVP with DNA.

IV.3.2.1 SEC-MALLS (Size Exclusion Chromatography- Multi Angle Laser Light Scattering)

Purified native SVP, 6xHis-SUMO-SVP and 6xHis-GST-SVP were analyzed by SEC-MALLS to determine its molecular mass in solution and thus derive its oligomeric state. SEC-MALLS was performed in collaboration with Mr. Roger Miras from the group of Biology of Metals, Biosciences and Biotechnology Institute of Grenoble, CEA. SEC-MALLS determines the molecular mass independent of the shape, flexibility and conformation of the protein (Wyatt, 1993; Folta-Stogniew, 2006; Folta-Stogniew and Williams, 1999; Wen et al., 1996). Approximately 2.5 mg/ml of native SVP, 12.5 mg/ml of 6xHis-SUMO-SVP and 4.5 mg/ml of 6xHis-GST-SVP were used for analysis. The estimated molecular weight of the constructs can be used to determine their oligomeric state. As observed in Figure IV-18 and Table IV-4 all the SVP constructs studied were predicted to be dimers.

Protein	MW from SEC-MALLS	Theoretical MW (monomer kDa)	Oligomeric state
SVP native	59	26.7	59/26.9=2.19 (dimer)
His-SUMO-SVP	75	39.5	75/39.5=1.89 (dimer)
His-GST-SVP	104	54.2	104/54.2=1.91 (dimer)

*Table IV-4: **Analysis of SEC-MALLS results.** The molecular weight estimates from SEC-MALLS (Figure IV-18) were used to determine the oligomeric state of SVP protein constructs. All three SVP constructs were predicted to be dimers.*

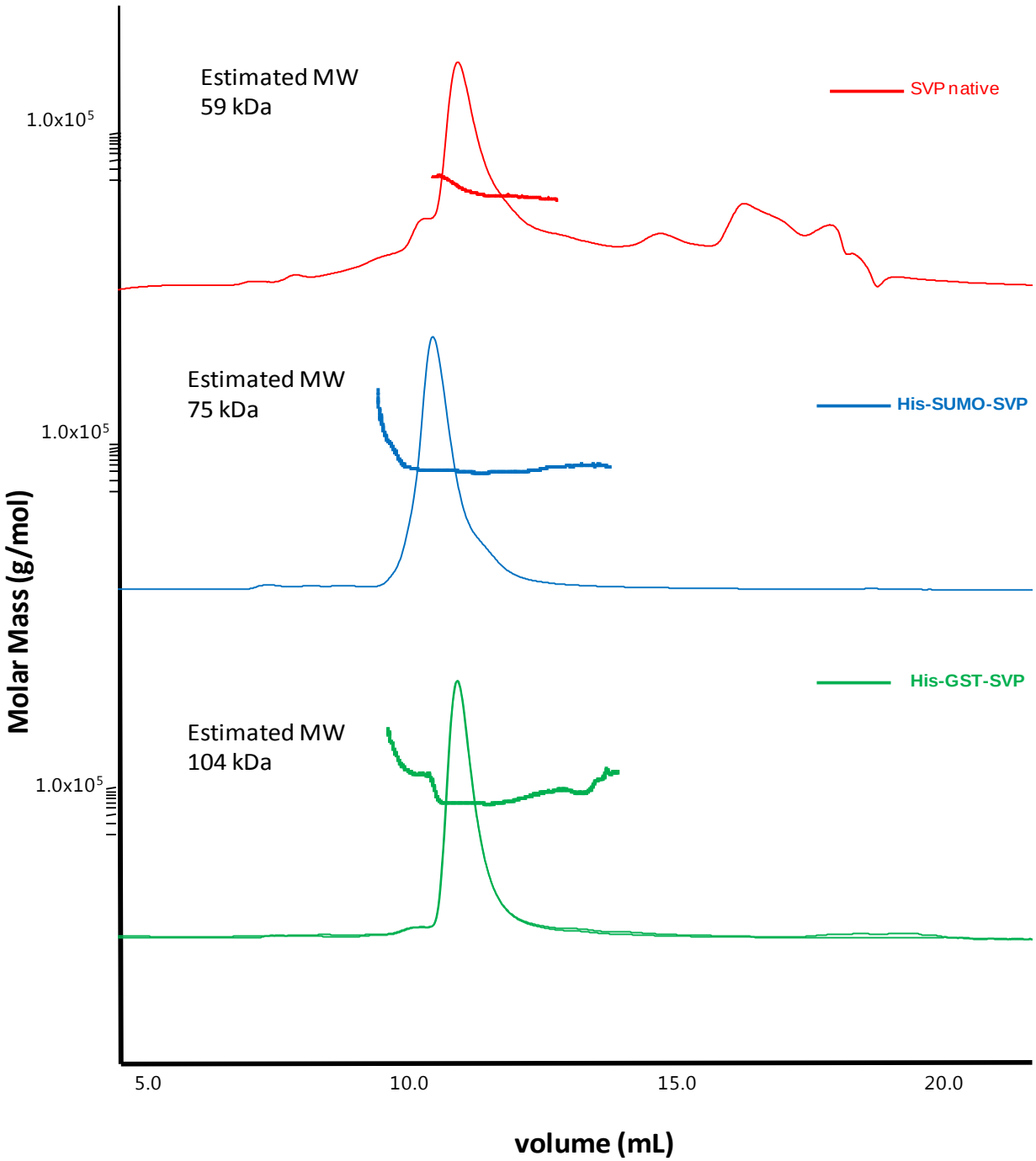


Figure IV-18: ***SEC MALLS for SVP constructs.*** *The SVP native, 6xHis-SUMO-SVP and 6xHis-GST-SVP were analyzed using SEC-MALLS. Single peak suggests presence of single species.*

IV.3.2.2 Crystallisation trials

A high-resolution crystal structure can provide information about the atomic level determinants of protein complex formation. The secondary structure prediction as shown in (Figure IV-1), suggests SVP contains a long unstructured C–terminal region making the protein a difficult target for crystallisation. The purified constructs however were found to be stable even after freeze thaw and retained their stability even after the tag was cleaved. High throughput crystallisation trials were therefore performed at the EMBL-Grenoble HTX facility using crystallisation robot (Cartesian PixSys 4200) at 20°C. The crystallisation trials yielded a few crystalline precipitates for 6xHis-SUMO-SVP (Figure IV-19). However no diffraction quality crystals were obtained. The disorder in the C-terminal region as predicted by IUPRED is a possible reason for the low propensity of crystallisation for SVP (Figure IV-20)

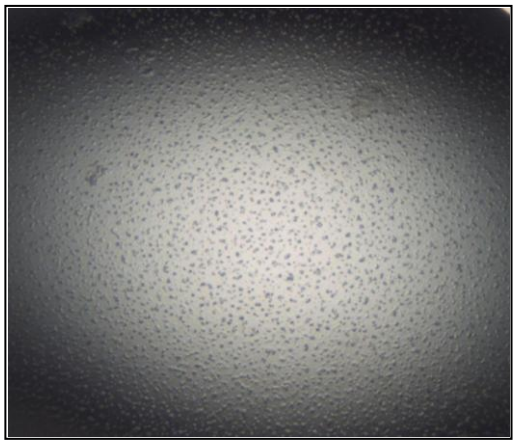


Figure IV-19: ***Crystallization trials for 6xHis-SUMO-SVP.*** *A high throughout crystallization screen was performed with 4 mg/ml of 6xHis-SUMO-SVP. Small crystalline precipitates were observed in condition consisting of 10% w/v PEG 1000.*

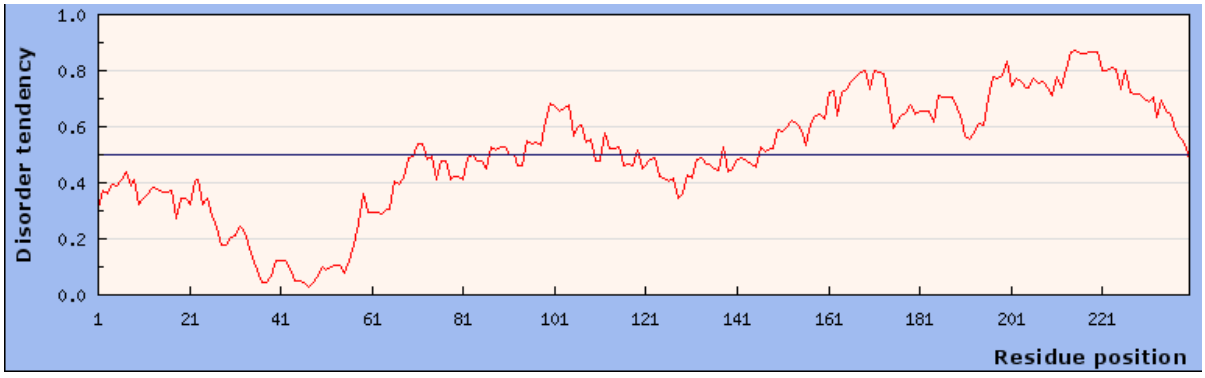


Figure IV-20: ***Disorder prediction for SVP using IUPRED*** (Dosztányi et al., 2005)***.*** *A high degree of disorder is observed in the C-terminal region of protein* (Ward et al., 2004).

IV.3.2.3 Small Angle X-ray Scattering

As crystallisation trials yielded no crystals suitable for X-ray diffraction studies, SAXS experiments were performed to study the solution structures of the SVP constructs obtained.

SAXS was used to study the solution conformation of native SVP and of the 6xHis-GST-SVP and 6xHis-SUMO-SVP protein constructs. SAXS data collection was performed in tandem with online HPLC system at ESRF beamline BM 29 (http://www.esrf.eu/home/UsersAndScience/Experiments/Beamlines/content/content/bm29.html) to ensure sample purity and homogeneity (Pernot et al., 2013; Round et al., 2013; Pernot et al., 2010). An analytical Superose 6 3.2/30 PC column (GE Healthcare) was used for online sample purification directly upstream of the SAXS data collection environment.

As SAXS is sensitive to change in sample environment and buffer conditions, the column was pre-equilibrated with the sample buffer. Use of the online-HPC system ensured the presence of a single species in solution and the use of a continuous flow through the capillary helped to reduce the effect of radiation damage (Pernot et al., 2010; Martel et al., 2012). The A_{280} chromatogram from the on-line HPLC purification for SVP, 6xHis-SUMO-SVP and 6xHis-GST-SVP tagged proteins is shown in Figure IV-21 and shows the presence of a single species. In the case of native SVP, a partial second peak was observed. This was due to technical issues during data collection- specifically a beam dump and restart of the HPLC system. However, as the data corresponding to the first peak was already collected, the presence of the second peak did not affect our measurements.

Initial data processing was performed automatically using the EDNA pipeline (Incardona et al., 2009). The R_g and I(0) derived during the experiment for the three protein constructs are shown in Figure IV-22 while Figure IV-23 shows a representative screen shot for the HPLC trace plot and calculated R_g from ISPyB (De Maria Antolinos et al., 2015). The regions of the HPLC peak with constant R_g were selected for further analysis. The intensity profile for a single frame as obtained from ISPyB is shown in Figure IV-23.

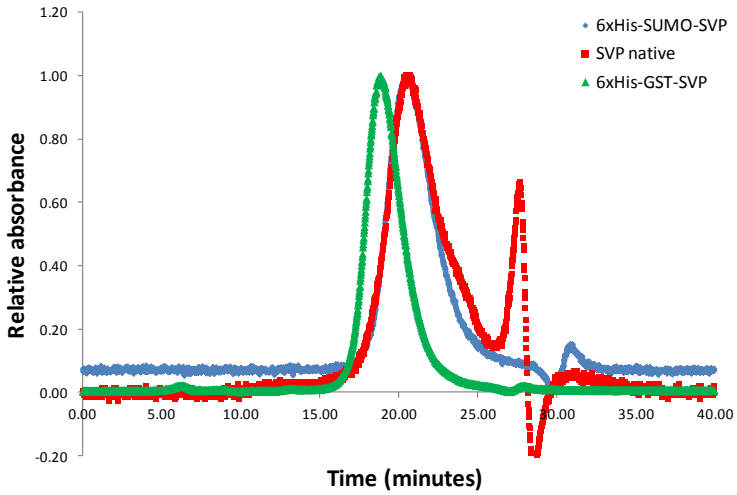


Figure IV-21: HPLC trace for three SVP constructs as obtained on Superose 6 3.2/ 30 PC during SAXS data collection. The three constructs are represented with different colours; red for SVP native, blue for 6xHis-SUMO-SVP and green for 6xHis-GST-SVP.

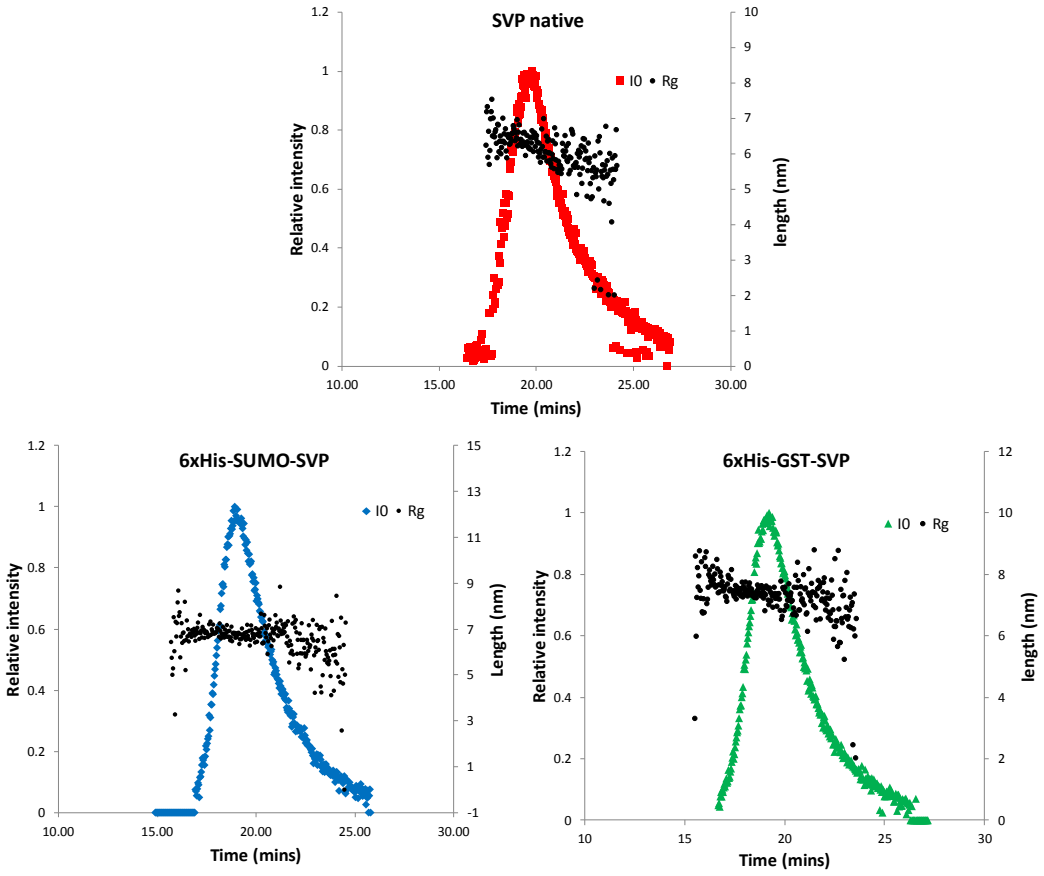


Figure IV-22: ***Rg and I0 trace for SVP constructs.*** *The three constructs SVP native (red), 6xHis-SUMO-SVP (blue) and 6xHis-GST-SVP show a single peak for the normalized scattering intensity (I0). The radius of gyration (Rg) is plotted on the secondary axis. The frames within a constant Rg were selected for further analysis..*

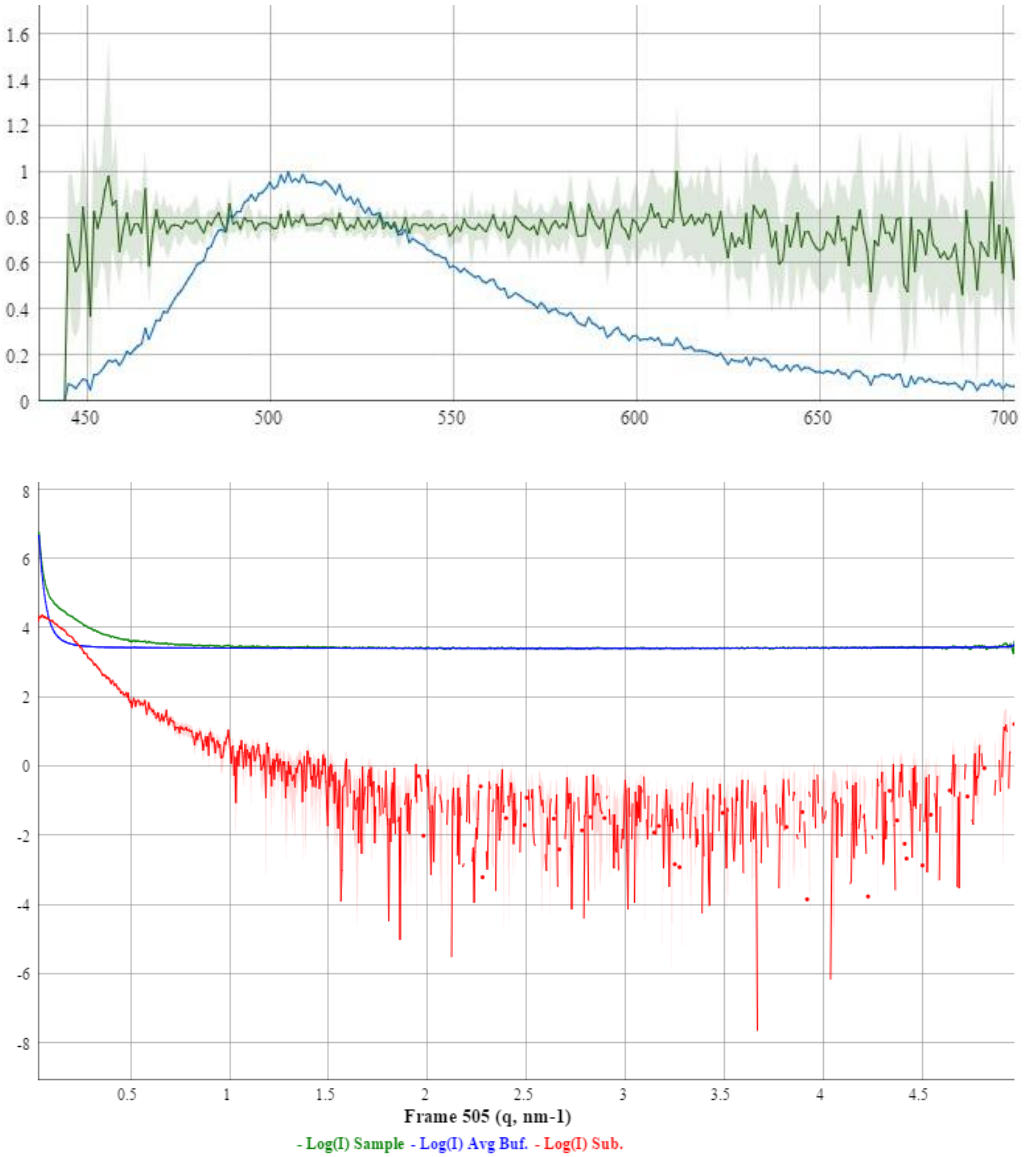


Figure IV-23: ***Figure (Top) Screen shot of HPLC A_{280} and calculated R_g for 6xHis-SUMO- SVP as extracted from ISPyB.*** *The green line represents R_g.;* ***(bottom)*** *intensity profile for frame 505 as extracted from ISPyB. Similar profiles were obtained for native and GST tagged protein. The green line corresponds to scattering from the sample, the blue line is buffer and the red line is subtraction of buffer from sample*

Model independent analysis was performed using tools from the ATSAS package (Petoukhov et al., 2012) and ScÅtter (Förster et al., 2010) from the BIOISIS package. The invariant parameters R_g (Guinier analysis) and D_{max} (GNOM) were calculated and the globularity flexibility (Kratky plot) assessed during this step. The Guinier analysis and scattering intensity profile also help to detect presence of aggregation and inter particle forces. Subsequently, model-fitting using *ab initio* programs were performed. The scattering curves obtained for the native SVP and both 6xHis-SUMO and 6xHis-GST tagged constructs are shown in the Figure IV-24.

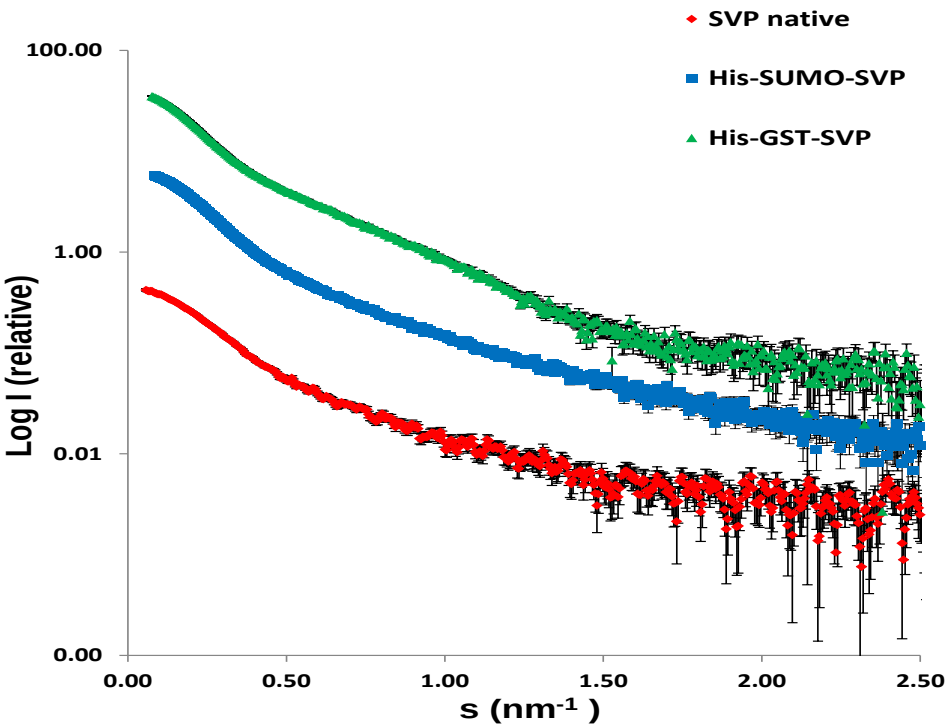


Figure IV-24: ***SAXS scattering curves for SVP constructs.*** *The plot displays relative Log (I) as a function of s for SVP native (red dots), 6xHis-SUMO-SVP (blue squares) and 6xHis-GST-SVP (green triangles).*

SAXS is very sensitive to interparticle effects. These interactions are evident at very low angles (Guinier, 1938; Chen and Bendedouch, 1986) and cause a characteristic "upturn" or "downturn" deviating from the ideal Guinier approximation which is indicative of aggregation or repulsion, respectively. The analysis of this Guinier region as observed in Figure IV-25, shows the absence of interparticle effects. The Guinier plot shown here represents values $s_{max}*R_g<1.3$. The radii of gyration obtained from the Guinier plots are 6.46 ± 0.17 nm, 6.86 ± 0.05 nm and 7.84 ± 0.09 nm for native SVP, 6xHis-SUMO-SVP and 6xHis-GST-SVP tagged constructs, respectively.

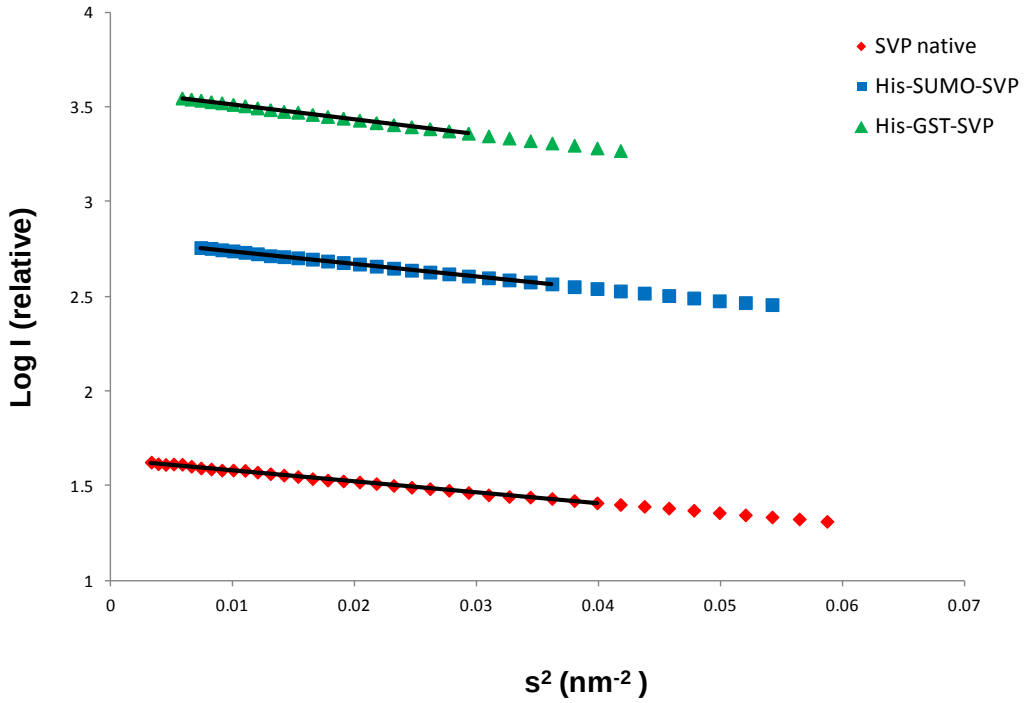


Figure IV-25: ***Guinier plots for SVP constructs obtained from SAXS data.*** *The plot shows relative Log(I) as a function of* s^2 *for SVP native (red), 6xHis-SUMO-SVP (blue) and 6xHis-GST-SVP (green). The upper range of the linear fit was limited by* s_{max}*Rg<1.3 and is indicated by a black line. The linear Guinier plot indicates absence of aggregation and inter particle effects.*

The molecular weight, volume, pair distribution function P(r) and maximum particle dimension, $D_{max,}$ values were estimated using GNOM (Svergun, 1992) from the one dimensional scattering curves. These parameters provide valuable information regarding the shape of the protein. As the distance distribution plot considers the whole scattering profile and not just the Guinier region, it is considered a more reliable set of parameters for the determination of particle shape. The distance distribution function plots for all the three constructs are shown in the Figure IV-26. They all possess a slightly elongated tails thus pointing towards an elongated shape. This elongated tail typically represents multi domain proteins with flexible linkers (Putnam 2007). The presence of two peaks for the GST tagged protein indicates the presence of multiple domains. Considering that the GST proteins possess a tendency to dimerize, a multi domain structure with a central core SVP dimer followed by an N-terminal GST dimer is likely. However, the P(r) plot suggests that this is unlikely to be the case for 6xHis-SUMO-SVP, as would be predicted since the N-terminal 6xHis-SUMO tag is likely to orient randomly with respect to a central SVP dimer.

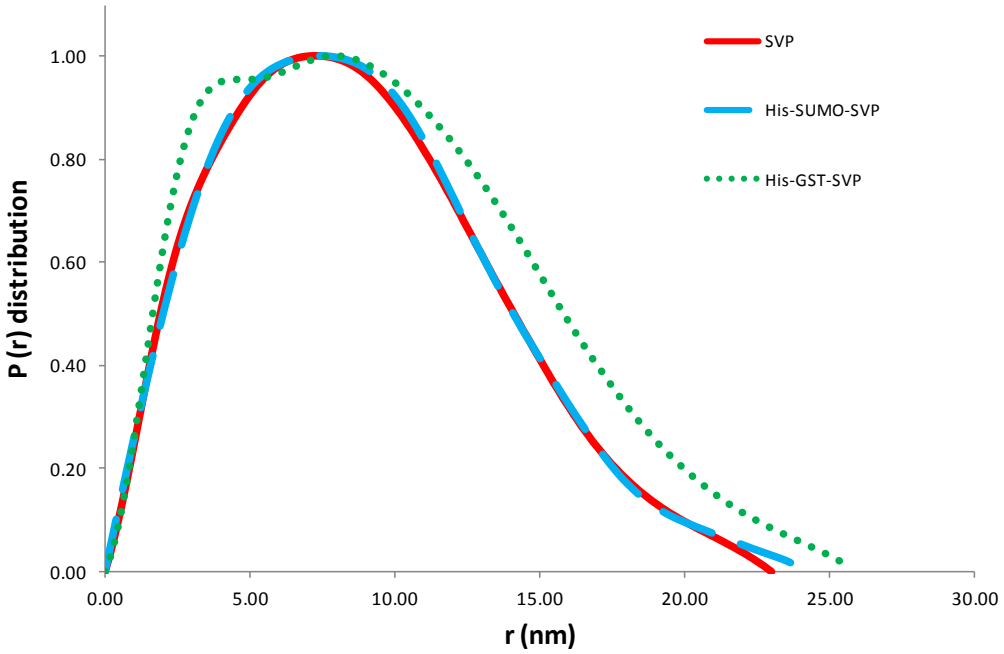


Figure IV-26: ***Normalized distance distribution function for SVP constructs.*** *The plot shows P(r) distribution for native SVP (red), Hi-SUMO-SVP (blue) and 6xHis-GST-SVP (green). The plots have been normalized for better visualization. All three profiles show an elongated tail conformation. Two peaks are observed for 6xHis-GST-SVP.*

Protein	Theoretical mol wt. (monomer) k Da	Radius of gyration (Rg) (nm)	D max (nm)
SVP	26.9	6.46 ± 0.17	22.19
His-SUMO-SVP	40	6.86 ± 0.05	23.17
His-GST-SVP	54.83	7.48 ± 0.09	25.58

Table IV-5: ***Parameters obtained from SAXS scattering for SVP protein constructs****. The R_g was obtained from Guinier plot while D_{max} was obtained from distance distribution function from GNOM. The R_g and D_{max} are observed to correlate with the change in theoretical molecular weight of the protein constructs.*

Kratky analysis qualitatively assesses the compactness of the molecule and enables differentiation between folded and unfolded proteins (Doniach, 2001; Kikhney and Svergun, 2015; Uversky and Dunker, 2010; Pérez et al., 2001; Bernadó and Blackledge, 2009). The normalized Kratky plots for globular proteins exhibit a peak maxima of 1.104 at $s*R_g = \sqrt{3}$ while for a random chain the curve gradually rises to form a plateau at around 2 (Durand et al., 2010; Receveur-Brechot and Durand, 2012). The normalized Kratky plot as shown in the Figure IV-27 indicates high flexibility for all three SVP constructs. Due to relatively low concentration for SVP native, the data exhibits more noise than for the other constructs.

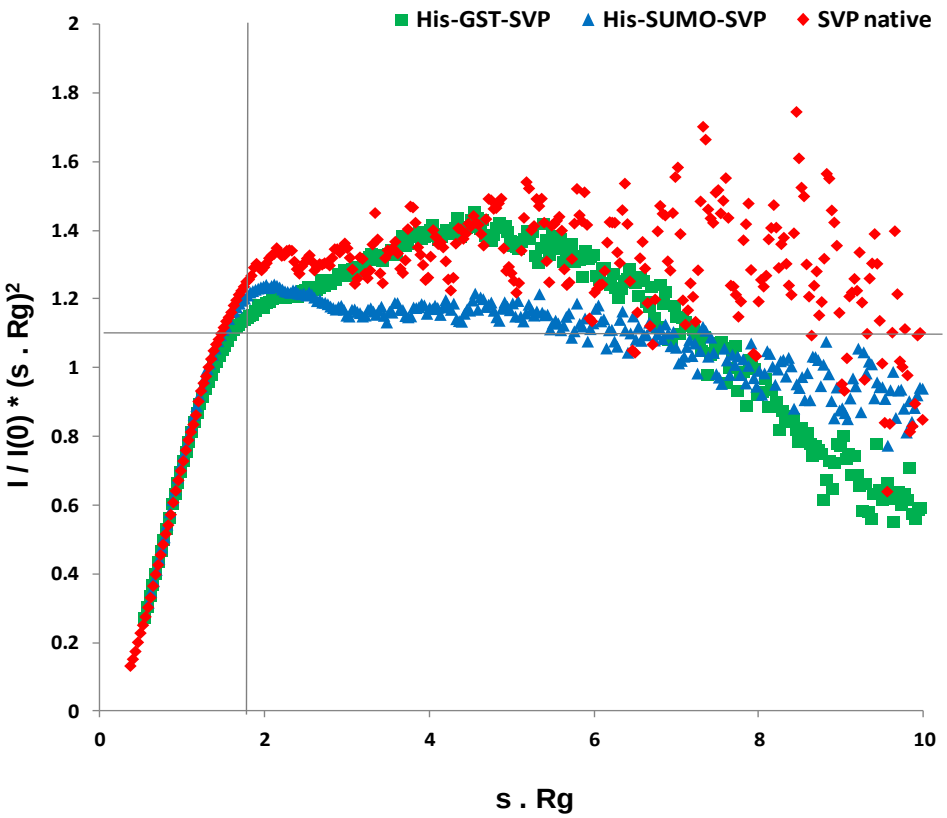


Figure IV-27: ***Dimensionless Kratky plot for native SVP (red), 6xHis-SUMO-SVP (blue) and 6xHis-GST-SVP (green).*** *The crosslines mark the position for a well folded monodomain protein. All three SVP constructs show a shift towards the right. This is indicative of a high degree of flexibility or a partial unfolding of the protein.*

The result of the SEC-MALLS experiments, shown in Figure IV-18, suggests that SVP constructs are dimers in solution even at relatively high protein concentrations. Structural studies of the DNA binding M domains of mammalian proteins MEF2A and SRF (Pellegrini et al., 1995; Wu et al., 2010) and yeast protein MCM1 (Tan and Richmond, 1998), shows an

obligate dimerisation for DNA binding. Based on homology of MADS TFs for the M domains and the K domains we can postulate a structural organisation for SVP with the M domain and K domain forming a dimer. The C-terminal region being unstructured possess a freedom of movement in solution (Figure IV-28).

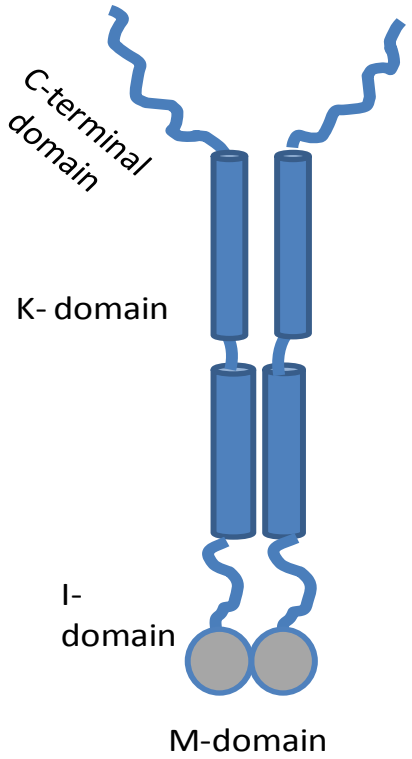


Figure IV-28: ***Postulated structural conformation for SVP.*** *The highly conserved M domain can be postulated to be dimeric similar to its mammalian and yeast homologues. The K domain being the main oligomerisation interface also dimerizes. The conformation of I domain cannot be predicted due to absence of structural information. The C-terminal domain is unstructured and can have random conformation.*

The second part of SAXS analysis usually consists of building a three dimensional model using the one-dimensional scattering data. The two approaches generally used are *ab initio* and rigid body modelling. However, as the Kratky plot suggested internal flexibility of the proteins, the rigid body modelling approach could not be successfully implemented. An attempt was made to utilize the ensemble approaches used for flexible proteins. However, Flexible Meccano (Ozenne et al., 2012) is not optimized for the arrangement of the putative SVP protein dimer as it was not possible within the program constraints to represent the configuration of SVP to allow minimisation. Currently, Ensemble Optimisation method (EOM) (Bernado et al 2007) for generating an ensemble of thousands of models for SVP is being investigated to address these issues, but is outside the scope of the thesis studies presented here.

IV.3.2.3.1 Ab-initio Model fitting

Ab-initio molecular envelopes of the solution structure of all three SVP constructs were generated using DAMMIF (Franke and Svergun, 2009). 50 models were generated without any symmetry restraints. The resultant models were averaged and filtered using DAMAVER (Volkov and Svergun, 2003). Superimposition of the averaged model generated by DAMAVER and the filtered model obtained from DAMFILT (Volkov and Svergun, 2003) is shown in Figure IV-29, Figure IV-30 and Figure IV-31. These models represent an average of all the conformations that exist in the solution. Three orthogonal views of the most representative bead models are also presented in the figures. The DAMMIF generated bead models are constructed such that the scattering pattern calculated from the model fits well with the experimental data.

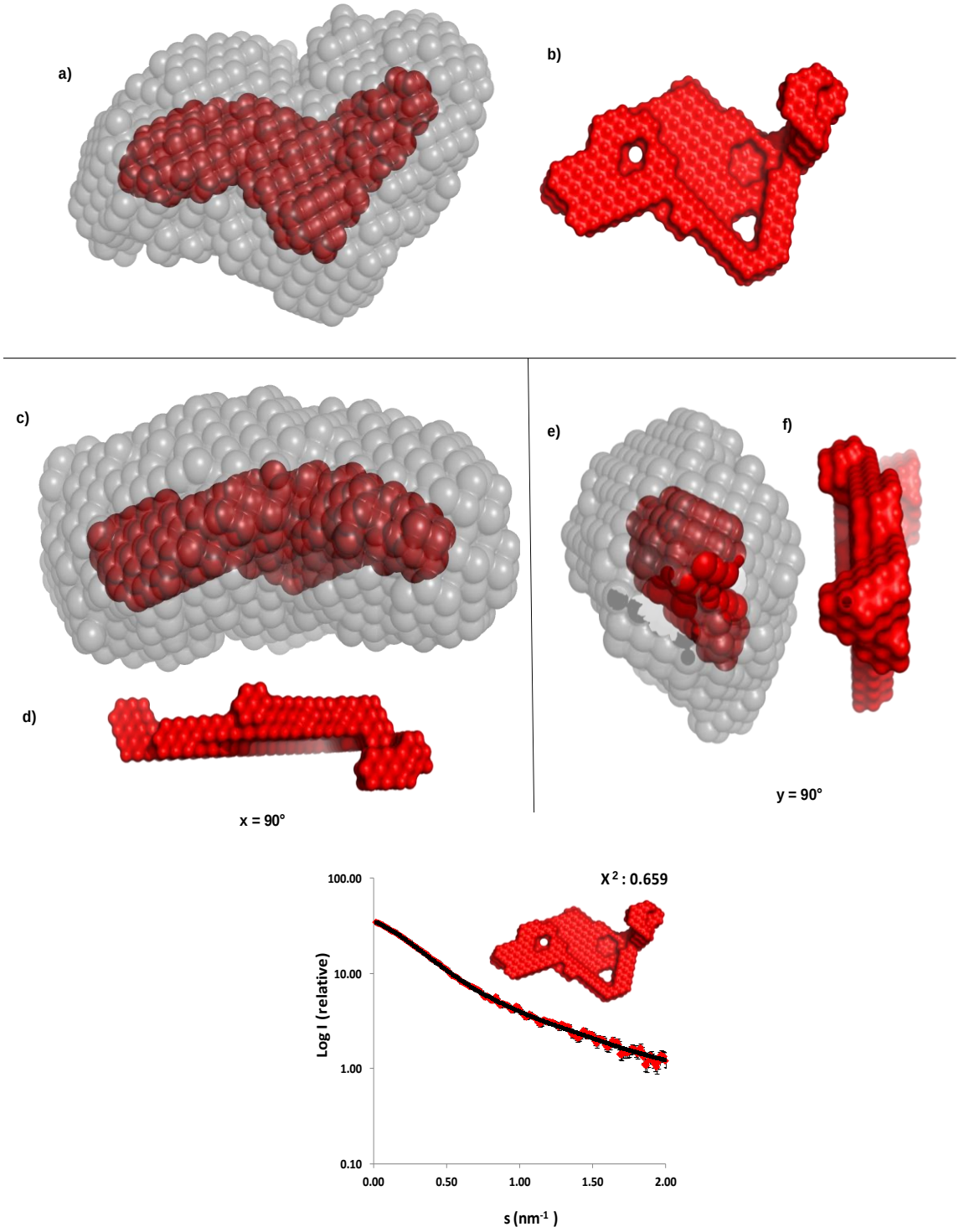


Figure IV-29: ***Ab-initio model for solution structure of native SVP.*** *(**top**)a) Overlay of average and filtered ab initio model from DAMFILT and that from DAMAVER for native SVP;* ***c,e)*** *rotation at x = 90° and y =90°;* ***b)*** *Most representative model obtained from DAMMIF****. d,f)*** *rotation at x = 90° and y =90°;* ***(bottom)*** *Comparison of theoretical and experimental scattering curves for most representative model.*

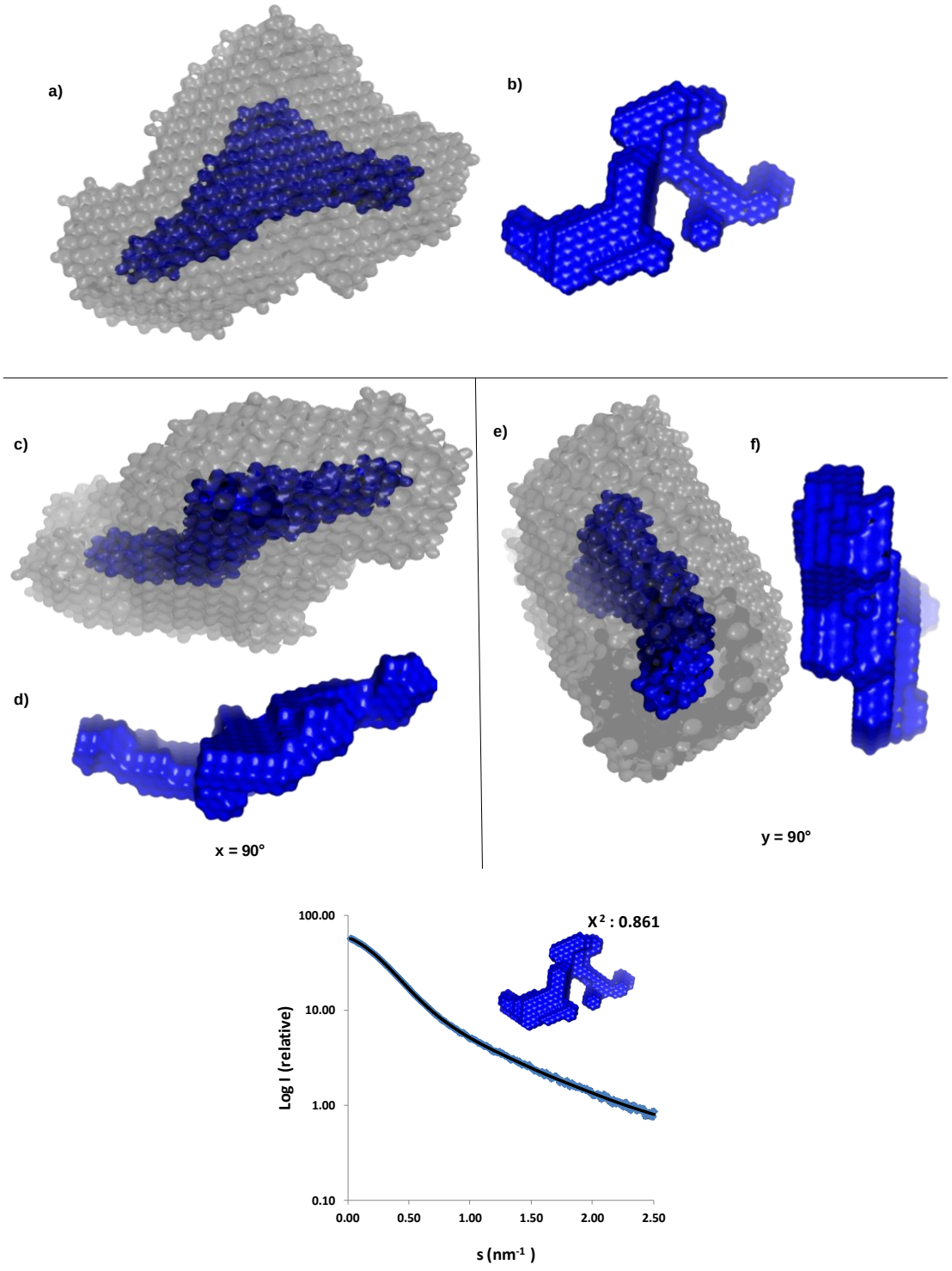


Figure IV-30: ***Ab-initio model for solution structure of 6xHis-SUMO-SVP construct. (top)a)****Overlay of average and filtered ab initio model from DAMFILT and that obtained from DAMAVER for 6xHis-SUMO-SVP;* ***c,e)*** *rotation at x = 90° and y =90°;* ***b)*** *Most representative model obtained from DAMMIF.* ***d,f)*** *rotation at x = 90° and y =90°;* ***(bottom)*** *Comparison of theoretical and experimental scattering curves for most representative model.*

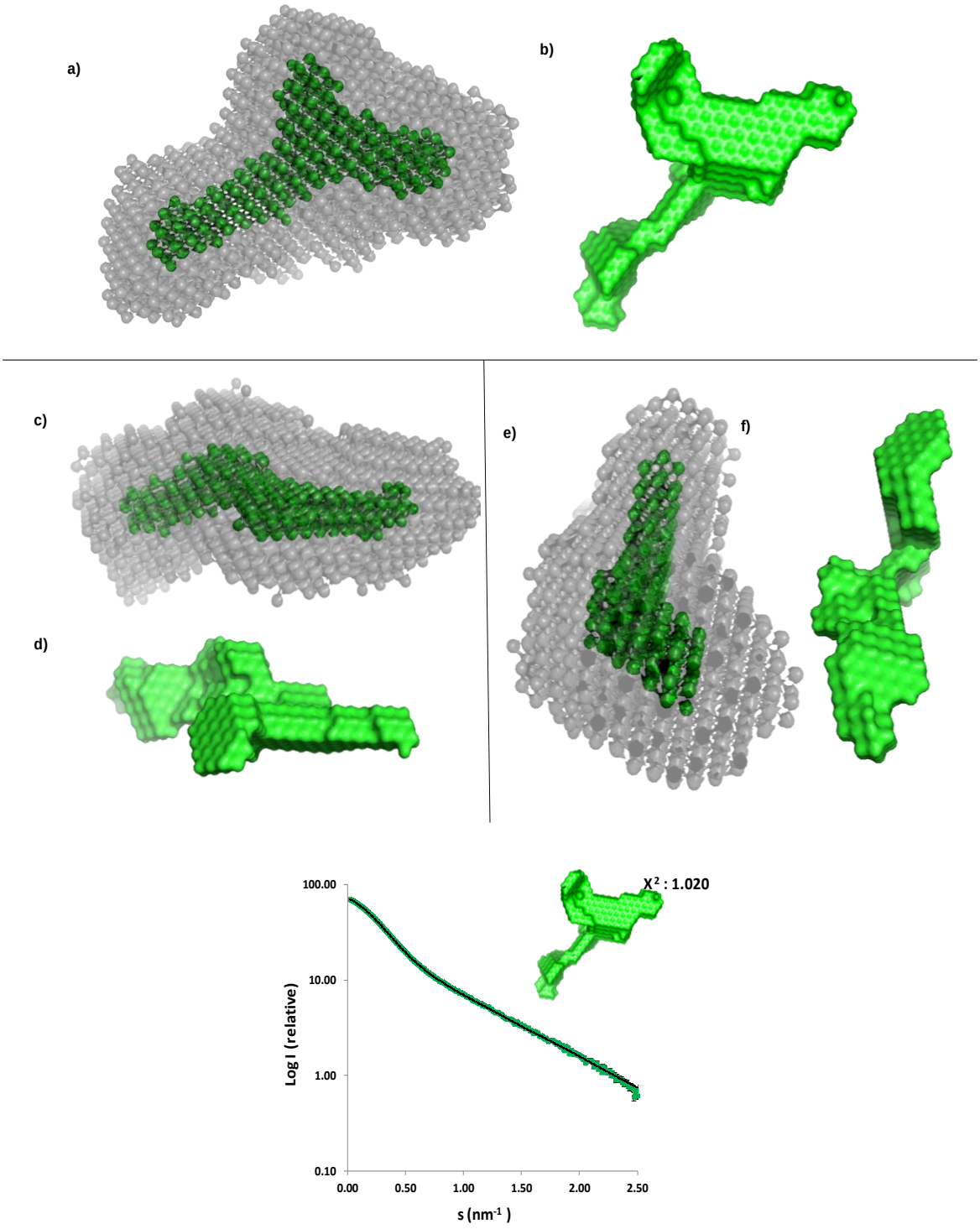


Figure IV-31: ***Ab-initio model for solution structure of 6xHis-GST-SVPconstruct (top) a)****Overlay of average and filtered ab initio model from DAMFILT and that obtained from DAMAVER for 6xHis-GST-SVP;* ***c,e)*** *rotation at x = 90° and y =90°;* ***b)*** *Most representative model obtained from DAMMIF.* ***d,f)*** *rotation at x = 90° and y =90°;* ***(bottom)*** *Comparison of theoretical and experimental scattering curves for most representative model.*

The *ab initio* models presented here demonstrate different lengths over the x and y-axis. This difference in the aspect ratio points towards a specific structural conformation. The slight elongated shape agrees with the data obtained from the pair distribution function as previously described. The theoretical scattering obtained for the models is in accordance with the experimental scattering pattern as observed from the fits and the chi $(\chi)^2$ values.

The initial scattering analysis for SVP constructs validates the data quality and confirms the absence of interparticle interactions. Although the *ab initio* models show good fit to the SAXS data, it is difficult to generate a unique model representing the three-dimensional structure using low-resolution one dimensional scattering data. The problem is further compounded in the case of flexible systems as the observed scattering can be a result of an ensemble of dynamic structures. The Kratky plot suggests internal flexibility for all three SVP constructs. Multiple models providing possible orientation of the protein were generated and validated by fitting with the experimental scattering data. Only the most representative models suggesting a potential orientation of the molecules in solution were displayed here.

As discussed in the previous chapters, MADS TFs are able to form tetramers and dimers. Based on our studies of the SEP3 and AG oligomerisation domains, I was able to perform a structure based sequence alignment and compare residues between SEP3 and SVP, which might explain why SVP is dimeric in solution whereas the SEP3 K domain formed a tetrameric species. In the chapter discussing SEP3, (section II.3.2.3, Figure II-27) mutagenesis studies revealed the effect of changes to different residues on the oligomeric state of protein. It has also been postulated that the absence of Gly-Pro motif is a major factor for disruption of tetramer and formation of dimer in the case of AG (Silva et al., 2016). An alignment of the SVP oligomerisation domain as observed in Figure IV-32 shows the absence of a Gly-Pro motif in the kink region between the two alpha helices of the K-domain. This might explain the dimerisation propensity of SVP. The absence of a proline residue to break the kink region may contribute towards leucine-zipper type interactions of the second K-domain alpha helix at the level of the dimer and not force the helices apart as is needed for efficient tetramerisation. The presence of a flexible kink region, which separates the K domain in two alpha helices (helix 1 and helix 2), may result in different oligomeric states depending on the interaction surface and partners.

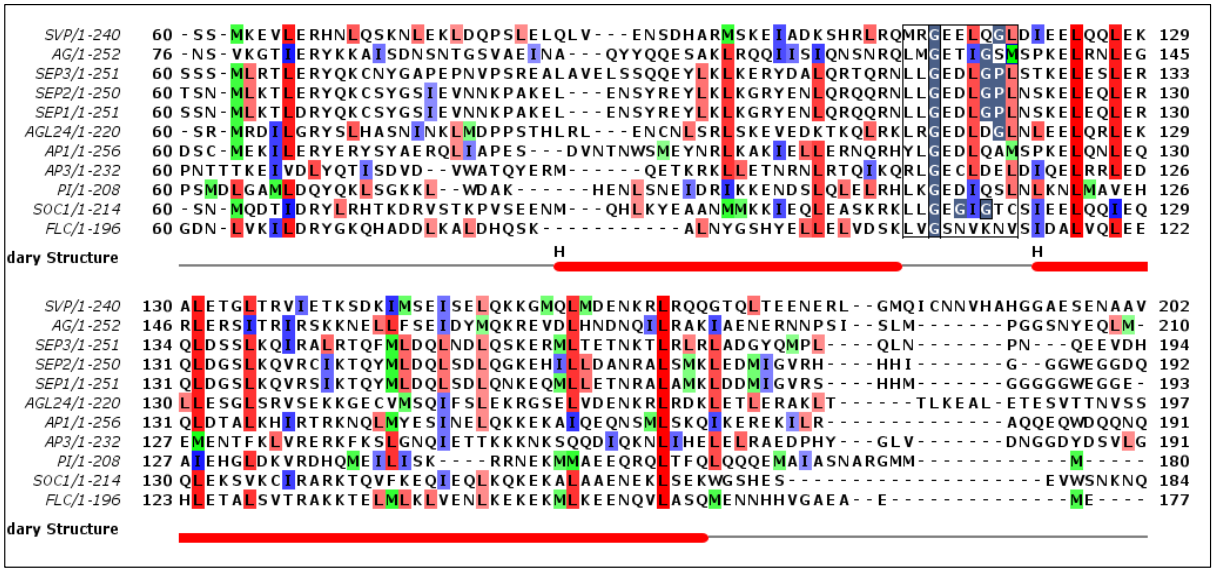


Figure IV-32: ***Analysis of conserved residues in MADS TFs oligomerisation domains.*** *Sequence alignment of representative MADS TFs from Arabidopsis spanning the oligomerisation domain. The structure obtained from crystallized construct SEP3$^{(75\text{-}178)}$ is represented below by H' followed by red bar representing alpha helix. The black box marks the kink region. The conserved glycines and prolines in the kink region are coloured grey. The conserved hydrophobic residues M (green), L (red) and I (blue) are coloured depending on degree of conservation, darker being highly conserved.*

Helix 1 of SVP exhibits several non-conserved residues as compared to other MADS TFs suggesting that SVP might interact in a slightly different manner as compared to other MADS TFs. Helix 2 possesses many conserved hydrophobic residues which might lead to the formation of a dimeric form by closing the second helix (Figure IV-33). However, the presence of a flexible kink region between the two helices likely contributes to alternate conformations of the dimer interface. Exposure of these hydrophobic residues might contribute towards interactions with other MADS TFs forming heterodimers and putatively heterotetramers. Thus, based on homology of MADS TFs and the structural conformations of the AG and SEP3 K domains as presented previously, the following structural organisation for SVP can be postulated. The bent conformation, as represented by SEP3 (PDB 4OX0, Figure II-24), and the possible extended conformation, of AG (Section III.3.2.2, Figure III-19), can be considered as the two extremes for the SVP structure. The SAXS data, however, points towards a tendency to form an elongated conformation. This might point towards preference of elongated dimeric configuration for SVP.

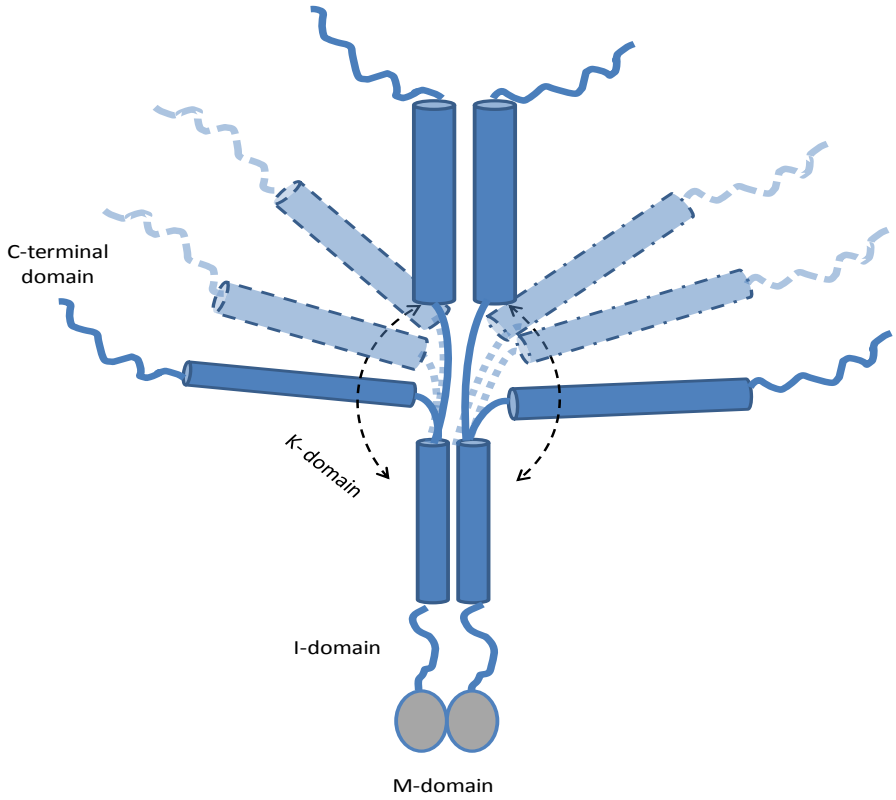


Figure IV-33: ***Postulated domain conformation for SVP.*** *The highly conserved M domain can be postulated to be dimeric similar to its mammalian and yeast homologues. The two helices of K domain are separated due to flexible kink region. The helix 1 dimerizes due to presence of hydrophobic residues while helix 2 can form an elongated and bent structure (dark coloured) or any intermediate conformations (light coloured). The conformation of I domain cannot be predicted due to absence of structural information. The C-terminal domain is unstructured and can have random conformation.*

Also, as observed in the Figure IV-34 SVP possesses interspersed hydrophobic leucine residues (L-71,76,79,84) in the C-terminal region of the I domain which are not conserved in other MADS TFs. The interface provided by these residues may lead to strong hydrophobic interactions with the coiled-coil domain of partner Intervening domains forming a very stable homodimer. The non-conserved residues in the Intervening domain may influence the differential binding of SVP to FLM variants FLM-β and FLM-δ. It has been observed that, even though SVP interacts with both FLM-β and FLM-δ, the resultant complexes differ in their DNA binding activity with FLM-β forming a DNA-binding competent complex while FLM-δ forms a non-DNA binding complex (Posé et al., 2013). Sequence alignment of two FLM variants (β and δ) reveals 100% sequence identity barring the region of 18 residues (D62-L81) in the I domain. Alignment of the FLM variants with SVP suggests that the

shorter I domain of the delta variant may destabilise the dimerisation of the adjacent M domain with the SVP M domain, resulting in a complex which is not able to efficiently bind DNA, as shown in previous EMSA studies (Pose et al, 2013)

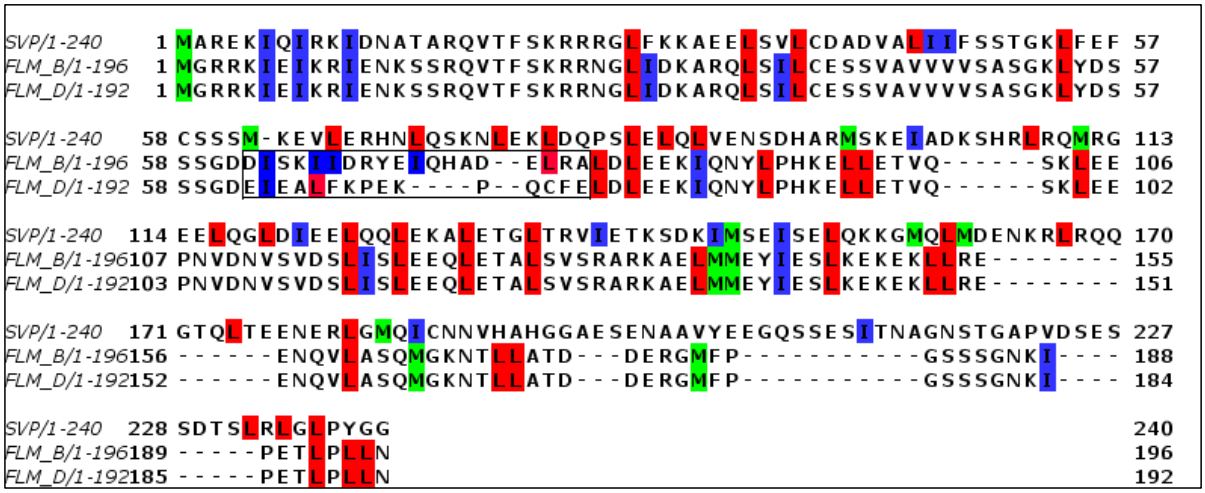


Figure IV-34: ***Sequence alignment of SVP with FLM variants (β and δ).*** *The sequence alignment shows high degree of conservation between FLM variants. The variable region is marked with black box. The conserved hydrophobic residues are coloured with green (methionine), red (leucine) and blue (isoleucine).*

IV.3.2.4 DNA- protein and interactions and protein oligomerisation

The production and purification of full-length recombinant SVP and its biophysical characterisation as described above is a breakthrough in the study of MADS TFs. However, the results obtained did not provide any information about the DNA binding characteristics of the protein which are crucial in determining the mechanism of transcription regulation and specificity of MADS TFs. Here, the results of EMSA, AFM and MST studies used to investigate the DNA binding properties of SVP are described.

The prerequisite of DNA binding studies is the selection of an appropriate DNA sequence. Two different DNA strands were selected from bioinformatic analysis of the ChIP-seq data (Gregis et al., 2013) and the DNA binding motifs available from the JASPAR database (Figure IV-35) (Mathelier et al., 2015). For AFM studies, DNA strands ~1000 bp in length from each of the two selected sequences were amplified using Arabidopsis genomic DNA as a template. The CArG boxes were organized asymmetrically for easier analysis of protein binding. The details of the primers and the PCR amplification process are given in the Materials and Methods section IV.5. The two DNA sequences were labelled *SVP2* and *SVP3* for convenience. The CArG box regions in the DNA sequence are printed in bold while the

Figure IV-36 shows the relative position and spacing of the CArG box binding sites and the approximate lengths in nanometers for the target DNA.

SVP2 (1033 bp)

CACTACAAACTAGAACAAATAAATGCCACGTGAAATCCTACACGACAGCTAAGTTGCGGAGTA
CTGGTGATATATGTGGTCCATTGGGCGTGAGATGAGAATCGGACGGCTTTGAGGCAATGTCGTGA
AGGGTCGAAGAAAGAATAAGAACCGTTGGATTAATCTG**CTAAAAATAG**ACGGACAGGGTCCA
CGTCAAGACGTAATGGGAAGGGGACCTCGTTAC**CCAAAATAGA**AAGATGGAGTTATCGTCAA
ACAAGTCACGTCTGGGTTTATTGAGTGGGGTTAAGTTTTTGGACCAGTCTCTCTACTTTTTTTTTCA
CCCCTCCTCTAATATATAGTAAGTTACTGTTGTCATAGACTCTTAGTCATTATGCATTTTAAAGGTT
TTCACTTTTCAGCGTGCAATGAAAAATTTGGTTCATACCTTATGCCGTAAACTACTATAATGGATTA
TTATGTTATTTAAAAACTCATGTAGATGTTATTTAAGTGTTGTATCACTTATGTACACACTAATAAA
TTTCAGATTCATAACTAATATATAAGCACTCAAACTAGTTAACATTTTGTCATTTGATTAAGATAGT
GACTGGGAGTAAATAACTTTCATCCTTGTAAGTTGTAAAACCTCATGACACCATTAAAACCATAGT
TTAGATGCCTAATCCAAGATAAGGATAGCATCAGAAGAGAATATATAGTCCATCGAACTAGATTT
GGAGGAGTGTTGAAGCTAGTCCATCGAACTAGACTTTTGACAGTTTGATCCACTTTTTTAAAGATT
TGAGATAGTAGAACATCGTAGTTGAGAGTGAATCTTGTTAATCTTGTCACCGTTTCACGCCAAAGT
TCGTAGATTGATGAATGCCAAAGCAGGTTGAGGATATAGCACCTGGCCTTAGGTATGTTTTGGGCC
GGGGCGAGATAACCAAGTACAAATGTACAATCATCCTCTTGACAAGAAATTTTGCTTAGATTTTTA
CATATATAGGCTGTTAGAAAAGAGAATGGAATAGGAAACCAACTCTTATAG

SVP3 (1004 bp)

CATCTCCATAGTCGTCCCTTGAATCTTGGTCGAACTCTGTAGCATCAAGGTTGAAATTAAACCT
AAATGACAATCGTGTGTACGTGTATAAATATACATATAAGTGATGATTACATAAACTCATAGTCAT
ATATGGGATCTATTATGGTTTTATGCTCACTCTATACGTCGGCCTAACGTGGCAAAAGTGGATATTT
CTTTGAATATAAAATAGAAAAACAAAAAGCAAATAAGCAAACATTGCAATAATA**CCAAAAGA**
GGAATGAAAACATATTTCTGTATTTGGAAGTAATGATCTTAGACTCCACGCGACTGTGTTCTTTAT
TGTTTTGTTTCGAAGAGCGGTGTGATATTCTTATTGGCTTGCAACCAAAATGGGCTC**CCAAAAA**
GAAAGCAAGCACCACCACTTCTTCTATCTGATGAGGGAACACTTTTAACATAAAAGTCACACAAC
CATAACATATATTCAAATTTCGGGTGATAACATAATGATAAAAATAGCACAACCCAATTAACTACA
TGTTTTGGTCATATTTTATTCATAGTTTATAGTTTTTTCTTTTTTTTTGGTAGGAGTAAAGAACACTG
TCGCGTGGAGTCTAAGATCATTACTTCCAAATACAGAAATTTGGTTTCATTCCTCATTTGGTATTAT
TACAATGTTTGCTTATTTGCTTTTTTTCCTACTTTTTTTTTTATAACATCTAGTTTATAGTTTTTGTTAT
ATATTCAATCTATTTAGTGTCTAAACAATACTGATGATATCAAATGATTAAAAGGCTTTTGAACTG
TAAAAATGACAAACAAATATCTATGATCTATGTCATAATTGAAAAAAAAAAGTTACCAATAAAGA
TTTTTTTTTATATAGGTAGTATTTCTTGGCTAATATGGACATGTATATAATTATATTTAGTCTTATCT
ATGTATGAAATTGGTTTGGATGTTGAAAAATGAATCATCCTTATCGTCAACCATTTTGGAAAGAGA
CAAATCTAGATTTTG

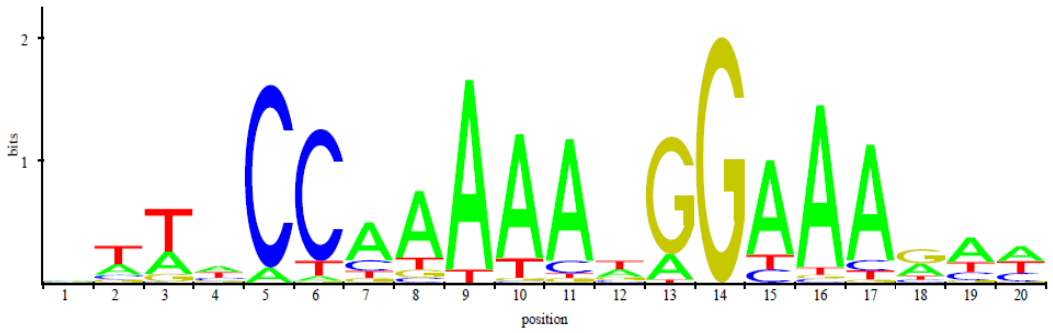


Figure IV-35: ***CArG box regions for SVP binding****. The sequence logo represents the SVP binding profile obtained from JASPAR database* (Mathelier et al., 2015)*. The height of nucleotides indicates its relative conservation at that position.*

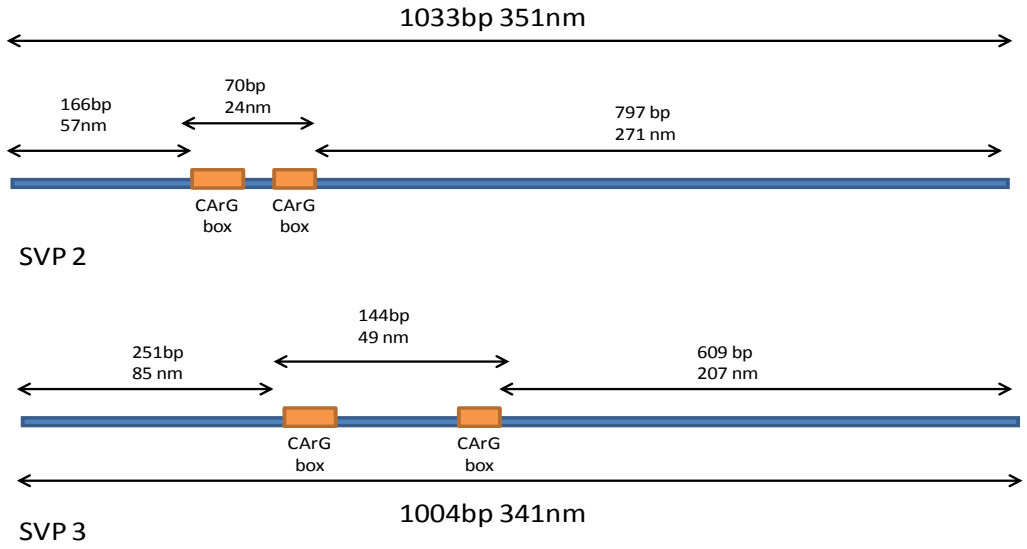


Figure IV-36: ***DNA strands generated for AFM scans.*** *The CArG regions are represented by brown boxes. The DNA sequence was PCR amplified from genomic DNA. The position of CArG boxes is asymmetrical for better analysis in AFM.*

For EMSA experiments, *SVP2* and *SVP3* DNA were used as templates for the amplification of shorter DNA constructs. *SVP2.1* and *SVP2.2* were amplified from *SVP2* and contain 1 CArG box and 2 CArG box sites respectively (Figure IV-37). *SVP2.1* with 1 CArG box binding site was also used for MST experiments. The forward PCR primer was labelled with TAMRA (Eurofins Genomics). The details of the PCR amplification and primers are provided in the materials and methods section IV.5.

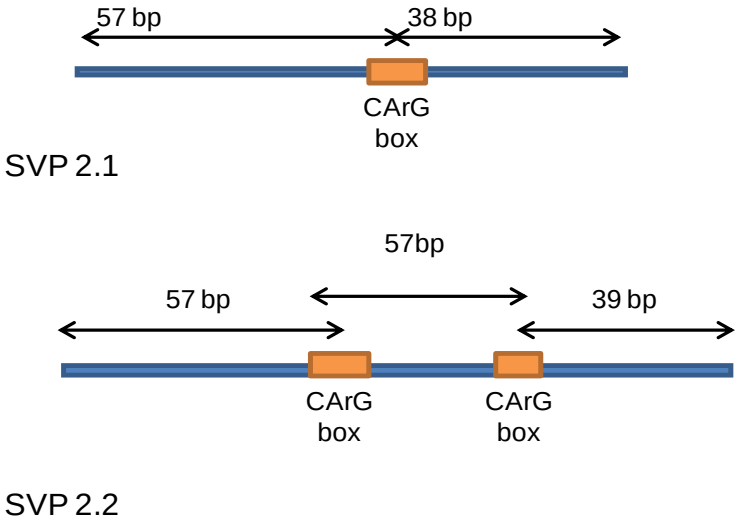


Figure IV-37: ***DNA strands designed for EMSA.*** *The DNA was amplified using the SVP2 DNA from Figure IV-36 as template. The target DNA SVP2.1 was also used for MST experiments*

IV.3.2.4.1 Micro Scale thermophoresis (MST)

MST was used in order to determine the stoichiometry and binding affinities of SVP complex formation. MST is a simple, fast and sensitive method that is used to quantitatively measure biomolecular interactions on the basis of their mobility in a temperature gradient (Jerabek-Willemsen et al., 2014, 2011). It is extremely sensitive to changes in the molecular properties such as size, charge and conformation. It uses a titration approach to gather information about the binding affinities of the interacting macromolecules. One of the main advantages of MST is its ability to be used in almost any buffer condition without immobilisation by utilizing very small quantity of samples. Figure IV-38 shows a schematic of a typical MST experiment.

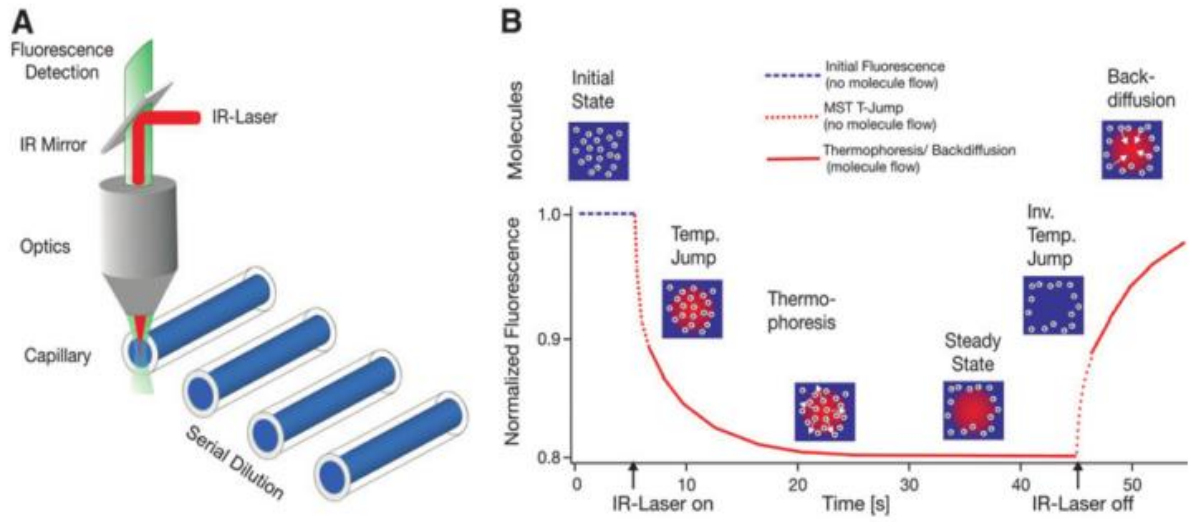


Figure IV-38: ***Schematic of typical MST experiment. a)*** *MST experiments are performed in glass capillaries with a volume of about 4µl. An IR laser is used to generate high temperature at microscopic region. The change in the base fluorescence is measured over a time scale and compared to initial fluorescence. b) A typical MST signal consists of an initial state with homogenous distribution of molecules followed by a fast temperature jump on switching on the laser, followed by thermophoresis. As the laser is switched off, the temperature change is observed accompanied by back diffusion of the molecules.* (Jerabek-Willemsen et al., 2011)

In the current project, MST was used to quantify the binding affinities for protein-protein interactions both in the presence and absence of partner DNA. The protein construct 6xHis-SUMO-SVP was used for the experiments. The fluorescent red dye NT-647 (MO-1004, Nano Temper Technologies Gmbh, Munich, Germany) was used for labelling the protein samples.

The assay was validated using capillary scans and analysing the sample-capillary interaction. Figure IV-39 shows a capillary scan for the 6xHis-SUMO-SVP protein-protein interaction in the presence and absence of *SVP2.1* DNA. The similar height of the peaks for all the capillaries confirms the constant initial fluorescence. The peaks with abnormal values were not considered for further analysis. The fluorescence intensity count is also observed between an optimal range of 200-1500 counts. The capillary shape analysis Figure IV-39 shows a single smooth peak thus confirming the absence of sample-capillary interactions, which might affect the fluorescence scans and subsequent data analysis.

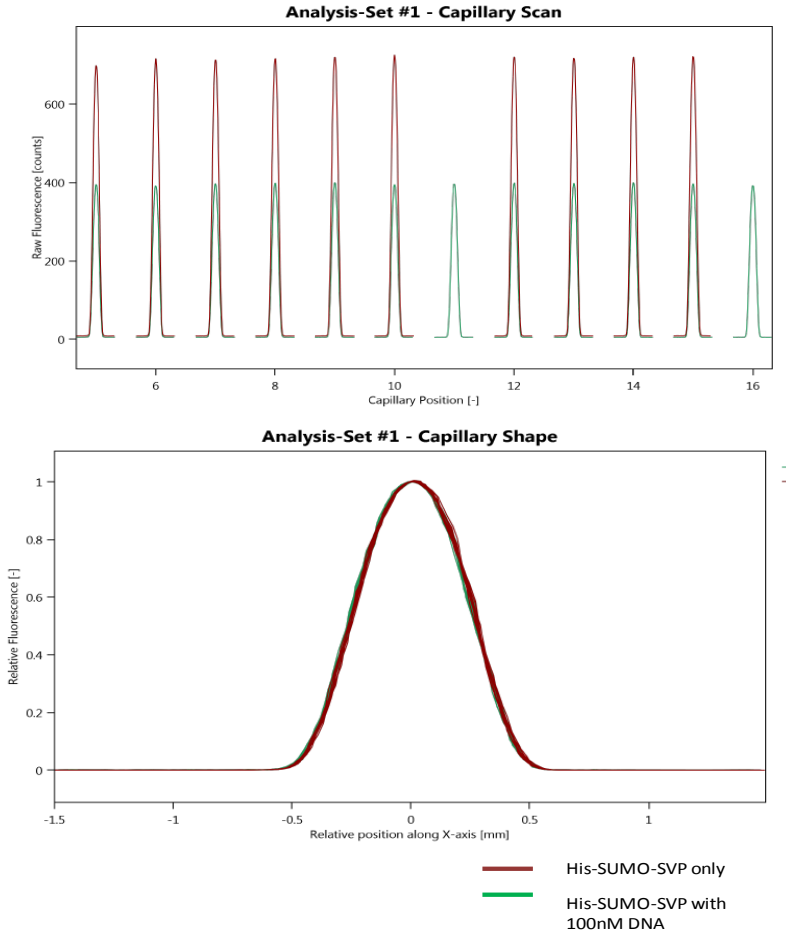


Figure IV-39: ***Validation of MST screen: capillary scan.*** *The concentration of fluorescently labelled (NT-647) 6xHis-SUMO-SVP protein construct was kept constant (10 nM) while non-labelled protein was varied between 10 µM- 0.25 nM. The assay was performed in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl2, 0,05 % Tween-20). After a short incubation, the samples were loaded in to MST NT.115 premium glass capillaries and capillary scan was performed using Monolith NT.115 (MST 40%, LED30%).* ***(top)*** *initial fluorescence is observed to be constant for the 6xHis-SUMO-SVP sample (brown) without DNA and with (green) 100nM DNA. The peaks with abnormal initial fluorescence are not used for further calculations.* ***(bottom)*** *The peak in the capillary shape analysis shows absence of capillary-sample interactions. Peaks for all the capillaries also seem to overlap. If a distortion in peak is observed, then the scan is repeated by changing the capillaries.*

Figure IV-40 shows the MST traces over a time period. The sample is observed to be homogenous and without aggregation. The capillary scan, capillary shape and the time traces performed validate the choice of sample, buffer and scanning parameters.

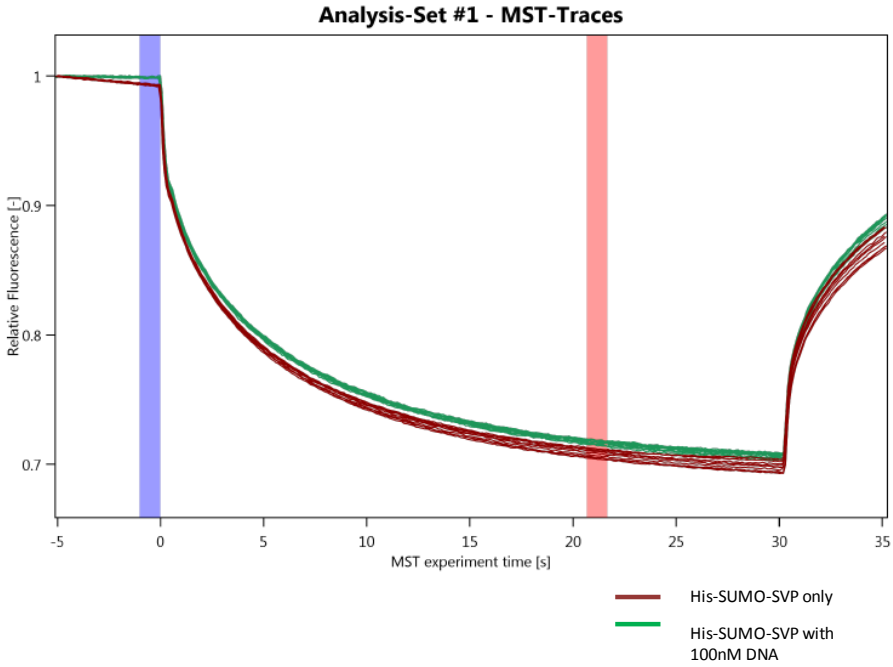


Figure IV-40: ***Validation of MST screen: MST traces.*** *The MST trace detects the change in fluorescence over period of time. The shape appears similar to typical MST trace as shown in Figure IV-38. The smooth traces confirm absence of aggregation. The change in thermophoresis is measured as change in normalized fluorescence over the initial and the excited state. The values for the initial state correspond to the average of the violet coloured region while the excited state corresponds to that from the pink region.*

The Figure IV-41 shows a titration curve of unlabelled protein 6xHis-SUMO-SVP with labelled protein in the presence and absence of *SVP2.1* DNA in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl2, 0,05 % Tween-20). A K_d of ~ 33 nM was measured for protein-protein interactions by titrating unlabelled 6xHis-SUMO-SVP protein against NT-647 labelled 6xHis-SUMO-SVP protein. Addition of 100 nM *SVP2.1* DNA (with one CArG box binding site) did not lead to significant changes in the measured K_d value and a value of 53nM was obtained for the protein-protein interaction. A stable K_d suggests no or little effect of the presence of DNA on the binding affinity for the protein-protein interaction.

Thus, MST was useful to quantify the binding affinities for protein-protein interactions. The results obtained show a very high affinity for protein-protein interaction with K_d of about 33nM with little effect due to the presence of DNA. The binding affinities in

the low nanomolar range suggest strong homodimerisation tendencies for SVP. DNA binding of SVP was qualitatively investigated by EMSA and AFM.

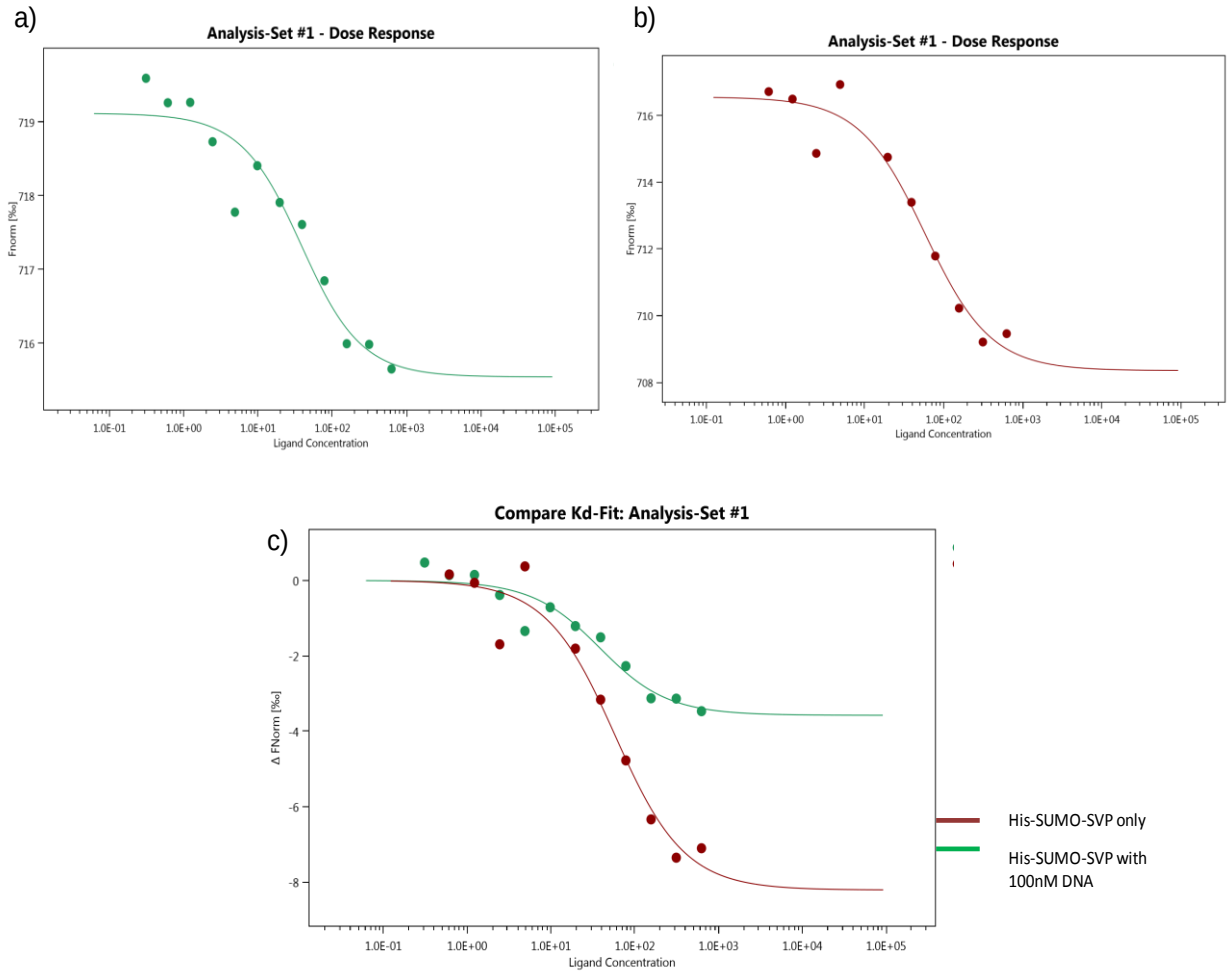


Figure IV-41: ***MST dose response curves.*** *The concentration of fluorescently labelled (NT-647) 6xHis-SUMO-SVP protein was kept constant (10nM) while non-labelled protein was varied between 10 µM – 0.25 nM. The assay was performed in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl2, 0,05 % Tween-20). After a short incubation, the samples were loaded in to NT.115 premium glass capillaries and MST analysis was performed using Monolith NT.115 (MST 40%, LED 30%).* ***a)*** *K_d of about 33 nM was obtained from the dose response curve for protein-protein interaction.* ***b)*** *100 nM SVP2.1 DNA was added to the sample and the experiment repeated. The dose response curve yielded a K_d of about 53 nM.* ***c)*** *A comparison of protein-protein interaction with (green) and without (red) DNA does not a show a significant change in binding affinities. Concentrations on the x-axis are plotted in nM*

IV.3.2.4.2 Electrophoretic Mobility Shift Assays (EMSA)

DNA binding activity of SVP was confirmed using EMSA. Increasing concentrations of native SVP was incubated with 10 nM DNA *SVP2.1* and *SVP2.2* at 4°C for 10-15 min. Band

shifts were analysed by running the complexes on a 5% polyacrylamide gel using Tris-borate/EDTA buffer under nondenaturing conditions at 4°C. Figure IV-42 shows a clear band shift for both DNA sequences. In the case of *SVP2.1* (*DNA 1*) a single band is observed even at high protein concentration (500 nm) while for *SVP2.2* (DNA 2) two bands are observed pointing towards the presence of two species. Through EMSA we can thus confirm the DNA binding activity of SVP. According to SEC-MALLS, SVP exists as a dimer even at high protein concentrations. Hence, it can be considered that, in the case of *SVP2.1* with one CArG-box binding site, the band shift occurs due to binding of an SVP dimer. As the two binding sites in *SVP2.2* are different they would most likely have different binding affinity. Thus in case of *SVP2.2*, the lower band likely represents binding of a single SVP dimer to one binding site while the upper band probably represents two SVP dimers bound to both CArG box binding sites.

While EMSA demonstrates DNA-protein interactions, it provides limited information about the mechanism of interaction. In order to probe the dynamics of protein-DNA interaction and visualize the process, AFM was used.

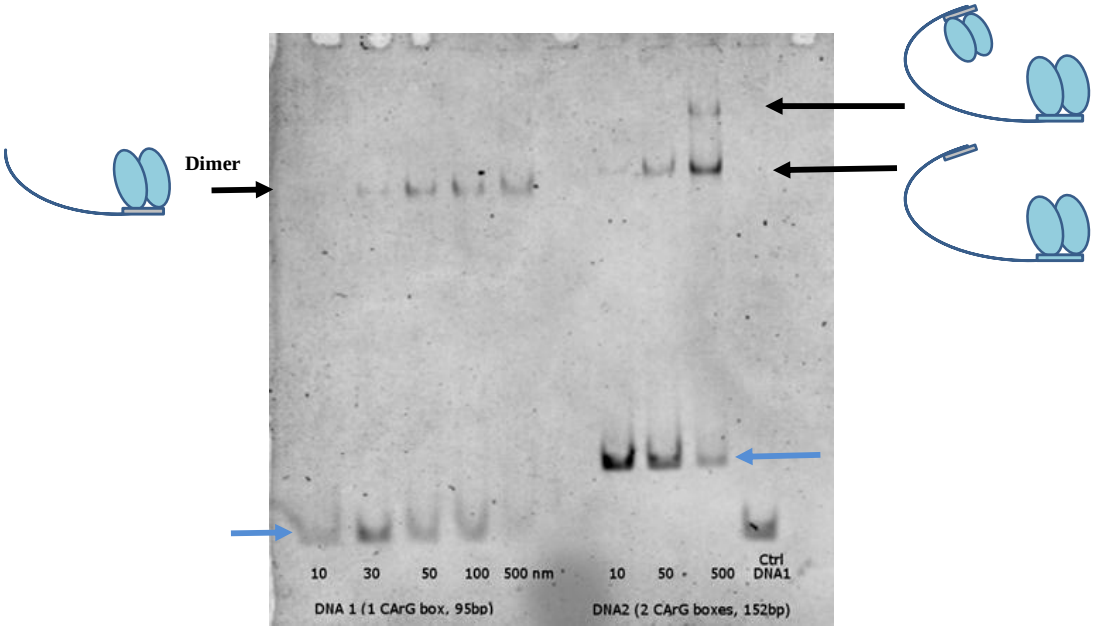


Figure IV-42: ***Electrophoretic Mobility Shift Assay (EMSA) for SVP with two different DNA constructs.*** *10 nM TAMRA labelled DNA was treated with increasing protein concentration from 10 nM to 500 nM. The DNA1 shows a band shift (black arrow) at 30 nM with almost complete shift at 500 nM. The single species observed can correspond to dimer as represented by the cartoon to the left. The DNA2 shows formation of two bands corresponding to two different species. The lower band (black arrow) corresponds to single dimer while the upper band (black arrow) likely corresponds to binding of two dimers on both the binding sites as represented by cartoon to the right. The free DNA is marked by blue arrow.*

IV.3.2.4.3 Atomic Force Microscopy

AFM was performed on protein-DNA complexes with two different DNA target sequences, *SVP2* and *SVP3*, shown in Figure IV-36. The AFM scans were performed as for SEPALLATA3 described in Chapter II using a Cypher S atomic force microscope (Asylum Research, Santa Barbara, CA) in tapping mode. A thorough optimisation process was necessary to obtain good quality images.

Figure IV-43 (left) shows interactions of 2-5 nM native SVP with ~ 2 nM *SVP2* DNA. The images show formation of higher order oligomeric structures represented by the bright spot. However determination of the number of monomers in this complex is beyond the scope of AFM imaging. The two CArG box binding sites are only separated by a distance of 70 bp, equivalent to ~24 nm. This short distance made it difficult to analyse the individual components of the complex and understand the binding pattern of SVP on its binding sites.

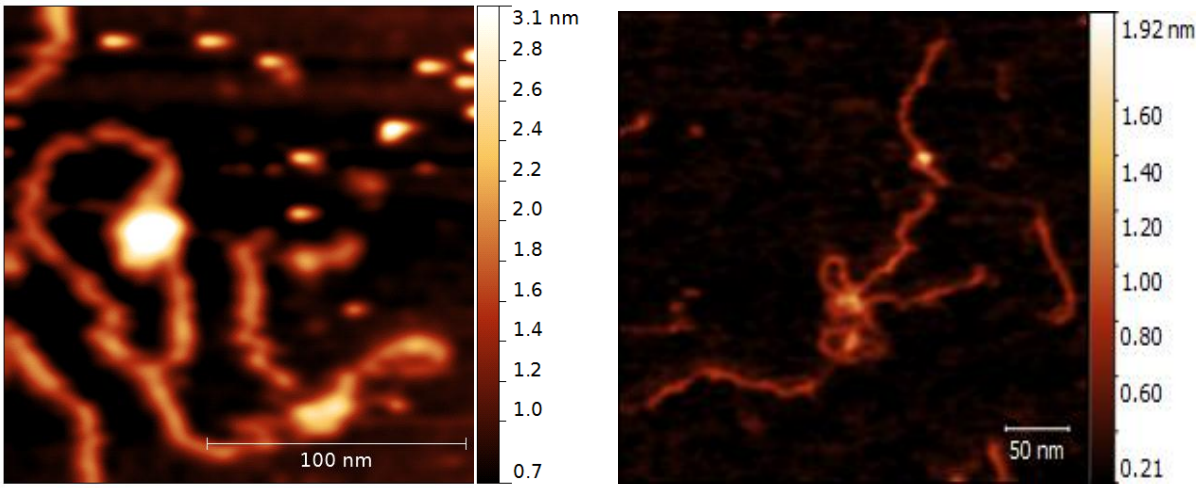


Figure IV-43: ***AFM scan for studying SVP-DNA interactions****. AFM images for SVP for 2-5 nM SVP protein with (left) 2 nM SVP2 DNA and (right) SVP3 DNA. SVP3 possess different CArG boxes at larger distances as compared to SVP2.*

AFM experiments were therefore repeated using *SVP3* DNA with a distance of 144 bp (~49 nm) between CArG box binding sites. Figure IV-43 (right) shows interaction of 2-5 nM SVP with ~2 nM *SVP3* DNA. Multiple DNA strands are observed interacting with SVP protein forming knot of loop-like structures. An isolated DNA strand with protein bound at its binding site is also seen. Several repeated attempts also yielded images showing structures with multiple DNA strands bound to a protein complex. The AFM results suggest that SVP may form higher oligomeric structures and loop DNA, similarly to what was observed for SEP3. The presence of two CArG box binding sites seems to change the dynamics of binding

interactions favouring a higher order oligomeric state. Due to the presence of multiple DNA strands, determination of the exact interaction mode and number of protein molecules was difficult. Although AFM allows us to study interaction of individual molecules, the samples are premixed before the scan and hence, the initial binding interactions and the dynamics of binding cannot be measured in real time. Moreover scanning rate of ~ 1 min/frame used during the scans is relatively slow to measure these dynamic interactions. Recently, a novel mode of AFM has been developed which can scan at rate of 1-2 s/ frame and is capable of visualizing dynamic interactions (Ando et al., 2001).

IV.3.2.4.4 High speed AFM (HS-AFM)

High speed AFM is a cutting edge technique that has been developed by the Ando group (Ando et al., 2001) and subsequently used to study various biological dynamic interactions in real time (Milhiet et al., 2010; Ando et al., 2008, 2014; Casuso et al., 2011). The high-speed AFM scans presented (Figure IV-44 & Figure IV-45) were performed at the Bio-AFM Frontier Research Centre, Kanazawa University, Japan in collaboration with Dr. Takahiro Nakayama. About 2 nM *SVP3* DNA was scanned and SVP protein was titrated in from 0 to 10 nM. A carefully tuned concentration of KCl and NaCl was used to maintain partial mobility of the DNA and protein while keeping them adhered to the surface for scanning. Changes in this concentration either lead to release of DNA from the surface or strong immotile adherence. The free movement of proteins allows them to locate and bind appropriate binding sites while maintaining the plasticity of DNA. As observed in the series of images shown in Figure IV-44 and Figure IV-45, the DNA strand *SVP3* is loosely attached to the surface of mica while the native SVP, represented by white spots, is freely motile. The images show the protein probing for its appropriate binding site in the liquid medium. This was the first time that DNA-protein interaction for MADS TFs was studied in real time. Although the visualisation of the entire interaction process has not yet been possible due to time constraints, this technique has a potential to reveal important information as to the binding mechanism. Other binding characteristics like rate, specificity and affinity can also be determined using HS-AFM. These initial experiments demonstrate that HS-AFM offers a technique for probing dynamic events at the single molecule level for SVP-DNA interactions.

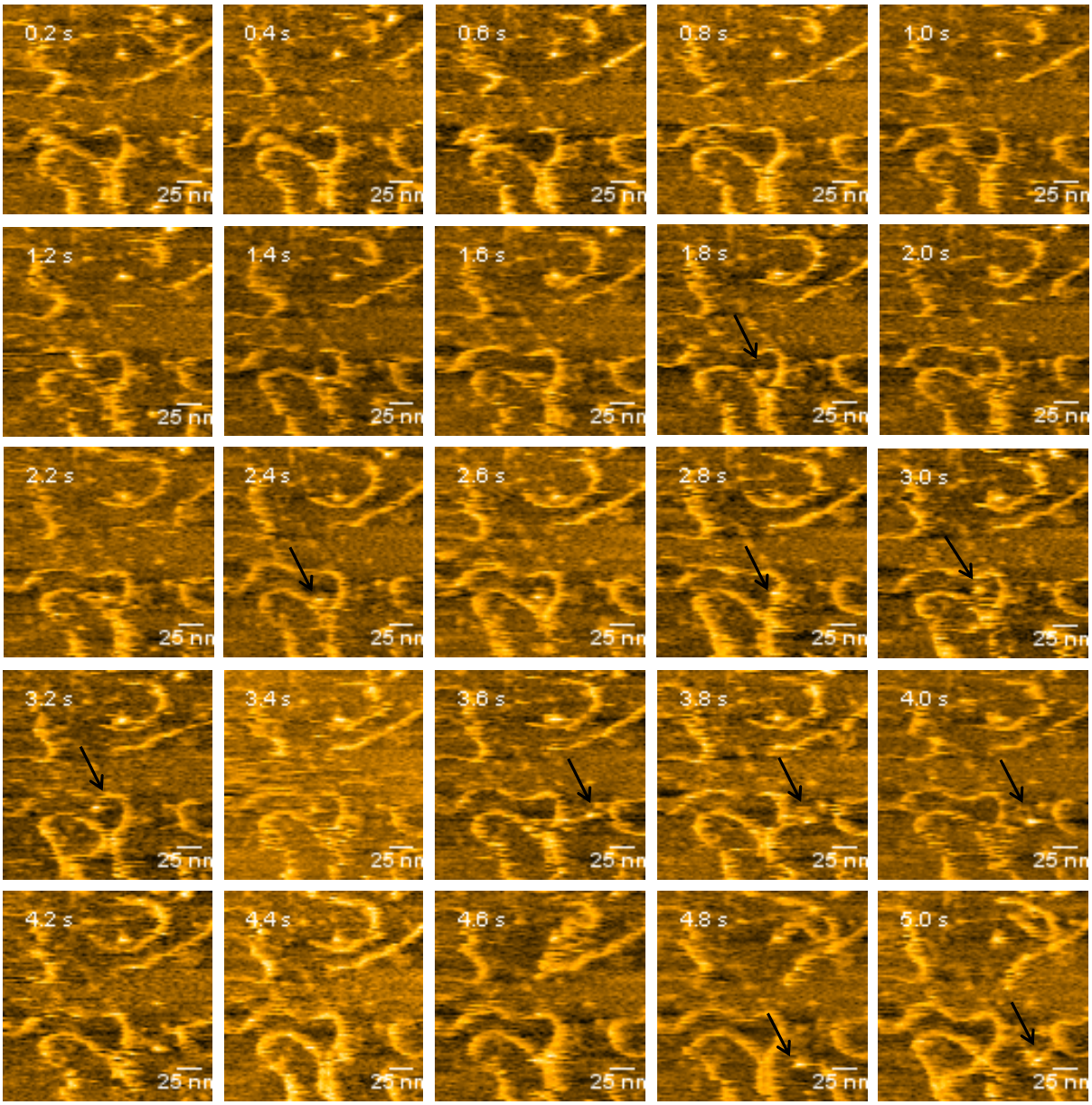


Figure IV-44: ***HS-AFM scans for studying SVP-DNA interactions.*** *HS-AFM images for 2 nM SVP3 DNA treated with 10 nM native SVP. The scan rate was 0.2 s/frame. The DNA strand is clearly visible. The protein represented by bright spot and labelled by black arrows is freely motile. The scans were performed using AFM developed in house at University of Kanazawa, Japan.*

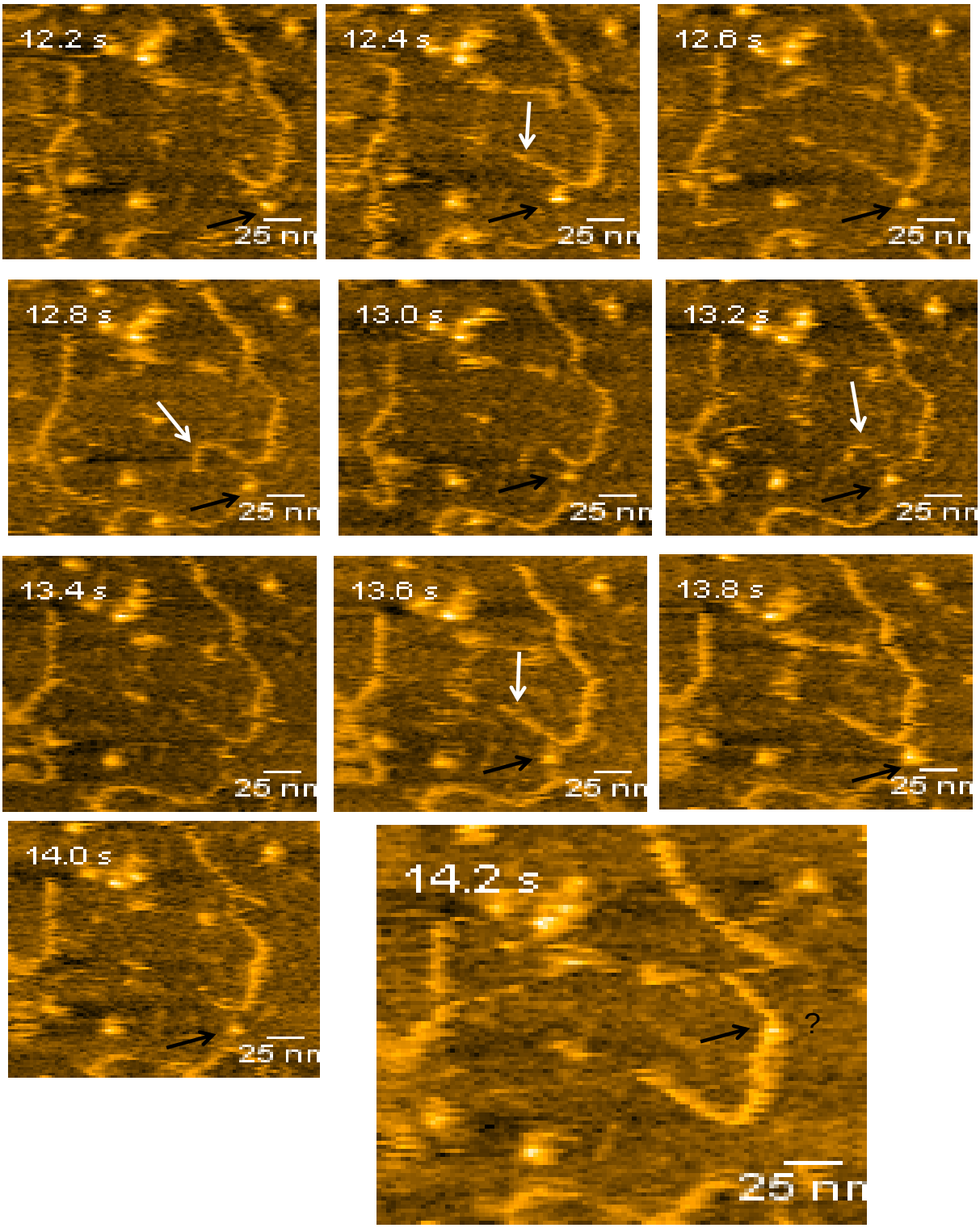


Figure IV-45: ***HS-AFM scan for studying SVP - DNA interactions*** *HS-AFM images for 2 nM SVP3 DNA treated with 10 nM native SVP. The scan rate was 0.2 s/frame. The white arrow marks the change in the structure of DNA strand while the black arrow marks the position of protein. The last image is enlarged to verify a possible DNA-protein binding. The scans were performed using AFM developed in house at University of Kanazawa, Japan.*

IV.4 Conclusions

The difficulties in obtaining purified protein has formed one of the main hurdles in studying of MADS TFs. Using high throughput approaches and different expression systems, I was able to establish a protocol for the purification of full-length native SVP as well as two constructs tagged with either 6xHis-GSTor 6xHis-SUMO. This opened up SVP to a wide range of biochemical and biophysical studies.

The SEC-MALLS analysis suggested a dimeric state for SVP while MST analysis shows low dissociation constants for the dimers suggesting strong protein-protein interactions. The SAXS data suggests that SVP prefers a more elongated configuration. It also displays a high level of flexibility, which can help in rapidly changing conformation in the presence of appropriate binding partners. As observed from the sequence alignment with other MADS TFs, the absence of a Gly-Pro motif may play a role in driving dimerisation rather than , as is seen for SEP3, tetramerization.

The interactions of SVP with DNA were studied using EMSA and AFM. EMSA shows a distinct gel shift with the formation of a single species for one CArG box containing DNA with two species observed in presence of DNA containing of two CArG box binding sites. AFM experiments show SVP bound to a single CArG binding site as well as the formation of higher order complexes and DNA loops when multiple DNA binding sites are present.

Being a single molecule technique, AFM enables us to observe interactions that might be difficult to visualize using bulk techniques like EMSA. The HS-AFM shows the motion of protein and its target DNA *in vitro* in real time. The optimisation of the conditions required for this has opened new vistas for further study. HS-AFM can be used to understand the mechanism and study the formation of protein-protein and protein-DNA complexes.

Overall SVP seems to be a dynamic structure which prefers a dimeric state in absence of DNA but can form higher order complexes to loop DNA as observed in AFM. SVP performs two main and very distinct functions and it has always been a mystery that how SVP controls two regulatory processes during different stages of flower development. We postulate that the flexibility and plasticity of the protein has a role in diverse functions of SVP. Through its K domain, SVP can form a homodimer and bind to DNA. However, in presence of an appropriate partner or co-factor, SVP may be able to form tetramers. However further research is necessary to address this issue.

IV.5 Materials and methods

IV.5.1 Construct design and purification

IV.5.1.1 Multi vector expression screen

SVP, *SEP3* and *AG* templates were cloned in range of vectors from pOPIN suite (e.g. pOPINF, Figure IV-46) using the In-Fusion™ (Takara- Clontech, CA, USA) cloning technique and verified by colony PCR. The list of the target genes and the expression plasmids used is provided in the Appendix I.

A parallel screening was performed in *E.coli*, insect cell and mammalian expression system. About 32 constructs for single expression and 36 for co-expression were designed for screening in *E.coli* and 22 single expressions were selected for insect cell and HEK 293 cell expression. In the case of bacterial expression screening, the positive clones verified by colony PCR were transformed in two expression strains Rosetta2 (DE3)placI and Lemo21 (DE3) using the heat shock transformation method (30 min ice, 30 s 45°C, 2 min ice) incubated at 37°C with SOC medium for 1.5 h in static incubator and plated on LB-agar in 24 well blocks supplemented with appropriate antibiotics. For small scale expression test, deep well blocks with 0.7 ml of Power broth (PB) (AthenES^{TM}) grown overnight at 37°C with appropriate antibiotics were used as a starter culture. 24 well blocks, with 3ml of PB and auto inducible medium were inoculated with 150 µl of starter culture for Lemo21 and 250µl for Rosetta2. The PB culture was grown at 240 rpm at 37°C for 3-4 h, induced with 1 mM IPTG at O.D 600 of ~0.5 and incubated overnight at 20°C at 240 rpm. The auto inducible medium culture was grown for similar time and temperature until O.D 600 was ~0.5 and then temperature was lowered to 25°C for overnight incubation. 1 ml of overnight expressed culture was harvested (6000 g for 10 min) and lysed by freezing at -80°C for 20 min followed by resuspension of the pellets in lysis buffer for 30 min. After a centrifugation at 6000 g for 30 min, the supernatant was applied to Ni-NTA magnetic beads (Qiagen), incubated for 30 min at room temperature and purified with Qiagen BioRobot 8000. The steps performed include two times washing with 200 µl wash buffer and elution with 50 µl elution buffer. The elute was analysed on 10% acrylamide gel with SDS-PAGE.

Lysis buffer: 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% v/v Tween 20, 1 mg/ml Lysozyme and 3units/ml Benzonase or 400 units/ml DNAse typeI.

Wash buffer: 50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM Imidazole, 0.05% v/v Tween 20

Elution buffer: 50 mM NaH_2PO_4 pH8.0, 300 mM NaCl, 250 mM Imidazole,0.05% Tween 20

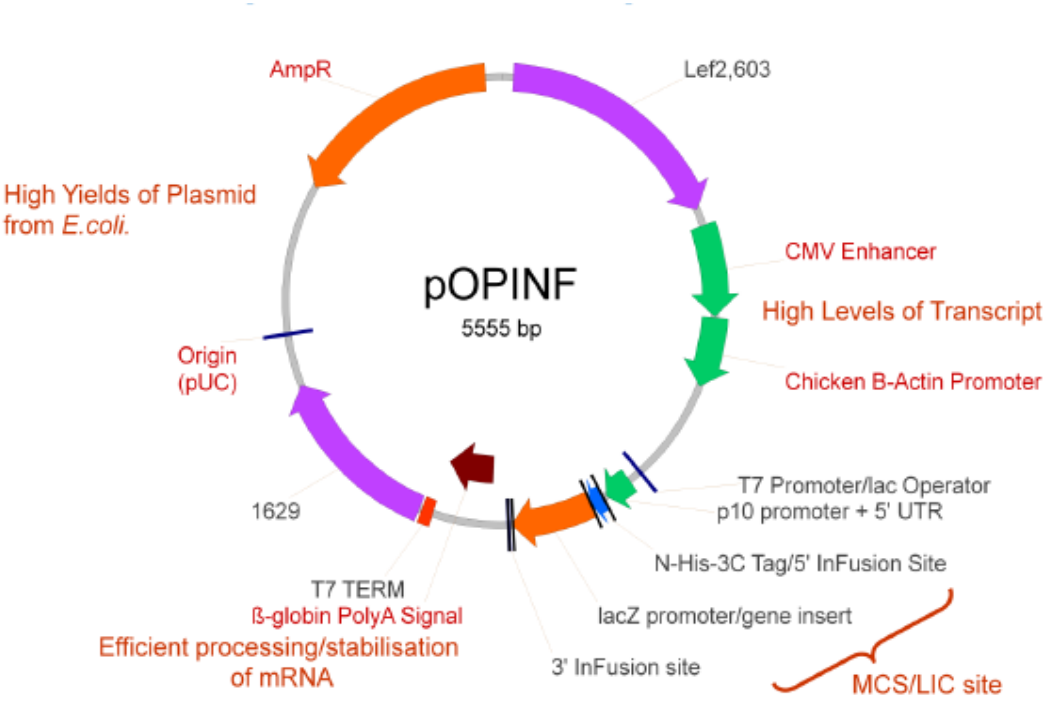


Figure IV-46: ***pOPINF vector for multi-vector expression screen.*** *pOPINF vector was modified to generate pOPINF family with an additional solubility tag after the 6xHis tag. The common restriction sites allow ease of shuttling of PCR product between different vectors. Image courtesy: OPPF, Oxford, UK*

The proteins were also simultaneously tested in insect cell expression system. Bacmid glycerol stock from Dr. Ian Jones was amplified and bacmid extracted using the Epicentre BacMax kit (Cambio Ltd., Cambridge, UK) (Zhao et al., 2003). Transfection was performed using Fugene HD transfection reagent (Promega) with 5-7 * 10^5 Sf9 cells /ml in Sf900II medium (Gibko, Life Technologies). Cells were grown in monolayer at 27°C and viruses harvested at 5-6 days. These viruses were later used for further infections and amplified in shaker flasks. 3 µl of 1st generation virus (P1) was used for infection and samples were grown for 72 h followed by SDS-PAGE analysis.

The 24 well mammalian expression screens were performed using Human embryonic kidney (HEK) 293 cell lines. A cocktail of plasmid DNA (1 µg) and GeneJuice transfection reagent (2 µl of 1.33 mg/ml) was made in 60 µl of serum-free DMEM (Dulbecco's Modified Eagle Medium) and incubated for 10 min at room temperature. The HEK 293 cells were

grown in a monolayer in a 24 well plate and covered with 1 ml of DMEM and 2% FCS (Foetal calf serum). The transfection is performed by adding the DNA/GeneJuice cocktail to the cells and incubating at 37°C in a 5% CO_2/95% air environment. After about 3 days, the supernatant is isolated, centrifuged at 6000 g for 15 min and ~ 20 µl is analyzed using western blot.

IV.5.1.2 Insect cell expression

6xHis-GST-SVP and 6xHis-SUMO-SVP were cloned in pAceBac1 using the primers given in Table IV-6. About 1 µg of plasmid was transformed in DH10EMBacY cells using heat shock transformation method (20 mins ice, 45s at 42°C, 2 min ice, 400 µl LB) and incubated overnight at 37°C in a shaker incubator. On the next day, the transformed cells were plated on 4 plates (Kanamycin / Tetracycline/ Gentamycin /IPTG/ BluOGal) in serial dilution and incubated overnight at 37°C in a static incubator. About 5 white colonies were picked and replated on agar plates with similar composition as above and incubated overnight at 37°C in a static incubator. Two white colonies were selected and used to inoculate 2 ml of LB media (Kanamycin/ Tetracycline/ Gentamycin) and incubated overnight at 37°C in a shaker incubator. This culture was later used for isolation of bacmid.

Bacmid isolation: The 2 ml culture is centrifuged at 2900 g for 10 min and the supernatant discarded. The pellet was resuspended in 300 µl of buffer P1 (Qiagen plasmid miniprep kit) followed by 300 µl of buffer P2. The sample is gently inverted; 300 µl of buffer N3 is added and centrifuged for 10 min in a bench top centrifuge at maximum speed. The supernatant is transferred to a fresh tube and about 700 µl of 100% isopropanol is added. After gentle mixing, the sample is again centrifuged for 10 min at maximum speed. The supernatant is carefully removed avoiding disturbing the pellet and 200 µl of 70% ethanol is slowly added. The sample is centrifuged for 5 min at maximum speed and the ethanol removed. 50 µl of 70% ethanol is again added slowly and tube transferred to sterile biological safety cabinet. In the biological safety cabinet, the ethanol is slowly removed and pellet dried for about 10 min. The pellet is resuspended in 20 µl of sterilized water.

Transfection: A transfection reagent mixture is prepared by adding 100 µl of Sf21 medium to 10 µl of XtremeGene transfection reagent. The DNA is diluted with 200 µl of Sf 900-II medium (Gibco- Life Technologies) followed by 100 µl of XtremeGene transfection reagent (Roche) mixture. The resultant solution is then added to 3 ml of Sf 21 cells (0.5-1 * 10^6 cells/ml) in a 6 well plates in duplicates. After about 48-60 h, the infected cells are observed to be larger in size. The supernatant is slowly removed and forms the P_0 (initial virus stock).

The 3 ml of virus P_0 is used for infecting 25 ml culture at a density of 0.5-0.8 * 10^6 cells/ml. The virus (P1-1st generation) is obtained from the supernatant about 24 h after the cells stop dividing (i.e day of proliferation arrest). The virus can be stored at 4 °C for few weeks.

Large scale purification: The large scale insect cells expression involved infection of about 50-100 ml of Sf21 cells at concentration of 0.7 * 10^6 cells/ml with 1-2 % of virus P1 (1st generation) and harvesting about 48-72 h post day of proliferation arrest. The cells were harvested by centrifugation at 1200 g and 4°C for ~ 10 min, resuspended in lysis buffer (30 mM Tris, pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 1 mM Tris (2-carboxyethyl)phosphine hydrochloride) (TCEP), 1x protease inhibitor (Roche, EDTA free) and thawed using two cycles of alternate liquid N_2 and 37° C water bath. The supernatant was isolated using centrifugation at 18,000 rpm at 4°C for 30 min and applied to a 1 ml Ni-NTA column (GE Healthcare) pre-equilibrated with lysis buffer. After a brief washing step with wash buffer (30 mM Tris pH8.0, 300 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, 30 mM Imidazole), the bound protein was eluted using the elution buffer (30 mM Tris pH8.0, 300 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, 300 mM Imidazole). The eluent was concentrated to ~ 500 µl and purified using size exclusion chromatography with Superose 6 10/300 GL (GE Healthcare) on AKTA purified system (GE Healthcare) pre-equilibrated with lysis buffer.

DNA target	Primer sequence
His-GST-SVP in pAcebac1	5'-TCACTTGTCGCGGCCGCATGGCACATCACCATCACCATCACAT-3' 5'-TGCAGGCTCTAGATTAGCCGCCATACGGCAGGCC-3'
His-SUMO-SVP in pAcebac1	5'-TCACTTGTCGCGGCCGCATGGCACACCATCACCACCAT-3' 5'-TGCAGGCTCTAGATTAGCCGCCATACGGCAGGCC-3'
***SVP2* for AFM**	5'-CACTACAAACTAGAACAAATAAATGCCACGTGAAATC-3' 5'-CTATAAGAGTTGGTTTCCTATTCCATTCTCTTTTC-3'
***SVP3* for AFM**	5'-CATCTCCATAGTCGTCCCTTGAATCTTGGTC-3' 5'-CAAAATCTAGATTTGTCTCTTTCCAAAATGGTTGAC-3'
***SVP2.1* for EMSA (Fwd-TAMRA labelled)**	5'-GCAATGTCGTGAAGGGTCG-3' 5'-CCTTCCCATTACGTCTTGACG-3'
***SVP2.2* for EMSA (Fwd-TAMRA labelled)**	5'-GCAATGTCGTGAAGGGTCG-3' 5'-CCCAGACGTGACTTGTTTGACG-3'
***SVP2.1* for MST**	5'-GCAATGTCGTGAAGGGTCG-3' 5'-CCTTCCCATTACGTCTTGACG-3'

Table IV-6: ***Primers used for DNA amplification.*** *The different primers used for amplification of DNA target sequences are provided here.*

IV.5.2 Biophysical characterization and DNA binding studies

IV.5.2.1 Crystallisation

Native SVP, 6xHis-SUMO-SVP and 6xHis-GST-SVP at a concentration of 1.5 mg/ ml, 3 mg/ml and 4.9 mg/ml respectively in 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol were subjected to high throughput crystallisation trials using the EMBL-Grenoble HTX facility (https://embl.fr/htxlab/). 200 nl sitting drops were set up using Cartesian PixSys 4200 (Genomic Solutions, UK) crystallisation robot using the Greiner CrystalQuick plates (flat bottom, untreated) and imaged with a Formulatrix Rock Imager (Formulatrix Inc.,USA) at 277 K (Dimasi et al., 2007). Commercial crystal screens from Hampton Research (Aliso Viejo, California, USA), Qiagen (Hilden, Germany) and Molecular Dimensions (Suffolk, UK) and Rigaku (Bainbridge island, WA, USA were used in robotic screening trials.

IV.5.2.2 Small Angle X-ray Scattering

SAXS data collection for all three SVP constructs was performed similar to that for AGAMOUS with an on line HPLC system at ESRF beamline (BM29). For each construct, the sample (50 µl) was injected onto the column (Superose 6 3.2/300 PC, GE Healthcare) equilibrated beforehand with 2-5 CV of gel filtration buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP). Multiple buffer tests were performed prior to data collection to verify column equilibration and absence of residual glycerol. The initial concentrations used were 4 mg/ml for 6xHis-GST-SVP, 1.5 mg/ml for 6xHis-SUMO-SVP and 1 mg/ml for native SVP constructs. Approximately 1800 frames (1 frame/sec) per HPLC run were collected. The data collection parameters and initial data processing followed the same pattern as for AGAMOUS. The initial data processing was performed automatically using the EDNA pipeline (Incardona et al., 2009). Frames with consistent R_g were merged using PRIMUSqt from the ATSAS package version 2.7.1. The peaks of interest were reprocessed manually to maximize the signal-to-noise ratio. 20 frames for SVP native, 60 for 6xHis-SUMO-SVP and 45 frames for 6xHis-GST-SVP corresponding to highest protein concentration (based on UV absorbance and I(0) values) were merged and used for all further data processing and model fitting.

IV.5.2.2.1 Data analysis

The data analysis was performed using different tools from the ATSAS package (Petoukhov et al., 2012) and ScÅtter (Förster et al., 2010) from the BIOISIS package. The data quality was assessed using the scattering intensity profile and the Guinier approximation. The molecular weight, volume and pair distribution function (P(r)), maximum particle dimension (D_{max}) values were estimated using PRIMUS (Konarev et al., 2003) from the one dimensional scattering curves. The Kratky plots were used for flexibility analysis.

IV.5.2.2.2 Ab-initio model fitting

Ab-initio models representing the solution structure were generated for all the three SVP constructs. Using the distance distribution function P(r) as input files, 50 models were generated using the program DAMMIF (Franke and Svergun, 2009). Models were generated without symmetry constraints. The resultant models were aligned, averaged and the most probable models were selected with the program suite DAMAVER (Volkov and Svergun, 2003). The final bead models generated were visualized using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

IV.5.2.3 Microscale thermophoresis

20 µM of 6xHis-SUMO-SVP was labelled with dye NT-647 (MO-1004, Nano Temper Technologies Gmbh, Munich, Germany) in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl2, 0.05% Tween-20) and purified using a desalt column. 10 µM of unlabelled protein was titrated against 10 nM (10 µl) of labelled protein and a serial dilution was performed for 16 samples making the total volume of 20 µl. The 16 samples were loaded in premium capillaries (MO-AK005, Nano Temper Technologies Gmbh, Munich, Germany) and a capillary scan was performed to assess the homogeneity and stability of the samples. After validation, a MST scan was performed with MST NT.115 (LED: 30%, MST: 40%). The procedure was repeated with an additional component of unlabelled *100 nM SVP 2.1* DNA.

IV.5.2.4 Electrophoretic Mobility Shift Assays (EMSA)

SVP2 and *SVP3* DNA amplified for AFM scans were used as templates for synthesis of DNA for EMSA. Two binding DNA strands *SVP2.1* and *SVP2.2* (named for convenience)

were amplified from SVP comprising 1 CArG box and 2 CArG box sites, respectively. The forward PCR primer was labelled with TAMRA (Eurofins Genomics). The details of the primers are given in Table IV-6. The PCR product was gel purified and protected from light using aluminium foil. Native SVP and DNA were incubated at 4°C for 10-15 min in a buffer of 30 mM Tris pH 8.0, 1 mM $MgCl_2$ and 5% glycerol. The DNA concentration was kept constant at 10 nm while the protein concentration was gradually increased. The complex was run on a 5% polyacrylamide gel using Tris/borate/EDTA buffer under nondenaturing conditions at 4°C then the gel scanned on a Typhoon scanner (GE Healthcare). The gel was pre-run for ~ 1-2 h at 4°C before loading the samples.

IV.5.2.5 Atomic force microscopy

The target DNA sequences *SVP2* (1033 bp) and *SVP3* (1004 bp) were PCR amplified from Arabidopsis genomic DNA using primers given in Table IV-6. A 50 µl reaction volume comprising 5x Phusion HF buffer, 10 mM dNTPs, 0.5 µl Phusion polymerase and 10µM forward and reverse primers (Table IV-6), 50ng genomic DNA (*Arabidopsis thaliana*) template was used to amplify the DNA target sequences. The PCR product was gel purified using QIAquick gel extraction kit (Qiagen Gmbh Germany).

The AFM scans were performed similar to that for SEPALLATA 3 described earlier using a Cypher S atomic force microscope (Asylum Research, Santa Barbara, CA). Images were obtained both at 256 * 256 pixels and 512 * 512 pixels with a scan size between 0.2- 2 um and processed using Gwyddion (Nečas and Klapetek, 2011). Image processing comprised filtering, flattening and masking in order to remove the background noise. The initial scans were performed using *SVP2* later followed with *SVP3*. A thorough optimisation process was necessary to obtain good quality images.

IV.5.2.6 High Speed AFM

The HS-AFM scans were performed at Bio-AFM Frontier research centre at University of Kanazawa, Kanazawa, Japan in collaboration with Dr. Takahiro Nakayama. The target DNA sequences *SVP2* and *SVP3* were PCR amplified and gel purified. The AFM scans were performed on the small round mica sheet (radius = 2 mm) attached to a glass stage. The instrument used for the experiment was built in-house at the University of Kanazawa. BioLever-fast BL-AC10DS-A2 (Olympus Co. Tokyo, Japan) probes with a cantilever resonant frequency of ~ 400 kHz in liquid was used for the scan. Approximately 2 nM of

SVP3 DNA was added on the freshly cleaved mica using adhesion buffer (10 mM $NiSO_4$, 10 mM HEPES, pH 8.0). Preliminary scans were performed to confirm proper DNA binding and distribution. During the scan, 30 mM KCl and 30 mM NaCl were added in order to provide appropriate conditions for enabling DNA partial mobility. This was followed by addition of ~ 2.5 µl of 250 nM native SVP to chamber (~10 nM final concentration). The scans were performed at multiple positions to analyze different DNA-protein interaction. Images were obtained both at 256 * 256 pixels and 512 * 512 pixels with a scan size between 0.2- 1 µm and scan rate of ~ 2 sec/frame. The imaged were pprocessed using Gwyddion (Nečas and Klapetek, 2011), Image J (Schneider et al., 2012) and Kodec4 (Sakashita et al.,2013).

Chapter V

CONCLUSION

V Conclusion

MADS TFs play a significant role in various processes involved in the plant life cycle, including flowering and floral organ differentiation. In this project, three important plant MADS TFs were examined, SEP3 and AG involved in floral organ development and SVP involved in flowering time. Although they have been highly studied genetically, very limited structural information is available, primarily due to difficulties in recombinant protein expression and purification. In this project, using high throughput library generation techniques and parallel expression trials in different expression systems, purifiable soluble protein constructs are identified. This opens up plant MADS TFs for a wide range of biochemical and biophysical studies.

A high-resolution crystal structure for the oligomerization (K) domain of SEP3 is presented here. The structure reveals that the SEP3 K domain is a tetramer composed of two alpha helices separated by hydrophobic kink region and aligned at an angle of 90° from each other. This arrangement provides a dynamic interaction surface for SEP3 allowing it to form homo and heterodimers and tetramers with other interaction partners. Mutagenesis studies have shown that the SEP3 tetramer is weak and perturbed by single point mutations in the hydrophobic region. However, the AFM studies presented here show the even weak tetramers are capable of binding and looping DNA. This may be a general characteristic of MADS TFs whose tetramerisation is dependent on DNA binding *in vivo*.

The AG K domain construct was generated using a homologous region from the SEP3 K domain. However, the AG K domain is dimeric as shown by SEC and SAXS studies. The comparison of the residues between SEP3 and AG suggests that the absence of a Gly-Pro motif in the kink region may be responsible for the preference of AG to form homodimers and not homotetramers. This suggests that even small differences between the amino acid sequences can lead to a major shift in the oligomeric state of MADS TFs. The Kratky plot suggested some degree of flexibility making it difficult to provide an accurate shape and conformation for the AG K domain. However, two likely models have been proposed here, with a bent and elongated conformation, representing extremes of potential AG dimer conformation.

In the case of SVP, a high yielding purification protocol for the full length protein was established which has opened up prospects for its biophysical and biochemical characterisation. SEC-MALLS reveals that SVP is a dimer in solution even at relatively high

protein concentration. SAXS studies show a high degree of flexibility of the protein in solution. MST experiments reveal a high binding affinity for protein-protein interactions. AFM experiments confirmed the DNA binding ability of the protein. Although solution studies of SVP suggested a dimer, some AFM experiments show formation of higher order oligomers when bound to DNA with two CArG box binding sites. Overall, it can be stated that SVP is a dynamic molecule, which can form a homodimer and bind to DNA. DNA binding may act as a driver for tetramer or higher order oligomer formation.

Thus, through these studies of selected MADS TFs, it seems likely that the K domain forms one of the major interaction surfaces and the flexible kink region assists in formation of diverse complexes. DNA-binding may act as a driver for oligomerisation, with only limited tetramerisation occurring in the absence of DNA binding. The MADS TFs form a dynamic interaction network to form complexes with different proteins to specifically regulate target genes thus controlling important physiological processes. The study presented here, provides valuable information regarding structural and functional aspects of MADS TFs and opens up new vistas for their study.

Bibliography

Abremski, K., Hoess, R., and Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: Evidence for topologically unlinked products following recombination. Cell **32**: 1301–1311.

Acajjaoui, S. and Zubieta, C. (2013). Crystallization studies of the keratin-like domain from Arabidopsis thaliana SEPALLATA 3. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **69**: 997–1000.

Airoldi, C. a, Bergonzi, S., and Davies, B. (2010). Single amino acid change alters the ability to specify male or female organ identity. Proc. Natl. Acad. Sci. U. S. A. **107**: 18898–902.

Alvarez-Buylla, E.R., Liljegren, S.J., Pelaz, S., Gold, S.E., Burgeff, C., Ditta, G.S., Vergara-Silva, F., and Yanofsky, M.F. (2000a). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. Plant J. **24**: 457–66.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martínez-Castilla, L., and Yanofsky, M.F. (2000b). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. U. S. A. **97**: 5328–5333.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K., and Toda, A. (2001). A high-speed atomic force microscope for studying biological macromolecules. Proc. Natl. Acad. Sci. U. S. A. **98**: 12468–72.

Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Miyagi, A., Taniguchi, M., and Yamashita, H. (2008). High-speed AFM and nano-visualization of biomolecular processes. Pflügers Arch. Eur. J. Physiol. **456**: 211–25.

Ando, T., Uchihashi, T., and Scheuring, S. (2014). Filming biomolecular processes by high-speed atomic force microscopy. Chem. Rev. **114**: 3120–3188.

Andrés, F. and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. **13**: 627–39.

Angenent, G.C. and Colombo, L. (1996). Molecular control of ovule development. Trends Plant Sci. **1**: 228–232.

Augusto, L., Davies, T.J., Delzon, S., and De Schrijver, a. (2014). The enigma of the rise

of angiosperms: Can we untie the knot? Ecol. Lett. **17**: 1326–1338.

Becker, A. and Ehlers, K. (2015). Arabidopsis flower development-of protein complexes, targets, and transport. Protoplasma.

Beloin, C., Jeusset, J., Revet, B., Mirambeau, G., Le Hégarat, F., and Le Cam, E. (2003). Contribution of DNA conformation and topology in right-handed DNA wrapping by the Bacillus subtilis LrpC protein. J. Biol. Chem. **278**: 5333–42.

Bemer, M., Wolters-Arts, M., Grossniklaus, U., and Angenent, G.C. (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. Plant Cell **20**: 2088–101.

van den Berg, B., Wain, R., Dobson, C.M., and Ellis, R.J. (2000). Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. EMBO J. **19**: 3870–5.

Berger, I. (2014). Complex Expressions by BEVS. In.

Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. Nat. Biotechnol. **22**: 1583–7.

Berger, I., Garzoni, F., Chaillet, M., Haffke, M., Gupta, K., and Aubert, A. (2013). The multiBac protein complex production platform at the EMBL. J. Vis. Exp.: e50159.

Bernadó, P. and Blackledge, M. (2009). A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys. J. **97**: 2839–45.

Berrow, N.S., Alderton, D., and Owens, R.J. (2009). The precise engineering of expression vectors using high-throughput In-Fusion PCR cloning. Methods Mol. Biol. **498**: 75–90.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: Expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. **37**: 49–57.

Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., Romier, C., Berger, P., Richmond, T.J., Steinmetz, M.O., and Berger, I. (2009). Automated unrestricted multigene recombineering for multiprotein complex production. Nat. Methods **6**: 447–50.

Bieniossek, C., Richmond, T.J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr. Protoc. Protein Sci. **Chapter 5**: Unit 5.20.

Bird, L.E. (2011). High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. Methods **55**: 29–37.

Bird, L.E., Rada, H., Flanagan, J., Diprose, J.M., Gilbert, R.J.C., and Owens, R.J.

(2014). Application of In-Fusion™ cloning for the parallel construction of E. coli expression vectors. Methods Mol. Biol. **1116**: 209–34.

Blazquez, M.A. (2000). Flower development pathways. J. Cell Sci. **113**: 3547–3548.

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouzé, P., Theissen, G., and Van de Peer, Y. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J. Mol. Evol. **56**: 573–86.

Boldon, L., Laliberte, F., and Liu, L. (2015). Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. Nano Rev. **6**: 25661.

Bossinger, G. and Smyth, D.R. (1996). Initiation patterns of flower and floral organ development in Arabidopsis thaliana. Development **122**: 1093–102.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in Arabidopsis. Plant Cell **1**: 37–52.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. Development **112**: 1–20.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (2012). The ABC model of flower development: then and now. Development **139**: 4095–4098.

Braga, P.C. and Ricci, D. (2004). Atomic Force Microscopy: Biomedical Methods and Applications (Springer Science & Business Media).

Burgess, R.R. (2009). Refolding solubilized inclusion body proteins. Methods Enzymol. **463**: 259–82.

Burley, S.K. (2000). An overview of structural genomics. Nat. Struct. Biol. **7 Suppl**: 932–4.

Casuso, I., Rico, F., and Scheuring, S. (2010). Biological AFM: where we come from--where we are--where we may go. J. Mol. Recognit. **24**: 406–13.

Casuso, I., Rico, F., and Scheuring, S. (2011). High-speed atomic force microscopy: Structure and dynamics of single proteins. Curr. Opin. Chem. Biol. **15**: 704–9.

Cerreta, A., Vobornik, D., and Dietler, G. (2013). Fine DNA structure revealed by constant height frequency modulation AFM imaging. Eur. Polym. J. **49**: 1916–1922.

Chandler, J.W. (2012). Floral meristem initiation and emergence in plants. Cell. Mol. Life Sci. **69**: 3807–18.

Chen, L., Haushalter, K. a, Lieber, C.M., and Verdine, G.L. (2002). Direct visualization of a DNA glycosylase searching for damage. Chem. Biol. **9**: 345–50.

Chen, S.H. and Bendedouch, D. (1986). Structure and interactions of proteins in solution studied by small-angle neutron scattering. Methods Enzymol. **130**: 79–116.

Coen, E.S. and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature **353**: 31–7.

Colombo, L., Battaglia, R., and Kater, M.M. (2008a). Arabidopsis ovule development and its evolutionary conservation. Trends Plant Sci. **13**: 444–50.

Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M.M., and Colombo, L. (2008b). AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. Plant J. **54**: 1037–48.

Crampton, N., Bonass, W. a., Kirkham, J., and Thomson, N.H. (2005). Formation of aminosilane-functionalized mica for atomic force microscopy imaging of DNA. Langmuir **21**: 7884–7891.

Dahlgren, Dahlgren, Lyubchenko, and Lyubchenko (2002). Atomic Force Microscopy Study of the Effects of Mg2+ and Other Divalent Cations of the End-to-End DNA Interactions. Biochemistry **41**: 11372–11378.

van Dijk, A.D.J., Morabito, G., Fiers, M., van Ham, R.C.H.J., Angenent, G.C., and Immink, R.G.H. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. PLoS Comput. Biol. **6**: e1001017.

Dimasi, N., Flot, D., Dupeux, F., and Márquez, J.A. (2007). Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **63**: 204–8.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr. Biol. **14**: 1935–40.

Doniach, S. (2001). Changes in biomolecular conformation seen by small angle X-ray scattering. Chem. Rev. **101**: 1763–1778.

Dornelas, M.C., Patreze, C.M., Angenent, G.C., and Immink, R.G.H. (2011). MADS: the missing link between identity and growth? Trends Plant Sci. **16**: 89–97.

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics **21**: 3433–4.

Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination (Elsevier).

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and

Fieschi, F. (2010). NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. J. Struct. Biol. **169**: 45–53.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L., and McCafferty, J. (2004). Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnol. **4**: 32.

Egea-Cortines, M., Saedler, H., and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. EMBO J. **18**: 5370–9.

Ellis, R.J. (2006). Molecular chaperones: assisting assembly in addition to folding. Trends Biochem. Sci. **31**: 395–401.

Endrizzi, K., Moussian, B., Haecker, A., Levin, J.Z., and Laux, T. (1996). The SHOOT MERISTEMLESS gene is required for maintenance of undifferentiated cells in Arabidopsis shoot and floral meristems and acts at a different regulatory level than the meristem genes WUSCHEL and ZWILLE. Plant J. **10**: 967–79.

Esposito, D. and Chatterjee, D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. Curr. Opin. Biotechnol. **17**: 353–8.

Evans, G. and Pettifer, R.F. (2001). CHOOCH : a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. J. Appl. Crystallogr. **34**: 82–86.

Fan, H.Y., Hu, Y., Tudor, M., and Ma, H. (1997). Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. Plant J. **12**: 999–1010.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M., and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. Plant Cell **15**: 2603–2611.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods **3**: 1021–32.

Folta-Stogniew, E. (2006). Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. Methods Mol. Biol. **328**: 97–112.

Folta-Stogniew, E. and Williams, K.R. (1999). Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. J. Biomol. Tech. **10**: 51–63.

de Folter, S. et al. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell **17**: 1424–33.

de Folter, S. and Angenent, G.C. (2006). trans meets cis in MADS science. Trends Plant Sci. **11**: 224–31.

Förster, S., Apostol, L., and Bras, W. (2010). Scatter: software for the analysis of nano- and mesoscale small-angle scattering. J. Appl. Crystallogr. **43**: 639–646.

Franke, D. and Svergun, D.I. (2009). DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. **42**: 342–346.

Friedman, W.E. (2009a). The meaning of Darwin's "abominable mystery." Am. J. Bot. **96**: 5–21.

Friedman, W.E. (2009b). The meaning of Darwin's "abominable mystery". Am. J. Bot. **96**: 5–21.

Friis, E.M., Pedersen, K.R., and Crane, P.R. (2010). Diversity in obscurity: fossil flowers and the early history of angiosperms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **365**: 369–382.

Gabadinho, J. et al. (2010). MxCuBE: a synchrotron beamline control environment customized for macromolecular crystallography experiments. J. Synchrotron Radiat. **17**: 700–7.

Gileadi, O., Burgess-Brown, N. a, Colebrook, S.M., Berridge, G., Savitsky, P., Smee, C.E. a, Loppnau, P., Johansson, C., Salah, E., and Pantic, N.H. (2008). High throughput production of recombinant human proteins for crystallography. Methods Mol. Biol. **426**: 221–246.

Gramzow, L., Ritz, M.S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. Trends Genet. **26**: 149–153.

Gramzow, L. and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. Genome Biol. **11**: 214.

Gregis, V. et al. (2013). Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. Genome Biol. **14**: R56.

Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2008). AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. Plant J. **56**: 891–902.

Gregis, V., Sessa, A., Dorca-Fornell, C., and Kater, M.M. (2009). The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral

homeotic genes. Plant J. **60**: 626–37.

Grow, R.J., Minne, S.C., Manalis, S.R., and Quate, C.F. (2002). Silicon nitride cantilevers with oxidation-sharpened silicon tips for atomic force microscopy. J. Microelectromechanical Syst. **11**: 317–321.

di Guana, C., Lib, P., Riggsa, P.D., and Inouyeb, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. Gene **67**: 21–30.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortin, J., Hart, D.J., and Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. **15**: 500–506.

Guinier, A. (1938). Examination of fine particles and colloidal suspensions by means of x-rays diffused at very small angles. Comptes Rendus: 206, 1374–6.

Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1994). Regulation of the arabidopsis floral homeotic gene APETALA1. Cell **76**: 131–143.

Hammarström, M., Hellgren, N., van Den Berg, S., Berglund, H., and Härd, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci. **11**: 313–21.

Hansma, H.G., Golan, R., Hsieh, W., Lollo, C.P., Mullen-Ley, P., and Kwoh, D. (1998). DNA condensation for gene therapy as monitored by atomic force microscopy. Nucleic Acids Res. **26**: 2481–2487.

Hansma, H.G., Kasuya, K., and Oroudjev, E. (2004). Atomic force microscopy imaging and pulling of nucleic acids. Curr. Opin. Struct. Biol. **14**: 380–5.

Hansma, H.G. and Laney, D.E. (1996). DNA binding to mica correlates with cationic radius: assay by atomic force microscopy. Biophys. J. **70**: 1933–1939.

Hansma, P.K. et al. (1994). Tapping mode atomic force microscopy in liquids. Appl. Phys. Lett. **64**: 1738.

Harper, S. and Speicher, D.W. (2011). Protein Chromatography D. Walls and S.T. Loughran, eds (Humana Press: Totowa, NJ).

Hart, D.J. and Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in Escherichia coli. Acta Crystallogr. Sect. D Biol. Crystallogr. **62**: 19–26.

Hartmann, U. et al. (2000). Molecular cloning of SVP : a negative regulator of the Floral transition in Arabidopsis. Plant J. **21**: 351–360.

Heddle, J.G., Mitelheiser, S., Maxwell, A., and Thomson, N.H. (2004). Nucleotide binding to DNA gyrase causes loss of DNA wrap. J. Mol. Biol. **337**: 597–610.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., and Theissen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Mol. Biol. Evol. **19**: 801–14.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409**: 525–529.

Hu, J., Wang, M., Frantz, P., Kolbe, W., Ogletree, D.F., and Salmeron, M. (1996). Imaging of Single Extended DNA Molecules on Flat (Aminopropyl) triethoxysilane - Mica by Atomic Force. Langmuir **12**: 1697–1700.

Huang, K., Louis, J.M., Donaldson, L., Lim, F.L., Sharrocks, A.D., and Clore, G.M. (2000). Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. EMBO J. **19**: 2615–2628.

Immink, R.G.H., Tonaco, I. a N., de Folter, S., Shchennikova, A., van Dijk, A.D.J., Busscher-Lange, J., Borst, J.W., and Angenent, G.C. (2009). SEPALLATA3: the "glue" for MADS box transcription factor complex formation. Genome Biol. **10**: R24.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J. Synchrotron Radiat. **16**: 872–9.

Ito, T. (2011). Coordination of flower development by homeotic master regulators. Curr. Opin. Plant Biol. **14**: 53–9.

Ito, T., Ng, K.-H., Lim, T.-S., Yu, H., and Meyerowitz, E.M. (2007). The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in Arabidopsis. Plant Cell **19**: 3516–29.

Ito, T., Wellmer, F., Yu, H., Das, P., Ito, N., Alves-Ferreira, M., Riechmann, J.L., and Meyerowitz, E.M. (2004). The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. Nature **430**: 356–60.

Jang, S., Torti, S., and Coupland, G. (2009). Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis. Plant J. **60**: 614–25.

Jerabek-Willemsen, M. et al. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. J. Mol. Struct. **1077**: 101–113.

Jerabek-Willemsen, M., Wienken, C.J., Braun, D., Baaske, P., and Duhr, S. (2011).

Molecular interaction studies using microscale thermophoresis. Assay Drug Dev. Technol. **9**: 342–53.

Jetha, K., Theißen, G., and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. Nucleic Acids Res. **42**: 1–16.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292**: 195–202.

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr **66**: 125–132.

Kantardjieff, K.A. and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. **12**: 1865–71.

Kapust, R.B. and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. **8**: 1668–1674.

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347**: 183–198.

Kaufmann, K., Muiño, J.M., Jauregui, R., Airoldi, C. a, Smaczniak, C., Krajewski, P., and Angenent, G.C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biol. **7**: e1000090.

Kelly, A.E. and Goulden, M.L. (2008). Rapid shifts in plant distribution with recent climate change. Proc. Natl. Acad. Sci. U. S. A. **105**: 11823–11826.

Khoury, G.A., Baliban, R.C., and Floudas, C.A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci. Rep. **1**.

Khow, O. and Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac. J. Trop. Biomed. **2**: 159–62.

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.

Koch, M.H., Vachette, P., and Svergun, D.I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003). The Polycomb-group protein MEDEA regulates seed development by controlling

expression of the MADS-box gene PHERES1. Genes Dev. **17**: 1540–53.

Köhler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat. Genet. **37**: 28–30.

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.

Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol. Biol. Evol. **29**: 293–302.

Larkin, M.A. et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23**: 2947–8.

Larquet, E., Le Cam, E., Fourcade, A., Culard, F., Furrer, P., and Delain, E. (1996). [Complementarity of microscopies in the structural analysis of DNA minicircles associated to protein MC1]. Comptes rendus l'Académie des Sci. Série III, Sci. la vie **319**: 461–71.

Lee, B. and Richards, F.M. (1971). The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. **55**: 379–IN4.

Lee, J.H., Ryu, H.-S., Chung, K.S., Posé, D., Kim, S., Schmid, M., and Ahn, J.H. (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science **342**: 628–32.

Lee, J.H.J.S., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.H.J.S., and Ahn, J.H. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes Dev. **21**: 397–402.

Lenhard, M., Bohnert, A., Jürgens, G., and Laux, T. (2001). Termination of Stem Cell Maintenance in Arabidopsis Floral Meristems by Interactions between WUSCHEL and AGAMOUS. Cell **105**: 805–814.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C. a, Ito, T., Meyerowitz, E., and Yu, H. (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Dev. Cell **15**: 110–20.

Li, M., Su, Z.-G., and Janson, J.-C. (2004). In vitro protein refolding by chromatographic procedures. Protein Expr. Purif. **33**: 1–10.

Liu, C., Xi, W., Shen, L., Tan, C., and Yu, H. (2009). Regulation of floral patterning by flowering time genes. Dev. Cell **16**: 711–22.

Liu, Z., Li, Z., Zhou, H., Wei, G., Song, Y., and Wang, L. (2005). Observation of the mica

surface by atomic force microscopy. Micron **36**: 525–31.

Lyubchenko, Y.L., Jacobs, B.L., and Lindsay, S.M. (1992). Atomic force microscopy of reovirus dsRNA: a routine technique for length measurements. Nucleic Acids Res. **20**: 3983–3986.

Lyubchenko, Y.L., Shlyakhtenko, L.S., and Ando, T. (2011). Imaging of nucleic acids with atomic force microscopy. Methods **54**: 274–83.

Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in Escherichia coli. Microbiol. Rev. **60**: 512–38.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. Nature **360**: 273–7.

Manske, M. (2006). GENtle, a free multi-purpose molecular biology tool.

Mara, C.D. and Irish, V.F. (2008). Two GATA transcription factors are downstream effectors of floral homeotic gene action in Arabidopsis. Plant Physiol. **147**: 707–18.

De Maria Antolinos, A. et al. (2015). ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments. Acta Crystallogr. Sect. D Biol. Crystallogr. **71**: 76–85.

Martel, A., Liu, P., Weiss, T.M., Niebuhr, M., and Tsuruta, H. (2012). An integrated high-throughput data acquisition system for biological solution X-ray scattering studies. J. Synchrotron Radiat. **19**: 431–4.

Masiero, S., Colombo, L., Grini, P.E., Schnittger, A., and Kater, M.M. (2011). The emerging importance of type I MADS box transcription factors for plant reproduction. Plant Cell **23**: 865–72.

Masiero, S., Imbriano, C., Ravasio, F., Favaro, R., Pelucchi, N., Gorla, M.S., Mantovani, R., Colombo, L., and Kater, M.M. (2002). Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. J. Biol. Chem. **277**: 26429–26435.

Mathelier, A. et al. (2015). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res.: gkv1176–.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. **33**: 491–7.

McCarthy, A.A., Brockhauser, S., Nurizzo, D., Theveneau, P., Mairs, T., Spruce, D., Guijarro, M., Lesourd, M., Ravelli, R.B.G., and McSweeney, S. (2009). A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. J. Synchrotron Radiat. **16**: 803–12.

Melzer, R. and Theissen, G. (2009). Reconstitution of “floral quartets” in vitro involving

class B and class E floral homeotic proteins. Nucleic Acids Res. **37**: 2723–36.

Melzer, R., Verelst, W., and Theissen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in "floral quartet"-like complexes in vitro. Nucleic Acids Res. **37**: 144–57.

Mendes, M.A., Guerra, R.F., Berns, M.C., Manzo, C., Masiero, S., Finzi, L., Kater, M.M., and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. Plant Cell **25**: 2560–72.

Mertens, H.D.T. and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. **172**: 128–41.

Milhiet, P.-E., Yamamoto, D., Berthoumieu, O., Dosset, P., Le Grimellec, C., Verdier, J.-M., Marchal, S., and Ando, T. (2010). Deciphering the Structure, Growth and Assembly of Amyloid-Like Fibrils Using High-Speed Atomic Force Microscopy. PLoS One **5**: e13240.

Mizukami, Y., Huang, H., Tudor, M., Hu, Y., and Ma, H. (1996). Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. Plant Cell **8**: 831–45.

Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001). Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. J. Mol. Biol. **314**: 495–506.

Moreno-Herrero, F., Colchero, J., and Baró, A.M. (2003). DNA height in scanning force microscopy. Ultramicroscopy **96**: 167–174.

Moreno-Herrero, F., Colchero, J., Gómez-Herrero, J., and Baró, A.M. (2004). Atomic force microscopy contact, tapping, and jumping modes for imaging biological samples in liquids. Phys. Rev. E **69**: 031915.

Mouradov, A., Cremer, F.F., and Coupland, G. (2002). Control of Flowering Time: Interacting Pathways as a Basis for Diversity. Plant Cell **14**: S111–30.

Muiño, J.M., Smaczniak, C., Angenent, G.C., Kaufmann, K., and van Dijk, A.D.J. (2013). Structural determinants of DNA recognition by plant MADS-domain transcription factors. Nucleic Acids Res.: 1–9.

Müller, D.J. and Dufrêne, Y.F. (2008). Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology. Nat. Nanotechnol. **3**: 261–9.

Nečas, D. and Klapetek, P. (2011). Gwyddion: an open-source software for SPM data analysis. Cent. Eur. J. Phys. **10**: 181–188.

Nilsson, O. (2013). Plant science. A pathway to flowering--why staying cool matters.

Science **342**: 566–7.

Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988). Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell **55**: 989–1003.

Nurrish, S.J. and Treisman, R. (1995). DNA binding specificity determinants in MADS-box transcription factors. Mol. Cell. Biol. **15**: 4076–85.

Ó'Maoiléidigh, D.S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling Arabidopsis thaliana flower development. New Phytol. **201**: 16–30.

Okusa, H., Kurihara, K., and Kunitake, T. (1994). Chemical modification of molecularly smooth mica surface and protein attachment. Langmuir **10**: 3577–3581.

ÓMaoiléidigh, D.S. et al. (2013). Control of Reproductive Floral Organ Identity Specification in Arabidopsis by the C Function Regulator AGAMOUS. Plant Cell **25**: 2482–2503.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M.R., Segard, S., Bernadó, P., Charavay, C., and Blackledge, M. (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics **28**: 1463–70.

Parcy, F., Nilsson, O., Busch, M.A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. Nature **395**: 561–566.

Parenicová, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., and Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell **15**: 1538–51.

Parot, P., Dufrêne, Y.F., Hinterdorfer, P., Le Grimellec, C., Navajas, D., Pellequer, J.-L., and Scheuring, S. (2010). Past, present and future of atomic force microscopy in life sciences and medicine. J. Mol. Recognit. **20**: 418–31.

Passmore, S., Maine, G.T., Elble, R., Christ, C., and Tye, B.K. (1988). Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. J. Mol. Biol. **204**: 593–606.

Pastré, D., Piétrement, O., Fusil, S., Landousy, F., Jeusset, J., David, M.-O., Hamon, L., Le Cam, E., and Zozime, A. (2003). Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study. Biophys. J. **85**: 2507–18.

Pelaz, S., Ditta, G.S., Yanofsky, M.F., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes.

Nature **405**: 200–3.

Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.L., and Yanofsky, M.F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. Plant J. **26**: 385–94.

Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA.pdf. Nature **376**: 490–8.

Pérez, J., Vachette, P., Russo, D., Desmadril, M., and Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. J. Mol. Biol. **308**: 721–43.

Pernot, P. et al. (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. J. Phys. Conf. Ser. **247**: 012009.

Pernot, P. et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. J. Synchrotron Radiat. **20**: 660–4.

Petoukhov, M. V and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J. **89**: 1237–50.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Crystallogr. **45**: 342–350.

Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S.J., Baumann, E., Wisman, E., and Yanofsky, M.F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature **424**: 85–88.

Piwarzyk, E., Yang, Y., and Jack, T. (2007). Conserved C-terminal motifs of the Arabidopsis proteins APETALA3 and PISTILLATA are dispensable for floral organ identity function. Plant Physiol. **145**: 1495–505.

Poethig, R.S. (2003). Phase change and the regulation of developmental timing in plants. Science **301**: 334–6.

Pollock, R. and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. **18**: 6197–6204.

Pollock, R. and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5**: 2327–2341.

Portereiko, M.F., Lloyd, A., Steffen, J.G., Punwani, J.A., Otsuga, D., and Drews, G.N. (2006). AGL80 is required for central cell and endosperm development in Arabidopsis. Plant Cell **18**: 1862–72.

Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G.H., and Schmid, M. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature **503**: 414–7.

Puranik, S. et al. (2014). Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in Arabidopsis. Plant Cell: 1–14.

Putman, C.A.J., Van der Werf, K.O., De Grooth, B.G., Van Hulst, N.F., and Greve, J. (1994). Tapping mode atomic force microscopy in liquid. Appl. Phys. Lett. **64**: 2454.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. **40**: 191–285.

Rambo, R.P. and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature **496**: 477–81.

Receveur-Brechot, V. and Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. Curr. Protein Pept. Sci. **13**: 55–75.

Riechmann, J.L., Krizek, B.A., and Meyerowitz, E.M. (1996). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. Proc. Natl. Acad. Sci. U. S. A. **93**: 4793–8.

Rivetti, C., Guthold, M., and Bustamante, C. (1999). Wrapping of DNA around the E.coli RNA polymerase open promoter complex. EMBO J. **18**: 4464–75.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. Front. Microbiol. **5**: 172.

Round, A., Brown, E., Marcellin, R., Kapp, U., Westfall, C.S., Jez, J.M., and Zubieta, C. (2013). Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. Acta Crystallogr. D. Biol. Crystallogr. **69**: 2072–80.

Rounsley, S.D., Ditta, G.S., and Yanofsky, M.F. (1995). Diverse roles for MADS box genes in Arabidopsis development. Plant Cell **7**: 1259–69.

Sahdev, S., Khattar, S.K., and Saini, K.S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol. Cell. Biochem. **307**: 249–64.

Santelli, E. and Richmond, T.J. (2000). Crystal structure of MEF2A core bound to DNA at 1.5 A resolution. J. Mol. Biol. **297**: 437–49.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25

years of image analysis. Nat. Methods **9**: 671–675.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*. Science **250**: 931–6.

Scipioni, A., Anselmi, C., Zuccheri, G., Samori, B., and De Santis, P. (2002). Sequence-dependent DNA curvature and flexibility from scanning force microscopy images. Biophys. J. **83**: 2408–18.

Seong, G.H., Kobatake, E., Miura, K., Nakazawa, A., and Aizawa, M. (2002). Direct atomic force microscopy visualization of integration host factor-induced DNA bending structure of the promoter regulatory region on the Pseudomonas TOL plasmid. Biochem. Biophys. Res. Commun. **291**: 361–6.

Severing, E.I., van Dijk, A.D.J., Morabito, G., Busscher-Lange, J., Immink, R.G.H., and van Ham, R.C.H.J. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. PLoS One **7**: e30524.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr. A. **64**: 112–122.

Shore, P. and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. Eur. J. Biochem. **229**: 1–13.

Sieburth, L.E., Running, M.P., and Meyerowitz, E.M. (1995). Genetic separation of third and fourth whorl functions of AGAMOUS. Plant Cell **7**: 1249–58.

Silva, C.S., Puranik, S., Round, A., Brennich, M., Jourdain, A., Parcy, F., Hugouvieux, V., and Zubieta, C. (2016). Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower. Front. Plant Sci. **6**.

Singh, S.M. and Panda, A.K. (2005). Solubilization and refolding of bacterial inclusion body proteins. J. Biosci. Bioeng. **99**: 303–10.

Skou, S., Gillilan, R.E., and Ando, N. (2014). Synchrotron-based small-angle X-ray scattering of proteins in solution. Nat. Protoc. **9**: 1727–39.

Smaczniak, C. et al. (2012a). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc. Natl. Acad. Sci. U. S. A. **109**: 1560–5.

Smaczniak, C., Immink, R.G.H., Angenent, G.C., and Kaufmann, K. (2012b). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development **139**: 3081–98.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in

Escherichia coli as fusions with glutathione S-transferase. Gene **67**: 31–40.

Smith, G.E., Summers, M.D., Fraser, M.J., G E Smith, M.D.S.M.J.F., Smith, G.E., Summers, M.D., Fraser, M.J., and G E Smith, M.D.S.M.J.F. (1983). Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol. Cell. Biol. **3**: 2156–2165.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in Arabidopsis. Plant Cell **2**: 755–67.

Soltis, D.E., Soltis, P.S., Albert, V. a, Claude, W., Frohlich, M.W., and Schiller, F. (2002). Missing links : the genetic architecture of flower and floral Floral Genome Project Research Group. Trends Plant Sci. **7**: 22–31.

Srikanth, A. and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. Cell. Mol. Life Sci. **68**: 2013–37.

Steeves, T.A. and Sussex, I.M. (1989). Patterns in plant development (Cambridge University Press: Cambridge).

Steffen, J.G., Kang, I.-H., Portereiko, M.F., Lloyd, A., and Drews, G.N. (2008). AGL61 Interacts with AGL80 and Is Required for Central Cell Development in Arabidopsis. PLANT Physiol. **148**: 259–268.

Stevens, R.C. (2000). Design of high-throughput methods of protein production for structural biology. Structure **8**: R177–R185.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. **41**: 207–34.

Sun, B. and Ito, T. (2010). Floral stem cells: from dynamic balance towards termination. Biochem. Soc. Trans. **38**: 613–6.

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.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. **25**: 495–503.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. **76**: 2879–86.

Svergun, D.I. and Koch, M.H.J. (2003). Small-angle scattering studies of biological macromolecules in solution. Reports Prog. Phys. **66**: 1735–1782.

Tan, S. and Richmond, T.J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature **391**: 660–666.

The UniProt Consortium (2014). UniProt: a hub for protein information. Nucleic Acids Res. **43**: D204–212.

Theissen, G. (2001). Development of floral organ identity: Stories from the MADS house. Curr. Opin. Plant Biol. **4**: 75–85.

Theissen, G., Kim, J.T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J. Mol. Evol. **43**: 484–516.

Theissen, G. and Saedler, H. (2001). Plant biology. Floral quartets. Nature **409**: 469–471.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. J. Struct. Biol. **172**: 45–54.

Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. Protein Expr. Purif. **28**: 1–8.

Tucker, M.R. and Laux, T. (2007). Connecting the paths in plant stem cell regulation. Trends Cell Biol. **17**: 403–10.

Uversky, V.N. and Dunker, A.K. (2010). Understanding protein non-folding. Biochim. Biophys. Acta **1804**: 1231–64.

Volkov, V. V. and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. **36**: 860–864.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. **337**: 635–45.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M. (1992). LEAFY controls floral meristem identity in Arabidopsis. Cell **69**: 843–59.

Wen, J., Arakawa, T., and Philo, J.S. (1996). Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal. Biochem. **240**: 155–66.

Willis, K. and McElwain, J. (2013). The Evolution of Plants (OUP Oxford).

Winn, M.D. et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D. Biol. Crystallogr. **67**: 235–42.

Wu, Y., Dey, R., Han, A., Jayathilaka, N., Philips, M., Ye, J., and Chen, L. (2010). Structure of the MADS-box/MEF2 Domain of MEF2A Bound to DNA and Its Implication for Myocardin Recruitment. J. Mol. Biol. **397**: 520–533.

Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules.

Anal. Chim. Acta **272**: 1–40.

Yang, Y., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. **33**: 47–59.

Yang, Y. and Jack, T. (2004). Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol. Biol. **55**: 45–59.

Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature **346**: 35–39.

Yumerefendi, H., Tarendeau, F., Mas, P.J., and Hart, D.J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. J. Struct. Biol. **172**: 66–74.

Zhao, Y., Chapman, D.A.G., and Jones, I.M. (2003). Improving baculovirus recombination. Nucleic Acids Res. **31**: E6–6.

Zheng, J., Li, Z., Wu, a, and Zhou, H. (2003). AFM studies of DNA structures on mica in the presence of alkaline earth metal ions. Biophys Chem **104**: 37–43.

Zhong, Q., Inniss, D., Kjoller, K., and Elings, V.B.B. (1993). Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy. Surf. Sci. **290**: 1–2.

Chapter VI

APPENDIX-I

VI Appendix I

The current project employed many high-throughput screening techniques for designing soluble constructs such as ESPRIT and multi-vector expression screen. Among them, ESPRIT has already been described in section II.3.1.4. This chapter presents the results obtained for multi-vector expression screen performed at OPPF, Oxford (http://www.oppf.rc-harwell.ac.uk/OPPF/) along with description of large scale insect cell expression performed at EMBL, Grenoble. The instrumentation used in AFM experiments and the theory behind biophysical techniques such as SAXS is also described here.

VI.1 Library generation and purification techniques

VI.1.1 Multi-vector expression screen

The bacterial expression system has been extensively used for production of recombinant proteins due to its fast, simple and economical nature. Using conventional approaches and library methods like ESPRIT, many soluble constructs of different target proteins were obtained as described in previous chapters. However, in many cases, studying full length protein is significant as different domains can affect the characteristics of protein binding. In such cases, use of fusion tags has been proved to increase the expression and solubility of the target protein (Esposito and Chatterjee, 2006). A large number of fusion tags can be used depending on the purification strategy, size of protein and downstream application. The multi-solubility tag screening at Oxford protein production facility (OPPF), UK specializes in parallel vector construction and expression screening in *E.coli*, mammalian and insect cells. A large number of fusion tags can be used depending on the purification strategy, size of protein and downstream application. .

The protein of interest is cloned in pOPIN suite of expression vectors (Figure VI-1) with different solubility tags like GST, MBP, SUMO using the In-fusion^{TM} cloning (Clontech Laboratories Inc., CA, USA) (Bird et al., 2014). This technique allows a high throughput cloning with no extra amino acids in final product. A common cloning site enables ease of high throughput cloning in multiple vectors. Most of pOPIN vectors are capable of protein expression in all three expression systems; bacterial, insect cells and mammalian. Details of commonly used pOPIN vectors are shown in Figure VI-2. The use of different antibiotic resistance allows co-expression of different proteins. A test expression screen is used to determine the optimal conditions for variables such as cell lines, growth media and

temperature. The final step includes Ni-NTA affinity purification of all the constructs using liquid handling robots and an SDS-PAGE analysis to determine the soluble ones. Expression tests and western blot analysis is used to determine protein expression in Sf9 and HEK cells.

Five full length proteins; SEP3, AG, AP3, PI, SVP and FLM proteins were screened using this technique as part of this project. The choice of vectors was based on earlier studies correlating the size of protein with the preferential tag that imparts maximum solubility (Bird, 2011). Co-expression tests were performed for AP3-PI and SEP3-AG in insect cell expression system. The results obtained from the SDS-PAGE analysis of the expression tests are provided here. The positive constructs obtained in these tests were further scaled up and optimized as explained in the previous chapters.

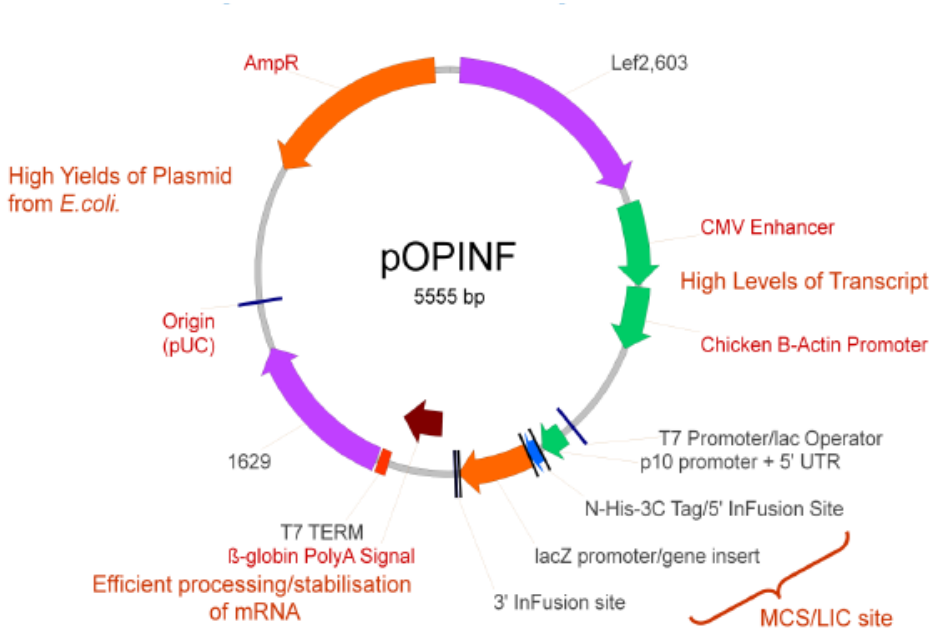


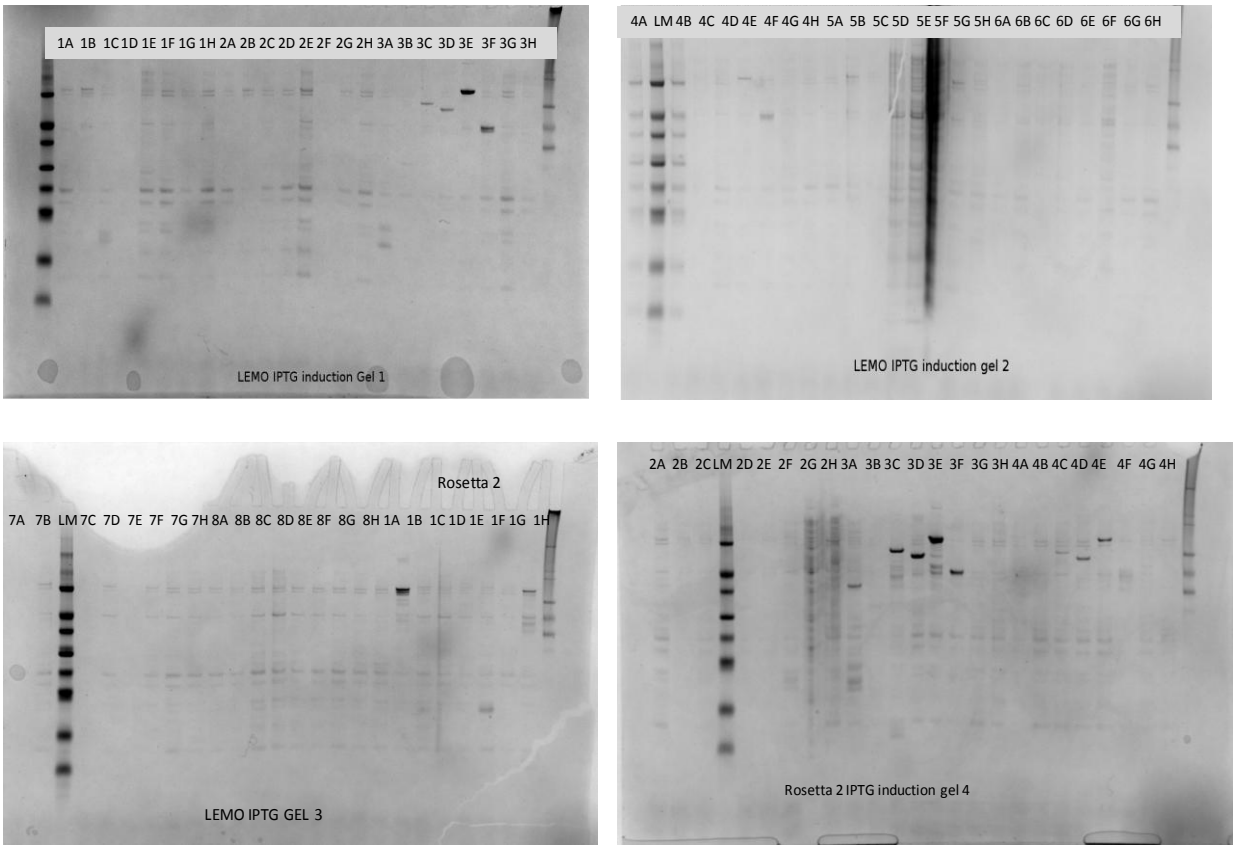
Figure VI-1: pOPINF plasmid vector with N-terminal his tag separated by 3C protease cleavage site. Image courtesy: OPPF, Oxford, UK

Vector	Fusion Tag	Parent Vector/ Antibiotic resistance	Forward Primer Extension	Reverse Primer Extension
pOPINA	...KHHHHHH tag	pET28a/Kanamycin	AGGAGATATACCATG	GTGGTGGTGGTGTTT
pOPINB	MGSSHHHHHHSSGLEVLFQ↓GP ... tag	pET28a/Kanamycin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPINE	...KHHHHHH tag	pTriEx2/Ampicillin	AGGAGATATACCATG*	GTGATGGTGATGTTT*
pOPINF	MAHHHHHHSSGLEVLFQ↓GP... tag	pTriEx2/Ampicillin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPINJ	MAHHHHHHSSG•*GST*•LEVLFQ↓GP... tag	pTriEx2/Ampicillin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPING	MGILPSPGMPALLSLVSLLSVLL MGCVA↑ETG... cleavable secretion leader and ...KHHHHHH tags	pTriEx2/Ampicillin	GCGTAGCTGAAACCGGC	GTGATGGTGATGTTT
pOPINH	MGILPSPGMPALLSLVSLLSVLLMGCV A↑ETMAHHHHHHSSGLEVLFQ↓GP ..... cleavable secretion leader and cleavable N•his tag	pTriEx2/Ampicillin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPINI	MAHHHHHHSSG... tag	pTriEx2/Ampicillin	ATCATCACAGCAGCGGC	ATGGTCTAGAAAGCTTTA
pOPINK	MAHHHHHHSSG•*GST*•LEVLFQ↓GP... tag	pET28a/Kanamycin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPINM	MAHHHHHHSSG•*MBP*•LEVLFQ↓GP... tag	pTriEx2/Ampicillin	AAGTTCTGTTTCAGGGCCCG‡	ATGGTCTAGAAAGCTTTA‡
pOPINS	MGSSHHHHHH•*SUMO*↓... tag	pET28a/Kanamycin	GCGAACAGATCGGTGGT	ATGGTCTAGAAAGCTTTA

Figure VI-2: ***pOPIN suite of expression vectors****. The plasmid vectors possess different solubility tags and antibiotic resistance. (OPPF-UK Standard Protocols (https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/index.jsp)*

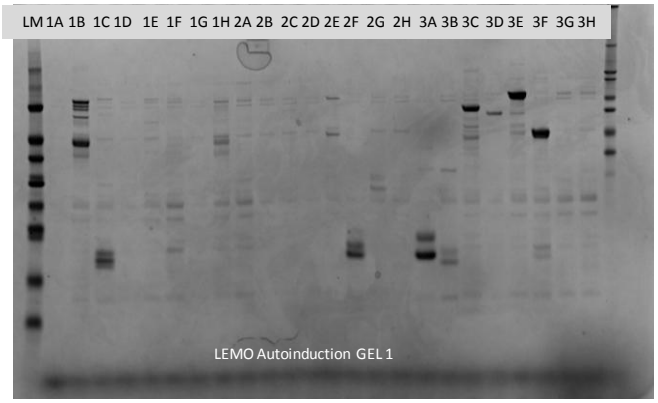
Well	Gene name	aa	Vector	Tag	Strain/s to screen E. coli
A1	**SEPALLATA 3**	251	pOPINCDF	HIS6-3C-POI	Lemo/Rosetta
B1	**SEPALLATA 3**	251	pOPINM	HIS6-MBP-3C-POI	Lemo/Rosetta
C1	**SEPALLATA 3**	251	pOPINS3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
D1	**SEPALLATA 3**	251	pOPINCDE	POI-KHIS6	Lemo/Rosetta
E1	**SEPALLATA 3**	251	pOPINCDJ	HIS6-GST-3C-POI	Lemo/Rosetta
F1	**Agamous**	234	POPINS3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
G1	**Agamous**	234	pOPINRSJ	HIS6-GST-3C-POI	Lemo/Rosetta
H1	**Agamous**	234	pOPINM	HIS6-MBP-3C-POI	Lemo/Rosetta
A2	**Agamous**	234	pOPINE	POI-KHIS6	Lemo/Rosetta
B2	**Agamous**	234	pOPINRSF	HIS6-3C-POI	Lemo/Rosetta
C2	**AP3**	232	pOPINCDF	HIS6-3C-POI	Lemo/Rosetta
D2	**AP3**	232	POPIN CDJ	HIS6-GST-3C-POI	Lemo/Rosetta
E2	**AP3**	232	pOPIN CDM	HIS6-MBP-3C-POI	Lemo/Rosetta
F2	**AP3**	232	pOPIN S3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
G2	**PI**	208	pOPIN RSJ	HIS6-GST-3C-POI	Lemo/Rosetta
H2	**PI**	208	pOPIN RSF	HIS6-3C-POI	Lemo/Rosetta
A3	**PI**	208	pOPIN S3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
B3	**FLM**	141	pOPINS3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
C3	**FLM**	141	pOPINM	HIS6-MBP-3C-POI	Lemo/Rosetta
D3	**SVP**	240	pOPINJ	HIS6-GST-3C-POI	Lemo/Rosetta
E3	**SVP**	240	pOPINM	HIS6-MBP-3C-POI	Lemo/Rosetta
F3	**SVP**	240	pOPINS3C	HIS6-SUMO-3C-POI	Lemo/Rosetta
G3	**SEPALLATA 3**	251	pOPINE	POI-KHIS6	Lemo/Rosetta
H3	**SEPALLATA 3**	251	pOPINF	HIS6-3C-POI	Lemo/Rosetta
A4	**SEPALLATA 3**	251	pOPINJ	HIS6-GST-3C-POI	Lemo/Rosetta
B4	**Agamous**	234	pOPINF	HIS6-3C-POI	Lemo/Rosetta
C4	**Agamous**	234	pOPINJ	HIS6-GST-3C-POI	Lemo/Rosetta
D4	**AP3**	232	pOPINF	HIS6-3C-POI	Lemo/Rosetta
E4	**AP3**	232	pOPINJ	HIS6-GST-3C-POI	Lemo/Rosetta
F4	**AP3**	232	pOPINM	HIS6-MBP-3C-POI	Lemo/Rosetta
G4	**PI**	208	pOPIN J	HIS6-GST-3C-POI	Lemo/Rosetta
H4	**PI**	208	pOPIN F	HIS6-3C-POI	Lemo/Rosetta

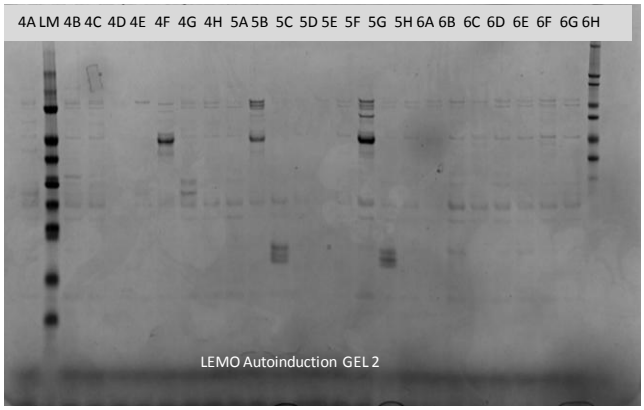
Figure VI-3: ***E.coli test expression screen*** *SEP3, AG, AP3, PI and FLM were cloned in vectors with different solubility tags and expressed in Lemo 21 and Rosetta 2 cells. Two induction methods, IPTG and auto induction system were tested. An SDS-PAGE was performed to check the presence of soluble constructs (Figure VI-4 & Figure VI-5)*

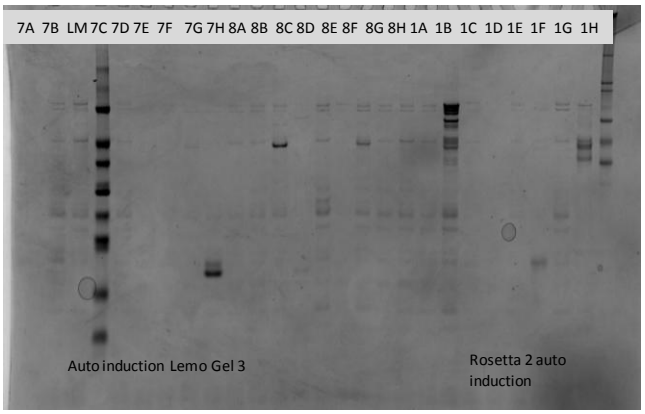


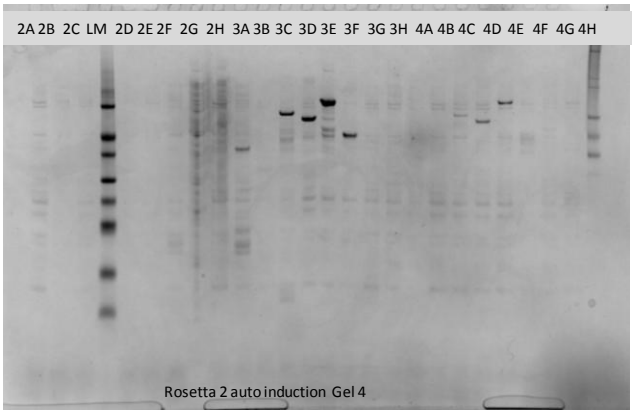
Cells	Well	Construct	Vector
Lemo	3C	FLM	pOPINM
	3D	SVP	pOPINJ
	3E	SVP	pOPINM
	3F	SVP	pOPINS3C
	4E	AP3	pOPINM
Rosetta2	3A	PI	pOPINS3C
	3C	FLM	pOPINM
	3D	SVP	pOPINJ
	3E	SVP	pOPINM
	3F	SVP	pOPINS3C
	4D	AP3	pOPINF
	4E	AP3	pOPINJ

Figure VI-4: ***SDS-PAGE analysis for IPTG induced cells.*** *The clones generated using pOPIN vectors shown in Figure VI-3 were expressed in Lemo 21 and Rosetta 2 cells and induced with 1 mM IPTG. The positive bands obtained in the SDS-PAGE analysis are annotated in the table.*









Cells	Well	Construct	Vector
Lemo	3C	FLM	pOPINM
	3D	SVP	pOPINJ
	3E	SVP	pOPINM
	3F	SVP	pOPINS3C
Rosetta2	3A	PI	pOPINS3C
	3C	FLM	pOPINM
	3D	SVP	pOPINJ
	3E	SVP	pOPINM
	3F	SVP	pOPINS3C
	4D	AP3	pOPINF
	4E	AP3	pOPINJ

Figure VI-5: ***SDS-PAGE analysis for cells grown in auto induction media.*** *The clones generated using pOPIN vectors shown in Figure VI-3 were expressed in Lemo 21 and Rosetta 2 cells and grown in auto induction media. The positive bands obtained in the SDS-PAGE analysis are annotated in the table.*

EXPT	Original Well	OPTIC	Gene name	aa_N	aa_C	OPPF No.	Vector	Clone	Approx MW.	Insect +3ul	Insect +30ul
A1	A3	14272	PI	2	208	16134	pOPIN S3C	1	35983	(+) dimerising	(+) dimerising
B1	B3	14273	FLM	2	141	16135	pOPINS3C	1	28613	(+)	(+)
C1	C3	14273	FLM	2	141	16136	pOPINM	1	58111	+	+
D1	D3	14274	SVP	2	240	16137	pOPINJ	1	54244	+	+
E1	E3	14274	SVP	2	240	16138	pOPINM	1	69001	+	+
F1	F3	14274	SVP	2	240	16139	pOPINS3C	1	39503	+	+
G1	G3	14269	SEPALLATA 3	2	251	16145	pOPINE	1	28469		
H1	H3	14269	SEPALLATA 3	2	251	16146	pOPINF	1	29658		
A2	A4	14269	SEPALLATA 3	2	251	16147	pOPINJ	1	55454	(+)	(+)
B2	B4	14270	Agamous	1	234	16148	pOPINF	1	27898		
C2	C4	14270	Agamous	1	234	16149	pOPINJ	1	53694	(+)	
D2	D4	14271	AP3	2	232	16150	pOPINF	1	27568		
E2	E4	14271	AP3	2	232	16151	pOPINJ	1	53364	(+)	
F2	F4	14271	AP3	2	232	16152	pOPINM	1	68121	(+)?	(+)small
G2	G4	14272	PI	2	208	16154	pOPIN J	1	50724		
H2	H4	14272	PI	2	208	16155	pOPIN F	1	24928		
A3	B1	14269	SEPALLATA 3	2	251	16119	pOPINM	2	70211	++(clipping)	++ (clipping)
B3	C1	14269	SEPALLATA 3	2	251	16120	pOPINS3C	2	40713	(+)	(+)
C3	F1	14270	Agamous	1	234	16140	POPINS3C	2	38953	(+)	(+)
D3	H1	14270	Agamous	1	234	16142	pOPINM	2	68451	(+)	(+)
E3	A2	14270	Agamous	1	234	16143	pOPINE	2	26709		
F3	F2	14271	AP3	2	232	16131	pOPIN S3C	2	38623		
G3											
H3											

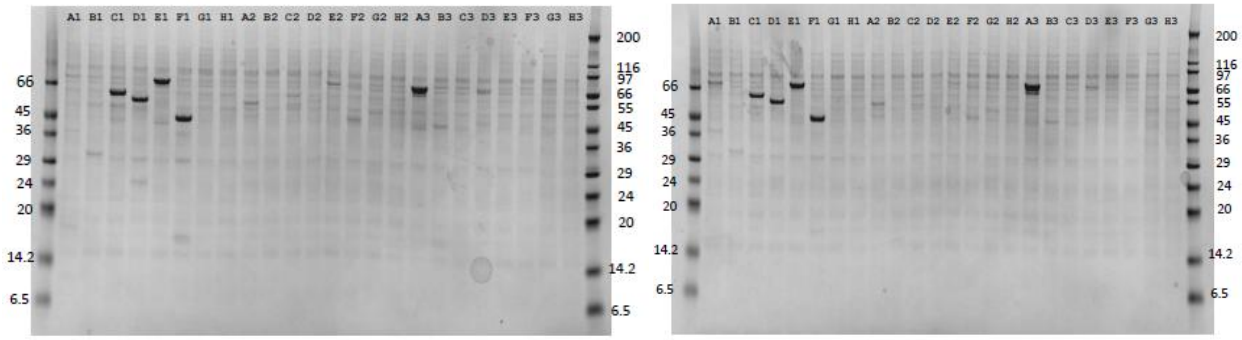


*Figure VI-6**: Insect cell expression screen.** The target proteins were cloned in different vectors and expressed in Sf9 cells. The SDS-PAGE analysis shows expression at 72 h for 3 ml of Sf9 cells @1x10^6 cells/ml infected with 3 µl (left) and 30 µl (right) of 1st generation virus. The lanes are annotated in the table above.*

		Component 1												Component 2										Co-Infection	
EXPT	Original Construct	OPTIC	Gene name	aa_N	aa_C	OPPF No.	Vector	Approx Insect MW. +9ul	Insect +90ul	Original Construct	OPTIC	Gene name	aa_N	aa_C	OPPF No.	Vector	Approx Insect MW.	Insect +9ul	Insect +90ul	Component 1	Component2
A1	B1	14288	SEPALLATA3	2	251	16119	pOPINM	70211	++ prot	F1	14270	Agamous	1	234	16142	pOPIN33C	38953	(+)		+	
B1	B1	14288	SEPALLATA3	2	251	16119	pOPINM	70211	++ prot	H1	14270	Agamous	1	234	16142	pOPINM	68451	(+)		+?	+?
C1	B1	14288	SEPALLATA3	2	251	16119	pOPINM	70211	++ prot	A2	14270	Agamous	1	234	16143	pOPINE	26705			++	(+)
D1	B1	14288	SEPALLATA3	2	251	16119	pOPINM	70211	++ prot	B4	14270	Agamous	1	234	16148	pOPINF	27898			(+)	(+)
E1	B1	14288	SEPALLATA3	2	251	16119	pOPINM	70211	++ prot	C4	14270	Agamous	1	234	16149	pOPINJ	53694	(+)		+	(+)
F1	C1	14288	SEPALLATA3	2	251	16120	pOPINS3C	40713	(+)	F1	14270	Agamous	1	234	16140	pOPIN33C	38953	(+)		(+)	?
G1	C1	14288	SEPALLATA3	2	251	16120	pOPINS3C	40713	(+)	H1	14270	Agamous	1	234	16142	pOPINM	68451	(+)		(+)	+
H1	C1	14288	SEPALLATA3	2	251	16120	pOPINS3C	40713	(+)	A2	14270	Agamous	1	234	16143	pOPINE	26705				
A2	C1	14288	SEPALLATA3	2	251	16120	pOPINS3C	40713	(+)	B4	14270	Agamous	1	234	16148	pOPINF	27898			(+)?	?
B2	C1	14288	SEPALLATA3	2	251	16120	pOPINS3C	40713	(+)	C4	14270	Agamous	1	234	16149	pOPINJ	53694	(+)			
C2	G3	14288	SEPALLATA3	2	251	16145	pOPINE	28469		F1	14270	Agamous	1	234	16140	pOPIN33C	38953	(+)		(+)	
D2	G3	14288	SEPALLATA3	2	251	16145	pOPINE	28469		H1	14270	Agamous	1	234	16142	pOPINM	68451	(+)		(+)?	(+)
E2	G3	14288	SEPALLATA3	2	251	16145	pOPINE	28469		A2	14270	Agamous	1	234	16143	pOPINE	26705				
F2	G3	14288	SEPALLATA3	2	251	16145	pOPINE	28469		B4	14270	Agamous	1	234	16148	pOPINF	27898				
G2	G3	14288	SEPALLATA3	2	251	16145	pOPINE	28469		C4	14270	Agamous	1	234	16149	pOPINJ	53694	(+)		(+)	
H2	H3	14288	SEPALLATA3	2	251	16146	pOPINF	29658		F1	14270	Agamous	1	234	16140	pOPIN33C	38953	(+)		(+)	
A3	H3	14288	SEPALLATA3	2	251	16146	pOPINF	29658		H1	14270	Agamous	1	234	16142	pOPINM	68451	(+)		+(+)	(+)
B3	H3	14288	SEPALLATA3	2	251	16146	pOPINF	29658		A2	14270	Agamous	1	234	16143	pOPINE	26705				
C3	H3	14288	SEPALLATA3	2	251	16146	pOPINF	29658		B4	14270	Agamous	1	234	16148	pOPINF	27898				
D3	H3	14288	SEPALLATA3	2	251	16146	pOPINF	29658		C4	14270	Agamous	1	234	16149	pOPINJ	53694	(+)			
E3	A4	14288	SEPALLATA3	2	251	16147	pOPINJ	55454	(+)	F1	14270	Agamous	1	234	16140	pOPIN33C	38953	(+)		(+)	(+)
F3	A4	14288	SEPALLATA3	2	251	16147	pOPINJ	55454	(+)	H1	14270	Agamous	1	234	16142	pOPINM	68451	(+)		(+)	(+)
G3	A4	14288	SEPALLATA3	2	251	16147	pOPINJ	55454	(+)	A2	14270	Agamous	1	234	16143	pOPINE	26705			(+)	
H3	A4	14288	SEPALLATA3	2	251	16147	pOPINJ	55454	(+)	B4	14270	Agamous	1	234	16148	pOPINF	27898			(+)	
A4	A4	14288	SEPALLATA3	2	251	16147	pOPINJ	55454	(+)	C4	14270	Agamous	1	234	16149	pOPINJ	53694	(+)		(+)?	(+)?
B4	F2	14271	AP3	2	233	16131	pOPIN 33C	38623		A3	14272	PI	2	208	16134	pOPIN 33C	35983	(+) dimerisin		(+) dimerisin	
C4	F2	14271	AP3	2	233	16131	pOPIN 33C	38623		G4	14272	PI	2	208	16154	pOPIN J	50724				(+)
D4	F2	14271	AP3	2	233	16131	pOPIN 33C	38623		H4	14272	PI	2	208	16155	pOPIN F	24923				
E4	D4	14271	AP3	2	233	16150	pOPINF	27568		A3	14272	PI	2	208	16134	pOPIN 33C	35983	(+) dimerisin		(+) dimerisin	
F4	D4	14271	AP3	2	233	16150	pOPINF	27568		G4	14272	PI	2	208	16154	pOPIN J	50724				(+)
G4	D4	14271	AP3	2	233	16150	pOPINF	27568		H4	14272	PI	2	208	16155	pOPIN F	24923				
H4	E4	14271	AP3	2	233	16151	pOPINJ	53364	(+)	A3	14272	PI	2	208	16134	pOPIN 33C	35983	(+) dimerisin		(+) dimerisin	
A5	E4	14271	AP3	2	233	16151	pOPINJ	53364	(+)	G4	14272	PI	2	208	16154	pOPIN J	50724				
B5	E4	14271	AP3	2	232	16151	pOPINJ	53364	(+)	H4	14272	PI	2	208	16155	pOPIN F	24923				
C5	F4	14271	AP3	2	233	16152	pOPINM	68121	(+)?	A3	14272	PI	2	208	16134	pOPIN 33C	35983	(+) dimerisin		(+) dimerisin	(+)
D5	F4	14271	AP3	2	233	16152	pOPINM	68121	(+)?	G4	14272	PI	2	208	16154	pOPIN J	50724			(+)?	
E5	F4	14271	AP3	2	232	16152	pOPINM	68121	(+)?	H4	14272	PI	2	208	16155	pOPIN F	24923			(+)?	

Figure VI-7: ***Insect cell co-expression screen.*** *SEP3-AG and AP3-PI were co-expressed in Sf9 insect cells. 3 ml of Sf9 cells @1x10^6 cells/ml were infected with 3 µl of each 1st generation virues. The SDS-PAGE analysis is shown in Figure VI-8. The table above describes the different combinations used and the expression obtained for each component.*

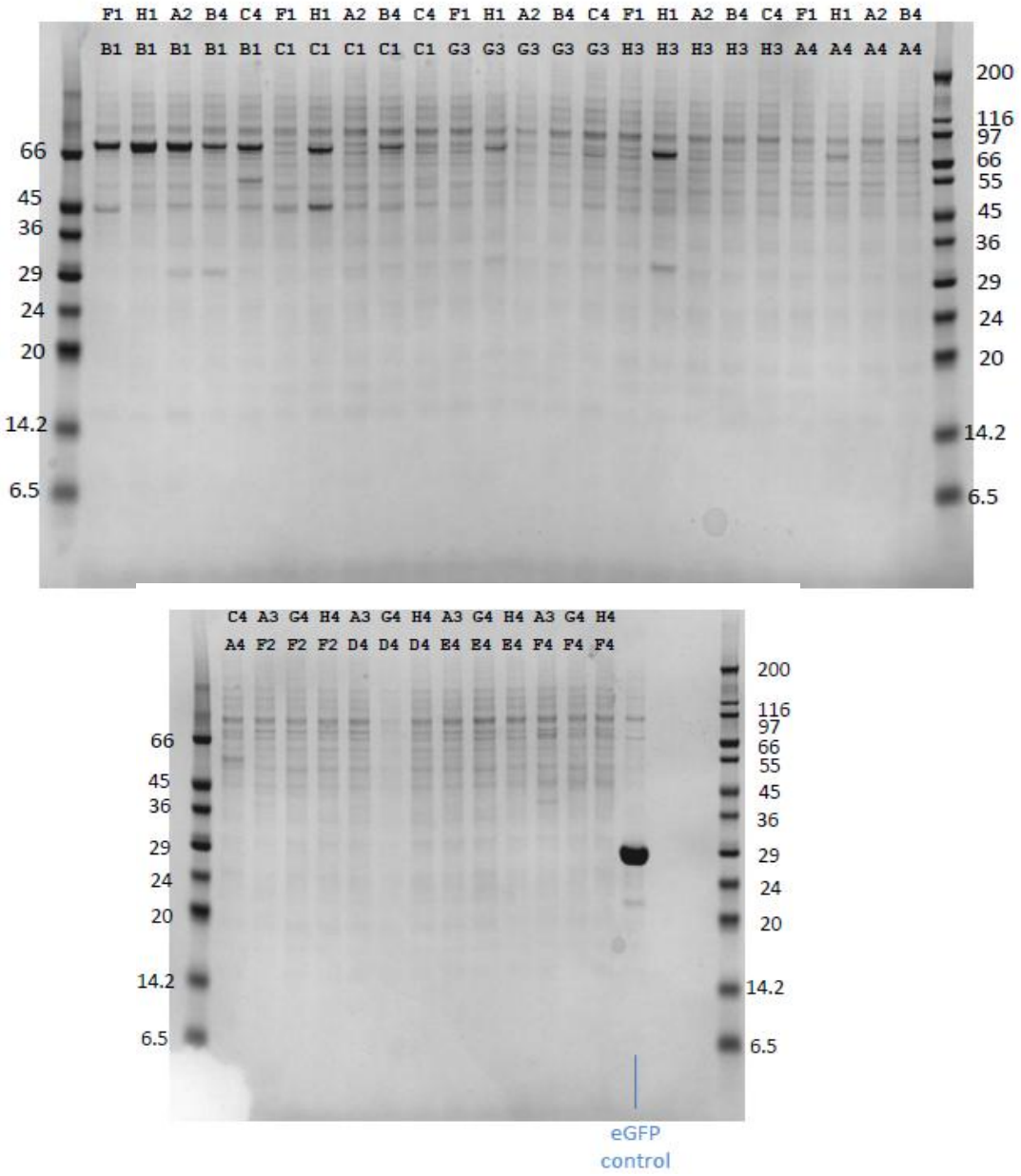
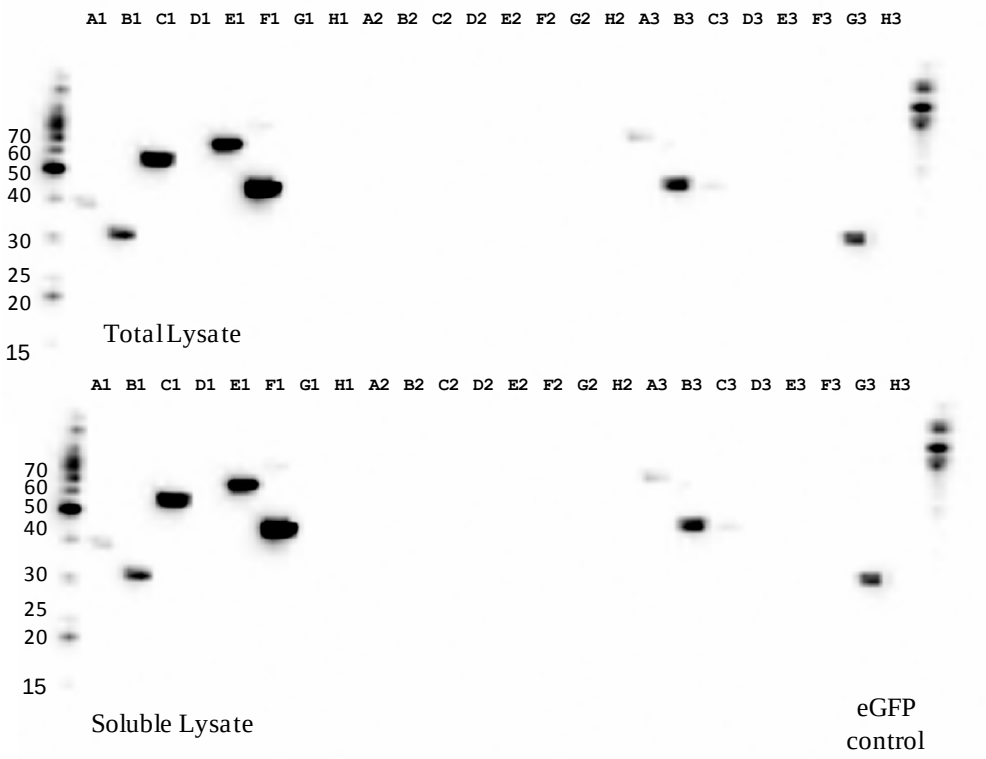


Figure VI-8: ***SDS-PAGE analysis for insect cell co-expression screen.*** *The target proteins were cloned in different vectors and co- expressed in Sf9 cells using the insect cell expression system. The lanes are annotated in the Figure VI-7.*



Well HEK screen	Construct well	Construct name	Vector	Mol wt (KDa)
A1 (low)	A3	PI	pOPIN S3C	35.9
B1	B3	FLM	pOPINS3C	28.6
C1	C3	FLM	pOPINM	58.1
D1	D3	SVP	pOPINJ	54.2
E1	E3	SVP	pOPINM	69.0
F1	F3	SVP	pOPINS3C	39.5
A3(low)	B1	SEP3	pOPINM	70.2
B3	C1	SEP3	pOPINS3C	40.7
C3(low)	F1	Agamous	pOPINS3C	38.9

Figure VI-9: ***Western blot for HEK 293 expression tests probed with anti-His antibody. (top)****The protein bands are visible in both total and soluble lysate.* ***(bottom)*** *The table shows the name of protein, the vector used for cloning, the solubility tag use and the theoretical molecular weight of the protein is provided.*

VI.1.2 Insect cell expression system at EMBL-Grenoble

Baculovirus expression system is used to produce eukaryotic proteins by utilizing recombinant baculovirus carrying heterologous genes to infect insect cells. Since its first use in 1983, it has been used extensively for recombinant protein expression (Smith et al., 1983). *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and *Bombyx mori* (silkworm) nuclear polyhedrosis virus (BmNPV) are the most common baculoviruses used in foreign gene expression. The Figure VI-10 provides an overview of the baculovirus expression system. It consist of a donor plasmid with the gene of interest cloned downstream of a promoter (polh or p10) and flanked by Tn7R and Tn7L sites for site specific recombination. The competent cells DH10Bac^{TM} (Thermo Fischer Scientific Inc.), possess a recombinant bacmids with an attachment site for transposition (attTn7). In presence of transposition proteins provided by the helper plasmid, the Tn7 transposon enzyme inserts the gene of interest in the bacmid. Blue-white screening verifies the presence of positive transposition by disruption of LacZα gene. The bacmid is transfected in insect cells and the recombinant virus is amplified to produce recombinant proteins. Over the years, new developments have made possible multigene transfer in the bacmid like Bac-To-Bac system and MultiBac system (Berger et al., 2004; Bieniossek et al., 2012).

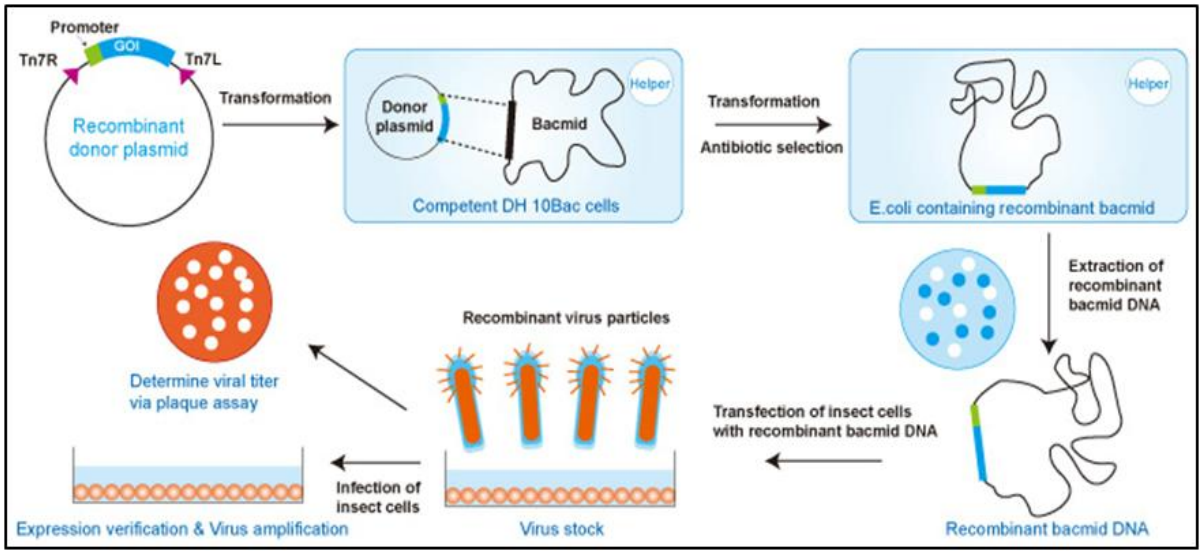


Figure VI-10: ***Overview or Baculovirus expression system.*** *The target gene is cloned in donor plasmid and shuttled in to a bacmid. The recombinant bacmid is transfected in insect cells and recombinant virus is amplified and further used for infecting more insect cells and production of proteins (www.genscript.com).*

MultiBac is a baculovirus expression system optimized for expression of protein complexes using an array of specialized DNA plasmids. It has undergone considerable development since its launch and the current system offers a range of recombination strategies for genes and a fluorescent marker for detecting their expression in *Spodoptera frugiperda* insect cells (Berger et al., 2004; Fitzgerald et al., 2006; Bieniossek et al., 2008, 2012; Berger et al., 2013; Berger, 2014). The plasmid vectors and the baculovirus genome used in MultiBac system are shown in the Figure VI-11.

One of the most important part of this system are its specially designed plasmids, which allow convenience of cloning and propagation in *E.coli* and a facility to generate mutli-protein complexes with a combination of Tn7, Cre-LoxP and homing endonuclease systems. The acceptors contain a standard oriColE1 while the donors possess an origin of replication which is activated in the presence of protein expressed by pir gene. Specially designed *E.coli* strains are inserted with pir gene to allow replication of donors (Fitzgerald et al., 2006; Bieniossek et al., 2008). Both plasmids are integrated with expression cassettes under control of baculovirus promoter (polh or p10) and strong eukaryotic polyhedron signals (SV40). Acceptors have gentamycin resistance while donors have either chloramphenicol, kanamycin or spectinomycin resistance. The donor and acceptor plasmids are combined using Cre-loxP system by mixing them in presence of Cre recombinase and then selecting the successful combinations by transforming in to pir- strain in appropriate antibiotics (Bieniossek et al., 2008; Abremski et al., 1983; Berger et al., 2004; Bieniossek et al., 2009). The homing endonuclease and BstXI compatible sites surrounding the expression cassettes can also be used to generate a multigene assembly. The transfer of gene cassette from the acceptors to the baculovirus genome is undertaken by the Tn7L and Tn7R present on the acceptor and the attTn7 attachment site present on the baculovirus genome. In MultiBac system, the genome has also been engineered to remove the genes which code for viral proteases and apoptic activities thus increasing the stability of protein production. Presence of yellow fluorescent protein as a marker serves as an indicator for protein expression.

In our study, we cloned GST and SUMO tagged SVP in pAceBac1 and expressed them in sf21 cell lines in sf9 medium. The protein constructs were well expressed and were stable upon purification and tag cleavage

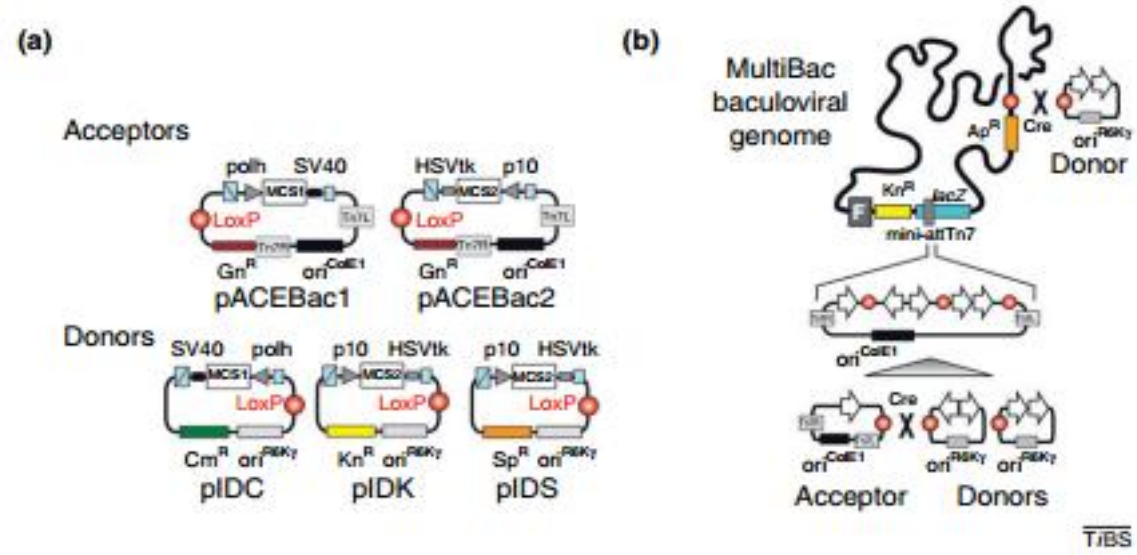


Figure VI-11: ***a)*** *Plasmids used in MultiBac system* ***b)*** *Combination of acceptor and donor plasmids and their integration in baculovirus genome.* (Bieniossek et al., 2012)

VI.2 Biophysical techniques

VI.2.1 Atomic force microscopy

VI.2.1.1 AFM instrumentation

MFP-3D and Cypher-S from Asylum Research (Santa Barbara, CA, USA) were used for AFM experiments. The MFP-3D is a classic closed loop AFM designed for high resolution imaging, accuracy and reproducibility. It has been used successfully in various fields from materials science and life science to nanolithography (MFP-3D manual, Asylum research). As shown in the Figure VI-12a, the MFP-3D is composed of a head containing the piezo and the laser-photodiode detection system; the X-Y scanner which has the NPS (Nano positioning system); and the base which has the optics for sample and tip alignment. The NPS helps in hysteresis and creep correction in the close loop system.

Cypher S is most advanced AFM from Asylum Research which is capable of high speed high resolution imaging. The integrated enclosure as shown in Figure VI-12 provides acoustic isolation and thermal control thus reducing the background noise and thermal drift (Cypher-S manual, Asylum research). The small laser spots and automatic laser alignment is significant for high speed scans with small cantilevers. The photothermal excitation using Bluedrive system removes the spurious frequencies observed while scanning in tapping mode making scanning in liquids much simpler and accurate.

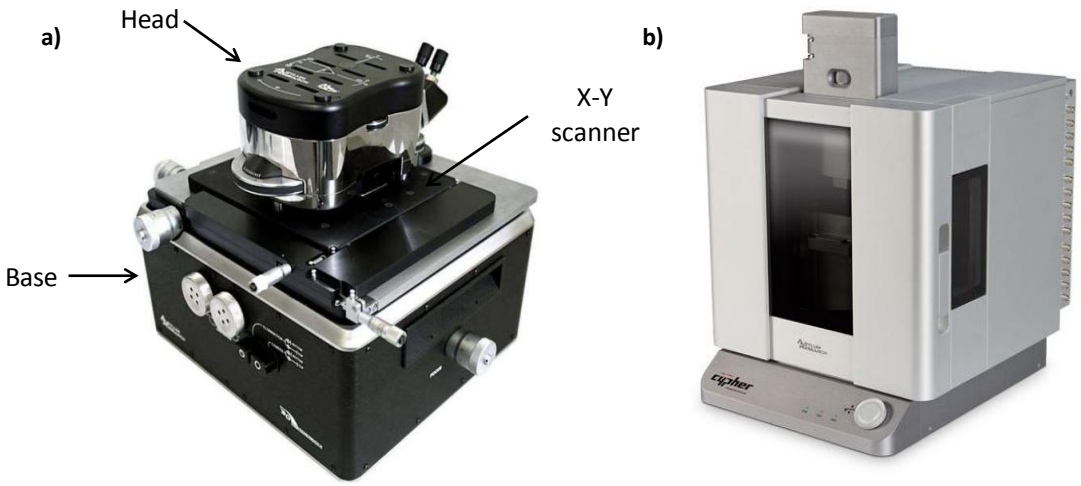


Figure VI-12: ***AFMs from Asylum research. a)*** *MFP-3D comprising of head, base and X-Y scanner* ***b)*** *Cypher-S composed of similar components enclosed for acoustic isolation and temperature maintenance.*

VI.2.1.2 AFM probes

An AFM probe consists of a cantilever with a sharp tip at its end. The length, material and stiffness of the cantilever and tip are influenced on its application. A sharp tip of high spring constant is ideal for imaging macromolecules while live cell imaging requires use of blunt rounded tips. Silicone nitride cantilevers with silicone tips are often used for biological samples as they are thin, sharp, flexible and with low spring constant (Grow et al., 2002). The Figure VI-13 shows image of the cantilevers and tips used in our study. MSNL-10 (Bruker) probe includes 6 high resolution silicone nitride cantilevers with a sharp tip of radius 2-12 nm. BioLever mini (BLAC40TS) (Olympus) are highly sensitive short cantilevers of around 40 µm with a tip radius of ~ 10 nm.

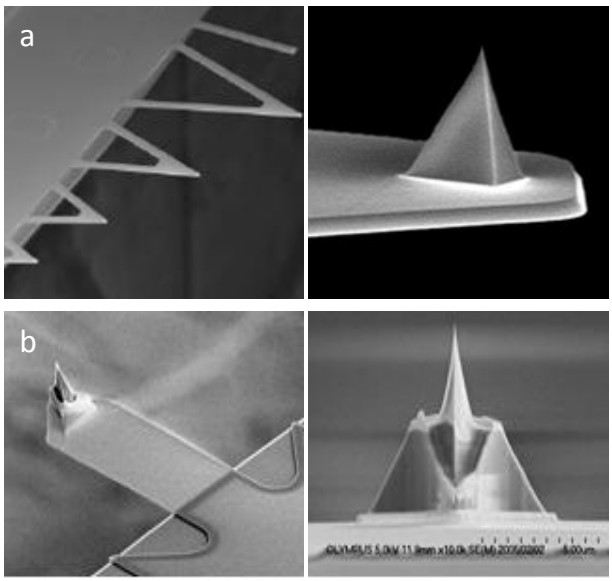


*Figure VI-13: C****antilever and tips used for AFM experiments a)*** *MSNL-10 from Bruker &* ***b)*** *Biolever AC40 TS from Olympus*

VI.2.1.3 Sample preparation

The sample preparation forms the crucial step in AFM imaging. In case of biological samples such as macromolecules, the samples are adsorbed on flat substrates like mica, glass or gold using either electrostatic or chemical interactions. Mica or muscovite mica is the preferred surface as a substrate due to its unique properties of being extremely smooth and atomically flat. The review by (Lyubchenko et al., 2011) elaborates different methods used for activating mica surface for sample adsorption. Previous studies have shown that mica is alternatively layered with potassium cations located between hydroxyl groups (Liu et al., 2005). Mica can be cleaved in to flat layers by either scotch tape or sharp razor blades. The cleavage results in the potassium ions to spread over both the surface giving a net negative charge to both the sheets. The electrostatic adsorption procedure involves coating the layer with divalent cations like Ni^{2+}, Mg^{2+} which imparts a net positive charge and has been proved as a robust and preferential method for binding of macromolecules (Pastré et al., 2003; Hansma et al., 2004; Dahlgren et al., 2002; Zheng et al., 2003). Other methods involving use of chemical modifications and binding agents like poly-lysine are also widely used (Hu et al., 1996; Hansma et al., 1998; Lyubchenko et al., 1992; Crampton et al., 2005; Okusa et al., 1994). While binding using divalent cations, the radii of the cations play an important role, with better binding for metals with ionic radius of 0.74 Å or less (Hansma and Laney, 1996). Their binding may be related to their ability to fit in to the cavities above the buried hydroxyl groups and their high enthalpies of radiation leading to formation of strong complexes. However, the exact mechanism of binding of macromolecules to the surface has not been properly understood. Monovalent cations like NaCl or KCl are added to impart mobility to DNA and protein as they weaken the binding between the mica surface and the samples. The concentrations of divalent and monovalent cations need to be optimized to obtain ideal condition in which the DNA molecules are bound to substrate but are still motile and accessible to the protein molecules. In our experiments 10 mM $NiSO_4$ was used for binding of macromolecules in appropriate buffer. As protein, DNA molecules were incubated before binding, addition of monovalent cations was not necessary during imaging.

VI.2.1.4 Analysis of AFM images

Being a single molecular technique, visualisation of macromolecules in AFM can provide important information. To ensure appropriate analysis, it is utmost important to distinguish the sample from the artefacts. Artefacts can be introduced due to damaged or contaminated

tip, excessive force applied on sample and impurities in buffer and sample (Braga and Ricci, 2004). The image is first flattened to correct the tilt which can be introduced due to slight bend in the mica surface followed by an analysis using the software Gwyddion (Nečas and Klapetek, 2011). Once the image is flattened and the background noise reduced, the height of DNA and proteins is measured. The measured height is compared with the theoretical DNA width of 2 nm to ensure that excessive force was not applied on the sample resulting in sinking or damaging of DNA. The position of protein on the DNA is measured by measuring the length of DNA in 'nm' and calculating the approximate DNA base pair (1 bp ~ 0.34 nm) using the image editing software ImageJ (Schneider et al., 2012). In this study as the DNA strands were designed in house, the protein binding could be verified by comparing the location of CArG boxes on the DNA strands.

VI.2.2 Small Angle X-ray Scattering

Small Angle X-ray scattering (SAXS) is a powerful structure biology technique used to study the structure and interaction of macromolecules at low resolution in solution. It is often used as complementary technique with X-ray crystallography to give more detailed three dimensional information. SAXS uses relatively less material, has easier sample preparation procedures and can yield much faster results as compared to X-ray crystallography, NMR and EM. Moreover it can characterize shape and conformation in solution for systems ranging from small proteins to large macromolecular complexes. The SAXS pipeline for structural determination consists of sample preparation, data collection and validation followed by structure determination using *ab initio* and rigid body modelling.

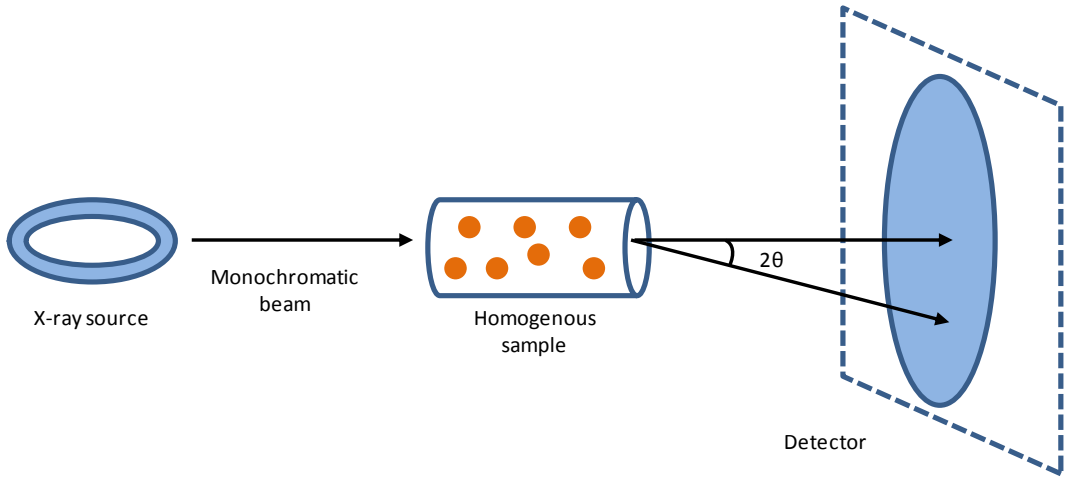


***Figure VI-14**: **Schematic representation of a typical SAXS experiment.** The X-rays are scattered by the solution sample and the scattering is measured on a detector.*

A typical SAXS experiment as shown in the Figure VI-14 consists of a X-ray source such as a synchrotron, a homogenous sample and a detector. The X-rays incident on a randomly oriented sample in solution are scattered by the macromolecules and an isotropic scattering pattern is generated on the detector. This scattering pattern is a function of scattering vector *q* and can be defined as:

$$q = \frac{4\pi \sin\,\theta}{\lambda}$$

where, 2θ is half the scattering angle and λ is the wavelength of the incident beam.

As the wavelength is kept constant (~ 1 Å) and the scattering angle is confined within low range (0.1-10°), the q range measured in SAXS experiment is approximately 0.5 $Å^{-1}$ (Koch et al., 2003). The scattering function of an ideal solution of molecules can be represented by

$$I(q) = I\,(q)\Omega = \mathrm{A(q)A*\,(q)}\Omega$$

Where, the scattering amplitude A(q) is the Fourier transform of the excess scattering length density and the scattering is averaged over all orientations (Ω). Following subtraction of the solvent scattering, the background corrected intensity I(q) is proportional to the scattering of a single particle averaged over all orientations.

SAXS is a contrast technique and thus the scattering signal is derived from the difference between the scattering intensity of the sample and that of solution. This difference for macromolecules is generally very small making it crucial to have minimum background interference. To avoid introduction of artefacts, the buffer composition of sample and solvent should be identical. This can be achieved by dialysing the sample in same batch of buffer. High concentration of sample usually in range of 1-5 mg/ml increases the signal from the sample thus giving better contrast. The prerequisite for obtaining a good SAXS data, is a homogenous sample. It is recommended to confirm the homogeneity of the sample using methods like DLS before data collection. Online HPLC system and chromatographic purification of sample just before data collection is useful to avoid problems caused by aggregation of sample. Generally high concentration of glycerol should be avoided in the buffer. SAXS can measure small changes in the sample environment such as pH, temperature, salt concentration and ligand/co-factor titration (Mertens and Svergun, 2010).

SAXS data analysis consists of derivation of initial parameters such as size, oligomeric state, shape from the scattering data followed by a three dimensional model building by combining complementary data using more advanced programs (Mertens and Svergun,

2010). These parameters include: the molecular mass (MM), radius of gyration (R_g), hydrated particle volume (Vp) and maximum particle diameter (D_{max}). The most important parameters obtained from SAXS experiment are the I(0) and the R_g both of which can be obtained from the Guinier plot (Guinier, 1938) (Figure VI-14). The Guinier approximation can be represented as

$$I(q) = I_0 \, exp\left(-\frac{1}{3} q^2 R_g^{\,2} \right)$$

This equation can be converted in to a linear equation

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

The R_g value is the square root of the averaged square distances of each scatterer from the particle centre (Putnam et al., 2007). The I(0) value is proportional to the number of electrons in the scatterer (Koch et al., 2003) and thus dependent on the concentration of the sample. The I(0) can be used to measure the molecular weight if the concentration of the sample is known. The Guinier approximation is only valid at low q values. At this low angle, the plot of ln(I(q)) versus q^2 results in straight line. R_g is determined from the slope of the curve while I(0) is determined by the y intercept value at zero angle (q=0) which can be obtained from extrapolation of the curve.

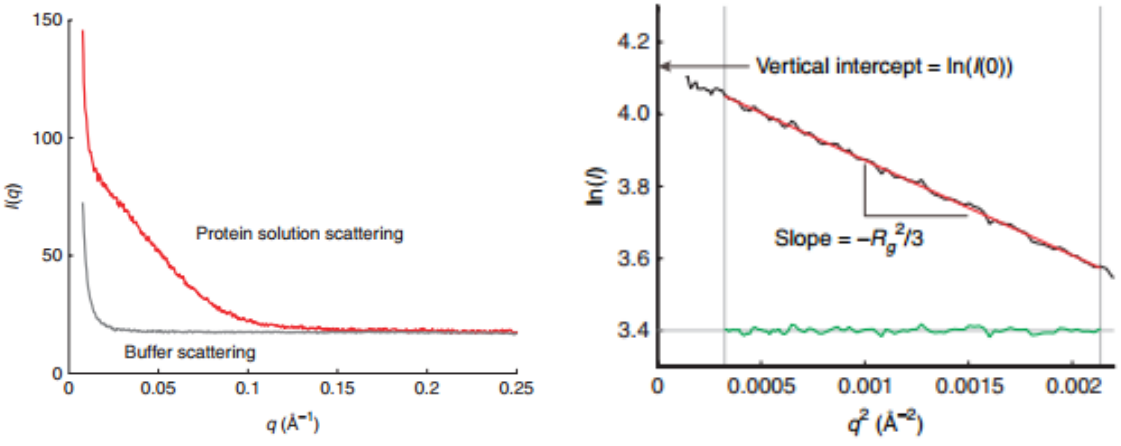


Figure VI-15: ***Typical scattering curves and Guinier plot. (left)*** *Typical scattering from protein solution (red) and the corresponding buffer (grey).* ***(right)*** *The red line is fitted to low-q region of a Guinier curve such that maximum q< 1.3/ R_g. R_g is derived from the slope, and I(0) is derived from the vertical intercept. The linearity of the fitted region is determined by the flatness of the residuals (green)* (Skou et al., 2014).

The Guinier plot generally provides information about the interparticle interaction such as aggregation as observed which is detected by the curvature in the Guinier plot as shown in the Figure VI-16. However, as the maximum q value is limited to $q*R_g<1.3$.

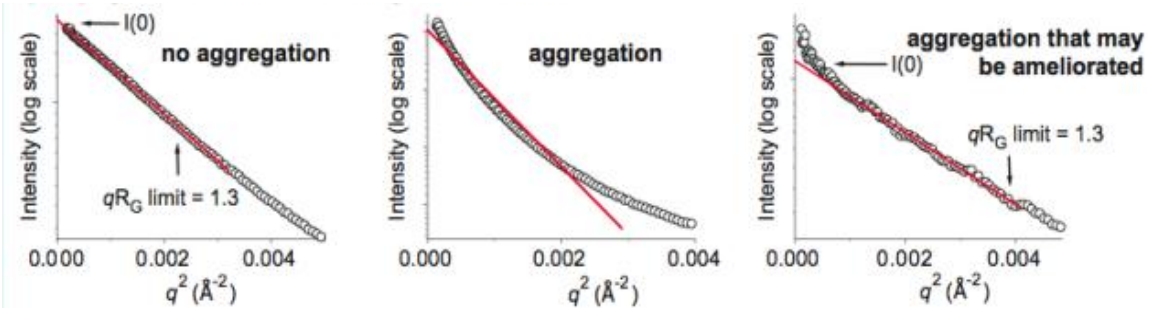


Figure VI-16: ***Calculation of R_g and I(0) from Guinier plot.*** *A linear log (I(q)) vs q^2 represents an ideal Guinier plot. The slope gives the R_g while the I(0) is obtained from the y intercept. A non-linear plot shows presence of aggregation.* (Putnam et al., 2007)

The next important parameter used for sample characterisation is the pair distribution function P(r). The P(r) function describes the paired-set of all distances between points within an object. It is the Fourier transform of the scattering intensities and is represented by

$$I_{molecule}(q) = \int p(r) \cdot \frac{\sin(q \cdot r)}{q \cdot r} \, dr$$

The P(r) function is useful tool for visibly detecting conformational changes within a macromolecule as observed in the Figure VI-17. The X-intercept provides value of the maximum dimension which can be used to compare the different models with the scattering data.

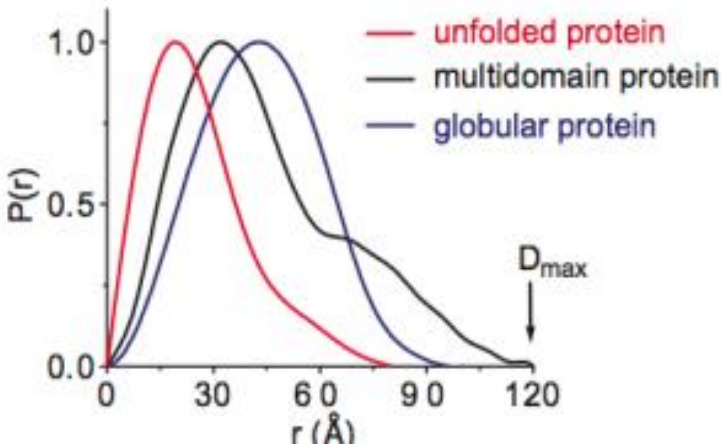


Figure VI-17: ***P(r) function analysis for proteins.*** *Globular proteins have a P(r) function with a single peak while elongated macromolecules have a longer tail at large r and can have multiple peaks* (Putnam et al., 2007).

The protein folding is analysed using Krakty plot (q^2 *I(q) vs q) (Figure VI-18 left). Folded globular proteins generally give a prominent peak at low angles while unfolded proteins show a continuous increase in q^2 * I(q). The flexible multi-domain proteins generally bear characteristics of both the globular and the unfolded. In the dimensionless Kratky plot (Figure VI-18 right) the q vector on the X-axis is multiplied by the radius of gyration (R_g) and the I(q) on the y axis is multiplied by ($q.R_g$)2. A well folded globular protein shows a peak maxima of 1.104 at $q*R_g = \sqrt{3}$ (Receveur-Brechot and Durand, 2012; Durand et al., 2010).

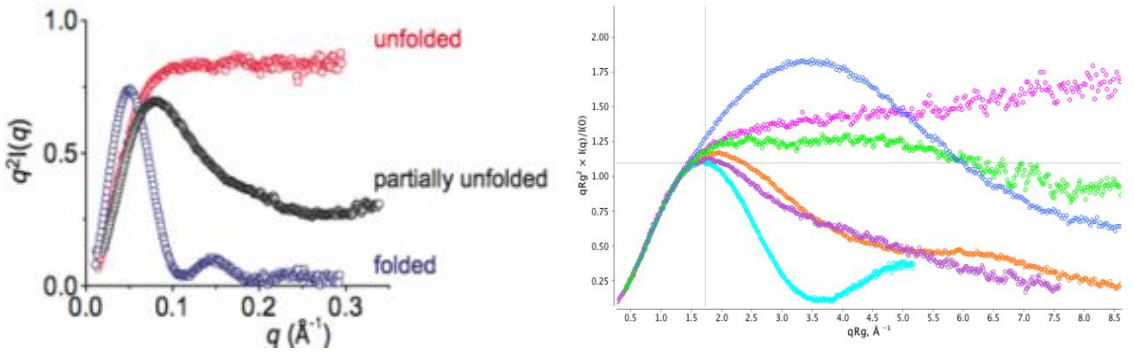


Figure VI-18: Kratky plot analysis for flexible proteins. (left) Globular molecules follow Porod's law and have bell shaped curves. Unfolded molecules lack this peak and show a plateau or slightly increasing curve at higher q range. (Putnam et al., 2007) ***(right)*** *Dimensionless Kratky shows similar tendency with a well folded protein showing a peak maxima of 1.104 at q*Rg =√3 (http://www.bioisis.net/)*

In addition to sample characterisation from extracted parameters, SAXS can also be used to model three dimensional structures from the one dimensional scattering data. Two approaches, *ab initio* modelling and rigid body modelling approaches are generally used for this depending on the availability of structural data for homologous regions. In ab initio method, the molecule is assigned as a collection of densely packed beads followed by shape reconstruction using simulated annealing procedures. At each step in the simulated annealing procedure, theassignment of a single bead is randomly changed leading to anew model and is evaluated by measuring the discrepancy (χ^2) between the experimental and calculated scattering intensities. The solution is constrained by penalties such as requiring that the beads must be connected and the model compact to ensure that physically feasible low-resolution structures are generated. Several constraints like specific symmetry can be imposed. The programs such as DAMMIN (Svergun, 1999) and DAMMIF (Franke and Svergun, 2009) are most commonly used for *ab initio* modelling. Generally a large number of models are generated using these tools and are filtered and averaged to select the most

representative one using tools such as DAMAVER. It is however necessary to critically evaluate the *ab initio* models as the generated models might not always represent the actual structure and may be affected by polydispersity of the sample.

The second approach uses data from complementary methods to establish a three dimensional structure for protein or a protein complex. In this approach, the theoretical scattering from the known structures or homology models is calculated using spherical harmonics and forms the basis of model adaptation. The different known structures of a protein complex or of individual domains of proteins are arranged in arbitrary positions and simulated annealing is employed. Various constraints such as symmetry, pairing of different domains can be used. Several programs such as SASREF and BUNCH (Petoukhov and Svergun, 2005) use this approach. Some programs such as CORAL (Petoukhov et al., 2012) also allow use of unstructured linkers between different domains or proteins representing regions of unknown structure.

Additional methods are used in case of flexible proteins where representation using single model is challenging. Thus SAXS data can be processed using different approaches to provide valuable information using the work flow as shown in Figure VI-19.

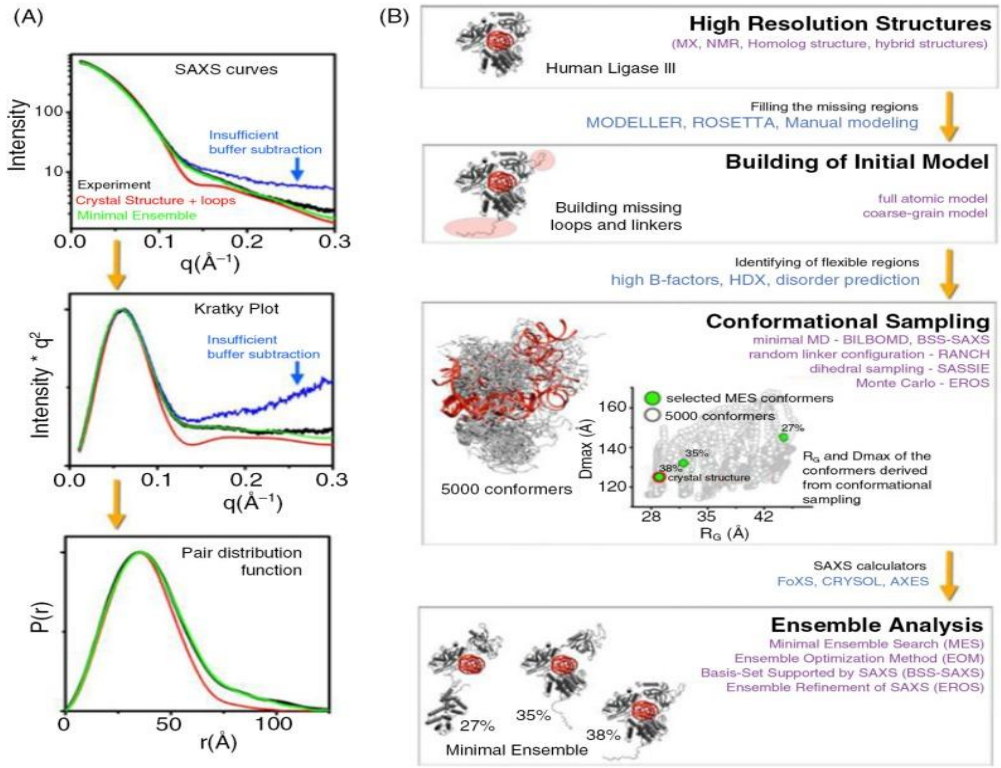


Figure VI-19: ***Typical work flow followed for SAXS data analysis.*** *The initial analysis using scattering curves, Kratky plot, Pair distribution function and Guinier provides information regarding the shape, size, flexibility and conformation of protein. It is followed by model building strategies composed of ab initio and rigid body modelling. Ensemble analysis is used in case of flexible proteins* (Boldon et al., 2015).

Chapter VII

APPENDIX-II

VII Appendix II – Collection of articles

This dissertation resulted in publication of two peer reviewed articles which are attached here.

The Plant Cell, Vol. 26: 3603–3615, September 2014, www.plantcell.org

Structural Basis for the Oligomerization of the MADS Domain Transcription Factor SEPALLATA3 in *Arabidopsis*[W]

Sriharsha Puranik,[a] Samira Acajjaoui,[a] Simon Conn,[b] Luca Costa,[a] Vanessa Conn,[b] Anthony Vial,[a] Romain Marcellin,[a,c] Rainer Melzer,[d] Elizabeth Brown,[a] Darren Hart,[e] Günter Theißen,[d] Catarina S. Silva,[f,g,h,i] François Parcy,[f,g,h,i] Renaud Dumas,[f,g,h,i] Max Nanao,[j,k] and Chloe Zubieta[f,g,h,i,1]

[a] European Synchrotron Radiation Facility, Structural Biology Group, 38042 Grenoble, France
[b] Centre for Cancer Biology, SA Pathology and the University of South Australia, Adelaide SA 5000, Australia
[c] Faculté des Sciences de Montpellier, place Eugène Bataillon, 34095 Montpellier, France
[d] Department of Genetics, Friedrich Schiller University, 07737 Jena, Germany
[e] Université Grenoble Alpes, CNRS, Integrated Structural Biology Grenoble, Unit of Virus Host Cell Interactions, Unité Mixte Internationale 3265 (CNRS-EMBL-UJF), UMS 3518 (CNRS-CEA-UJF-EMBL), 38042 Grenoble, France
[f] CNRS, Laboratoire de Physiologie Cellulaire and Végétale, UMR 5168, 38054 Grenoble, France
[g] Université Grenoble Alpes, Laboratoire de Physiologie Cellulaire et Végétale, F-38054 Grenoble, France
[h] Commissariat à l'Energie Atomique, Direction des Sciences du Vivant, Institut de Recherches en Technologies et Sciences pour le Vivant, Laboratoire de Physiologie Cellulaire et Végétale, F-38054 Grenoble, France
[i] INRA, Laboratoire de Physiologie Cellulaire et Végétale, USC1359, F-38054 Grenoble, France
[j] European Molecular Biology Laboratory, Grenoble Outstation, 38042 Grenoble, France
[k] Unit for Virus Host-Cell Interactions, Université Grenoble Alpes-EMBL-CNRS, 38042 Grenoble, France

In plants, MADS domain transcription factors act as central regulators of diverse developmental pathways. In *Arabidopsis thaliana*, one of the most central members of this family is SEPALLATA3 (SEP3), which is involved in many aspects of plant reproduction, including floral meristem and floral organ development. SEP3 has been shown to form homo and heterooligomeric complexes with other MADS domain transcription factors through its intervening (I) and keratin-like (K) domains. SEP3 function depends on its ability to form specific protein-protein complexes; however, the atomic level determinants of oligomerization are poorly understood. Here, we report the 2.5-Å crystal structure of a small portion of the intervening and the complete keratin-like domain of SEP3. The domains form two amphipathic alpha helices separated by a rigid kink, which prevents intramolecular association and presents separate dimerization and tetramerization interfaces comprising predominantly hydrophobic patches. Mutations to the tetramerization interface demonstrate the importance of highly conserved hydrophobic residues for tetramer stability. Atomic force microscopy was used to show SEP3-DNA interactions and the role of oligomerization in DNA binding and conformation. Based on these data, the oligomerization patterns of the larger family of MADS domain transcription factors can be predicted and manipulated based on the primary sequence.

INTRODUCTION

The astonishing diversity of all complex organisms relies on the evolutionary co-option of developmental pathways present in simpler ancestral phyla. Changes in genes that regulate development, and the transcription factors (TFs) they encode, are at the advent of this new functionality. Events such as gene duplications, deletions, mutations, domain swapping, fusions, and fixation via selection or random drift change the activity of the encoded TFs, resulting in alterations in downstream pathways, thus providing a basis for morphological diversity and increased complexity. The MADS box genes are an example of a family of developmental regulatory genes present in all eukaryotes that have dramatically diversified during evolution and have undergone a particularly large lineage expansion in plants (Münster et al., 1997; Alvarez-Buylla et al., 2000; Becker et al., 2000; Theissen et al., 2000; Soltis et al., 2002; Becker and Theissen, 2003; De Bodt et al., 2003a, 2003b; Gramzow et al., 2010; Melzer et al., 2010). Diversification of plant MADS domain TF function has been achieved by adding dimerization and tetramerization domains to the basic DNA binding machinery, allowing the precise regulation of a plethora of distinct developmental processes.

The MADS box family, with representatives in protists, fungi, animals, and plants, is named for the founding members: MCM1 from yeast (*Saccharomyces cerevisiae*), AGAMOUS (AG) from *Arabidopsis thaliana*, DEFICIENS from snapdragon (*Antirrhinum majus*), and SRF from *Homo sapiens* (Schwarz-Sommer et al., 1990). Data from whole-genome sequencing and computational homology searching suggest that the MADS domain evolved from a coding region of DNA-topoisomerase II via a duplication event in the lineage that led to the most recent common ancestor

[1] Address correspondence to chloe.zubieta@cea.fr.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Chloe Zubieta (chloe.zubieta@cea.fr).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.114.127910

of extant eukaryotes (Gramzow et al., 2010). Based on sequence homology and preferred DNA sequence and conformation, the MADS domain TFs fall into two distinct lineages: type I (SRF-like) and type II (MEF2-like) (Alvarez-Buylla et al., 2000). Both types recognize a similar CArG-box consensus sequence, $CC(A/T)_6GG$ (Pollock and Treisman, 1990) and $CTA(A/T)_4TAG$ (Pollock and Treisman, 1991), respectively, with additional specificity due to the flanking regions of the consensus sequence.

In plants, the type I MADS box genes are compartmentalized into one or two exons encoding the MADS DNA binding domain and an ancillary and highly variable C-terminal domain. The type I TFs do not have well-defined, plant-specific domains, and relatively little is known about their dimerization and DNA binding specificity in planta. In contrast, the type II genes comprise an average of seven exons and contain three plant-specific domains that are seminal for their expanded role in plant development (Rounsley et al., 1995; Theissen et al., 1996; Egea-Cortines et al., 1999). In addition to the MADS DNA binding (M) domain, the type II TFs contain the intervening (I) domain, keratin-like coiled-coil (K) domain, and C-terminal (C) domain (Theissen et al., 1996; Kaufmann et al., 2005) (Figure 1A). The I domain plays a role in dimer formation and specificity (Masiero et al., 2002), the K domain is important for both dimerization and tetramerization (Yang et al., 2003; Yang and Jack, 2004), and the C domain, a highly variable and largely unstructured domain based on secondary structure prediction, is important in some proteins for transactivation and higher order complex formation (Egea-Cortines et al., 1999; van Dijk et al., 2010). The addition of these ancillary domains, which are not present in protist, animal, or fungal MADS TFs, allows the plant type II MADS TFs (also called MIKC-type after their conserved domain structure) to form different homo- and heterodimeric and tetrameric complexes with other MADS domain proteins. The choice of partners and the cellular context of these complexes are responsible for triggering specific developmental processes. The functional consequence of this can be seen, for example, in the class A, B, C, D, and E floral homeotic genes whose encoded MADS domain TFs determine the correct formation of sepals, petals, stamens, ovules, and carpels (Theissen and Saedler, 2001).

Floral organ development depends of the combinatorial activity of class A-E MADS box genes whose overlapping expression patterns determine the identity of all the floral organs. This is postulated to occur via the assembly of organ-specific tetrameric MADS domain protein complexes ("floral quartets") that are able to bind two DNA sites in the regulatory regions of target genes, causing a DNA loop and resulting in target gene expression or repression and thus determining developmental fate (Theissen and Saedler, 2001; Melzer and Theissen, 2009; Smaczniak et al., 2012). As revealed by extensive genetic experiments, the class E genes are necessary for the formation of all floral organs (Melzer et al., 2009). The most promiscuous member of the E class in terms of interaction propensity is SEPALLATA3 (SEP3); based on yeast two-hybrid screening, it has been shown to form over 50 different complexes, including complexes with all other homeotic type II MADS domain TFs (Immink et al., 2009). However, the atomic level determinants for complex formation and specificity are not well understood.

In order to elucidate the rules governing MADS domain TF complex formation, structural characterization of the oligomerization domains of the proteins is critical. Here, we report the 2.5-Å crystal structure of a small portion of the I domain and complete K domain from *Arabidopsis* SEP3, mutagenesis studies of the tetramerization interface of the SEP3 K domain, and atomic force microscopy (AFM) experiments demonstrating looping of target DNA by the full-length SEP3 protein.

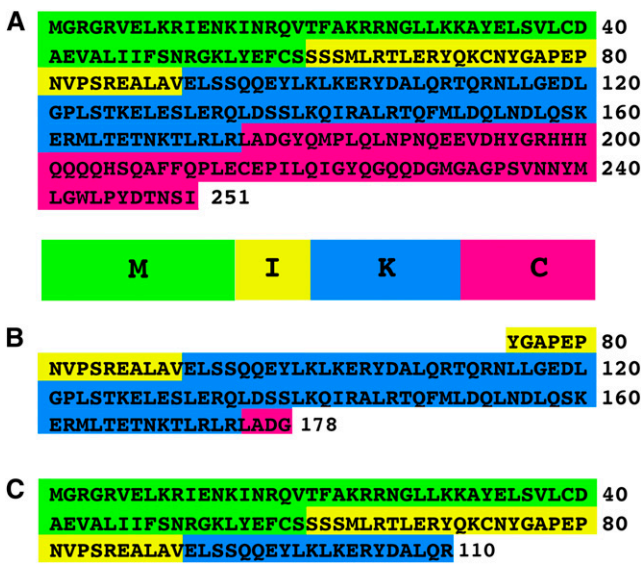


Figure 1. Amino Acid Sequence of SEP3 and Truncation Constructs.

(A) SEP3 sequence colored by domain, with the M domain in green, the I domain in yellow, the K domain in blue, and the C domain in pink. The domain structure is depicted schematically below the amino acid sequence.

(B) Sequence of the $SEP3^{75-178}$ construct used for all crystallization studies spanning a portion of the I domain, the complete K domain, and a portion of the C domain.

(C) Sequence of the $SEP3^{1-110}$ construct used in the AFM studies.

RESULTS

In order to find soluble and well-expressing constructs of the MADS domain TF, SEP3, we performed library screening of ~3000 constructs using the ESPRIT random library method, which identifies well-expressing soluble domain constructs in poorly annotated regions (Tarendeau et al., 2007; Yumerefendi et al., 2010). The construct comprising residues 75 to 178 ($SEP3^{75-178}$) was selected for further studies (Acajjaoui and Zubieta, 2013) (Figure 1B). This construct contained the complete K domain (91 to 173) and overlapped a portion of the I domain (residues 75 to 90) and the C domain (residues 174 to 178).

The protein was purified via affinity chromatography and gel filtration as a mixture of tetramer and dimer. The protein exhibited a small degree of concentration-dependent oligomerization, with the higher molecular weight peak corresponding predominantly to a tetrameric species. This peak was concentrated and used for all crystallization trials. While it is likely that the protein reequilibrated to a mixture comprising dimers and tetramers, this did not impede crystallization. $SEP3^{75-178}$ crystals grew in space group $P2_12_12$ with diffraction to 2.5 and 3.2 Å for the native and seleno-methionine

derivatized protein, respectively. Final data collection and refinement statistics are summarized in Table 1. The protein crystallized with four monomers per asymmetric unit, with the tetrameric biological unit, a dimer of dimers, formed via a crystallographic 2-fold rotation. Residues 83 to 175 (monomer A), 83 to 177 (monomer B), 88 to 178 (monomer C), and 93 to 175 (monomer D) were clearly visible in the electron density (Figure 2). Disordered N- and C-terminal residues were not modeled. Each monomer folded into two long amphipathic alpha helices (helices 1 and 2) with a kink of ~90° between helices (Figure 3A). Helices 1 and 2 comprise leucine zipper-like heptad repeats (abcdefg) with hydrophobic residues at the a and d positions and charged residues at the e and g positions. Each monomer in the asymmetric unit associates with a partner via interactions mediated by the C-terminal portion of helix 2. Helix 1 intermolecular interactions occur upon a 2-fold rotation and provide an extensive interface comprising all of helix 1 and the N-terminal residues of helix 2. The kink region between helices 1 and 2 breaks the canonical heptad repeats, preventing a single leucine zipper from forming (Figure 3B). This glycine- and proline-rich kink region (residues 117 to 127; Gly-117, Gly-121, and Pro-122) forces the two helices apart and is stabilized by extensive intramolecular hydrophobic interactions of multiple leucine residues (Leu-115, -120, -123, -128, -131, and -135). Further stabilization of the kink region is provided by hydrogen bonding interactions between Glu-127 and Ser-124 and a salt bridge between residues Arg-113 and Glu-118. This configuration, comprising both hydrogen bonding interactions and hydrophobic packing, not only impedes self-association into a monomeric intramolecular coiled-coil but also hinders the dimerization of monomers into a single leucine zipper fold. Thus, each monomer presents two distinct amphipathic helices that are able to act independently during oligomerization.

Table 1. Data Collection and Refinement Statistics $SEP3^{75-178}$

Data Collection	
Space group	P21212
Cell dimensions	
a, *b*, *c* (Å)	123.1, 143.2, 48.77
α,β,γ (°)	90, 90, 90
Resolution (Å)	60-2.49 (2.55-2.49)*
R_{sym} or R_{merge} (%)	6.1 (40.1)
$I/\sigma(I)$	17.6 (4.3)
Completeness (%)	77.2 (20.2)
Redundancy	5.9 (6.1)
Refinement	
Resolution (Å)	28.1-2.49
No. reflections	23,723
R_{work}/R_{free} (%)	27.4/23.0
No. atoms	3,362
Protein	2,970
Ligand/ion	0
Water	392
B-factors (Å^2)	
Protein	69.4
Ligand/ion	–
Water	56.8
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.2

The asterisk refers to the highest resolution shell. R.m.s., root mean square.

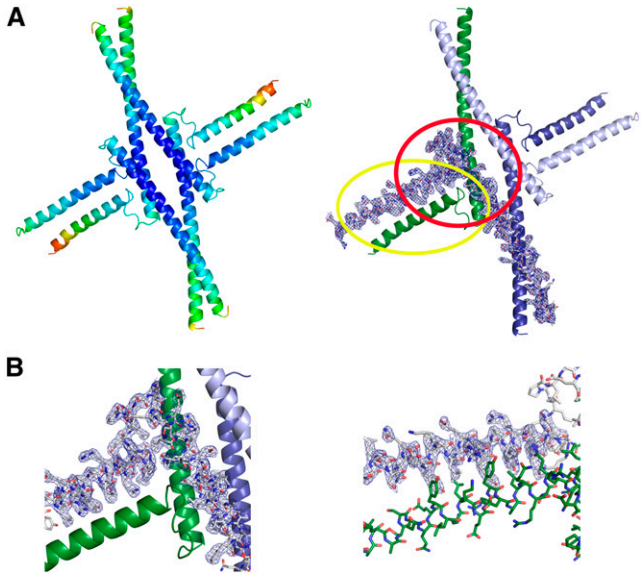


Figure 2. Overview of Structural Quality.

(A) At left, SEP3 tetramer depicted as a cartoon and colored by temperature factor (B-factor) with dark blue (lowest) and red (highest). The average B-factor for the structure was 69 Å^2. At right, view as per left with each monomer colored uniquely and one monomer displayed with 2Fo-Fc electron density contoured at 1.5 sigma. The loop region is circled in red and the dimerization region in yellow.

(B) At left, close-up of the loop region corresponding to the red circled region in **(A)**, right. At right, electron density for the dimerization region corresponding to the yellow circled region in **(A)**, right. Based on the quality of the electron density map, the protein backbone and side chains could be positioned unambiguously.

Dimer Interface

Interactions between the N-terminal helices (helix 1) of two partner monomers result in the formation of a left-handed coiled-coil, with the C-terminal helices (helix 2) oriented 180° apart, precluding intramolecular association of these regions from the same dimer. The coiled-coil of helix 1 comprises two heptad repeats of tyrosine and leucine 98-YxxLxxxYxxLxxx-111 forming a large hydrophobic interaction surface. In addition, three pairs of salt bridges are formed between partner monomers of helix 2 comprising residues Glu-129/Arg-146, Glu-132/Arg-146, and Asp-136/Arg-143 (Figure 3C), further stabilizing the dimer interface which buries over 3000 Å^2 (17% of the total accessible surface area of the dimer), as calculated with AREAiMOL (Lee and Richards, 1971; Winn et al., 2011).

The DNA binding MADS domain (residues 1 to 58) and a short portion of the I domain (residues 59 to 74) were removed in the construct used for crystallization. Based on homology to the structurally characterized mammalian (Pellegrini et al., 1995; Santelli and Richmond, 2000) and yeast (Tan and Richmond, 1998) MADS

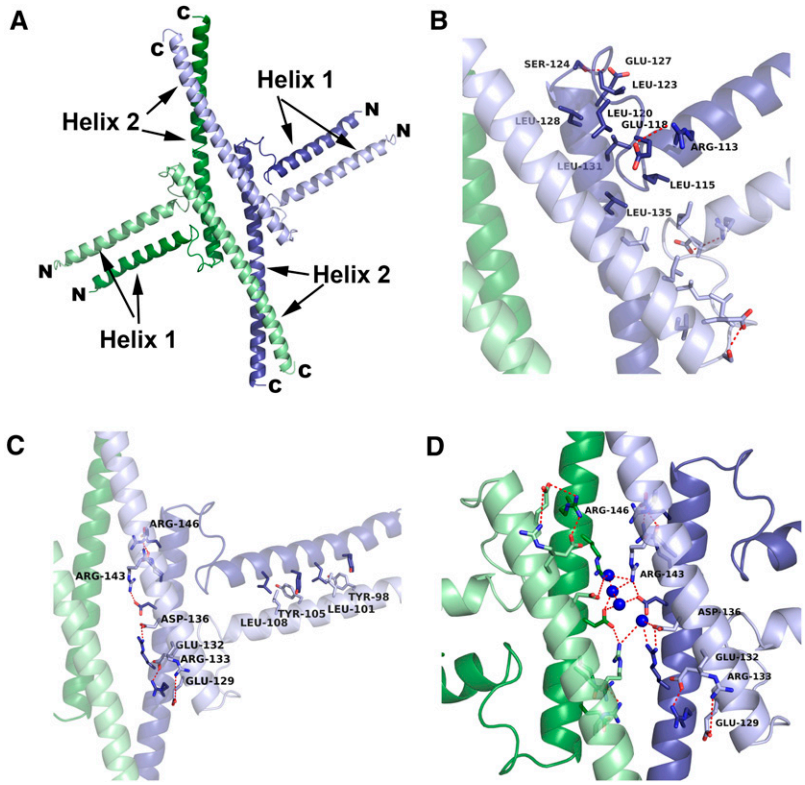


Figure 3. Structure of SEP3 Oligomerization Domains.

(A) SEP3 tetramer depicted as a cartoon, with each monomer A to D colored uniquely in light green, dark green, light purple, and dark purple, respectively, with the N and C termini labeled. Helix 1 and helix 2 are labeled and indicated by arrows.
(B) The hydrophobic kink region is shown for one monomer with the view as per **(A)**. Residues are labeled and drawn as sticks colored by atom. Hydrogen bonds are shown as dashed red lines.
(C) Dimerization of SEP3 is mediated by leucines and tyrosines in helix 1 and intermolecular salt bridges via the N-terminal portion of helix 2. Residues are labeled and depicted as sticks colored by atom. Hydrogen bonds are drawn as dashed red lines. For clarity, residues are labeled for one monomer.
(D) View down the 2-fold crystallographic axis that forms the tetramerization interface. The intermolecular water-mediated hydrogen bonding network is shown. Residues are depicted as sticks and colored by atom, water molecules are in dark blue, and residues labeled for a single monomer for clarity.

domains, the M domain of SEP3 would fold into a functional and intertwined dimer in a similar manner, as these domains are evolutionarily highly conserved across kingdoms. Indeed, the M domain is only competent to bind DNA as a dimer. The presence of the MADS domain adjacent to the coiled-coil of helix 1 would act to further stabilize the dimeric conformation seen in the crystal structure described here.

Tetramer Interface

In the dimeric arrangement formed via the coiled-coil interactions of helix 1, the glycine-proline kink between helices 1 and 2 forces the two C-terminal helices 180° apart and requires the presence of a second MADS dimer with the same arrangement of amphipathic helices to form the tetramer (Figures 3 and 4). Thus, two homodimers of SEP3 K domain are able to associate into a tetramer primarily due to hydrophobic interactions of helix 2. Residues (150-MxxxLxxLxxxxxxLxxxxxxL-171) form an interacting hydrophobic surface and bury a total of 2700 Å^2 (~9% of the total surface area of the tetramer). A salt bridge between Lys-160 and Glu-161 and a hydrogen bond between Thr-167 and Asn-168 of its partner help lock the helices 2 together, although the tetramerization interface formed via helix 2 is much less extensive than that of the dimer formed via helix 1 (Figure 4). In addition to the helix 2 interactions, the 2-fold crystallographic axis of the dimer-dimer interface is rich in charged residues and forms an extensive water-mediated hydrogen bonding network that may further stabilize the tetramer (Figure 3D).

In order to investigate the importance of the hydrophobic residues of the tetramerization interface formed by helix 2, site-directed mutagenesis was performed. Interestingly, the protein was highly sensitive to even a single alanine mutation in this region. Point mutations dramatically abrogated tetramer formation, as observed by gel filtration experiments, with M150A, L154A, and L171A all showing impaired tetramerization versus the wild-type construct, with the dimeric species dominating even at relative high protein concentrations of 10 to 13 mg/mL (Figure 5). In addition to the methionine and leucine mutants, a truncation mutant comprising residues 75 to 149 (Sep3$^{75\text{-}149}$) showed no tetramer formation, as would be predicted by the removal of helix 2.

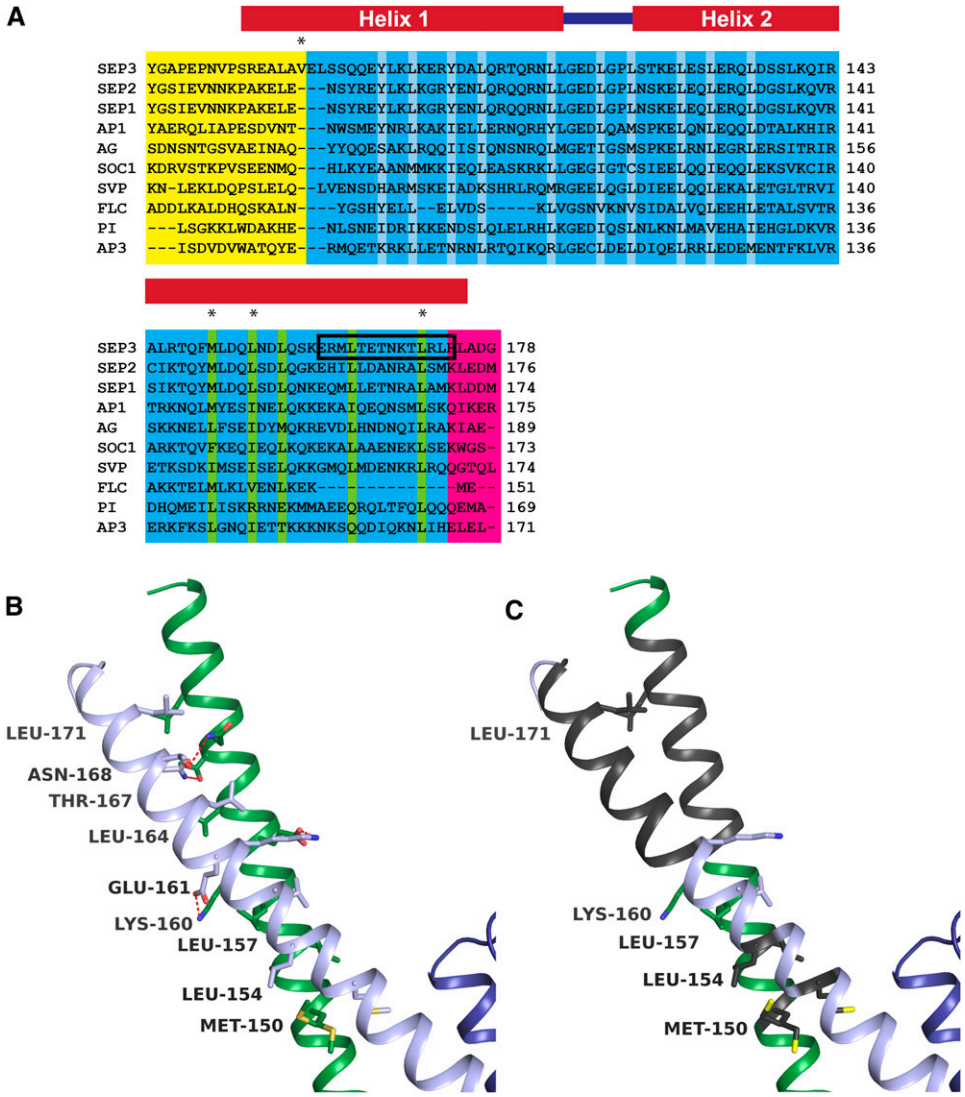


Figure 4. Tetramerization Interface of SEP3.

(A) Sequence alignment of representative MADS TFs spanning the sequence of the crystallized SEP3$^{75\text{-}178}$ construct described here with the I domain in yellow, the K domain in blue, and the C-terminal domain in pink. Residues involved in dimerization and tetramerization are highlighted in light blue and light green, respectively. Mutants are marked with a star, and residues corresponding to deletion mutant SEP3$^{\Delta 161\text{-}174}$ are boxed in black. All proteins are from *Arabidopsis* with SEP3, SEP2, SEP1, AP1, AG, SOC1, SVP (SHORT VEGETATIVE PHASE), FLC, PI, and AP3.
(B) Close-up of the tetramerization interface of SEP3. Interacting residues are depicted as sticks and colored by atom with monomers colored uniquely. Hydrogen bonds are show as dashed red lines. Residues from the green monomer are labeled.
(C) Residues that are deleted in SEP3$^{\Delta 161\text{-}174}$ are shown in cartoon colored gray. Point mutants that affect tetramerization are depicted as sticks and colored gray. Labels are as per **(B)**.

We also investigated three natural alternate splice variants of SEP3 based on contemporary TAIR annotations (www.arabidopsis.org), despite a previous report identifying only two SEP3 splice variants (Severing et al., 2012). These splice variants differ in the K domain. The additional splice variants deviate from the wild-type sequence (crystallized in this study) due to an alternate 3′ splicing donor site producing a valine deletion in helix 1 at position 90 (SEP3$^{\Delta V90}$) or a 14-amino acid deletion in helix 2 from skipping of exon 6 (SEP3$^{\Delta 161\text{-}174}$). Gel filtration experiments showed little change in oligomerization state between the wild-type and splice variant SEP3$^{\Delta V90}$, with the ΔV90 protein forming primarily tetrameric species in solution. This would be expected as V90 is in a region of the crystallized construct that does not contribute to the dimer or tetramer interfaces. However, splice variant SEP3$^{\Delta 161\text{-}174}$ was completely dimeric, as determined by gel filtration and confirmed by light scattering experiments (Figure 5). Based on these mutagenesis experiments and the characterization of natural splice variants, tetramerization is easily perturbed by changes in the hydrophobic

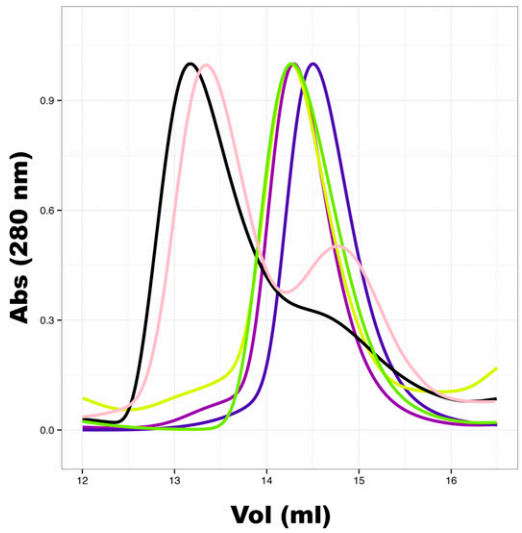


Figure 5. Size-Exclusion Chromatograms of Wild-Type and Mutant SEP3 Proteins.

SEP3$^{75\text{-}178}$ (wt) is in black, SEP3$^{\Delta V90}$ is in pink, SEP3$^{\Delta 161\text{-}174}$ is in dark blue, M150A is in green, L171A is in purple, and L154A is in yellow. The oligomerization state of the point mutants and SEP3$^{\Delta 161\text{-}174}$ was predominantly dimer as confirmed by multiangle laser light scattering. M150A gave a molar mass of 20,560 g mol^{-1} (±7.3%), L154A 22,990 g mol^{-1} (±9.5%), and SEP3$^{\Delta 161\text{-}174}$ 18,790 g mol^{-1} (±3.8%), all corresponding to predominantly a dimeric species in solution (calculated molecular mass of the dimer ~24 kD). L171A was not measured with multiangle laser light scattering; however, its elution profile was the same as the other point mutants. The wild type and SEP3$^{\Delta V90}$ eluted as a mixture of tetramer and dimer as shown in the chromatograms. All chromatograms were overlaid and the maximum absorbance at 280 nm normalized to 1.

residues in helix 2 of the K domain. This sensitivity may have physiological relevance due to the presence of a natural splice variant that lacks the capacity to homotetramerize and would likely have impaired heterotetramerization based on the highly conserved nature of repeating hydrophobic residues in helix 2 for representative homeotic MADS TFs (Figure 4A). Conservative changes in helix 1 N-terminal to the dimerization interface do not have the same dramatic effect on oligomerization, likely due to the more extensive protein-protein contacts, as noted for the SEP3$^{\Delta V90}$ variant with its similar elution profile to that of the wild-type construct. However, as the construct used in crystallization lacks the M domain and the complete I domain, additional interactions in the I domain may be present, which further stabilize the dimer interface and cannot be ruled out.

DNA Binding Domain Models

The crystallized construct starts after the N-terminal DNA binding M domain (residues 1 to 58) and lacks a portion of the I domain (residues 59 to 74). Structural data are available for the M domain of the mammalian proteins Myocyte-specific enhancer factor 2A (MEF2A) (Perry et al., 2009; Wu et al., 2010; He et al., 2011) and serum response factor (SRF) (Pellegrini et al., 1995; Hassler and Richmond, 2001; Mo et al., 2001) and the fungal protein Minichromosome maintenance protein 1 (MCM1) (Tan and Richmond, 1998), which are all obligate dimers. Based on homology to the structures of MEF2A and SRF M domains (sequence identity of 58 and 47%, respectively, over residues 1 to 58), composite structures of SEP3 encompassing the MIK domains were modeled (Figure 6). The available structural data for the MADS TFs includes residues C-terminal to the M domain, the MEF domain (for MEF2A, residues 60 to 89) (He et al., 2011), and the SAM (SRF/Arg80/MCM1) domain (for SRF, residues 198 to 227) (Pellegrini et al., 1995). The I domain present in SEP3 bears little sequence similarity to the MEF or SAM domains, but these domains have approximately the same number of amino acids as the I domain and, in the case of the MEF domain, are intrinsically folded into the M domain. The crystal structure of MEF2A (PDB code 3KOV) in complex with DNA has a third beta strand that extends the beta sheet of the M domain and an additional alpha helix. This fold positions the N and C termini of the MADS/MEF domain on the same face as the bound DNA. The MADS/MEF-fold modeled onto SEP3 would force a configuration in which the DNA is clamped between helices 1 of the dimer in a conformation resembling a bZIP TF (Hurst, 1995; Vinson et al., 2006). However, helix 1 does not have a series of basic residues to help stabilize the DNA. As shown in Figure 6B, left, steric clashes with the bound DNA would force the coiled-coil of helix 1 apart, disrupting the coiled-coil and, in order for the DNA to bind, at least a partial unfolding of the structure would be necessary, making this conformation unlikely. The SRF structure, which includes 25 residues C-terminal to the M domain (residues 198 to 222), lacks the third beta strand present in the MEF2A structure and terminates with an alpha helix positioned opposite the DNA binding surface. This structure is compatible with the K domain fold of SEP3 as shown in the composite model, with the two DNA binding domains distally oriented on the rigid helical arms of the IK domains via a random coil. Secondary structure predictions (Jones, 1999) for residues 59 to 74 of the I domain predict an alpha helical stretch (63 to 73) followed by a random coil, which is similar to the SRF structural model (Figure 6B, right).

AFM Studies of SEP3

An important aspect of the putative activity of SEP3, and the plant MADS TFs in general, is their ability to form tetramers in the context of DNA binding. Tetramer formation is able to purportedly loop DNA, which has been shown indirectly through gel shift assays (Melzer and Theissen, 2009; Melzer et al., 2009) and more recently through tethered particle motion (Mendes et al., 2013). In order to determine more directly whether the homotetramerization observed in the crystal structure for SEP3 could occur in the context of protein-DNA complexes, AFM experiments were performed using the full length GST-SEP3 protein and a truncated version (SEP3$^{1\text{-}110}$), which has the DNA binding domain but lacks the K-domain necessary for tetramerization (Figure 1C). The inclusion of a 6-His/glutathione *S*-transferase (GST) tag for the full-length protein resulted in a more easily purified construct that showed a lower propensity for aggregation versus the cleaved protein. The truncated SEP3$^{1\text{-}110}$ did not show the same propensity for aggregation and the 6-His tag was

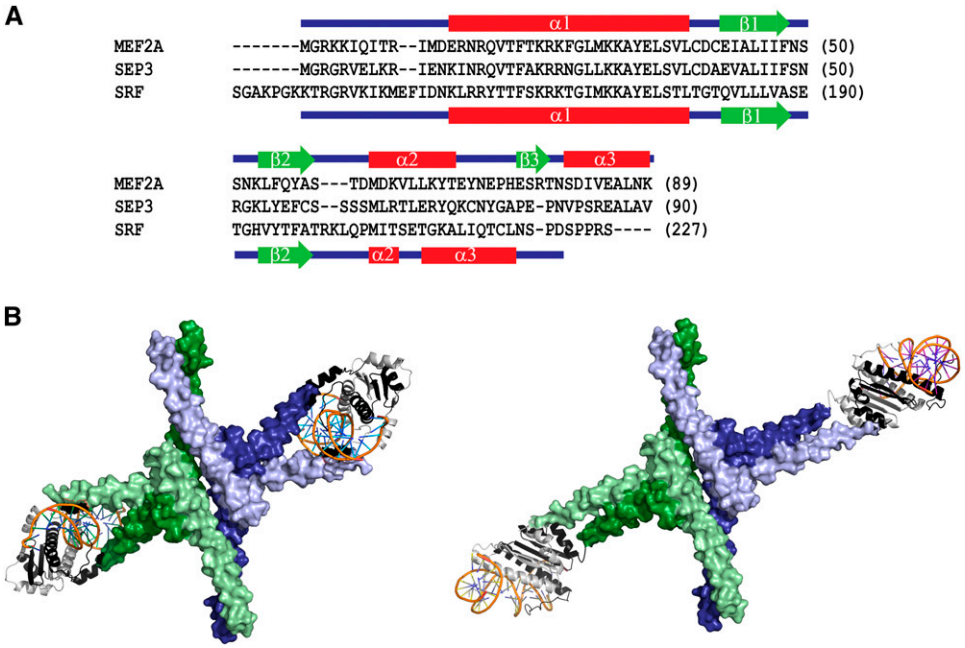


Figure 6. Comparison of MEF2A, SRF, and SEP3.

(A) Partial sequence alignment of SEP3 from *Arabidopsis*, MEF2A from *H. sapiens*, and SRF from *H. sapiens*. The M domains span residues 1 to 58 of SEP3, residues 1 to 59 of MEF2A, and residues 141 to 197 of SRF. The SEP3 I domain (59 to 90), MEF domain (residues 60 to 89), and SRF SAM domain (198 to 227) were included in the structure-based sequence alignment. Helices are depicted as red cylinders, random coils as blue lines, and beta sheets as green arrows with the MEF2A secondary structure elements above and the SRF secondary structure elements below.

(B) Alternate composite models of SEP3 MIK domains using the structure of MEF2A residues 1 to 89 (PDB 3KOV), left, and the structure of SRF structure residues 141 to 227 (PDB 1SRS), right. The DNA binding site is located at the distal extremes of the tetramer based on the SEP3$^{75-178}$ (I and K domains) structure determined here. The SEP3 structure is displayed as a surface colored by monomer and the MEF2A (left) or SRF (right) structure as a cartoon with protein in light and dark gray and DNA in orange and blue. The model with MEF2A requires an opening of helix 1 of the SEP3 K domain to accommodate the DNA.

cleaved prior to AFM studies. Based on previous studies predicting the binding of a SEP3 homotetramer to two adjacent CArG boxes in the *SUPPRESSOR OF CONSTANS1* (*SOC1*) promoter region (SEP3 is able to act as a repressor of *SOC1* expression) (Muiño et al., 2014), we used a 1-kb fragment from this promoter comprising two CArG boxes that are separated by 93 bp. Depending on the concentration of SEP3 used (~2 to 5 nM or 10 to 15 nM), the protein was able to bind either 1 or 2 CArG box sites. At higher protein concentrations with two CArG boxes bound, intramolecular DNA looping was observed as well as intermolecular associations, offering direct in vitro evidence of DNA looping by SEP3 homotetramers (Figure 7). Attempts to form DNA loops with longer spacing between CArG boxes were unsuccessful, suggesting an optimum spacing of binding sites is necessary for looping to occur. As a control, SEP3$^{1-110}$, which lacks the keratin-like domain and is thus unable to tetramerize, was used to test whether DNA looping was due to tetramer formation. Even at high protein concentrations (25 nM), in which nonspecific protein-DNA binding occurred, no DNA looping was observed (Figure 7; Supplemental Figures 1B to 1D). Attempts to purify the tetramerization point mutants and splice variants for the full-length construct with a variety of tags were unsuccessful due to poor recombinant overexpression and extensive aggregation of the proteins. As GST is known to dimerize, we cannot exclude that the tag helps stabilize the M domain in the GST full-length construct. However, it is unlikely that the N-terminal GST tag interferes with the tetramerization interface of SEP3 and consequently will not play a role in the tetramer-dependent DNA looping reported here.

DISCUSSION

The plant type II MADS domain TFs (MIKC-type proteins) have acquired an ancient DNA binding domain and elaborated on it through the addition of a coiled-coil domain, the keratin-like domain. Coiled-coil domains are ubiquitous modules that allow proteins to increase their functionality by providing protein-protein interaction surfaces (Mason and Arndt, 2004). By presenting two amphipathic alpha helices that are capable of forming intermolecular coiled-coils, the MADS TFs have obtained a versatile oligomerization interface, which allows for homo- and heterodimers and tetramers with other K domain-containing MADS TFs. Through a hydrophobic kink region that orients helices 1 and 2 ~90° apart, self-association is prevented and each helix can act independently during oligomerization, thus increasing the potential diversity of complexes formed.

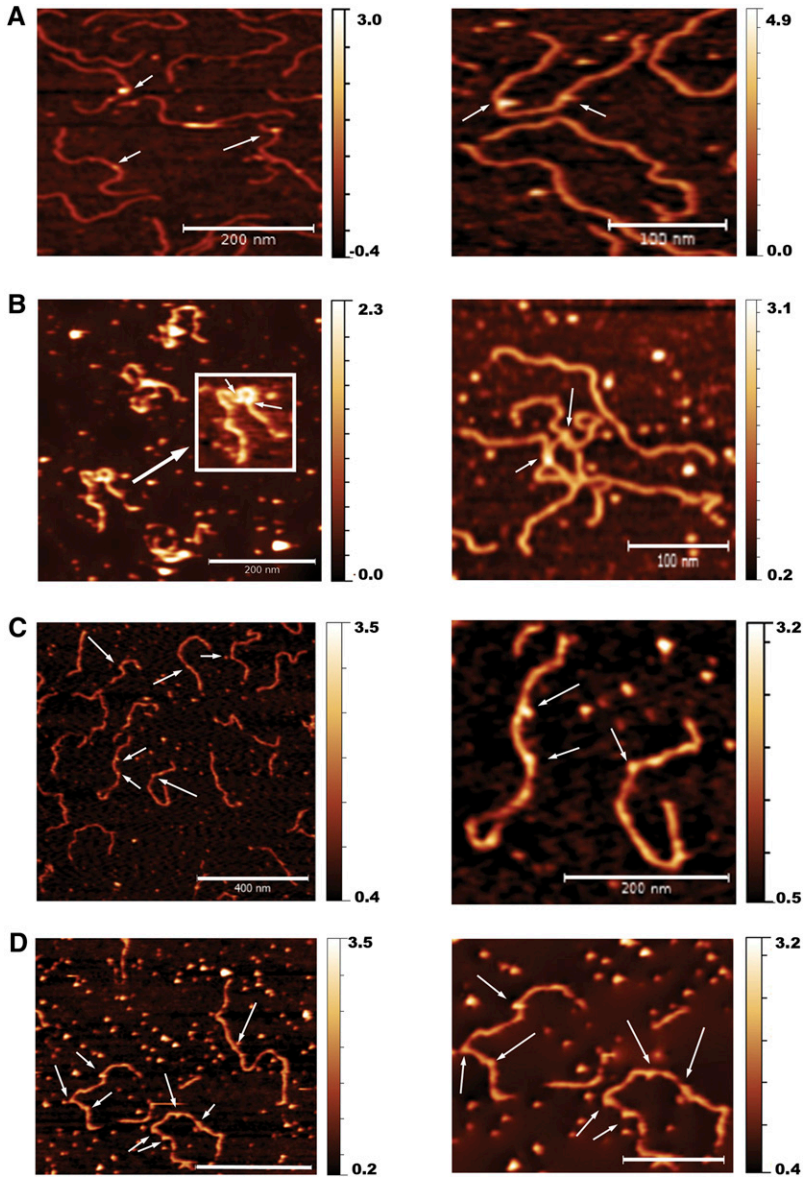


Figure 7. Atomic Force Micrographs of SEP3 in Complex with *SOC1* Promoter DNA.

Heights are color coded in nanometers at right of each image.
(A) Full-length SEP3 in complex with a 1-kb DNA comprising two putative SEP3 binding sites. Arrows indicate bound SEP3 proteins. The bar at left is 200 nm and at right is 100 nm. DNA-protein complexes were formed at 2 to 5 nM protein and DNA and diluted to ~1 nM for imaging.
(B) Complex of SEP3 and DNA as per **(A)** with the complex formed at 10 to 15 nM protein and 5 nM DNA before dilution to 1 nM DNA concentration for imaging. Arrows indicate looping of DNA. Inset highlights the SEP3-DNA complex. The bar is 200 nm (left). Intermolecular interactions of SEP3 and DNA under the same conditions were observed (right). The bar is 100 nm.
(C) SEP3$^{1-110}$ lacking the K domain in complex with DNA. The complex was formed at 5 nM protein and DNA and diluted to 1 nM for imaging. Arrows show protein bound to DNA. No inter- or intramolecular looping of DNA was observed. Image at right is a close-up view. Bars are 400 nm (left) and 200 nm (right).
(D) SEP3$^{1-110}$ in complex with DNA as per **(C)** with a protein concentration of 25 to 5 nM DNA before dilution to 1 nM DNA concentration for imaging. Image at right is a close-up view with image masking to remove tailing. Bars are 200 nm (left) and 100 nm (right). Proteins bound to DNA are indicated by arrows.

The oligomerization patterns of the MADS TFs have profound implications in downstream developmental processes. As high-level regulators, the MADS TFs trigger the expression or repression of thousands of target genes. For example, based on ChiP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) studies, over 4000 binding sites and over 3000 potential target genes have been identified for SEP3 (Kaufmann et al., 2009). These targets are not only SEP3 homooligomer targets, but rather targets of SEP3 and all its possible multiprotein complexes. The activity of SEP3, and by extension other MADS

domain TF family members, is due largely to its ability to interact with different partners through a highly adaptable K domain. Helices 1 and 2 have repeating series of hydrophobic residues that provide hydrophobic interaction surfaces for dimer and tetramer formation. These residues are well conserved in the SEP class of MADS domain TFs but are more variable in other MADS domain TF classes (Figure 4A). The differences in hydrophobic amino acids help to account for the promiscuous activity of the SEP proteins by providing a versatile and plastic interaction surface for different partners. Alterations in this pattern will potentially destabilize protein-protein interactions, resulting in changes in heterocomplex formation.

The dimerization and tetramerization interfaces identified in the structural studies presented here are likely conserved in the formation of heterooligomers of the MADS TFs and determine the relative stability of the final MADS protein complex. For example, based on amino acid substitutions Y98K and L115R in APETALA3 (AP3) (Figure 4A), homodimerization would be disfavored. AP3 forms an obligate heterodimer with PISTILLATA (PI), and this is likely due to compensatory amino acid substitutions, specifically Y98I (numbering for SEP3), which would be able to accommodate the aliphatic portion of the AP3 lysine residue. A bulky tyrosine residue present in SEP3 would be disfavored at this position, leading to preferential AP3/PI heterodimerization versus a SEP3/AP3 heterodimer. Likewise, PI has a number of leucine-to-isoleucine and valine substitutions in the hydrophobic residues predicted to be important for coiled-coil formation. This may lead to subtle destabilization of a PI homodimer and the preferential formation of an AP3/PI heterodimer. Additional interactions, such as hydrogen bonding or salt bridge formation between residues flanking the hydrophobic positions, may also play a significant role in complex stabilization; however, these interactions are more difficult to identify based on the limited structural data for SEP3. Thus, the MADS TFs likely form very dynamic complexes that exist in complex equilibria. Relatively weak hydrophobic interactions, which are highly sensitive to even conservative amino acid substitutions, lead to sufficient stabilization or destabilization of functionally important complexes resulting in downstream gene activation or repression. By examining the changes in amino acids at the positions identified from the structure as critical for oligomerization, the likelihood of direct interaction between MADS TFs in planta, assuming they are expressed in the same temporal and spatial manner, can be predicted. These data identify "hot spots" of dimerization and tetramerization interactions and will allow highly targeted point mutations to critical residues. Mutagenesis studies, in conjunction with yeast two-hybrid screening and forward genetics, will be critical for fully elucidating how the complicated oligomerization patterns of the MADS TFs are determined at the amino acid level. This structural work provides an important foundation for these further studies.

While the dimeric arrangement of the MADS domain TFs buries much more surface area and encompasses both helix 1 and a portion of helix 2 of the K domain, the tetramerization interface is limited to the C-terminal portion of helix 2 and a water-mediated hydrogen bonding network along the dimer 2-fold axis. Based on gel filtration studies of the isolated oligomerization domain, homotetramerization of SEP3 is relatively weak and easily perturbed by point mutations in the hydrophobic patches of helix 2 (M150A, L154A, and L171A). This suggests three possible explanations: (1) heterotetramerization is much stronger than the homotetramerization observed here, (2) tetramerization of the MADS domain TFs in general is relatively weak and of functional consequence only in the context of DNA binding, or (3) additional cofactors are necessary to stabilize MADS tetrameric complexes in planta. Due to the observed dimer-tetramer equilibrium even at high micromolar protein concentrations, the limited tetramerization interface and the sensitivity of tetramerization to even single alanine mutations, it is doubtful that tetramers of SEP3 are the dominant species present at physiological concentrations. Based on sequence alignments of MADS domain TF K domains (Figure 4A), it is unlikely that heterotetramerization will lead to greater tetramer stability as the residues involved in homo- and heterotetramerization are relatively well conserved. However, at high protein concentrations used in crystallization studies or when bound at appropriately spaced sites on DNA, homo- and/or heterotetramerization is much more liable to be significant. As shown in the AFM experiments described above, SEP3 is able to form tetrameric complexes and loop DNA when bound on adjacent sites on the same DNA strand or when higher protein concentrations are used during protein-DNA complex formation. Thus, strong tetramerization may not be required for in planta function. When two MADS dimers are located near each other either on adjacent DNA binding sites or on distal sites that come into close contact depending on chromatin conformation, for example, tetramerization could occur. This would enable the MADS TFs to act as a dynamic interaction network, exploiting chromatin events that temporarily bring distal bound MADS dimers together to enable tetramer formation, as well as forming tetramers when bound on adjacent sites on the DNA. This raises the intriguing possibility that multiple CArG boxes in the promoter regions of different genes could facilitate the formation of transcription factories (Sutherland and Bickmore, 2009) with the MADS domain TFs playing a role in recruiting different genes to these sites of transcription. The ability of the MADS TFs to regulate different developmental processes is likely contingent upon this dynamic oligomerization, with high affinity for dimerization and DNA binding and lower affinity for tetramerization, which is able to occur only under certain conditions. Additional cofactors may stabilize higher order complex formation and evidence suggests that the MADS TFs interact with other TF families as well as chromatin remodelers (Kaufmann et al., 2005; Smaczniak et al., 2012). Thus, higher order complex formation may help to stabilize tetrameric MADS complexes and requires further investigation.

In addition to heterooligomerization, TFs increase their functional diversity through alternative splice variants. We demonstrate the dramatic in vitro effect of a natural splice variant, $SEP3^{\Delta 161\text{-}174}$, with its inability to homotetramerize. This may have significant in planta effects. Recently, temperature-sensitive phenomenon due to alternate splicing events was demonstrated for the MADS gene *MADS-AFFECTING FLOWERING2* (*MAF2*), a close relative of the key floral repressors *FLOWERING LOCUS C* (*FLC*) and *FLOWERING LOCUS M* (Rosloski et al., 2013). MAF2 full-length protein and splice variants with truncations in the K domain show functional differences with respect to flowering time under cold conditions. Based on alignment with the structure of SEP3, splice variants of MAF2 with K domain truncations would retain the ability to bind DNA and

dimerize but would be unable to tetramerize. Thus, it is probable that alternate splicing in the MADS family profoundly affects many downstream processes and may play a key role in determining developmental fate. Indeed, the majority of characterized splice variants cluster to the I and K domains, suggesting a general method for increasing functional diversity via alterations in oligomerization state and/or oligomerization partners (Severing et al., 2012). The structural data presented here help to provide a putative molecular basis for the observed phenotypes due to MADS splice variants. These data tie changes in in vivo function to changes in MADS TF oligomerization patterns due to alterations in primary sequence and provide a structural template for understanding and predicting oligomerization propensity of different MADS splice variants affected in the K domain.

Plants have dramatically expanded the family of MADS box genes and co-opted these developmental regulatory genes for diverse processes. By fusing the I and K oligomerization domains to the DNA binding M domain, the repertoire of interacting partners is concomitantly increased in plants versus other eukaryotes. The formation of diverse homo- and heterodimers and tetramers is key to the in vivo function of the plant MADS TFs. Thus, the elucidation of the determinants of oligomerization is crucial for understanding the biological complexity and the evolution of plant species. This work provides the structure of the oligomerization domain of the MADS domain TF SEP3 and demonstrates in vitro DNA looping mediated by the full-length protein via tetramer formation. These data provide a foundation for understanding the molecular level determinants of dimerization and tetramerization in the larger family of plant MADS domain TFs.

METHODS

Strains and Plasmids

SEP3$^{75-178}$ and SEP3$^{75-149}$ were cloned into the expression vector pESPRIT2 (Hart and Tarendeau, 2006; Guilligay et al., 2008) using the *Aat*II and *Not*I sites. The plasmid contains an N-terminal 6-His tag followed by a TEV protease cleavage site. SEP3$^{1-251}$ was cloned into the expression vector pETM-30 using the *Nco*I and *Xho*I sites. The plasmid has an N-terminal 6-His/GST tag followed by a TEV protease cleavage site. The SEP3$^{1-110}$ construct was cloned into the expression vector pET15b, which contains an N-terminal 6-His tag followed by a thrombin cleavage site. All mutants were generated from the SEP3$^{75-178}$ construct in the pESPRIT2 vector, and oligonucleotides for the mutants used for PCR are given in Supplemental Table 1. Mutants based on SEP3$^{75-178}$ were generated according to the manufacture's protocol using Phusion polymerase. All SEP3 proteins were overproduced in *Escherichia coli* BL21 (DE3) CodonPlus RIL (Agilent Technologies) except the GST-SEP3$^{1-251}$ construct, which was overproduced in Rosetta 2 (Novagen).

Protein Expression and Purification

Cells were grown in Luria-Bertani medium in the presence of 50 μg mL^{-1} kanamycin and 35 μg mL^{-1} chloramphenicol at 37°C and 180 rpm. At an OD_{600nm} of 0.8, the temperature was lowered to 20°C and protein expression was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min. Cells were resuspended in 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol (β-ME), and 5% glycerol (v/v). Cells were lysed by sonication and the cell debris pelleted at 25,000 rpm and 4°C for 30 min. The supernatant containing His-tagged SEP3$^{75-178}$ or His-tagged SEP3$^{75-149}$ was applied to a 5 mL Ni-NTA column preequilibrated with 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, and 5% glycerol. The bound protein was washed with 30 mM Tris, pH 8.0, 1 M NaCl, 5 mM β-ME, and 5% glycerol and eluted with 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, 5% glycerol, and 250 mM imidazole. Fractions of interest were pooled and dialyzed against 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, and 5% glycerol in the presence of TEV protease overnight at 4°C to remove the polyhistidine tag. After depletion of the TEV protease and uncut protein over the same Ni-NTA column, the cleaved protein was concentrated to ~5 mg/mL and applied to a size-exclusion FPLC column (Superdex 200 10/300 GL; GE Healthcare) preequilibrated with 30 mM Tris, pH 8.0, 300 mM NaCl, 2 mM TCEP, and 5% glycerol. SEP3$^{75-178}$ eluted as a mixture of tetramer and dimer and SEP3$^{75-149}$ eluted as a dimer based on gel filtration. Fractions of interest were pooled and concentrated to 10 to 15 mg/mL for crystallization. Seleno-methionine-derived protein was produced according to standard protocols (Doublié, 1997) and purified as above. All mutants were expressed and purified under the same conditions as wild-type protein.

Purification of GST-SEP3FL

GST-tagged, full-length SEP3 was grown in Rosetta 2 cells as above. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min. Cells were resuspended in lysis buffer containing 50 mM Tris, pH 8.0, 1 M NaCl, 5 mM β-ME, 5% glycerol, 20% sucrose, 1× protease inhibitors (Roche), lysozyme (1 mg/mL), and benzonase (1 μg/mL). Cells were lysed by sonication, the cell debris removed via centrifugation at 25,000 rpm and 4°C for 15 min, and the supernatant applied to a 3-mL IDA column (Macherey-Nagel) preequilibrated with lysis buffer. The column was washed with 15 column volumes of wash buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, and 5 mM β-ME) and the protein eluted with wash buffer plus 300 mM imidazole. Fractions of interest were applied to a heparin column preequilibrated in buffer A (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, and 2 mM TCEP). The protein was eluted using a linear gradient of 0 to 100% buffer B (50 mM Tris, pH 8.0, 1.2 M NaCl, 5% glycerol, and 2 mM TCEP). The protein eluted at ~30% buffer B. Fractions of interest were pooled, concentrated to 0.3 mg/mL, and applied to a size-exclusion FPLC column (Superdex 200 10/300 GL) preequilibrated with 50 mM Tris, pH 8.0, 1.2 M NaCl, 5% glycerol, and 2 mM TCEP.

Purification of SEP3$^{1-110}$

SEP3$^{1-110}$ was grown in *E. coli* BL21 (DE3) CodonPlus RIL cells as above. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min. Cells were resuspended in lysis buffer containing 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, 5% glycerol, 1× protease inhibitors (Roche), lysozyme (1 mg/mL), and benzonase (1 μg/mL). Cells were lysed by sonication, the cell debris removed via centrifugation at 25,000 rpm and 4°C for 15 min, and the supernatant applied to a 3-mL IDA column (Macherey-Nagel) preequilibrated with lysis buffer. The column was washed with 15 column volumes of wash buffer (30 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, and 5 mM β-ME) and the protein eluted with wash buffer plus 300 mM imidazole. Fractions of interest were applied to a heparin column preequilibrated in buffer A (30 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, and 2 mM TCEP). The protein was eluted using a linear gradient of 0 to 100% buffer B (30 mM Tris, pH 8.0, 1 M NaCl, 5% glycerol, and 2 mM TCEP). Fractions of interest were pooled and dialyzed against 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, and 5% glycerol in the presence of thrombin protease overnight at 4°C to remove the polyhistidine tag. After depletion of the thrombin protease using benzamidine sepharose and uncut protein over a Ni-NTA column, the cleaved protein was concentrated to ~10 mg/mL and applied to a size-exclusion FPLC column (Superdex 200 10/300 GL) preequilibrated with 30 mM Tris, pH 8.0, 300 mM NaCl, 2 mM TCEP, and 5% glycerol.

Electrophoretic Mobility Shift Assays

DNA binding activity for SEP3 constructs was confirmed via EMSA assays. Briefly, a 150-bp oligomer comprising two CArG boxes from the *SOC1* promoter DNA was PCR amplified using the following oligonucleotides: 5′-CGTGTCTAAAGAGGCATTTG-3′ and 5′-CGATTAACAATTTTATCTCC-3′ using the 1-kb *SOC1* fragment from AFM studies as template. The forward PCR primer was labeled with TAMRA (Eurofins Genomics). The TAMRA-labeled DNA was run on a 1% agarose gel and purified using a gel purification kit (Qiagen). Protein and DNA were incubated at room temperature for 10 min in a buffer of 30 mM Tris, pH 8.0, and 300 mM NaCl. DNA concentration was held constant at 100 nM and protein concentration varied from 500 nM, 1 μM, 2 μM, and 4 μM for $SEP3^{1-110}$ and 500 nM, 1, 2, 3, 4, 5, 8, and 10 μM for full-length SEP3. DNA-protein complexes (full-length SEP3 or $SEP3^{1-110}$) were run on a 5% polyacrylamide gel using Tris/borate/EDTA buffer under nondenaturing conditions at 4°C then the gel scanned on a Typhoon scanner (GE Healthcare) (Supplemental Figure 1A).

Multiangle Laser Light Scattering

The oligomerization state of $SEP3^{\Delta 161-174}$, M150A, and L154A was determined by MALLS. Fifty microliters of the purified protein at a concentration of 5 to 10 mg/mL was loaded onto an S200 size-exclusion column (Superdex 200 10/300 GL) at a flow rate of 0.5 mL/min. The column was preequilibrated with 30 mM Tris, pH 8.0, 300 mM NaCl, 5 mM β-ME, and 5% glycerol and connected to a multiangle laser light scattering detector (DAWN HELEOS II; Wyatt Technology) and a refractive index detector (Optilab T-rEX; Wyatt Technology). The data were processed with ASTRA 6.0 software (Wyatt Technology). A theoretical molecular weight of 12 kD for the monomer was later used as reference for calculation of the oligomeric state.

Protein Crystallization

$SEP3^{75-178}$ at a concentration of ~10 mg/mL in 30 mM Tris, pH 8.0, 300 mM NaCl, 2 mM TCEP, and 5% glycerol was crystallized in 20 to 25% ethylene glycol using hanging drops at 4°C. Crystals of the selenomethionine-derivatized protein were grown under the same conditions as the native protein with micro seeding of the crystallization drops using crystals of the native protein. Crystals grew to dimensions of ~200 × 200 × 100 μm over 2 to 3 days. Crystals were harvested and cryo-cooled without further cryoprotection for data collection at 100K.

Data Collection, Processing, and Refinement

A native diffraction data set to 2.5 Å was collected at 100K on beamline ID14-4 of the ESRF in Grenoble, France. Indexing was performed using EDNA (Incardona et al., 2009) and the default optimized oscillation range and collection parameters used for data collection. The data set was integrated and scaled using the programs *XDS* and *XSCALE* (Kabsch, 2010). Data for the seleno-methionine derivative was collected at the peak absorbance (1.07 Å, ID23-1, ESRF) to 3.18 Å, processed as for the native, and phasing was performed using SHELXD/E (Sheldrick, 2010). Based on the obtained phases, a partial structure was built and used as a molecular replacement model for the higher resolution native data. The selenomethionine structure was not further refined. All data sets collected exhibited a high degree of anisotropy with reflections along the c-axis showing the poorest diffraction. Based on this, the data were processed through the UCLA MBI anisotropy server (Strong et al., 2006) and the optimized direction dependent resolution limits were used during refinement. This anisotropy accounts for the relatively low overall completeness of the data at 2.5 Å (77%), as a resolution cutoff of 3.5 Å was applied to the c-axis reflections. The map quality, $R_{working}$, and R_{free} statistics were dramatically improved after applying the corrected resolution limits based on analysis of the anisotropy. All refinements were performed using BUSTER (Bricogne et al., 2011), and model building was performed in Coot (Emsley et al., 2010). Final Ramachandran statistics were 97.73% preferred, 1.13% allowed, and 1.13% outlier. The outlier residues corresponded to Leu-116 in all monomers, which is in a highly kinked region of the protein.

AFM Measurements

A 1-kb linear fragment containing *SOC1* promoter DNA was PCR amplified from *Arabidopsis thaliana* genomic DNA (ecotype Columbia) using the following oligonucleotides: 5′-CCTGTGAGTAATACAACTATATTGG-3′ and 5′-GCGAAAATTAGATTAGTTTATATGATTATGTAC-3′. The DNA comprised two noncanonical CArG boxes (CTATTTTTGG and CTTTTTTGG) separated by 93 bp. GST-tagged, full-length SEP3 and *SOC1* promoter DNA were mixed at two different protein concentrations (2 to 5 nM and 10 to 15 nM) to 5 nM DNA in a buffer comprising 10 mM HEPES and incubated on ice for 15 min. For $SEP3^{1-110}$/*SOC1* protein-DNA complexes, protein concentrations were 5 and 25 nM to 5 nM DNA prior to dilution for imaging. The complexes were diluted in adsorption buffer (10 mM $NiSO_4$ and 10 mM HEPES, pH 7.0) to obtain a final DNA concentration of ~0.5 to 1 nM and deposited on freshly cleaved mica (Agar Scientific). The complex was adsorbed to the surface for ~10 min. The mica sheet was rinsed two to three times with imaging buffer (10 mM HEPES, pH 7.0) to remove unbound material and scanned under 200 μL of imaging buffer. MFP 3D and Cypher S atomic force microscopes (Asylum Research) were used with MSNL E and F (Bruker) and Biolever mini BL-AC40TS (Olympus) silicon nitride cantilevers, respectively. All images were acquired in tapping mode under liquid to minimize the friction force applied to the sample. The nominal tip radius for the MSNL probes was 2 and 9 nm for the BL-AC40TS probes. The resonance frequencies of the cantilevers in liquid were 8 to 9 kHz for the MSNL-E, 30 to 31 kHz for the MSNL-F, and 40 kHz for the BL-AC40TS. The cantilevers were excited with a conventional dither piezoelectric excitation imposed at the cantilever base. The free oscillation amplitude of the tip was set to 20 nm, and the images were acquired with an amplitude set point of 12 nm. Images were obtained at 256 × 256 pixels and 512 × 512 pixels with a scan size between 0.2 to 2 μm. The scan speed was set to one to two lines/s for the MSNL and three to four lines/s for the BL-AC40TS cantilevers. All images were processed using Gwyddion.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL libraries under the following accession numbers: SEP3 (accession O22456), SEP2 (accession P29384), SEP1 (accession P29382), AP1 (accession P35631), AG (accession P17839), SOC1 (accession O64645), SHORT VEGETATIVE PHASE (accession Q9FVC1), FLC (accession Q5Q9J1), PI (accession P48007), and AP3 (accession P35632). Data necessary to validate protein structure determinations and modeling can be found in the Protein Data Bank under the following accession numbers: 4OX0 for $SEP3^{75-178}$, 3KOV for MEF2A, and 1SRS for SRF.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. SEP3-DNA Complexes.

Supplemental Table 1. Oligonucleotides Used for PCR Mutagenesis.

ACKNOWLEDGMENTS

We thank Philippe Mas for extensive technical assistance in construct library design and screening using the ESPRIT platform and the AFM

platform of the Surface Science Laboratory at the ESRF for AFM measurements. We thank the ESRF beamline staff of 23-1 and 14-4 for their support during the experiments. The ESPRIT platform is supported by EU FP7 Contracts P-CUBE [227764] and BioStruct-X [283570]. This work used the platforms of the Grenoble Instruct Centre (Integrated Structural Biology Grenoble; Unité Mixte de Service 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). This work was supported by ATIP-Avenir (to C.Z.).

AUTHOR CONTRIBUTIONS

F.P., R.D., and C.Z. designed the research. S.P., S.A., L.C., A.V., R.M., E.B., R.M., V.C., S.C., and C.S.S. performed research. D.H. contributed new molecular biology tools and constructs. S.P., L.C., S.C., M.N., G.T., F.P., R.D., and C.Z. analyzed data. C.Z. wrote the article.

Received May 21, 2014; revised August 20, 2014; accepted August 29, 2014; published September 16, 2014.

REFERENCES

Acajjaoui, S., and Zubieta, C. (2013). Crystallization studies of the keratin-like domain from Arabidopsis thaliana SEPALLATA 3. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **69:** 997–1000.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martínez-Castilla, L., and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97:** 5328–5333.

Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. **29:** 464–489.

Becker, A., Winter, K.U., Meyer, B., Saedler, H., and Theissen, G. (2000). MADS-Box gene diversity in seed plants 300 million years ago. Mol. Biol. Evol. **17:** 1425–1434.

Bricogne, G.B.E**., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P., Sharff A., Smart O.S., Vonrhein C., and Womack T.O.** (2011). BUSTER version 2.10.0. (Cambridge, UK: Global Phasing Ltd.).

De Bodt, S., Raes, J., Van de Peer, Y., and Theissen, G. (2003a). And then there were many: MADS goes genomic. Trends Plant Sci. **8:** 475–483.

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouzé, P., Theissen, G., and Van de Peer, Y. (2003b). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J. Mol. Evol. **56:** 573–586.

Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol. **276:** 523–530.

Egea-Cortines, M., Saedler, H., and Sommer, H. (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. EMBO J. **18:** 5370–5379.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. **66:** 486–501.

Gramzow, L., Ritz, M.S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. Trends Genet. **26:** 149–153.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W., Ortin, J., Hart, D.J., and Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. **15:** 500–506.

Hart, D.J., and Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*. Acta Crystallogr. D Biol. Crystallogr. **62:** 19–26.

Hassler, M., and Richmond, T.J. (2001). The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. EMBO J. **20:** 3018–3028.

He, J., Ye, J., Cai, Y., Riquelme, C., Liu, J.O., Liu, X., Han, A., and Chen, L. (2011). Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. Nucleic Acids Res. **39:** 4464–4474.

Hurst, H.C. (1995). Transcription factors 1: bZIP proteins. Protein Profile **2:** 101–168.

Immink, R.G., Tonaco, I.A., de Folter, S., Shchennikova, A., van Dijk, A.D., Busscher-Lange, J., Borst, J.W., and Angenent, G.C. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. Genome Biol. **10:** R24.

Incardona, M.F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J. Synchrotron Radiat. **16:** 872–879.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292:** 195–202.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. **66:** 125–132.

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347:** 183–198.

Kaufmann, K., Muiño, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P., and Angenent, G.C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biol. **7:** e1000090.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. **55:** 379–400.

Masiero, S., Imbriano, C., Ravasio, F., Favaro, R., Pelucchi, N., Gorla, M.S., Mantovani, R., Colombo, L., and Kater, M.M. (2002). Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. J. Biol. Chem. **277:** 26429–26435.

Mason, J.M., and Arndt, K.M. (2004). Coiled coil domains: stability, specificity, and biological implications. ChemBioChem **5:** 170–176.

Melzer, R., and Theissen, G. (2009). Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Res. **37:** 2723–2736.

Melzer, R., Verelst, W., and Theissen, G. (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. Nucleic Acids Res. **37:** 144–157.

Melzer, R., Wang, Y.Q., and Theissen, G. (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. Semin. Cell Dev. Biol. **21:** 118–128.

Mendes, M.A., Guerra, R.F., Berns, M.C., Manzo, C., Masiero, S., Finzi, L., Kater, M.M., and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. Plant Cell **25:** 2560–2572.

Mo, Y., Ho, W., Johnston, K., and Marmorstein, R. (2001). Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. J. Mol. Biol. **314:** 495–506.

Muiño, J.M., Smaczniak, C., Angenent, G.C., Kaufmann, K., and van Dijk, A.D. (2014). Structural determinants of DNA recognition by plant MADS-domain transcription factors. Nucleic Acids Res. **42:** 2138–2146.

Münster, T., Pahnke, J., Di Rosa, A., Kim, J.T., Martin, W., Saedler, H., and Theissen, G. (1997). Floral homeotic genes were recruited from

homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc. Natl. Acad. Sci. USA **94:** 2415–2420.

Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. Nature **376:** 490–498.

Perry, R.L., Yang, C., Soora, N., Salma, J., Marback, M., Naghibi, L., Ilyas, H., Chan, J., Gordon, J.W., and McDermott, J.C. (2009). Direct interaction between myocyte enhancer factor 2 (MEF2) and protein phosphatase 1alpha represses MEF2-dependent gene expression. Mol. Cell. Biol. **29:** 3355–3366.

Pollock, R., and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res. **18:** 6197–6204.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. Genes Dev. **5** (12A)**:** 2327–2341.

Rosloski, S.M., Singh, A., Jali, S.S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of Arabidopsis thaliana. Plant Mol. Biol. **81:** 57–69.

Rounsley, S.D., Ditta, G.S., and Yanofsky, M.F. (1995). Diverse roles for MADS box genes in Arabidopsis development. Plant Cell **7:** 1259–1269.

Santelli, E., and Richmond, T.J. (2000). Crystal structure of MEF2A core bound to DNA at 1.5 A resolution. J. Mol. Biol. **297:** 437–449.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science **250:** 931–936.

Severing, E.I., van Dijk, A.D., Morabito, G., Busscher-Lange, J., Immink, R.G., and van Ham, R.C. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. PLoS ONE **7:** e30524.

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. **66:** 479–485.

Smaczniak, C., et al. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc. Natl. Acad. Sci. USA **109:** 1560–1565.

Soltis, D.E., Soltis, P.S., Albert, V.A., Oppenheimer, D.G., dePamphilis, C.W., Ma, H., Frohlich, M.W., and Theissen, G., Floral Genome Project Research Group (2002). Missing links: the genetic architecture of flowers [correction of flower] and floral diversification. Trends Plant Sci. **7:** 22–31, 31–34.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA **103:** 8060–8065.

Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: gene expression in unions? Nat. Rev. Genet. **10:** 457–466.

Tan, S., and Richmond, T.J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature **391:** 660–666.

Tarendeau, F., et al. (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat. Struct. Mol. Biol. **14:** 229–233.

Theissen, G., and Saedler, H. (2001). Plant biology. Floral quartets. Nature **409:** 469–471.

Theissen, G., Kim, J.T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J. Mol. Evol. **43:** 484–516.

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winter, K.U., and Saedler, H. (2000). A short history of MADS-box genes in plants. Plant Mol. Biol. **42:** 115–149.

van Dijk, A.D., Morabito, G., Fiers, M., van Ham, R.C., Angenent, G.C., and Immink, R.G. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. PLOS Comput. Biol. **6:** e1001017.

Vinson, C., Acharya, A., and Taparowsky, E.J. (2006). Deciphering B-ZIP transcription factor interactions in vitro and in vivo. Biochim. Biophys. Acta **1759:** 4–12.

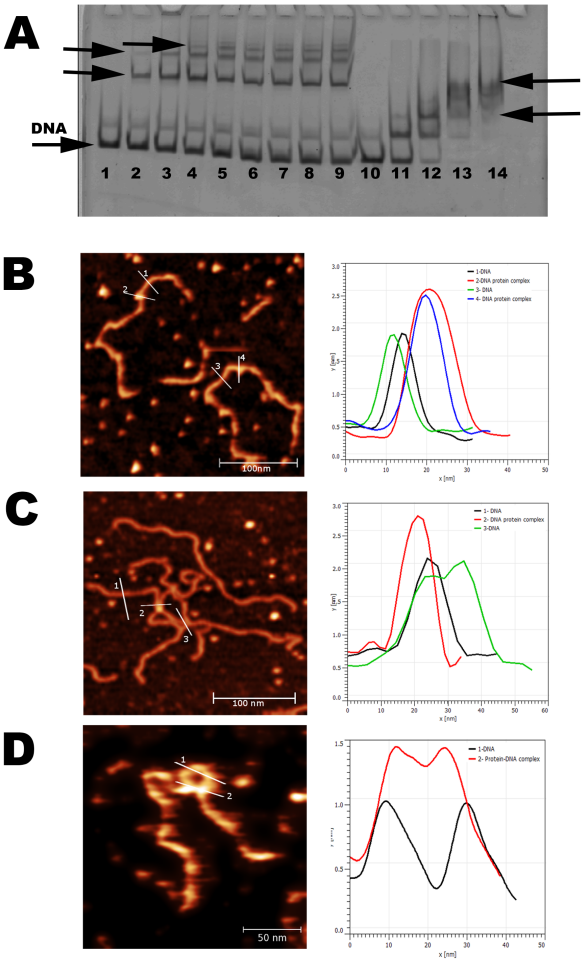
Winn, M.D., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. **67:** 235–242.

Wu, Y., Dey, R., Han, A., Jayathilaka, N., Philips, M., Ye, J., and Chen, L. (2010). Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for myocardin recruitment. J. Mol. Biol. **397:** 520–533.

Yang, Y., and Jack, T. (2004). Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. Plant Mol. Biol. **55:** 45–59.

Yang, Y., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. Plant J. **33:** 47–59.

Yumerefendi, H., Tarendeau, F., Mas, P.J., and Hart, D.J. (2010). ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. J. Struct. Biol. **172:** 66–74.



Supplemental Figure 1. SEP3–DNA complexes.
A. Full length SEP3 and SEP3$_{1-110}$ were incubated with a 150 bp fragment of the *SOC1* promoter comprising two CArG boxes and labelled with TAMRA. DNA concentration was 100 nM and protein concentration varied from 500 nM, 1, 2, 3, 4, 5, 8 and 10 µM for SEP3 full length (lanes 2-9) and from 500 nM, 1,2 and 4 $_{110}$ µM for SEP3$_{1-}$ (lanes 11-14). Lanes 1 and 10 correspond to DNA alone. Arrows indicate protein–DNA complexes. The lower band corresponds to one CArG box bound and the upper bands to 2 CArG boxes bound. SEP3 full length shows higher order complex formation (uppermost band). B. SEP3$_{1-110}$ in complex with 1 kb DNA from the *SOC1* promoter measured by AFM as in Figure 7. The heights of the DNA and DNA–protein complex are shown graphically at right. C. SEP3 full length in complex with 1 kb DNA from the *SOC1* promoter measured by AFM as in Figure 7. The heights of crossed DNA strands and the DNA-protein complex are shown graphically at right. D. SEP3 full length showing DNA looping measured by AFM as in Figure 7. The heights of the DNA and DNA–protein complex are shown graphically at right. Heights were measured and graphs generated using Gwyddion software.

Supplemental Data. Puranik et al. Plant Cell (2014) 10.1105/tpc.114.127910

Supplemental Table 1. Oligonucleotides used for mutagenesis

M150A	5’-CTCTCAGGACACAGTTTGCGCTTGACCAGCTCAAC-3’ 5’-GTTGAGCTGGTCAAGCGCAAACTGTGTCCTGAGAG-3’
L154A	5' –CAGTTTATGCTTGACCAGGCGAACGATCTTCAGAGTAAGG- 3' 5' –CCTTACTCTGAAGATCGTTCGCCTGGTCAAGCATAAACTG- 3'
L171A	5' -CTGACTGAGACAAATAAAACTGCAAGACTAAGGTTAGCTGATGG- 3' 5' -CCATCAGCTAACCTTAGTCTTGCAGTTTTATTTGTCTCAGTCAG- 3'
$SEP3^{\Delta V90}$	5’-GAGAGGCCTTAGCAGAACTTAGTAGCCAGC-3’ (forward) 5’-GCTGGCTACTAAGTTCTGCTAAGGCCTCTC-3’ (reverse)
$SEP3^{\Delta 161-174}$	5’- CCAGCTCAACGATCTTCAGAGTAAGCTAGCTGATGGATGAGAGACAAAT AAAACTCTAAGACTAAGG-3’ (forward) 5’- CCTTAGTCTTAGAGTTTTATTTGTCTCTCATCCATCAGCTAGCTTACTCTG AAGATCGTTGAGCTGG-3’ (reverse)

frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 January 2016
doi: 10.3389/fpls.2015.01193

Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower

Catarina S. Silva[1,2,3,4†], Sriharsha Puranik[5†], Adam Round[6,7,8], Martha Brennich[5], Agnès Jourdain[1,2,3,4], François Parcy[1,2,3,4], Veronique Hugouvieux[1,2,3,4] and Chloe Zubieta[1,2,3,4]*

[1] CNRS, Laboratoire de Physiologie Cellulaire & Végétale, UMR 5168, Grenoble, France, [2] Laboratoire de Physiologie Cellulaire & Végétale, University of Grenoble Alpes, Grenoble, France, [3] Commissariat à l'Energie Atomique et aux Energies Alternatives, Direction des Sciences du Vivant, Laboratoire de Physiologie Cellulaire & Végétale, Institut de Recherches en Technologies et Sciences pour le Vivant, Grenoble, France, [4] Laboratoire de Physiologie Cellulaire & Végétale, Institut National de la Recherche Agronomique, Grenoble, France, [5] European Synchrotron Radiation Facility, Structural Biology Group, Grenoble, France, [6] European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France, [7] Unit for Virus Host-Cell Interactions, University of Grenoble Alpes-EMBL-CNRS, Grenoble, France, [8] Faculty of Natural Sciences, Keele University, Keele, UK

Understanding the evolutionary leap from non-flowering (gymnosperms) to flowering (angiosperms) plants and the origin and vast diversification of the floral form has been one of the focuses of plant evolutionary developmental biology. The evolving diversity and increasing complexity of organisms is often due to relatively small changes in genes that direct development. These "developmental control genes" and the transcription factors (TFs) they encode, are at the origin of most morphological changes. TFs such as LEAFY (LFY) and the MADS-domain TFs act as central regulators in key developmental processes of plant reproduction including the floral transition in angiosperms and the specification of the male and female organs in both gymnosperms and angiosperms. In addition to advances in genome wide profiling and forward and reverse genetic screening, structural techniques are becoming important tools in unraveling TF function by providing atomic and molecular level information that was lacking in purely genetic approaches. Here, we summarize previous structural work and present additional biophysical and biochemical studies of the key master regulators of plant reproduction – LEAFY and the MADS-domain TFs SEPALLATA3 and AGAMOUS. We discuss the impact of structural biology on our understanding of the complex evolutionary process leading to the development of the bisexual flower.

Keywords: evolution, SEPALLATA3, AGAMOUS, LEAFY, protein crystallography, small angle X-ray scattering, homology modeling

OPEN ACCESS

Edited by:
Rainer Melzer, University College Dublin, Ireland

Reviewed by:
Marcelo Carnier Dornelas, Universidade Estadual de Campinas, Brazil
William Oki Wong, Institute of Botany – Chinese Academy of Sciences, China
Florian Ruempler, Friedrich Schiller University Jena, Germany

***Correspondence:**
Chloe Zubieta
chloe.zubieta@cea.fr

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Plant Evolution and Development, a section of the journal Frontiers in Plant Science

Received: *28 August 2015*
Accepted: *11 December 2015*
Published: *06 January 2016*

Citation:
Silva CS, Puranik S, Round A, Brennich M, Jourdain A, Parcy F, Hugouvieux V and Zubieta C (2016) Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower. Front. Plant Sci. 6:1193. doi: 10.3389/fpls.2015.01193

INTRODUCTION

The evolution of streptophytes (green plants), chronicled by the fossil record, follows a trajectory from simple green algae, to the earliest land plants (mosses, hornworts, liverworts), to free-sporing vascular plants (lycopsids including extant clubmosses, quillworts and spike mosses and monilophytes such as ferns and horsetails) and finally culminating with more complex seed plants (**Figure 1**). As the climate changed and became less favorable to spore-forming lycophtyes and monilophytes, spermatophytes (seed plants) were able to supplant these spore-forming vascular plants to become the majority of land plant species. The radiation of seed plants was due in large part to their ability to reproduce without the necessity of water for the dispersal of pollen or successful fertilization, as in the case of mosses and ferns. The reproductive adaptations in seed plants acted as a driver for terrestrial colonization and played a key role in their radiation across a wide range of habitats.

Extant seed plants are further divided into two sister groups, the gymnosperms and the angiosperms. Gymnosperms have naked seeds unprotected by a carpel and generally develop as the result of a single fertilization event. Exceptions exist as is the case of the genus *Ephedra* and *Gnetum* (Friedman, 1990; Friedman and Carmichael, 1996). In contrast, angiosperm seeds are enclosed and protected by the carpel and result from a double fertilization event that ensures the simultaneous development of the zygote and nutritive tissues, the endosperm (Lord and Russell, 2002). In addition to these variations in fertilization and seed development, the most striking difference between gymnosperms and angiosperms is the evolutionary innovation of the angiosperm flower. This novel arrangement joins the male and female organs into one reproductively competent structure. While the evolution of green plants from algae to seed plants follows a relatively smooth path in the fossil record, the evolution of the flower in angiosperms represents an evolutionary leap lacking an extensive step-wise fossil record. Since the time of Charles Darwin, the "abominable mystery" of flower origins and the unprecedented explosive radiation of angiosperm species have been the subject of extensive study and speculation (Burkhardt et al., 1985; Friedman, 2009).

In contrast to gymnosperm cones, which are unisexual and lack an enveloping perianth (sterile outer organs), angiosperm flowers have both male and female reproductive organs on a single axis surrounded by sepals and petals. A typical angiosperm flower is composed of four organs arranged in four concentric whorls. The outermost whorl contains the green protective sepals, followed by a whorl of petals involved in flower opening and pollinator attraction, the next whorl contains the stamens that produce pollen and constitute the male gametophyte, and finally the inner most whorl comprising the pistil, composed of one or more carpels, that contain the ovules. This basic floral architecture can vary across angiosperms. For example, basal angiosperms may contain tepals, sterile outer organs that cannot be differentiated into distinct sepals and petals. In addition, the number of flower parts and their arrangement around the central axis of the flower may vary as in orchids where the male and female organs are fused. However, the essential characteristic of the flower, co-localized male and female organs, is retained across all angiosperm species and acts as a defining trait.

Angiosperm and Gymnosperm Evolution

One of the central questions in plant evolutionary developmental biology is how the flower, a bisexual compacted reproductive structure, evolved and what were the underlying molecular mechanisms for this dramatic morphological change. Extant gymnosperms and angiosperms separated ∼300 Mya (Zhang et al., 2004), with angiosperms quickly achieving an unprecedented level of species dominance, with over 350,000 extant species, in a dramatically short evolutionary timescale. However, simple morphological comparisons between gymnosperm cones and angiosperm flowers offer limited insight into flower evolution (Bateman et al., 2006; Frohlich and Chase, 2007). An understanding of the abrupt appearance of the flower from gymnosperm cones requires not only a fossil record to probe the changing morphologies of plant reproductive structures, but also a molecular basis derived from genome sequencing, molecular biology and structural biology. Impressive progress has been made in understanding the gene networks that regulate plant reproduction in angiosperms and, albeit to a lesser extent, also in gymnosperms. Due to extensive forward and reverse genetic studies (Coen and Meyerowitz, 1991; Saedler et al., 2001; Theissen and Saedler, 2001; Krizek and Fletcher, 2005) and whole genome sequencing in model plants such as thale cress (*Arabidopsis thaliana*), snapdragon (*Antirrhinum majus*) and petunia (*Petunia x hybrida*), as well as the large scale gene sequencing initiatives such as the 1000 plant genomes project (https://sites.google.com/a/ualberta.ca/onekp/home) and the complete sequencing and annotation of the first gymnosperm genome from Norway spruce (Nystedt et al., 2013), many of the genes which regulate the transition from vegetative to reproductive growth in angiosperms and gymnosperms have been identified.

Gene Regulatory Networks Controlling Plant Reproductive Development

Despite the morphological difference between angiosperm and gymnosperm reproductive structures, a comparison of the genes responsible for male and female organ development demonstrates a high degree of conservation. Based on studies in angiosperm model plants such as *Arabidopsis*, development is switched from a vegetative to a reproductive program based on exogenous environmental and endogenous developmental signals such as plant age. This switch is orchestrated by the high level regulator of reproductive development, *LEAFY* (*LFY*), a gene that is conserved in gymnosperms and angiosperms (Vazquez-Lobo et al., 2007; Moyroud et al., 2010) and which has recently been identified in green algae, suggesting ancestral functions predating land plants (Sayou et al., 2014). Interestingly, while existing primarily as a single copy gene in most angiosperms, gymnosperms have two paralogous *LFY*-like genes-*LFY* and *NEEDLY* (*NLY*; Frohlich and Meyerowitz, 1997; Vazquez-Lobo et al., 2007), the only known exception being the gymnosperm genus *Gnetum* where *NLY* is absent (Frohlich and

Meyerowitz, 1997; Frohlich and Parker, 2000). In addition to conservation of *LFY*, the genes that determine the identity of male and female reproductive organs, the MADS-box genes, are also present in both angiosperms and gymnosperms (Gramzow et al., 2010, 2014; Melzer et al., 2010; Wang et al., 2010). However, in contrast to gene loss in angiosperms as observed for *NLY*, the MADS-box genes have undergone multiple duplication events, leading to a more extensive gene network in angiosperms versus their sister gymnosperms (**Figure 1**). *LFY*, *NLY* and the MADS-box genes all encode transcription factors (TFs). These TFs act as master regulators and are able to direct extensive downstream gene networks. Recent work examining the function of LFY, NLY

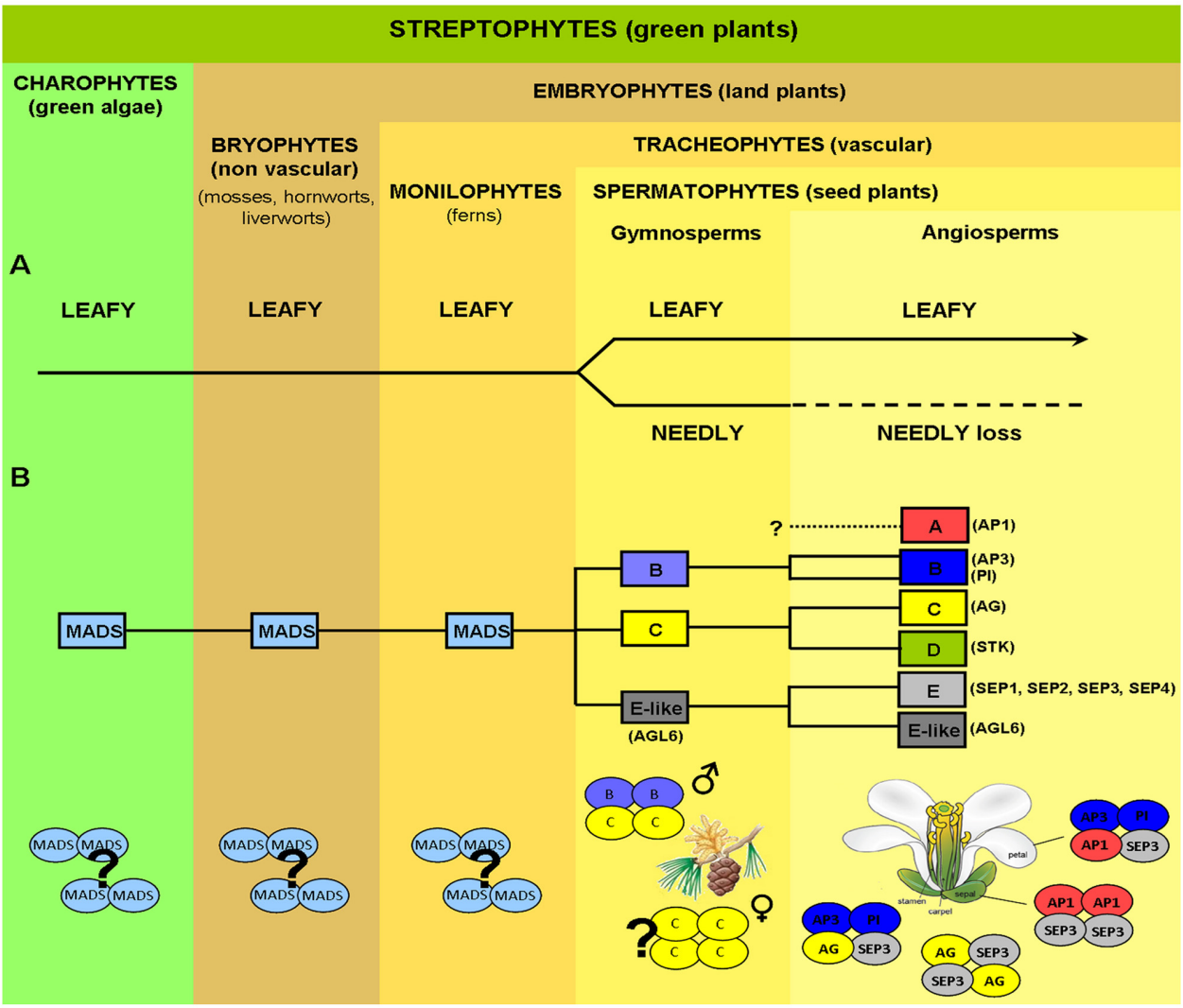


FIGURE 1 | Evolution of key genes controlling plant reproductive development. (A) Evolution of *LEAFY* (*LFY*) from green algae to angiosperms. *LFY* exists mostly as a single-copy gene in all streptophytes (green plants), with the exception of gymnosperms where a *LFY*-like paralog, *NEEDLY* (*NLY*), originated after a major duplication event (the only possible exception being the genus *Gnetum*). In gymnosperms, *LFY* and *NLY* are consistently expressed in both male (pollen-bearing) and female (seed-bearing) cones, in a spatiotemporal coordinated manner. In the angiosperm lineage, *NLY* was subsequently lost, with *LFY* now regulating the expression of genes responsible for both the male and female organs in the unified bisexual flower. **(B)** MADS-box homeotic gene family. MADS-box genes are present in the most simple green algae and, as plants became more complex, the MADS-box gene family expanded via multiple duplication and specification events. Putative orthologs of class B, C, and E-like (*AGL6*) floral homeotic genes have been isolated from different gymnosperms (conifers, gnetophytes, ginkgophytes, and cycads) as shown schematically by yellow and blue colored ovals. In contrast, *SEP*-like genes, the second subfamily conferring E-class function, as well as A-class genes, seem to be absent in extant gymnosperms but are present in all angiosperms. In gymnosperms, expression patterns of putative B and C-class gene orthologs resemble those of B and C-class genes in angiosperms, with B-class genes being expressed on male reproductive organs, whereas C-class genes are expressed in both male and female organs. In gymnosperms C-class proteins alone or C and B-class proteins together seem capable of forming tetrameric complexes (without any additional partners), which define, respectively, the female and male organs in these organisms as indicated. In angiosperms tetramer formation is dependent on the SEPALATTA (E-class) TFs which act as hubs by mediating interactions among proteins from different floral homeotic classes, strictly determining floral organ identity. Question marks indicate uncertainty as to physiological oligomerisation state, AP1, APETALA1; AP3, APETALA3; PI, PISTILLATA; AG, AGAMOUS; STK, SEEDSTICK; SEP, SEPALLATA; AGL6, AGAMOUS LIKE 6.

and the MADS TFs at the protein level has greatly advanced our understanding of how relatively small changes in a few key regulatory TFs can result in large differences at the morphological level of the organism. Current hypotheses point to changes in a few key genes, and the TFs they encode, as determining factors in the evolution of plant reproduction and the formation of the flower (Theissen, 2000, 2005; Zahn et al., 2005a,b; Theissen and Melzer, 2007; Melzer et al., 2010).

The Role of LFY and LFY/NLY in Angiosperms and Gymnosperms

In angiosperms such as *Arabidopsis* or *Antirrhinum*, the switch to reproductive growth involves the conversion of the shoot apical meristem (SAM) to an inflorescence meristem (IM). The IM will in turn generate the floral meristem (FM) on its flanks. The development of a FM can be divided into two main steps (1) the formation of a specific zone within the IM, called the anlage, from which the FM will arise and (2) the growth of the FM primordia and subsequent differentiation into the floral organs. It is a balance between inflorescence identity genes such as *TERMINAL FLOWER 1* (*TFL1*) and FM identity genes such as *LFY* that determines the acquisition of flower identity. *TFL1* is predominantly expressed in the IM and acts as a repressor, preventing *LFY* and the MADS-box gene, *APETALA1* (*AP1*), expression (Liljegren et al., 1999). Increasing levels of LFY act as a committing step in FM identity, with LFY repressing expression of *TFL1* and inducing the expression of FM identity genes such as *AP1* (Parcy et al., 1998; Wagner et al., 1999; Kaufmann et al., 2010; Moyroud et al., 2010; Winter et al., 2011).

In gymnosperms, *LFY* and *NLY* expression patterns overlap in male and female cones early in development with expression patterns diverging later into mutually exclusive but complementary domains, resulting in higher *LFY* expression levels in male cones and higher *NLY* expression in female cones (Shindo et al., 2001; Dornelas and Rodriguez, 2005; Vazquez-Lobo et al., 2007). Originally, the *NLY* gene was thought to exclusively specify gymnosperm female reproductive structures (seed-bearing cone) in *Pinus radiata* (Mouradov et al., 1998), whereas its paralogous gene *LFY* appeared restricted to the male pollen-carrying cones (Mellerowicz et al., 1998). However, subsequent findings of *LFY* orthologs being expressed in female cones of gnetophytes and congeneric conifers (Carlsbecker et al., 2004; Dornelas and Rodriguez, 2005), demonstrated concurrent expression of both genes in male and female reproductive structures. Thus, *LFY* and *NLY* from gymnosperms are both necessary to act as regulators of male and female cone development, likely fulfilling a similar critical role in plant reproduction as the single copy angiosperm *LFY*.

The Roles of the MADS-Box Genes and MADS TFs in Organ Identity

Once the FM is specified, LFY activates additional floral organ identity genes including the MADS-box genes *AP3*, *AG,* and *SEP3* (Weigel and Meyerowitz, 1993; Busch et al., 1999; Wagner et al., 1999; Lamb et al., 2002; Lohmann and Weigel, 2002; Winter et al., 2011). To date there is no direct evidence that gymnosperm LFY or NLY directly regulate MADS-box genes in gymnosperms as LFY does in angiosperms, although this is possible and warrants study. Once expressed, the overlapping patterns of the MADS-box genes will specify floral organ identity as outlined in the ABC(D)E model (Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991) and for review see (Sablowski, 2010). In essence, the MADS-box genes can be divided into classes A−E with A+E genes necessary for sepal development, A+B+E genes specifying petals, B+C+E genes specifying stamen, C+E genes specifying carpels and D+E specifying ovules (Theissen and Saedler, 1995; Theissen, 2000; Honma and Goto, 2001; Ng and Yanofsky, 2001; Theissen and Saedler, 2001; Favaro et al., 2003). In *Arabidopsis* the class A genes are *APETALA1* (*AP1*) and *APETALA2* (*AP2*), class B genes are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), class C is *AGAMOUS* (*AG*) and class E are *SEPALLATA1,2,3,4 (SEP1,2,3,4)*. Except for *AP2,* all the floral homeotic genes in the ABC(D)E model encode MADS-domain TFs. The molecular mechanism of action of these proteins is explained by the floral quartet model, in which the A-E class genes encode TFs which are able to homo and heterotetramerise in specific combinations, resulting in the activation or repression of distinct downstream target genes and thus specifying floral organ identity (Honma and Goto, 2001; Theissen, 2001).

Gymnosperms possess B- and C-like MADS-box genes with their expression patterns resembling B- and C- class genes in angiosperms (Tandre et al., 1995; Sundstrom et al., 1999; Becker et al., 2002, 2003; Jager et al., 2003; Melzer et al., 2010; Wang et al., 2010; Gramzow et al., 2014). Indeed, several studies have described the expression of C-like genes in both male and female cones, while B-like gene expression appeared to be restricted to male cones (Sundstrom and Engstrom, 2002; Wang et al., 2010). Complementation studies have demonstrated that B and C homologs are well-conserved between gymnosperms and angiosperms as B and C genes from gymnosperms can nearly fully restore a wild type flower phenotype (Winter et al., 2002; Zhang et al., 2004). In addition, the gymnosperm MADS-domain TFs from the B and C class appear competent to form homo and heterotetramers, similarly to their angiosperm orthologs (**Figure 1**; Wang et al., 2010). Interestingly, the *SEP* subfamily members are absent in gymnosperms but are present in all major lineages of extant angiosperms (Zahn et al., 2005a). Based on phylogenetic analysis, the closest relative of the *SEP* subfamily is the *AGL6* subfamily, which is found in both angiosperms and gymnosperms (Becker and Theissen, 2003; De Bodt et al., 2003; Martinez-Castilla and Alvarez-Buylla, 2003; Nam et al., 2003; Zahn et al., 2005a). Similarly to class E *SEP* genes in angiosperms, *AGL6*-like genes are predominantly expressed in reproductive tissues in gymnosperms (for review see, Melzer et al., 2010) and represent the closest homologs to the *SEP* subfamily. Changes in the regulation of B and C class genes during evolution coupled with the appearance of the *SEP*-like genes and the dependence on the SEP TFs to form tetrameric MADS protein complexes, have been proposed to be crucial for the appearance of the bisexual flower. By requiring the SEP TFs to form transcriptionally active complexes

with other homeotic MADS TFs, male and female organ identity may have become more easily co-regulated due to the multiple roles of the SEPs in specifying all reproductive organs.

The gene regulatory networks directing plant reproduction in gymnosperms and angiosperms are becoming more well-defined and the changes in key genes in gymnosperms and angiosperms which may be at the nexus of flower origins have been identified based on genetics studies in angiosperms and large scale sequencing initiatives in most plant lineages. However, only recently has the structure-function relationship of the proteins encoded by these key genes been determined. Here, we summarize available structural studies and provide new data to show how changes at the protein level in the key regulators LFY, NLY, and MADS-domain TFs potentially result in new functionality. Using biophysical data as a foundation, we probe the molecular mechanisms underlying the emergence and evolution of the novel reproductive architecture of the angiosperm flower and discuss how biochemistry and structural biology can provide new insights into evolutionary developmental biology.

MATERIALS AND METHODS

Sequence Alignments

Sequence alignments were performed using the server NPS@ (Network Protein Sequence Analysis; Combet et al., 2000). Sequences were aligned with ClustalW (Thompson et al., 1994) using the default parameters for both pairwise alignment and multiple alignment sections. Where appropriate, secondary structure predictions were carried out with PREDATOR (DSSP) using the NPS@ server. Protein sequences used were obtained from GenBank and the 1000 Plants (1KP) initiative (http://www.onekp.com). Resulting alignments and secondary structure predictions were rendered with ESPript (Robert and Gouet, 2014).

For the LFY/NLY sequence alignments (**Figures 2** and **3**) the sequences used are as follows: AtLFY (*A. thaliana* LFY, AED97525.1), OsLFY (*Oryza sativa japonica* LFY, RFL, AHX83808.1), AmtLFY (*Amborella trichopoda* LFY, AmboLFY, AGV98899.1), PrLFY (*Pinus radiata* LFY, PRFLL, AAB51587.1), GbLFY (*Ginkgo biloba* LFY, ADD64700.1), WmLFY (*Welwitschia mirabilis* LFY, AAF23870.1), PrNLY (*P. radiata* NLY, AAB68601.1), PaNLY (*Pinus armandii* NLY, ADO33969.1), GbNLY (*G. biloba* NLY, AAF77074.1), and WmNLY (*W. mirabilis* NLY, AAD38872.1). For MADS-domain TFs sequence alignments (**Figures 4** and **5**) the sequences used are: *A. thaliana* SEP3 (AEE30503.1), SEP1 (AED92208.1), SEP2 (AEE73791.1), AP3 (AEE79216.1), PI (AED92817.1), AP1 (AEE34887.1), AG (AEE84111.1), AGL6 (AEC10582.1), SOC1 (AEC10583.1), SVP (AEC07320.1), and FLC (AED91498.1); *Gnetum gnemon* GGM2 (CAB44448.1), GGM3 (CAB44449.1), GGM15 (CAC13991.1), GGM9 (CAB44455.1), and GGM11 (CAB44457.1); *Picea abies* DAL11 (AAF18373.1), DAL12 (AAF18375.1), DAL13 (AAF18377.1), DAL2 (CAA55867.1),

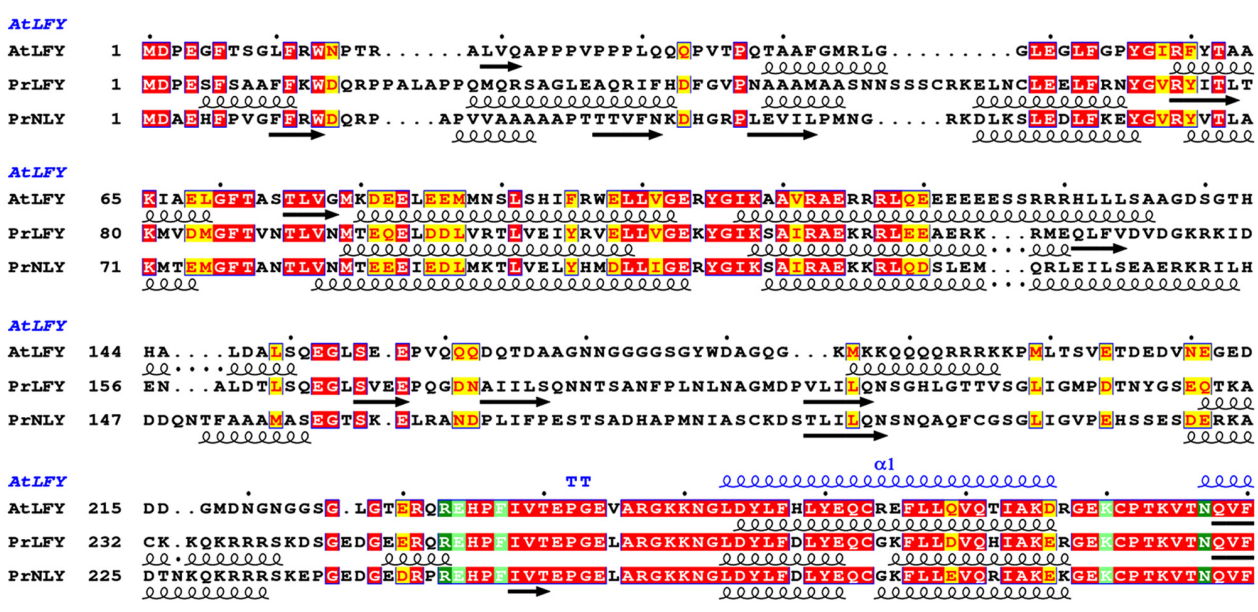


FIGURE 2 | Sequence alignment and secondary structure prediction of AtLFY, PrLFY, and PrNLY N-terminal domains, linker and start of conserved C-terminal DBD. Aligned amino acid sequences of AtLFY (*Arabidopsis thaliana* LFY), PrLFY (*Pinus radiata* LFY, PRFLL) and PrNLY (*Pinus radiata* NLY), with respective secondary structure prediction indicated in black below each respective sequence; α-helices are represented by spirals and β-strands by arrows, all other regions are predicted to be unstructured. Sequence numbering is shown on the left and dots mark every 10th residue for the first sequence. Highly conserved regions are boxed, with similar residues represented in red against a yellow background, invariant residues represented against a red background and non-conserved residues indicated in black. Partial AtLFY DBD secondary structure, as derived from its X-ray structure (PDB 2VY1), is shown in blue above the sequences. Alignment was prepared with NPS@ (Combet et al., 2000).

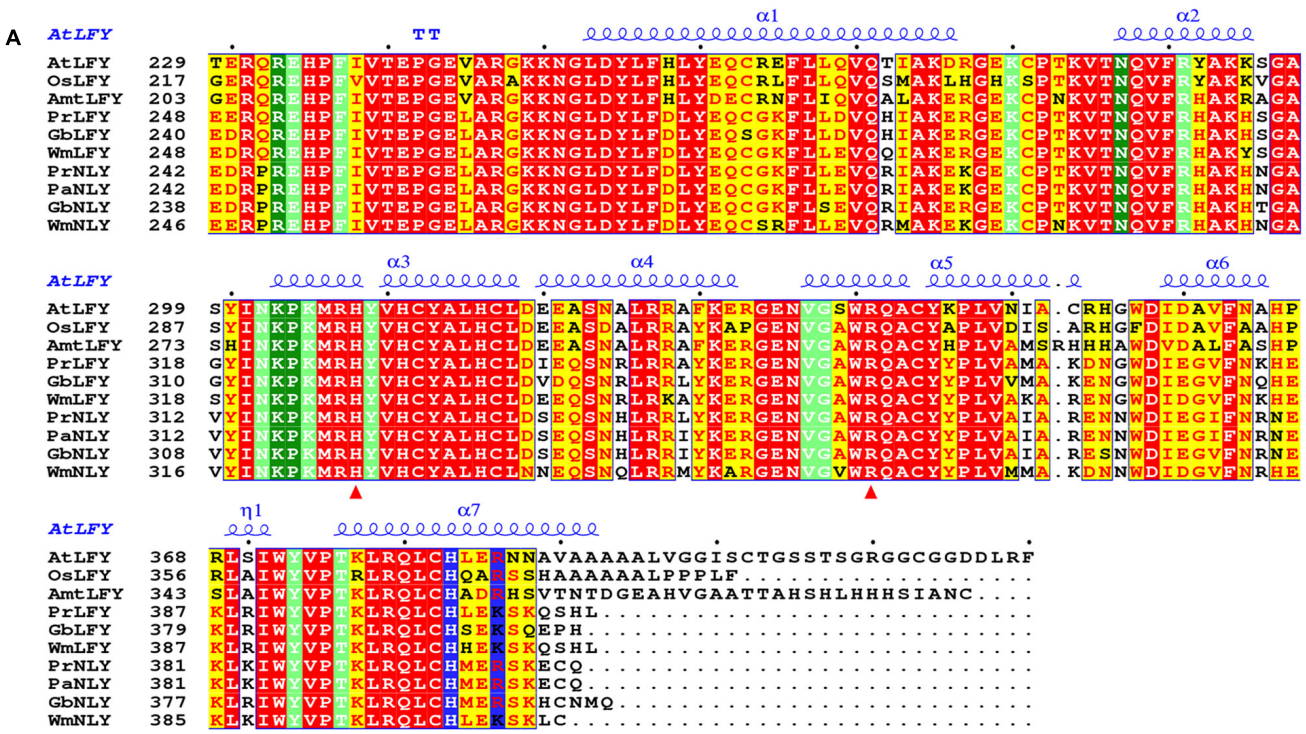


FIGURE 3 | Continued

FIGURE 3 | Continued

Sequence alignment and homology models of the DNA binding domain (DBD) of LFY and NLY. (A) Sequence alignment of LEAFY (LFY) and NEEDLY (NLY) DBDs. Aligned C-terminal DBD amino acid sequences of AtLFY (*A. thaliana* LFY, GenBank AED97525.1), OsLFY (*Oryza sativa japonica* LFY, RFL), AmtLFY (*Amborella trichopoda* LFY, AmboLFY), PrLFY (*Pinus radiata* LFY, PRFLL), GbLFY (*Ginkgo biloba* LFY), WmLFY (*Welwitschia mirabilis* LFY), PrNLY (*Pinus radiata* NLY), PaNLY (*Pinus armandii* NLY), GbNLY (*G. biloba* NLY), and WmNLY (*W. mirabilis* NLY). All sequences are numbered and dots mark every tenth residue above the sequences. Highly conserved regions are boxed, with similar residues represented in red against a yellow background, invariant residues represented against a red background and non-conserved residues indicated in black. The secondary structure annotation of AtLFY DBD, as derived from its three-dimensional X-ray structure (PDB 2VY1), is depicted in blue on top of the aligned sequences [alpha helices (α); strict β-turn (TT); 3_{10}-helix (η)]. Residues involved in interactions with the DNA are highlighted in dark-green (direct contact with DNA bases) and light-green (sugar phosphate backbone contacts); residues involved in dimerisation are depicted in blue. Red triangles indicate residues important for determining DNA half-site specificity. The AtLFY protein sequence (AED97525.1) differs from the AtLFY sequences in Hames et al. (2008) and (Sayou et al., 2014; AAA32826) by a four residue deletion after resdiue 152 resulting in a -4 sequence shift. **(B)** Homology model of *Pinus radiata* NEEDLY (PrNLY) DBD based on AtLFY DBD X-ray structure (PDB 2VY1). Monomers are represented in green and blue as cartoons with a partial transparent surface; bound DNA is represented in orange and gold. PrNLY DBD adopts the same seven α-helix fold, contacting the DNA through both the minor and major grooves with complete conservation of all DNA-binding amino acid residues determined for AtLFY DBD. **(C)** Close-up view of the dimerisation interface of PrNLY. Monomers are colored as per **(B)** and side chain residues involved in putative hydrogen bonding interactions are shown and labeled. **(D)** Close-up view of the dimerisation interface of PrLFY. Colors and residues as per **(C)**.

DAL1 (CAA56864.1), and DAL14 (AGR53802.1). Numbers indicated correspond to GenBank accession numbers.

Homology Modeling

The homology model of the DNA-binding domain (DBD) of *Pinus radiata* NLY (PrNLY) and LFY (PrLFY) proteins were built using the SWISS-MODEL server (Arnold et al., 2006; Biasini et al., 2014, swissmodel.expasy.org). Based on the sequence alignment between PrNLY and AtLFY DBDs the PrNLY partial sequence [E242-Q404] was fed to the server, as well as the AtLFY DBD PDB structure (PDB 2VY1, GenBank accession AAA32826). The homology model of PrNLY DBD comprises residues [R246-K401]. The same procedure was applied to PrLFY for which the partial sequence [Q251-H410] was fed to the server; the PrLFY homology model comprises residues [R252-K407]. Each of the models was superimposed on the AtLFY DBD structure (Hames et al., 2008) using COOT (Emsley et al., 2010); the DNA coordinates added to the composite homology models were taken from the AtLFY structure. The cartoon model representation was made using the program Pymol (The PyMOL Molecular Graphics System, 2010).

SEP3$^{(75-178)}$ Mutagenesis, Expression, and Purification

SEP3$^{(75-178)}$ construct (wild type) was cloned into the expression vector pESPRIT002 (Hart and Tarendeau, 2006; Guilligay et al., 2008) using the AatII and NotI restriction sites. The plasmid contains an N-terminal 6x-His tag followed by a TEV protease cleavage site. All mutants produced were generated using the SEP3$^{(75-178)}$ construct as the template and using Phusion polymerase (NEB) according to the manufacturer's protocol. The oligonucleotides used for mutagenesis are provided in **Table 1**.

SEP3$^{(75-178)}$ and all the tetramerisation mutant constructs were overproduced in *Escherichia coli* BL21 (DE3) CodonPlusRIL (Agilent Technologies; Puranik et al., 2014); all dimerisation mutant constructs were overproduced in *E. coli* Rosetta2 (DE3) pLysS cells (this study). Cells were grown at 37°C in Luria-Bertani (LB) culture medium supplemented with kanamycin (50 mg/mL) and chloramphenicol (37 mg/mL), until an OD_{600} of 0.7–0.8

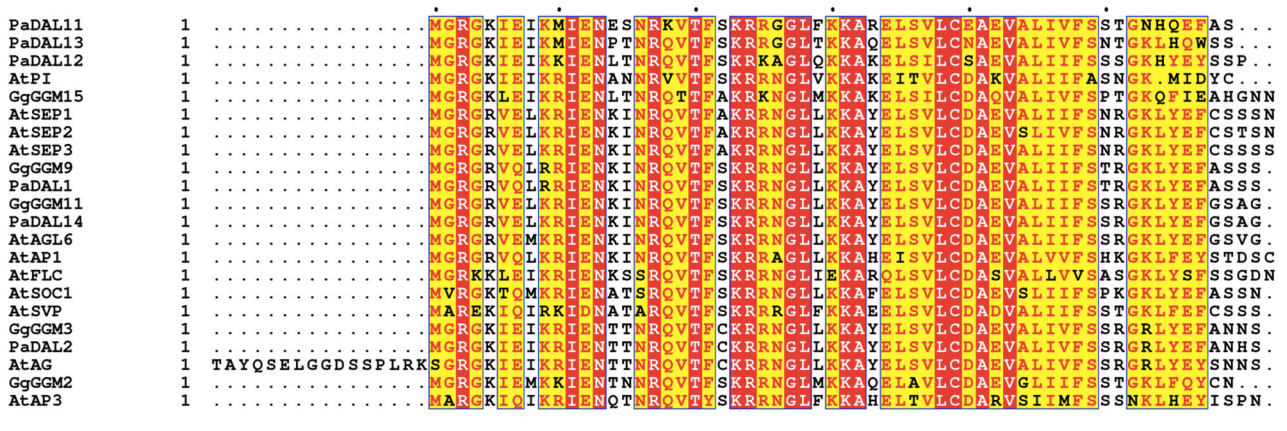


FIGURE 4 | Sequence alignment of MADS TFs M-domain. Aligned M-domain amino acid sequences of *A. thaliana* SEP3, SEP1, SEP2, AP3, PI, AP1, AG, AGL6, SOC1, SVP, and FLC; the gymnosperm *Gnetum gnemon* GGM2, GGM15 (AP3/PI-like), GGM3 (AG-like), and GGM9, GGM11 (AGL6-like); and the gymnosperm *Picea abies* DAL11, DAL12, DAL13 (AP3/PI-like), DAL2 (AG-like), and DAL1, DAL14 (AGL6-like) proteins. Sequence numbering is indicated on the left, with every tenth residue marked by black dots above the sequences. Highly conserved regions are boxed, with similar residues represented in red against a yellow background, invariant residues represented against a red background and non-conserved residues indicated in black.

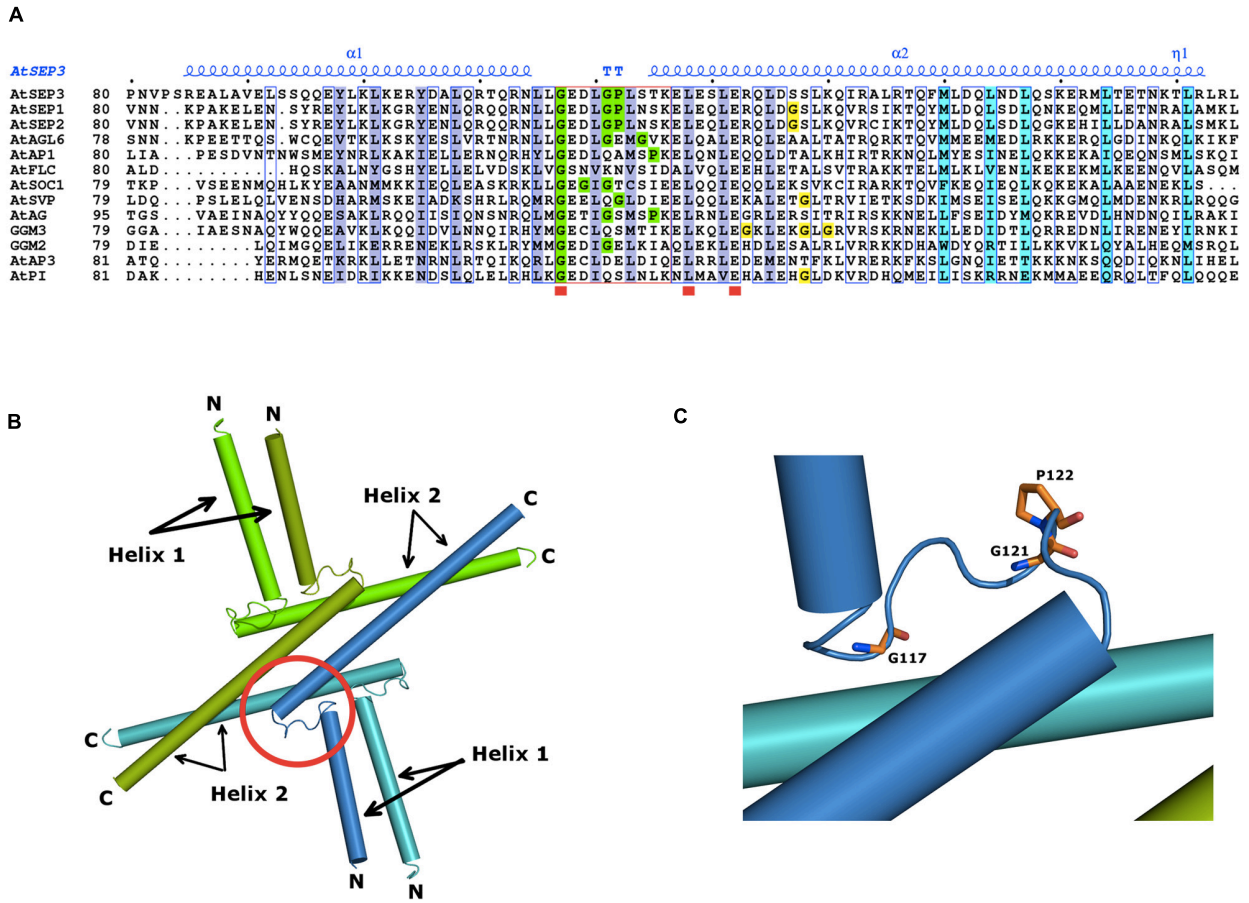


FIGURE 5 | MADS TFs oligomerisation domain. (A) Sequence alignment spanning the SEP3 crystallographic structure (PDB 4OX0). All sequences are from *A. thaliana* SEP3, SEP1, SEP2, AP3, PI, AP1, AG, AGL6, SOC1, SVP, and FLC and the gymnosperm *G. gnemon* GGM2 and GGM3 proteins. Numbering is indicated at the start of the sequences and every tenth residue indicated by a black dot above the SEP3 sequence. Highly conserved regions are boxed in blue with a white background; strictly conserved residues are depicted by a red square below the sequence. The secondary structure elements of AtSEP3 K-domain (PDB 4OX0), are shown in blue above the sequences (α; TT; η). Residues involved in dimerisation and tetramerisation in SEP3 K-domain structure are highlighted in violet and cyan, respectively. The kink region between helices 1 and 2 is framed in red; Gly and Pro residues present within the kink regions are highlighted in green; Gly residues in the N-terminal region of helix 2 are highlighted in yellow. **(B)** Structure of SEP3 K-domain (PDB 4OX0). The oligomerisation domains of SEP3 are represented as cylinders; each monomer, composed of two distinct helices (helices 1 and 2), is colored uniquely in green, dark green, blue, and light blue. N and C-terminal regions are indicated. **(C)** Close-up of the SEP3 kink between helices. Glycine and proline residues are depicted as sticks colored by atom with carbons in orange.

was reached. At this point, protein expression was induced by addition of 1 mM isopropyl-β-D-galactopyranoside (IPTG) and the temperature reduced to 20°C; expression was continued for 16 h (overnight). Cells were harvested by centrifugation at 6000 rpm for 30 min at 4°C and then resuspended in Buffer A [30 mM Tris pH 8.0, 300 mM NaCl, 5% (v/v) glycerol, 2 mM TCEP] to which benzonase (Sigma) and protease inhibitors (Roche EDTA-free) were added. Cells were disrupted by sonication, followed by centrifugation at 25000 rpm for 40 min at 4°C, to remove cell debris. The cell lysate was then passed onto a column containing 1 mL of Ni-Sepharose High-Performance resin (GE-Healthcare), previously equilibrated with Buffer A. Bound protein was washed in two steps: high salt (30 mM Tris pH 8.0, 1 M NaCl, 5% glycerol, 2 mM TCEP) and low imidazole concentration (buffer A + 20 mM Imidazole); and subsequently eluted with Buffer B (30 mM Tris pH 8.0, 300 mM NaCl, 5% glycerol, 250 mM Imidazole, 2 mM TCEP). Fractions of interest were pooled and dialysed overnight at 4°C, against Buffer A and in the presence of 2% (w/w) TEV protease, in order to cleave the 6xHis tag. The protein sample was passed over the same 1 mL Ni-Sepharose column, in order to deplete the His-tagged TEV protease and remove uncut protein from the cleaved protein sample. The purified protein was then concentrated and applied onto a size exclusion Superdex 200 10/300 GL column (GE Healthcare), pre-equilibrated with Buffer A. $SEP3^{(75-178)}$ and all mutants were purified following this same protocol.

EMSA Experiments

AG, SEP3 full length wild type, and SEP3 mutants (L171A, L115R, $SEP3^{\Delta C}$) were cloned into a pSPUTK plasmid and used for *in vitro* transcription translation (Promega SP6 High Yield

Expression System). $SEP3^{\Delta C}$ contains residues 1–160 with a –LADG-stop terminating sequence corresponding to a complete truncation of the C-terminal domain. Protein expression was performed as per the manufacturer's protocol and used without further purification. *SOC1* promoter DNA (121 bp *SOC1* specific DNA) comprising two CArG boxes was used as per (Kaufmann et al., 2009). Two *SOC1* promoter DNA fragments were generated with either the first or second CArG box mutated. Mutations were generated using a 1 kb *SOC1* promoter DNA as the template (inserted into pCR blunt vector) and using Phusion polymerase (NEB) according to the manufacturer's protocol. CArG-box 1 was mutated with the forward primer 5′-CGTGTCTAAAGAGGCATTTGACATATGACGTCCCTCG GATTACTAAAG-3′ and the reverse primer 5′-CTTTAG TAATCCGAGGGACGTCATATGTCAAATGCCTCTTTAGA CACG (CArG-box 1 mutation is underlined); and CArG-box 2 mutated with the forward primer 5′-GTGGCA CCAAAAAAATATACATATGACGAGATAAAATTGTTAATC G-3′ and the reverse primer 5′- CGATTAACAATTTT ATCTCGTCATATGTATATTTTTTTGGTGCCAC-3′ (CArG-box 2 mutation is underlined). Final 145 bp mutated *SOC1* DNA fragments were then PCR amplified using the primers 5′-CTAAAGAGGCATTTGACATATGACGTCCCTCG (fwd) and 5′-GATTAACAATT TTATCTCCAAAAAAGGATATTTTTTTG G (rev) for CArG-box 1, and 5′-CTAAAGAGGCATTTG CTATTTTTGGTCCCTCG (fwd) and 5′-GATTAACAATTTTA TCTCGTCATATGTATATTTTTTTGG (rev) for CArG-box 2 mutated DNAs, respectively. *SOC1* DNA labeled with DY-682 (Dyomics GmbH, wild type) or Cy5 (Eurofins, mutated CArG-boxes) was used at a concentration of approximately 5–10 nM for all reactions in a protein binding buffer containing 7 mM HEPES, pH 7.0, 1 mM BSA, 1 mM EDTA, 1 mM DTT, 2.5% CHAPS, 6% glycerol, 0.06 mg/ml salmon sperm DNA, 1.3 mM spermidine. 4 μl of TnT protein mix was added directly without purification to the binding buffer to a final volume of 20 μl.

TABLE 1 | Oligonucleotides used for $SEP3^{(75-178)}$ mutagenesis.

Mutant	Oligonucleotides
M150A	5′-CTCTCAGGACACAGTTTGCGCTTGACCAGCTCAAC-3′ 5′-GTTGAGCTGGTCAAGCGCAAACTGTGTCCTGAGAG-3′
L154A	5′-CAGTTTATGCTTGACCAGGCGAACGATCTTCAGA GTAAGG-3′ 5′-CCTTACTCTGAAGATCGTTCGCCTGGTCAAGCATAAA CTG-3′
L171A	5′-CTGACTGAGACAAATAAAACTGCAAGACTAAGGTTA GCTGATGG-3′ 5′-CCATCAGCTAACCTTAGTCTTGCAGTTTTATTTGTCTC AGTCAG-3′
$Splice^{\Delta 161-174}$	5′-CCAGCTCAACGATCTTCAGAGTAAGCTAGCTGATGGA TGAGAGACAAATAAAACTCTAAGACTAAGG-3′ (forward) 5′-CCTTAGTCTTAGAGTTTTATTTGTCTCTCATCCATCAG CTAGCTTACTCTGAAGATCGTTGAGCTGG-3′ (reverse)
Y98K	5′-CTTAGTAGCCAGCAGGAGAAGCTCAAGCTTAAGGA GCG-3′ 5′-CGCTCCTTAAGCTTGAGCTTCTCCTGCTGGCTACTA AG-3′
Y105N	5′-CTCAAGCTTAAGGAGCGTAATGACGCCTTACAAAGA ACCC-3′ 5′-GGGTTCTTTGTAAGGCGTCATTACGCTCCTTAAGCT TGAG-3′
L115R	5′-CAAAGAACCCAAAGGAATAGGTTGGGAGAAGATCTT GGACCTC-3′ 5′-GAGGTCCAAGATCTTCTCCCAACCTATTCCTTTGGG TTCTTTG-3′
L131V	5′-CTAAGTACAAAGGAGCTTGAGTCAGTTGAGAGACAG CTTGATTC-3′ 5′-GAATCAAGCTGTCTCTCAACTGACTCAAGCTCCTTT GTACTTAG-3′
L135M	5′-CTTGAGTCACTTGAGAGACAGATGGATTCTTCCTT GAAGC-3′ 5′-GCTTCAAGGAAGAATCCATCTGTCTCTCAAGT GACTCAAG-3′
L135A	5′-GTCACTTGAGAGACAGGCTGATTCTTCCTTGAAGC-3′ 5′-GCTTCAAGGAAGAATCAGCCTGTCTCTCAAGTGAC-3′

AG Expression and Purification

$AG^{(74-173)}$ was cloned into a pESPRIT002 vector using NotI and AatII restriction sites. The construct contained an N-terminal TEV protease cleavable poly-histidine tag (Hart and Tarendeau, 2006; Guilligay et al., 2008). The protein was overexpressed in *E. coli* BL21 Star (DE3)pLysS cells (Life Technologies). Cells were grown in Luria Bertani medium in the presence of 50 mg/ml kanamycin and 35 mg/ml chloramphenicol at 37°C and 180 rpm to an optical density $A_{600} = 0.8$ after which time the temperature was lowered to 20°C and 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for induction. After 16 h, the cells were harvested by centrifugation at 6000 rpm and 4°C for 15 min and resuspended in lysis buffer containing 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5%(v/v) glycerol, 20%(w/v) sucrose and 1x protease inhibitors (Roche EDTA-free). Cells were lysed by sonication and the insoluble fraction pelleted by centrifugation at 25000 rpm and 4°C for 30 min. The pellet was resuspended in denaturation buffer [30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol, 8 M Urea] and incubated for 1 h at room temperature. The solubilized fraction was applied to a 5 ml Ni-NTA column pre-equilibrated with denaturation buffer, followed by a wash with 10 CV of wash buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 30 mM imidazole) and eluted with 3 CV of elution buffer (30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol, 8 M Urea, 300 mM Imidazole). The eluted fraction was dialysed step-wise against 6, 4, and 2 M urea plus 30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% glycerol. After the final dialysis step, the protein was applied to a size exclusion chromatography column (Superdex 75 10/300 GL, GE Healthcare) pre-equilibrated with gel filtration buffer [30 mM Tris pH 8.0, 300 mM NaCl, 1 mM TCEP, 5% (v/v) glycerol]. The purity of the final fractions was assessed using SDS-PAGE. Fractions of interest were pooled and incubated overnight with TEV protease to remove the poly-histidine tag. After depletion of TEV and uncleaved protein over a 5 ml Ni-NTA column, the cleaved $AG^{(74-173)}$ was loaded onto a Superdex S75 10/300 GL column as a final purification step and the fractions of interest pooled and concentrated to approximately 4 mg/ml for SAXS studies.

SAXS Data Collection

An on-line hplc system (Viscotek, Malvern Instruments) was attached directly to the sample inlet valve of the BM29 sample changer (European Synchrotron Radiation Facility, bioSAXS bending magnet beamline 29; Pernot et al., 2013; Round et al., 2015). The protein sample (50 μl) was injected onto the column (Superdex 75 3.2/300 PC, GE Healthcare) after column equilibration. Buffers were degassed prior to the run and a flow rate of 0.1 ml/min at room temperature was used. Buffers used were as described above. All data from the run was collected using a sample to detector (Pilatus 1 M Dectris) distance of 2.86 m corresponding to an s range of 0.04–4.9 nm^{-1}. Approximately 1800 frames (1 frame/sec) per hplc run were collected. Initial data processing was performed automatically using the EDNA pipeline (Incardona et al., 2009), generating radially integrated, calibrated, and normalized 1-D profiles for each frame. All frames were compared to the initial frame and matching frames were merged to create the reference buffer. Any subsequent frames which differed from the reference buffer were subtracted and then processed within the EDNA pipeline using tools from the EMBL-HH ATSAS suite (Petoukhov and Svergun, 2007). The invariants calculated by the ATSAS autoRg tool were used to select a subset of frames from the peak scattering intensity. The 49 frames corresponding to the highest protein concentration were merged manually and used for all further data processing and model fitting. Molecular weight for the protein was estimated based on the correlated volume (Rambo and Tainer, 2013). The approximate molecular weight was 21 kDa, corresponding to a dimer. The volume of 36 nm^3 was calculated using the GNOM interface of the cross platform version of PRIMUS for the ATSAS software suite.

AG Model Fitting

Homology models for $AG^{(74-173)}$ were generated based on the SEP3 structure (PDB 4OX0; Puranik et al., 2014). For the elongated conformation, the kink between helices 1 and 2 was removed, the helices superposed and residues corresponding to the flexible region between the helices built in manually using COOT with idealized geometry and no secondary structure restraints. The model for the bent conformation was generated by threading the sequence of $AG^{(74-173)}$ directly onto the SEP3 dimer (4OX0). Structures corresponding to two different dimer conformations (bent and elongated) were used to calculate theoretical scattering curves. These curves were compared with the experimental data using CRYSOL (Svergun et al., 1995).

RESULTS AND DISCUSSION

LEAFY and NEEDLY Structure and Function-Homology Modeling of the DBDs

The angiosperm *LFY* gene is most often found as a single copy (Brunkard et al., 2015), however, gymnosperms possess two paralogous genes- *LFY* and *NLY,* born from an ancient duplication which occurred before the divergence of the angiosperm and extant gymnosperm lineages. Examination of the genomes of gymnosperms available through the 1000 plant genomes project as well as all partial deposited sequences reveals that *LFY* and *NLY* are present in all gymnosperm genomes characterized to date, with the exception of the genus *Gnetum* where *NLY* is absent. The proteins the *LFY* and *NLY* genes encode comprise two distinct domains, a partially conserved N-terminal domain (**Figure 2**) important for complex formation and a highly conserved C-terminal DBD (70% sequence identity between AtLFY and WmNLY, for example; **Figure 3A**), with connecting regions presenting a higher degree of variability. In order to probe the function of these proteins, we first aligned the DBDs of LFY and NLY using ClustalW in order to assess conservation of DNA-binding specificity (**Figure 3A**). We observed that the DBDs of LFY and NLY are highly conserved in all seed plants based on sequence alignment. To investigate any potential changes in quaternary structure or putative alterations in the DNA-binding interface, the crystal structure of the DBD of LFY from *A. thaliana* (Hames et al., 2008) was used as a homology model to generate 3D models of the DBDs of gymnosperm LFY and NLY (**Figure 3B**) using SWISS-MODEL with default parameters. Comparison of the primary sequences with secondary, tertiary and quaternary structure derived from the crystallographic data revealed that the DBDs are structurally identical and all amino acids involved in direct contacts with DNA are completely conserved between angiosperm LFY (aLFY), gymnosperm LFY (gLFY), and NLY. In addition, the dimerisation interface recently described as a key component in DNA binding specificity (Sayou et al., 2014) is also highly conserved between gymnosperms and angiosperms as shown in **Figure 3**. However, while AtLFY His383 is almost completely conserved in both angiosperms and gymnosperms, based on all available sequence data, the residue at position AtLFY 386 varies as either an arginine in aLFY and NLY (Arg399 in PrNLY, **Figure 3C**) or by substitution as a lysine in gLFY (Lys405 in PrLFY, **Figure 3D**). Arginine and lysine fulfill similar structural roles and can substitute for one another due to the conserved positive charge and hydrogen bonding ability of the primary ε-amine and guanidine group for lysine and arginine, respectively (Sokalingam et al., 2012). However the higher pKa and longer size of the arginine side chain may affect the hydrogen bonding interaction with the carbonyl oxygen of residue 276 (AtLFY; residue 289, PrNLY, **Figure 3C**) and cannot be ruled out as affecting dimer stability, and possibly conformation (relative positioning of the monomers). Overall, the high degree of sequence identity between the DBD of aLFY, gLFY, and NLY implies a likely conserved recognition of cognate DNA sequences. Recent studies by Sayou et al. (2014) have demonstrated the evolutionary trajectory of LFY from green algae to moss to angiosperms based on structural and biochemical studies of several DBDs including those of *Klebsormidium subtile* LFY (algae), *Physcomitrella patens* LFY (moss), and *Arabidopsis* LFY (Sayou et al., 2014). The distantly related LFY from algae, moss and angiosperms were shown to bind different DNA motifs due to small changes in the LFY dimerisation interface, as well as in two other key amino acids (AtLFY His308 and Arg341) that determine the DNA half-site sequence recognized (**Figure 3A**), as previously determined through a combination of structural and

SELEX experiments (Sayou et al., 2014). However, the SELEX motif for the DBD of gymnosperm *G. biloba* LFY (GbLFY) is almost identical to the SELEX motif for the DBD of AtLFY (Sayou et al., 2014). As the dimerisation interface and all residues directly contacting the DNA are highly conserved in angiosperms and gymnosperms for aLFY, gLFY and NLY, this suggests the proteins are able to bind the same or very similar DNA motifs. Thus, it is probable that the DBD of LFY/NLY in higher plants became fixed, with conservation of DNA binding and dimerisation motifs. While bound DNA sequences and DNA binding matrices are available for LFY from *A. thaliana* based on multiple ChIP-seq and SELEX studies, no such data is available for gLFY or NLY with the exception of the GbLFY motif. Additional data would be important to confirm that there are no subtle allosteric effects that may tune the DNA binding specificity of these different paralogs, a possibility that cannot be excluded based on available data.

Functional Implications of Complex Formation- the Role of the N-terminal Domain in LFY and NLY Function

Interestingly, functional studies do not show full complementation of a *lfy* mutant in *A. thaliana* by either *gLFY* (from *P. radiata*) or *NLY* (from *W. mirabilis*; Maizel et al., 2005). If the DBDs are able to recognize the same DNA sequences, why do gLFY and NLY less efficiently complement the *Arabidopsis lfy* mutant? One explanation relies on complex formation with ternary factors that may tune DNA binding specificity, for example through multi-site binding of different adjacent *cis*-elements. This suggests that differences in target gene regulation for aLFY, gLFY, and NLY likely rests on the structure and function of the N-terminal non-conserved regions of the LFY and NLY proteins. While the DBDs are virtually identical, the sequence conservation in the N-terminal regions of aLFY, gLFY, and NLY is much lower (**Figures 2** and **3A**).

The ability to interact with specific partners and form different ternary complexes changes the ability of a TF to regulate downstream genes. By retaining the core DBD and the essential DNA-binding functionality, the N-terminal region of the protein could vary, thus leading to relatively smooth changes in gene regulation over the course of evolution by simply tuning the interactions with ternary partners and thus modulating interactions with cognate DNA without requiring altering the DBD itself. The N-terminal ∼200 residues of LFY have been shown to be important for dimerisation (Siriwardana and Lamb, 2012) and can possibly play a role in the formation of higher order complexes with chromatin remodelers and other TFs (Wu et al., 2012). Indeed, unfolded, flexible loops, and low-complexity regions exhibit greater variability and tolerance for mutations, as they do not affect the overall fold of the macromolecule. In addition, these regions often have important functions and act as protein–protein interaction surfaces (Dyson and Wright, 2005). While alpha-helices are relatively disfavored as protein–protein interaction interfaces, exposed beta strands, hydrophobic patches and long loops are more likely to play a role in complex formation (Jones and Thornton, 1996; Neuvirth et al., 2004). These structural motifs are able to create relatively planar surfaces which are often correlated with protein–protein interactions (Hoskins et al., 2006). Few mapping studies of LFY have been performed and only a small number of interaction surfaces with partner proteins have been determined (Chae et al., 2008; Souer et al., 2008; Pastore et al., 2011; Siriwardana and Lamb, 2012; Wu et al., 2012). From the limited data available, however, it seems that several partners interact with the N-terminal region of the protein (Souer et al., 2008; Siriwardana and Lamb, 2012). Structural characterization of the N-terminal domain of LFY would allow determining whether its properties might have changed during evolution.

Due to the loss of NLY in the angiosperm lineage, aLFY likely assumed additional functions, fusing the functionality of NLY, a key regulator of female organ development, and gLFY, an important primary regulator of male cone development, into one fully competent regulator of plant reproduction. As has been recently shown for several conifers (*Picea abies, Podocarpus reichei,* and *Taxus globosa*), LFY and NLY have overlapping expression patterns (Vazquez-Lobo et al., 2007; Carlsbecker et al., 2013). This would mitigate any deleterious effects of NLY loss during the gymnosperm/angiosperm split by allowing more facile compensation for NLY function by LFY, as LFY was already present in the same tissues, possessed the same DBD, and likely recognized very similar cognate DNA sequences. Thus, aLFY compensation for NLY/gLFY during reproductive development would not necessitate extensive reprogramming of LFY expression patterns nor require any changes to the gene coding sequence of the DBD, important factors in the successful compensation due to gene loss of NLY in the angiosperm lineage.

MADS-Domain TFs and Their Role in Floral Organ Development

The homeotic class A-E MADS-box genes direct the specification of all the floral organs and as such are central players in flower evolution and development. In gymnosperms, orthologs to the B and C class MADS-box genes (*AP3/PI* and *AG* in *Arabidopsis*) are also present and play important roles in male and female organ development. While the MADS-box gene family has expanded in all land plants, this is most striking in angiosperms due to extensive duplication events giving rise to the class E *SEPALLATA* genes, which are not present in extant gymnosperms (Zahn et al., 2005a). The SEPALLATA (SEP) proteins have acquired new functionality and act as mediators of interactions between class A, B, and C MADS-domain TFs as shown by yeast two and three hybrid studies, EMSA experiments and *in vivo* studies (Pelaz et al., 2000; Honma and Goto, 2001; Kaufmann et al., 2005; Malcomber and Kellogg, 2005; Theissen and Melzer, 2007; Immink et al., 2009; Mendes et al., 2013). The SEP proteins form heteromeric complexes with other MADS TFs and all putative floral organ-specifying tetrameric MADS complexes contain at least one SEP protein leading to the specification of the different floral organs (Theissen, 2001; Theissen and Saedler, 2001). Indeed, *sep123* mutants are sterile and unable to produce male or female organs, with the flower converted to a collection of sepaloid-like structures, illustrating the requirement of the SEP proteins for proper reproductive

organ formation (Pelaz et al., 2000). Examination of the B and C class MADS TFs in gymnosperms such as *G. gnemon* suggests that tetramerisation can occur and is necessary for male and female organ development (**Figure 1**). This tetramerisation takes place without the obligatory mediation of the class E-like AGL6 proteins (Wang et al., 2010). However, angiosperms are dependent on the class E SEPs for tetramer formation, as the B and C class TFs have lost their ability to directly interact. Current hypothesis suggest that the changing interaction patterns of the MADS TFs, in particular the requirement of the SEPs to mediate tetramer formation in angiosperms is at the nexus of flower origins (Melzer et al., 2010; Wang et al., 2010). The evolution of the bisexual flower thus requires an understanding, at the protein level, of the MADS TFs, particularly how the SEPs are able to mediate the formation of tetrameric complexes which are critical to the development of all the floral organs.

Our recent crystallographic data of the oligomerisation domain of SEP3 (Puranik et al., 2014), together with mutagenesis studies, sequence alignments and biophysical characterization of the C-class MADS TF AGAMOUS (this study) help to explain the molecular function of the MADS TFs and contribute to our understanding of flower evolution. All MADS homeotic TFs are characterized by a four domain arrangement consisting of a highly conserved DBD "M" domain (∼60 amino acid MADS domain, **Figure 4**), an "I" domain (linker Intervening domain) important for dimerisation, a "K" domain (alpha helical Keratin-like domain) critical for dimerisation and tetramerisation, and a "C" domain (highly variable C-terminal domain) important for different functions including transactivation and higher order complex formation (Kaufmann et al., 2005). Based on the crystal structure of a portion of the I and the full K domain of SEP3 (Puranik et al., 2014) and extensive mutagenesis studies, the dimerisation and tetramerisation interfaces of the MADS-domain TFs can be mapped at the amino acid level (**Figures 5A,B**). Different amino acids along the dimer and tetramer interface were targeted for mutagenesis studies in order to probe the mechanisms of oligomerisation and stability (**Table 2**). Mutation of any residue making a direct contact with its partner along the dimer (Leu115, Leu131, Leu135, Tyr98, Tyr105; this study) or tetramer (Met150, Leu154, Leu171; Puranik et al., 2014) interface in SEP3 had a striking effect on oligomerisation, with even a single point mutation greatly destabilizing the complex as determined by size exclusion chromatography and comparison with the wild type protein. This suggests that subtle differences in the amino acids at the dimerisation and tetramerisation surface will shift the oligomerisation equilibrium to favor certain complexes when multiple MADS TFs are present. Examining structure based sequence alignments for the homeotic MADS-domain TFs demonstrates a conservation of hydrophobic residues at the oligomerisation interface, but the size and shape of these residues varies, which will help mediate protein–protein interactions (**Figure 5A**).

Based on the structure of the SEP3 homotetramer and mutagenesis studies, we probed the formation of hetero-oligomers using electrophoretic mobility shift assays (**Figure 6**). EMSA experiments and identification of putative complexes were performed according to previously published work (Smaczniak et al., 2012). SEP3 dimerisation and tetramerisation mutants were tested for DNA binding with AG, all expressed using *in vitro* transcription translation due to the difficulties in producing folded full length MADS TFs using standard recombinant bacterial expression. Sufficient heterodimers and tetramers were produced and a gel shift assay was performed using DNA corresponding to the *SOC1* promoter containing two CArG-box MADS TF binding sites (**Figure 6A**) and the *SOC1* promoter sequence with either the first or the second CArG-box mutated (**Figure 6B**). A SEP3 dimerisation-interface mutant, $SEP3^{L115R}$, was dramatically impaired in its ability to oligomerise based on studies of the K-domain alone (**Table 2**), however, it was able to bind DNA as a homodimer and heterotetramer with AG, albeit with less efficiency than the wild type SEP3. The $SEP3^{L115R}$ mutant was designed to mimic the sequence of AtAP3, which is unable to form homodimers but still retains the ability to interact with partners such as AtPI (Riechmann et al., 1996; Winter et al., 2002; Yang et al., 2003). Both AtPI and AtAG have a leucine residue at position 115, which is likely able to accommodate the arginine side chain during hetero-oligomer formation. AGAMOUS alone exhibited poor binding to the *SOC1* DNA due either to lower protein production in the *in vitro* transcription translation reaction or non-optimal sequences of the DNA, however, AG heterodimers with SEP3 were able to bind the *SOC1* sequences, suggesting differences in sequence specificity are important for AG homo and heteromer DNA binding interactions. Tetramerisation interface mutants $SEP3^{L171A}$ and a truncation mutant ($SEP3^{\Delta C}$) showed greatly impaired heterotetramerisation with AG, as expected. Altogether, these data provide strong evidence that the homotetramerisation interface observed in the crystal structure of SEP3 is conserved in the formation of heterotetramers.

Changes in the tetramerisation interface in SEP partner MADS proteins also has an effect on oligomer formation. For example, studies of the C-class genes *PLENA (PLE)* and

TABLE 2 | Effect of point mutations on SEP3 oligomerisation.

	$SEP3^{(75-178)}$ constructs	Oligomerisation state
Tetramerisation interface	Wild type	**Tetramer**/dimer
(Puranik et al., 2014)	M150A	**Dimer**
	L154A	**Dimer**
	L171A	**Dimer**
	$Splice^{\Delta 161-174}$	**Dimer**
Dimerisation interface	Y98K	**Dimer**/monomer
(this study)	Y105N	**Tetramer**/monomer
	L115R	Unstable complex
	L131V	Unstable complex
	L135M	**Dimer**/monomer
	L135A	Unstable complex

Point mutations targetting the highly conserved residues involved in the putative dimerisation and tetramerisation interface of SEP3 were chosen for mutational analysis. Oligomeric state was determined by size exclusion chromatography. Where two states exist, the predominant species is marked in bold. "Unstable complex" is used to denote a complex mixture of species between monomeric, dimeric, and tetrameric states with no predominance for a particular oligomerisation state.

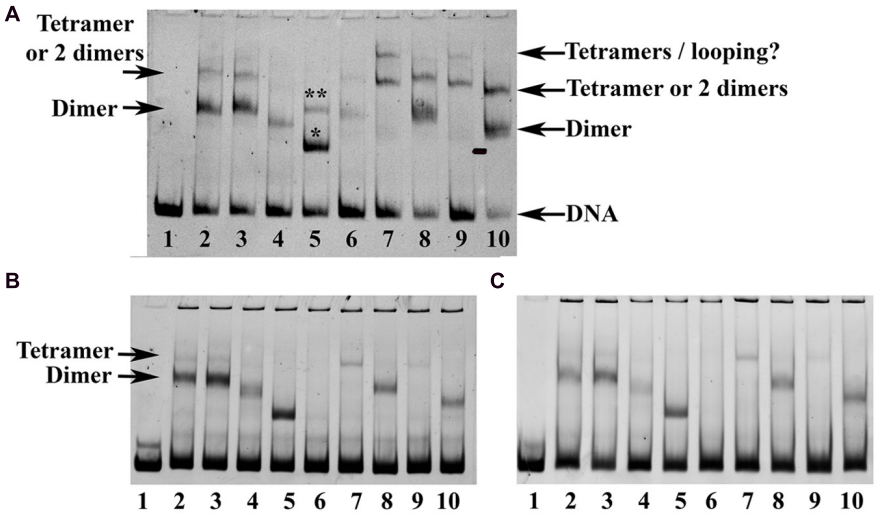


FIGURE 6 | Electrophoretic mobility shift assay (EMSA) for SEP3 and AG. (A) Comparison of the oligomerisation state of SEP3 wild type, dimerisation and tetramerisation mutants with AG using a 121 bp DNA fragment of the SOC1 promoter comprising two CArG boxes. Lane 1 corresponds to DNA alone, lane 2 SEP3 wild type, lane 3 $SEP3^{L171A}$, lane 4 $SEP3^{L115R}$, lane 5 $SEP3^{\Delta C}$ truncation mutant, lane 6 AG, lane 7 SEP3 + AG, lane 8 $SEP3^{L171A}$ + AG, lane 9 $SEP3^{L115R}$ + AG, lane 10 $SEP3^{\Delta C}$ + AG. Putative tetramer and dimer are indicated by arrows. The truncation mutant in lane 5 likely has one or two dimers bound to DNA as indicated above the bands by a single (one dimer bound) or double (two dimers bound) asterix. A faint highly retarded band corresponding to bound tetramers or tetramer-induced DNA-looping was noted in lanes 7 and 9 and indicated. **(B)** EMSAs run with either CArG box 1 (left) or CArG box 2 (right) mutated (see Materials and Methods). Proteins are as per **(A)**. A faint band for SEP3 wild type and $SEP3^{L171A}$ was noted running as per the tetramer band in **(A)**, suggesting homotetramer formation on a single CArG box site. All proteins were produced via *in vitro* transcription translation using equivalent amounts of template DNA and equivolumes of the reaction mixture were added to the final binding reaction. DNA was approximately 5–10 nM and labeled with DY-682 **(A)** or Cy5 **(B)** for imaging.

FARINELLI (FAR) from *A. majus* demonstrate that a single amino acid change was responsible for neofunctionality of these duplicated genes with *FAR* able to specify only male organs and *PLE* able to specify both male and female organs in a complementation assay in *Arabidopsis*. This activity was due to a single amino acid insertion in the K domain that altered the oligomerisation capabilities of PLE and FAR with the SEPALLATA proteins (Airoldi et al., 2010). An amino acid insertion shifts the hydrophobic pattern of all amino acids in the leucine zipper tetramerisation interface, thus modulating the hydrophobic protein–protein interface of the putative tetrameric complexes formed by PLE and FAR with their SEP partners.

In addition to the hydrophobic dimer and tetramer interface acting as a driver for oligomerisation, a key component of the MADS TFs oligomerisation propensity is the presence of a kink in between alpha helices 1 and 2 of the K domain (**Figures 5B,C**). Based on sequence alignments of the MADS homeotic TFs, this kink region is highly variable in the family with a tight turn predicted for SEP1, SEP2 and SEP3 due to the presence of a GlyPro motif (**Figure 5A**). Prolines act as "breakers" in an alpha helix due to their inability to form the appropriate hydrogen bonding interactions between the carbonyl backbone and amide proton due to the presence of the proline side chain. Glycine residues exhibit a high degree of conformational flexibility and have been shown to lead to kinks in alpha helices in soluble and membrane proteins (Wilman et al., 2014). These residues result in the formation of a tight turn and, in the case of SEP3, an approximately 90° bend between alpha helices 1 and 2 (**Figure 5B**). Examination of the sequences of other MADS TFs show scattered glycine and/or proline residues between helices 1 and 2, but not a conservation of the GlyPro motif observed in SEP1, SEP2 and SEP3. In order to investigate whether the presence of a GlyPro motif is required for complete opening of helices 1 and 2, we recombinantly overexpressed and purified the K-domain of AG.

The $AG^{(74-173)}$ construct, spanning the complete AG K domain, was designed based on both secondary structure predictions using PSIPRED (Jones, 1999) and homology modeling with the SEP3 structure using SWISS-MODEL. This protein was used in small angle X-ray scattering (SAXS) studies to determine oligomerisation state and conformational flexibility of the AG K domain in solution. The $AG^{(74-173)}$ construct was expressed in *E. coli*, purified from inclusion bodies under denaturing conditions and subsequently refolded. Protein monodispersity and purity were assayed by size exclusion column chromatography (SEC) and SDS-PAGE prior to SAXS experiments. In order to avoid any bias due to protein aggregation or the presence of multiple oligomeric species, the $AG^{(74-173)}$ construct was purified on-line and the complete elution profile measured directly in the X-ray beam (**Figures 7A–C**). The stable radius of gyration (R_g) across the eluted protein peak corresponding to the highest protein concentration demonstrates that there is one species in solution as the particle size is constant (**Figure 7A**). In contrast to the SEP3 K-domain, which is predominantly tetrameric in solution (Puranik et al., 2014), the AG K-domain is dimeric. Volume

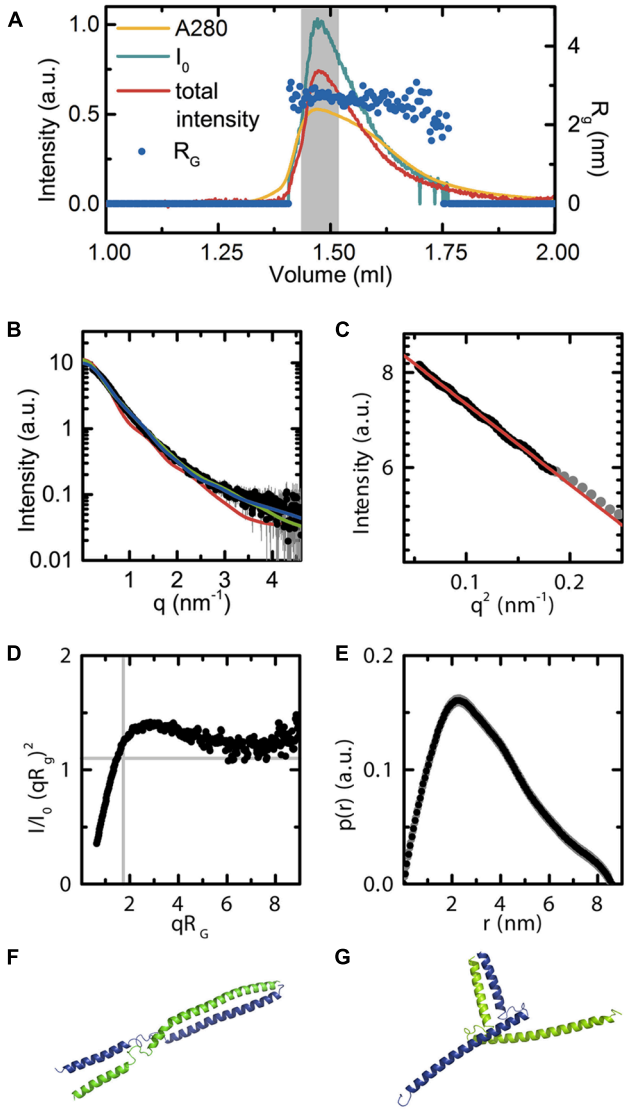


FIGURE 7 | Small angle X-ray scattering data for $AG^{(74-173)}$. **(A)** Experimental data showing the UV absorbance (yellow), X-ray scattering intensity (blue), and total intensity after buffer subtraction (red) for all collected frames. The radius of gyration across frames corresponding to the eluted peak are shown as dark blue dots ($R_g = 2.7$). The broad peak and slight variation in R_g corresponds to conformational flexibility of the protein in solution. The region of frames integrated for further analysis are highlighted in gray. Axes are as labeled. **(B)** Scattering curve in black for the integrated frames. CRYSOL fits for the bent and elongated dimer conformations, as well as the tetrameric SEP3 structure. Chi squared values were 5.6 for the elongated model (blue curve), 2.0 for the bent model (green curve) and 38.4 for the tetrameric model (red curve). **(C)** Close-up of the Guinier region. The linear fit demonstrates no evidence of aggregation of the protein. **(D)** Normalized Kratky plot calculated using the integrated frames. The shape of the curve is indicative of a flexible particle. **(E)** P(r) function. The calculated Porod volume for the particle is 36 nm^3. Based on the Porod volume, the molecular mass of the particle is approximately 21 kDa. **(F)** Elongated homology model for $AG^{(74-173)}$. The homology model was based on the SEP3 K domain (4OX0) and secondary structure predictions for AG. Each monomer is depicted as a cartoon and colored blue and green. **(G)** Bent homology model for $AG^{(74-173)}$. Each monomer is depicted as a cartoon and colored blue and green. The homology models **(F,G)** were used for fitting the data using CRYSOL as shown in **(B)**.

calculations based on the histogram of interatomic distances for the particle give a volume of 36 nm^3, corresponding to a molecular mass of approximately 21 kDa, the molecular mass of an $AG^{(74-173)}$ dimer (**Figure 7E**). $AG^{(74-173)}$ exhibits a great deal of flexibility based on the Kratky plot (**Figure 7D**), which is characteristic of a highly flexible and/or partially disordered protein in solution. In order to further investigate the possible conformations of the $AG^{(74-173)}$ dimer, homology models based on the structure of SEP3 (4OX0) were generated in an elongated and bent conformation (**Figures 7F,G**). CRYSOL fits (**Figure 7B**) were relatively consistent with either particle shape giving chi-squared values of 5.6 and 2.0 for the elongated and bent conformations, respectively. In contrast, the tetrameric SEP3 structure is inconsistent with the recorded data, giving a chi-squared of 38.4 (**Figure 7B**). The R_g for both dimeric homology models (3.1 nm for the bent and 3.6 nm for the elongated model) was slightly bigger than the calculated R_g of 2.7 nm for the measured data. This variation is attributable to disorder, multiple unmodeled conformations and/or partial unfolding at the termini of the protein. Contamination by a tetramer or soluble aggregates is considered highly unlikely as these species would elute prior to the measured peak and there is no evidence for this in the UV trace or X-ray scattering of the sample.

While possessing glycine residues in the kink region between helices 1 and 2, AG lacks the GlyPro motif seen in SEP1, SEP2 and SEP3. Although it is well-established that AG can form tetrameric complexes, these complexes usually contain a SEP partner. Indeed almost characterized floral organ tetrameric complexes of homeotic MADS TFs from angiosperms to date rely on at least one SEP protein for tetramer formation (Honma and Goto, 2001; Theissen and Saedler, 2001). Thus, the SEPs are able to act as hubs of tetramer formation for other MADS TFs. Because the GlyPro motif forces open helix 2 exposing hydrophobic surfaces, we postulate that the SEP proteins are able to preferentially form tetramers with themselves or other MADS TF proteins and this exposed hydrophobic surface on helix 2 acts as an entropic driving force for oligomerisation.

Some gymnosperm B and C-class MADS TFs are postulated to form tetramers when bound to DNA. *In vitro* studies of GGM2 (*G. gnemon* B-like) and GGM3 (*G. gnemon* C-like) demonstrate that GGM2 can form heterotetramers with GGM3 and that GGM3 is additionally able to homotetramerise when bound to DNA (Wang et al., 2010). Examination of the kink region between helices 1 and 2 as determined from secondary structure predictions and sequence alignments for GGM2 and GGM3 reveals the presence of two glycine residues for GGM2 but no proline. GGM3 has scattered glycines in both the kink region and in the N-terminal portion of helix 2 (**Figure 5A**). We speculate that these glycines will destabilize helix 2 and increase the conformational space the protein is able to sample. Indeed, GGM3 was shown to homotetramerise on DNA with non-optimally spaced binding sites, suggesting additional flexibility of the protein and the tetramerisation interface (Wang et al., 2010). It is likely that the combination

of helix destabilization in GGM3 and the relatively plastic kink region in GGM2 is sufficient to allow the formation of tetrameric complexes when the local concentration of the proteins is relatively high as would be the case when bound to adjacent regions of DNA. Further experiments to probe these interactions and extensive mutagenesis studies would be required to fully determine the rules governing tetramerisation. Nascent tetramerisation capabilities are present in at least some species of gymnosperm MADS TFs, though whether tetramerisation is required for proper gene regulation is less clear. However, in angiosperms, interactions mediated by the SEP class of MADS TFs is required for male and female organ specification and reproductive development. The gene duplication event giving rise to the SEPALLATA class of MADS TFs and their central role in organizing the homeotic MADS TFs into functional tetrameric complexes we hypothesize to be a key component in flower origins and evolution.

Taken together, these data suggest that the interaction surfaces and oligomerisation of the MADS TFs is both variable and highly sensitive to even small alterations in amino acid sequence which would allow for the fast evolution of different interactions within the family. By retaining the core DBD, the essential function of the MADS TFs- DNA binding to specific cognate sequences- would be preserved, but mutations in the auxiliary I, K, and C domains would allow for functional plasticity by changing the identity or altering the affinity of protein interaction partners. This model is very similar to what is observed for aLFY, gLFY, and NLY in which the C-terminal DBD is conserved and the auxiliary N-terminal region involved in protein–protein interactions is allowed to vary, likely changing ternary complex formation and tuning downstream gene regulation.

CONCLUSION

Small changes in TFs that do not directly affect the DBD can trigger very striking evolutionary developmental changes in an organism. LFY and the MADS TFs illustrate how small changes at the genetic level lead to dramatic alterations and novel functions at the protein level. While the evolutionary origins of the bisexual angiosperm flower are still unclear, major genetic changes - the loss of NLY and the duplication event resulting in the SEPALLATA genes in angiosperms- likely play key roles. How these genetic changes were able to result in morphological changes requires an integrated study incorporating detailed examination of protein structure and biochemistry. By exploring the protein structure-function relationship, particularly for TFs whose activity impacts entire downstream networks, we can begin to understand the molecular basis for evolution. Structural biology offers an important perspective in probing this relationship for the master regulators, LFY and the MADS TFs, and provides a foundation for understanding how alterations in protein structure lead to the evolution of new functions and new morphologies at the organismal level.

ACKNOWLEDGMENTS

We would like to acknowledge Darren Hart and Philippe Mas for the use of the pESPRIT002 vector, Kerstin Kaufmann and Cezary Smaczniak for EMSA protocols and pSPUTK SEP3 and AG TnT vectors and Renaud Dumas for discussions and critical reading of the manuscript. We are grateful to the ESRF for provision of beamtime and support for the experiments. This work was supported by an ATIP-Avenir (CZ).

REFERENCES

Airoldi, C. A., Bergonzi, S., and Davies, B. (2010). Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18898–18902. doi: 10.1073/pnas.1009050107

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201. doi: 10.1093/bioinformatics/bti770

Bateman, R. M., Hilton, J., and Rudall, P. J. (2006). Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *J. Exp. Bot.* 57, 3471–3503. doi: 10.1093/jxb/erl128

Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M. A., Saedler, H., et al. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics* 266, 942–950. doi: 10.1007/s00438-001-0615-8

Becker, A., Saedler, H., and Theissen, G. (2003). Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev. Genes Evol.* 213, 567–572. doi: 10.1007/s00427-003-0358-0

Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258. doi: 10.1093/nar/gku340

Brunkard, J. O., Runkel, A. M., and Zambryski, P. C. (2015). Evolution. Comment on "A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity." *Science* 347, 621. doi: 10.1126/science.1255437

Burkhardt, F., Sydney, S., David, K., and William, M. (1985). *The Correspondence of Charles Darwin*. Cambridge: Cambridge University Press.

Busch, M. A., Bomblies, K., and Weigel, D. (1999). Activation of a floral homeotic gene in *Arabidopsis*. *Science* 285, 585–587. doi: 10.1126/science.285.5427.585

Carlsbecker, A., Sundstrom, J. F., Englund, M., Uddenberg, D., Izquierdo, L., Kvarnheden, A., et al. (2013). Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* 200, 261–275. doi: 10.1111/nph.12360

Carlsbecker, A., Tandre, K., Johanson, U., Englund, M., and Engstrom, P. (2004). The MADS-box gene DAL1 is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.* 40, 546–557. doi: 10.1111/j.1365-313X.2004.02226.x

Chae, E., Tan, Q. K., Hill, T. A., and Irish, V. F. (2008). An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development* 135, 1235–1245. doi: 10.1242/dev.015842

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147–150. doi: 10.1016/S0968-0004(99)01540-6

De Bodt, S., Raes, J., Van De Peer, Y., and Theissen, G. (2003). And then there were many: MADS goes genomic. *Trends Plant Sci.* 8, 475–483. doi: 10.1016/j.tplants.2003.09.006

Dornelas, M. C., and Rodriguez, A. P. (2005). The rubber tree (*Hevea brasiliensis* Muell. Arg.) homologue of the LEAFY/FLORICAULA gene is preferentially expressed in both male and female floral meristems. *J. Exp. Bot.* 56, 1965–1974. doi: 10.1093/jxb/eri194

Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., et al. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611. doi: 10.1105/tpc.015123

Friedman, W. E. (1990). Double fertilization in ephedra, a nonflowering seed plant: its bearing on the origin of angiosperms. *Science* 247, 951–954. doi: 10.1126/science.247.4945.951

Friedman, W. E. (2009). The meaning of Darwin's 'abominable mystery.' *Am. J. Bot.* 96, 5–21. doi: 10.3732/ajb.0800150

Friedman, W. E., and Carmichael, J. S. (1996). Double fertilization in gnetales: implications for understanding reproductive diversification among seed plants. *Int. J. Plant Sci.* 157, S77–S94. doi: 10.1086/297405

Frohlich, M. W., and Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* 450, 1184–1189. doi: 10.1038/nature06393

Frohlich, M. W., and Meyerowitz, E. M. (1997). The search for flower homeotic gene homologs in basal angiosperms and Gnetales: a potential new source of data on the evolutionary origin of flowers. *Int. J. Plant Sci.* 158, S131–S142. doi: 10.1086/297513

Frohlich, M. W., and Parker, D. S. (2000). The Mostly Male theory of flower evolutionary origins:from genes to fossils. *Syst. Bot.* 25, 155–170. doi: 10.2307/2666635

Gramzow, L., Ritz, M. S., and Theissen, G. (2010). On the origin of MADS-domain transcription factors. *Trends Genet.* 26, 149–153. doi: 10.1016/j.tig.2010.01.004

Gramzow, L., Weilandt, L., and Theissen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* 114, 1407–1429. doi: 10.1093/aob/mcu066

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., et al. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* 15, 500–506. doi: 10.1038/nsmb.1421

Hames, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F., et al. (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.* 27, 2628–2637. doi: 10.1038/emboj.2008.184

Hart, D. J., and Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in *Escherichia coli*. *Acta Crystallogr. D Biol. Crystallogr.* 62, 19–26. doi: 10.1107/S0907444905036097

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083

Hoskins, J., Lovell, S., and Blundell, T. L. (2006). An algorithm for predicting protein-protein interaction sites: abnormally exposed amino acid residues and secondary structure elements. *Protein Sci.* 15, 1017–1029. doi: 10.1110/ps.051589106

Immink, R. G., Tonaco, I. A., De Folter, S., Shchennikova, A., Van Dijk, A. D., Busscher-Lange, J., et al. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol.* 10, R24. doi: 10.1186/gb-2009-10-2-r24

Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* 16, 872–879. doi: 10.1107/S0909049509036681

Jager, M., Hassanin, A., Manuel, M., Le Guyader, H., and Deutsch, J. (2003). MADS-box genes in Ginkgo biloba and the evolution of the AGAMOUS family. *Mol. Biol. Evol.* 20, 842–854. doi: 10.1093/molbev/msg089

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202. doi: 10.1006/jmbi.1999.3091

Jones, S., and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13–20. doi: 10.1073/pnas.93.1.13

Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014

Kaufmann, K., Muino, J. M., Jauregui, R., Airoldi, C. A., Smaczniak, C., Krajewski, P., et al. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biol.* 7:e1000090. doi: 10.1371/journal.pbio.1000090

Kaufmann, K., Wellmer, F., Muino, J. M., Ferrier, T., Wuest, S. E., Kumar, V., et al. (2010). Orchestration of floral initiation by APETALA1. *Science* 328, 85–89. doi: 10.1126/science.1185244

Krizek, B. A., and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat. Rev. Genet.* 6, 688–698. doi: 10.1038/nrg1675

Lamb, R. S., Hill, T. A., Tan, Q. K., and Irish, V. F. (2002). Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development* 129, 2079–2086.

Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S., and Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* 11, 1007–1018. doi: 10.1105/tpc.11.6.1007

Lohmann, J. U., and Weigel, D. (2002). Building beauty: the genetic control of floral patterning. *Dev. Cell* 2, 135–142. doi: 10.1016/S1534-5807(02)00122-3

Lord, E. M., and Russell, S. D. (2002). The mechanisms of pollination and fertilization in plants. *Annu. Rev. Cell Dev. Biol.* 18, 81–105. doi: 10.1146/annurev.cellbio.18.012502.083438

Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229

Malcomber, S. T., and Kellogg, E. A. (2005). SEPALLATA gene diversification: brave new whorls. *Trends Plant Sci.* 10, 427–435. doi: 10.1016/j.tplants.2005.07.008

Martinez-Castilla, L. P., and Alvarez-Buylla, E. R. (2003). Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13407–13412. doi: 10.1073/pnas.1835864100

Mellerowicz, E. J., Horgan, K., Walden, A., Coker, A., and Walter, C. (1998). PRFLL–a *Pinus radiata* homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta* 206, 619–629. doi: 10.1007/s004250050440

Melzer, R., Wang, Y. Q., and Theissen, G. (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin. Cell Dev. Biol.* 21, 118–128. doi: 10.1016/j.semcdb.2009.11.015

Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., et al. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *Plant Cell* 25, 2560–2572. doi: 10.1105/tpc.112.108688

Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S., et al. (1998). NEEDLY, a *Pinus radiata* ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6537–6542. doi: 10.1073/pnas.95.11.6537

Moyroud, E., Kusters, E., Monniaux, M., Koes, R., and Parcy, F. (2010). LEAFY blossoms. *Trends Plant Sci.* 15, 346–352. doi: 10.1016/j.tplants.2010.03.007

Nam, J., DePamphilis, C. W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20, 1435–1447. doi: 10.1093/molbev/msg152

Neuvirth, H., Raz, R., and Schreiber, G. (2004). ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol.* 338, 181–199. doi: 10.1016/j.jmb.2004.02.040

Ng, M., and Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nat. Rev. Genet.* 2, 186–195. doi: 10.1038/35056041

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584. doi: 10.1038/nature12211

Parcy, F., Nilsson, O., Busch, M. A., Lee, I., and Weigel, D. (1998). A genetic framework for floral patterning. *Nature* 395, 561–566. doi: 10.1038/26903

Pastore, J. J., Limpuangthip, A., Yamaguchi, N., Wu, M. F., Sang, Y., Han, S. K., et al. (2011). LATE MERISTEM IDENTITY2 acts together with LEAFY to activate APETALA1. *Development* 138, 3189–3198. doi: 10.1242/dev.063073

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103

Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., et al. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* 20, 660–664. doi: 10.1107/S0909049513010431

Petoukhov, M. V., and Svergun, D. I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* 17, 562–571. doi: 10.1016/j.sbi.2007.06.009

Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., et al. (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in *Arabidopsis*. *Plant Cell* 26, 3603–3615. doi: 10.1105/tpc.114.127910

Rambo, R. P., and Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477–481. doi: 10.1038/nature12070

Riechmann, J. L., Krizek, B. A., and Meyerowitz, E. M. (1996). Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl. Acad. Sci. U.S.A.* 93, 4793–4798. doi: 10.1073/pnas.93.10.4793

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., et al. (2015). BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallogr. D Biol. Crystallogr.* 71, 67–75. doi: 10.1107/S1399004714026959

Sablowski, R. (2010). Genes and functions controlled by floral organ identity genes. *Semin. Cell Dev. Biol.* 21, 94–99. doi: 10.1016/j.semcdb.2009.08.008

Saedler, H., Becker, A., Winter, K. U., Kirchner, C., and Theissen, G. (2001). MADS-box genes are involved in floral development and evolution. *Acta Biochim. Pol.* 48, 351–358.

Sayou, C., Monniaux, M., Nanao, M. H., Moyroud, E., Brockington, S. F., Thevenon, E., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648. doi: 10.1126/science.1248229

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931

Shindo, S., Sakakibara, K., Sano, R., Ueda, K., and Hasebe, M. (2001). Characterization of a FLORICAULA/LEAFY homologue of Gnetum parvifolium and its implications for the evolution of reproductive organs in seed plants. *Int. J. Plant Sci.* 162, 1199–1209. doi: 10.1086/323417

Siriwardana, N. S., and Lamb, R. S. (2012). A conserved domain in the N-terminus is important for LEAFY dimerization and function in *Arabidopsis thaliana*. *Plant J.* 71, 736–749. doi: 10.1111/j.1365-313X.2012.05026.x

Smaczniak, C., Immink, R. G., Muino, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1560–1565. doi: 10.1073/pnas.1112871109

Sokalingam, S., Raghunathan, G., Soundrarajan, N., and Lee, S. G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* 7:e40410. doi: 10.1371/journal.pone.0040410

Souer, E., Rebocho, A. B., Bliek, M., Kusters, E., De Bruin, R. A., and Koes, R. (2008). Patterning of inflorescences and flowers by the F-Box protein DOUBLE TOP and the LEAFY homolog ABERRANT LEAF AND FLOWER of petunia. *Plant Cell* 20, 2033–2048. doi: 10.1105/tpc.108.060871

Sundstrom, J., Carlsbecker, A., Svensson, M. E., Svenson, M., Johanson, U., Theissen, G., et al. (1999). MADS-box genes active in developing pollen cones of Norway spruce (*Picea abies*) are homologous to the B-class floral homeotic genes in angiosperms. *Dev. Genet.* 25, 253–266. doi: 10.1002/(SICI)1520-6408(1999)25:3<253::AID-DVG8>3.0.CO;2-P

Sundstrom, J., and Engstrom, P. (2002). Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J.* 31, 161–169. doi: 10.1046/j.1365-313X.2002.01343.x

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. doi: 10.1107/S0021889895007047

Tandre, K., Albert, V. A., Sundas, A., and Engstrom, P. (1995). Conifer homologues to genes that control floral development in angiosperms. *Plant Mol. Biol.* 27, 69–78. doi: 10.1007/BF00019179

The PyMOL Molecular Graphics System (2010). *The PyMOL Molecular Graphics System*. New York, NY: Schrödinger, Llc.

Theissen, G. (2000). Evolutionary developmental genetics of floral symmetry: the revealing power of Linnaeus' monstrous flower. *Bioessays* 22, 209–213. doi: 10.1002/(SICI)1521-1878(200003)22:3<209::AID-BIES1>3.3.CO;2-A

Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/S1369-5266(00)00139-4

Theissen, G. (2005). Birth, life and death of developmental control genes: new challenges for the homology concept. *Theory Biosci.* 124, 199–212. doi: 10.1007/BF02814484

Theissen, G., and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* 100, 603–619. doi: 10.1093/aob/mcm143

Theissen, G., and Saedler, H. (1995). MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Curr. Opin. Genet. Dev.* 5, 628–639. doi: 10.1016/0959-437X(95)80032-8

Theissen, G., and Saedler, H. (2001). Plant biology. Floral quartets. *Nature* 409, 469–471. doi: 10.1038/35054172

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Vazquez-Lobo, A., Carlsbecker, A., Vergara-Silva, F., Alvarez-Buylla, E. R., Pinero, D., and Engstrom, P. (2007). Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus*, and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evol. Dev.* 9, 446–459. doi: 10.1111/j.1525-142X.2007.00182.x

Wagner, D., Sablowski, R. W., and Meyerowitz, E. M. (1999). Transcriptional activation of APETALA1 by LEAFY. *Science* 285, 582–584. doi: 10.1126/science.285.5427.582

Wang, Y. Q., Melzer, R., and Theissen, G. (2010). Molecular interactions of orthologues of floral homeotic proteins from the gymnosperm *Gnetum gnemon* provide a clue to the evolutionary origin of 'floral quartets.' *Plant J.* 64, 177–190. doi: 10.1111/j.1365-313X.2010.04325.x

Weigel, D., and Meyerowitz, E. M. (1993). Activation of floral homeotic genes in *Arabidopsis*. *Science* 261, 1723–1726. doi: 10.1126/science.261.5129.1723

Wilman, H. R., Shi, J., and Deane, C. M. (2014). Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins* 82, 1960–1970. doi: 10.1002/prot.24550

Winter, C. M., Austin, R. S., Blanvillain-Baufume, S., Reback, M. A., Monniaux, M., Wu, M. F., et al. (2011). LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell* 20, 430–443. doi: 10.1016/j.devcel.2011.03.019

Winter, K. U., Saedler, H., and Theissen, G. (2002). On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm Gnetum. *Plant J.* 31, 457–475. doi: 10.1046/j.1365-313X.2002.01375.x

Wu, M. F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S. K., Li, Z., et al. (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3576–3581. doi: 10.1073/pnas.1113409109

Yang, Y., Fanning, L., and Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* 33, 47–59. doi: 10.1046/j.0960-7412.2003.01473.x

Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., et al. (2005a). The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223. doi: 10.1534/genetics.104.037770

Zahn, L. M., Leebens-Mack, J., DePamphilis, C. W., Ma, H., and Theissen, G. (2005b). To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J. Hered.* 96, 225–240. doi: 10.1093/jhered/esi033

Zhang, P., Tan, H. T., Pwee, K. H., and Kumar, P. P. (2004). Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* 37, 566–577. doi: 10.1046/j.1365-313X.2003.01983.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The Guest Associate Editor Rainer Melzer declares that, despite having collaborated with the author Chloe Zubieta, the review process was handled objectively.