

High-resolution structure of the SAM domain homodimer of the murine adapter protein SLY1

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Laura Katharina Kukuk
aus Düsseldorf

Düsseldorf, Mai 2018

aus dem Institut für physikalische Biologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichtersteller:

1. Prof. Dr. Dieter Willbold

2. PD Dr. Bernd W. König

Tag der mündlichen Prüfung: 26.06.2018

Für Linda.

Abstract

The murine adapter protein SLY1 (SH3 domain containing protein expressed in lymphocytes-1) is 380 amino acids in length and preferentially expressed in lymphocytes. SLY1 plays a role in the adaptive immune system and is required for the maturation and correct function of T-lymphocytes, B-lymphocytes and NK cells. The domain architecture of SLY1 comprises an N-terminal bipartite nuclear localisation signal (NLS), a Src homology 3 (SH3) domain and a sterile alpha motif (SAM) domain. Both, the SH3 and SAM domains, are interaction motifs that function by facilitating complex formation with biological macromolecules. Although data indicate that SLY1 interacts with 14-3-3 proteins and multiple ribosomal proteins, the function of, and biological targets that interact with the SH3 and SAM domains remain unresolved. SAM domains have been shown to bind proteins, nucleic acids and lipids. A prominent feature of SAM domains is their ability to self-associate in a variety of homo- and heterotypic interactions. These features make the assessment of SAM domain function in the context of a SAM domain containing protein challenging without detailed knowledge of the SAM domain structure and oligomerisation state. In this study, the three-dimensional structure and oligomeric state of a SLY1 SAM construct that comprised the folded core of the SLY1 SAM domain (P254–Y316) was determined by biophysical methods. Nuclear magnetic resonance (NMR) spectroscopy and analytical ultracentrifugation studies of SLY1 SAM revealed that the domain exists in a monomer-dimer equilibrium with an equilibrium dissociation constant in the lower micromolar range. Moreover, extension of the N-terminus to include a positively charged region of SLY1 increased the affinity 6–8-fold, making the SLY1 SAM dimer one order of magnitude more stable than previously studied SAM domain homodimers. Initial NMR analysis of the SLY1 SAM domain revealed severe line-broadening of signals because of chemical exchange due to the monomer-dimer equilibrium, which prohibited structure determination. To eliminate this chemical exchange process, S320 at the flexible C-terminal end of the SAM construct was exchanged for a cysteine residue to enable cross-linking of the two SAM monomers. The structure of the cross-linked SLY1 SAM dimer was solved by multidimensional heteronuclear NMR spectroscopy.

The solution structure of the SLY1 SAM domain adopts the canonical five-helix fold observed for other SAM domains. Interestingly, the SLY1 SAM domain forms a symmetric homodimer through a novel dimer interface. This interface is formed

primarily between the two long C-terminal helices $\alpha 5$ and $\alpha 5'$ of the domains packing against each other. Additional intermolecular contacts between amino acids in helices $\alpha 5$ and $\alpha 1'$ as well as loop $\alpha 1'/\alpha 2'$, and also between residues in the two C-termini further stabilise the dimer. The solution structure of the SLY1 SAM homodimer was corroborated by the structure of the wild-type SLY1 SAM determined by X-ray crystallography. The structure was solved to a resolution of 2.05 Å. Comparison of the solution and crystal structures yielded a C^α root-mean-square-deviation of 1.13 Å, indicating a high global fold similarity of the SAM dimer. The structure of the SLY1 SAM dimer presents a typical SAM domain interaction motif on each monomer that may provide a scaffold for SLY1 mediated protein interactions with other SAM domain containing proteins.

Zusammenfassung

Das Protein *SH3 domain containing protein expressed in lymphocytes-1* (SLY1) ist ein 380 Aminosäure großes Adapterprotein, das bevorzugt in Lymphozyten produziert wird. SLY1 spielt eine Rolle in der adaptiven Immunantwort des Körpers und ist von Bedeutung für die Reifung und die korrekte Funktion von B- und T-Lymphozyten und NK-Zellen. Es enthält ein N-terminales zweigeteiltes Kernlokalisierungssignal (*nuclear localisation signal*; NLS), eine *Src homology 3-* (SH3) und eine *sterile alpha motif-* Domäne (SAM). SH3- und SAM-Domänen sind Interaktionsdomänen, die die Komplexbildung mit anderen Biomakromolekülen ermöglichen. Experimentelle Daten weisen darauf hin, dass SLY1 mit Proteinen der 14-3-3-Familie und verschiedenen ribosomalen Proteinen interagiert. Allerdings sind die Rollen und die biologischen Zielbereiche der SH3- und SAM-Domänen noch ungeklärt. SAM-Domänen sind bekanntermaßen in der Lage, Proteine, Nucleinsäuren und Lipide zu binden. Außerdem haben viele SAM-Domänen die Eigenschaft mit sich selbst oder weiteren SAM-Domänen Homo- und Hetero-Oligomere zu bilden. Aufgrund dieser Besonderheiten von SAM-Domänen ist es schwierig, die Rolle, die eine SAM-Domäne im Kontext eines spezifischen Proteins spielt, ohne Kenntnis der Struktur und des Oligomerisationszustands der SAM Domäne zu beurteilen.

In dieser Arbeit wurde die dreidimensionale Struktur und der Oligomerisationszustand der SAM-Domäne von SLY1 mit biophysikalischen Methoden unter Verwendung eines SLY1-Konstruktes (P254–E321) bestimmt, welches die gesamte gefaltete SAM-Domäne (P254–Y316) enthielt. Mit Hilfe von Kernmagnetresonanz-Spektroskopie (*nuclear magnetic resonance*; NMR) und analytischen Ultrazentrifugations-Experimenten konnte gezeigt werden, dass SLY1 SAM sich in einem Monomer-Dimer-Gleichgewicht befindet, welches durch eine Gleichgewichts-Dissoziationskonstante im niedrigen mikromolaren Bereich charakterisiert ist. Die Verlängerung des N-terminus des SLY1-Konstruktes um eine Region positiv geladener Aminosäuren der flankierenden SLY1-Sequenz resultiert in einer Erhöhung der Bindungsaffinität der Dimerisierung um das Sechs- bis Achtfache. Damit ist das SLY1 SAM-Homodimer um eine Größenordnung stabiler als alle anderen bisher beschriebenen SAM-Homodimere. Chemischer Austausch auf Grund des Monomer-Dimer-Gleichgewichtes führte zu starken Linienverbreiterungen der Signale in den anfänglichen NMR-Spektren. Daher konnte zunächst keine Strukturaufklärung des SAM-Dimers mittels NMR-

Spektroskopie erfolgen. Der Austausch von S320 am flexiblen C-terminus des Konstrukts gegen ein Cystein ermöglichte es, die beiden SAM-Untereinheiten des Dimers zu verbrücken. Die Struktur des verbrückten SLY1 SAM-Dimers wurde mittels multidimensionaler, heteronuklearer NMR-Spektroskopie bestimmt.

Die Lösungsstruktur des SAM-Monomers von SLY1 weist die für SAM-Domänen charakteristische Fünf-Helix-Faltung ($\alpha 1$ – $\alpha 5$) auf. Interessanterweise bildet die SAM-Domäne von SLY1 ein Homodimer unter Verwendung einer für SAM-Domänen bisher noch nicht beschriebenen Interaktionsfläche. Diese Grenzfläche wird hauptsächlich durch die beiden langen C-terminalen Helices $\alpha 5$ und $\alpha 5'$ der Domänen gebildet, welche gegeneinander gepackt sind. Zusätzlich wird das Dimer durch intermolekulare Kontakte zwischen Aminosäureresten in Helices $\alpha 5$ und $\alpha 1'$ und der $\alpha 1'/\alpha 2'$ -Schleife sowie zwischen Aminosäureresten der beiden C-Termini stabilisiert. Die Lösungsstruktur des SLY1 SAM-Homodimers wird durch die Röntgenkristallstruktur der unverbrückten Wildtyp SLY1 SAM-Domäne bekräftigt, welche mit einer maximalen Auflösung von 2,05 Å bestimmt wurde. Der Vergleich der beiden Dimer-Strukturen mittels Superpositionierung ergab eine mittlere quadratische Abweichung der C ^{α} -Positionen von 1,13 Å, was auf eine sehr hohe Ähnlichkeit der globalen Faltung beider SAM-Dimere hinweist.

In der hier bestimmten Struktur des SLY1 SAM-Dimers präsentieren beide SAM-Untereinheiten ein für SAM-Domänen typisches Interaktionsmotiv auf der zugänglichen Oberfläche. Das SAM-Dimer könnte mit Hilfe dieser Interaktionsmotive ein Gerüst für SLY1 vermittelte Proteinwechselwirkungen mit anderen SAM-Domänen-enthaltenden Proteinen bilden.

Contents

| | |
|--|------------|
| Abstract | I |
| Zusammenfassung | III |
| Contents | V |
| Tables | IX |
| Figures | X |
| Abbreviations | XII |
| 1 Introduction | 1 |
| 1.1 Adaptor proteins | 1 |
| 1.2 The adaptor protein SLY1 | 1 |
| 1.3 SAM domains | 3 |
| 1.4 SAM domain of SLY1 | 9 |
| 1.5 Nuclear magnetic resonance spectroscopy | 9 |
| 1.5.1 Nuclear spins and nuclear spin angular momentum | 10 |
| 1.5.2 Basic principles of NMR experiments | 11 |
| 1.5.3 Multidimensional NMR spectroscopy | 13 |
| 1.5.4 Chemical shift | 15 |
| 1.5.5 Scalar couplings | 16 |
| 1.5.6 Nuclear Overhauser effect | 18 |
| 1.5.7 Dynamic processes and chemical exchange | 20 |
| 1.6 Structure calculation from NMR-derived experimental restraints | 22 |
| 1.7 Structure determination of symmetric homodimers by NMR spectroscopy | 24 |
| 1.8 Aims | 24 |
| 2 Materials | 27 |
| 2.1 Equipment | 27 |
| 2.2 Chemicals, enzymes and media components | 28 |
| 2.3 Columns and resins | 29 |
| 2.4 Kits | 29 |
| 2.5 Commercially available crystallisation screens | 30 |
| 2.6 Bacterial strains | 30 |
| 2.7 Plasmids | 31 |
| 2.8 Oligonucleotides | 31 |
| 2.9 Software and databases | 32 |
| 3 Methods | 35 |
| 3.1 Microbiological methods | 35 |
| 3.1.1 Bacterial growth media | 35 |
| 3.1.2 Cultivation of bacteria | 36 |
| 3.1.3 Preparation of chemically competent <i>E. coli</i> | 37 |
| 3.1.4 Transformation of chemically competent <i>E. coli</i> with plasmid DNA | 38 |

| | |
|--|-----------|
| 3.1.5 Storage of bacterial cells | 38 |
| 3.2 Genetic methods | 38 |
| 3.2.1 Preparation of plasmid-DNA..... | 38 |
| 3.2.2 Agarose gel electrophoresis..... | 39 |
| 3.2.3 DNA extraction from agarose gels | 39 |
| 3.3 Polymerase Chain Reaction | 39 |
| 3.3.1 Site-specific mutagenesis by QuikChange-PCR | 40 |
| 3.3.2 Molecular cloning | 41 |
| 3.3.3 DNA sequencing | 42 |
| 3.4 Preparation of protein samples | 42 |
| 3.4.1 Purification of SLY1 SAM variants from the cell extract | 42 |
| 3.4.2 Preparation of the fluorescent labelled SLY1 SAM domain | 45 |
| 3.4.3 Buffer exchange by dialysis | 46 |
| 3.4.4 Raising the sample concentration by ultrafiltration..... | 46 |
| 3.5 Analytical methods | 47 |
| 3.5.1 SDS-polyacrylamide gel electrophoresis..... | 47 |
| 3.5.2 Analytical ultracentrifugation..... | 48 |
| 3.5.3 Quantitation of analyte concentrations in solution by UV/Vis spectrophotometry | 49 |
| 3.5.4 Quantitation of the protein concentration using the Bradford method | 49 |
| 3.5.5 Microscale thermophoresis..... | 50 |
| 3.6 Nuclear magnetic resonance spectroscopy | 52 |
| 3.6.1 NMR sample preparation | 54 |
| 3.6.2 Assignment of resonances | 54 |
| 3.6.3 Chemical shift perturbation analysis | 56 |
| 3.6.4 Determination of torsion angle restraints | 56 |
| 3.6.5 Hydrogen bond coupling experiments..... | 59 |
| 3.6.6 NOE-derived distance restraints..... | 59 |
| 3.6.7 Surface mapping using paramagnetic relaxation enhancers | 60 |
| 3.6.8 Semiautomatic NOE resonance assignment and structure calculation | 60 |
| 3.6.9 Structure calculation using restrained molecular torsion angle dynamics simulations..... | 64 |
| 3.6.10 Analysis and evaluation of the NMR derived structure ensembles | 65 |
| 3.7 X-ray crystallography | 66 |
| 3.7.1 Sample preparation and crystallisation..... | 66 |
| 3.7.2 Data collection..... | 67 |
| 3.7.3 Data processing | 67 |
| 3.7.4 Model building and refinement | 68 |
| 3.7.5 Analysis and evaluation of the X-ray structure | 69 |
| 4 Results | 71 |

| | | |
|----------|--|------------|
| 4.1 | Molecular cloning of long and short SAM constructs | 71 |
| 4.1.1 | Molecular cloning of pGEX-6p-2_SLY1SAM _{Ig_C} | 72 |
| 4.1.2 | Molecular cloning of pGEX-6p-2_SLY1SAM _{wt} and pGEX-6p-2_SLY1SAM _C | 74 |
| 4.2 | SAM domain production and purification | 74 |
| 4.3 | Characterisation of the SAM _{Ig} dimerisation behaviour | 76 |
| 4.3.1 | 2D ¹ H, ¹⁵ N-HSQC spectra of SAM _{Ig} show concentration-dependent line broadening due to chemical exchange | 76 |
| 4.3.2 | Determination of the equilibrium dissociation constant of the SAM _{Ig} dimerisation by MST .. | 79 |
| 4.4 | Characterisation of the SAM _{Ig_C} single-cysteine mutant | 81 |
| 4.5 | SLY1 SAM _{wt} and SAM _C | 84 |
| 4.5.1 | Analytical ultracentrifugation verifies SAM _{wt} dimerisation | 84 |
| 4.5.2 | Determination of the equilibrium dissociation constant of SAM _{wt} dimerisation by MST | 85 |
| 4.5.3 | 2D ¹ H, ¹⁵ N-HSQC spectra of SAM _{wt} and SAM _C | 87 |
| 4.5.4 | Sequential assignment of backbone resonances of cross-linked SAM _C | 88 |
| 4.5.5 | Sequential assignment of backbone resonances of SAM _{wt} and non-cross-linked SAM _C | 90 |
| 4.5.6 | Differences in the ¹ H ^N and ¹⁵ N chemical shifts between SAM _{wt} and SAM _C | 92 |
| 4.5.7 | Assignment of side chain resonances of cross-linked SAM _C | 92 |
| 4.6 | Determination of structural restraints from NMR data for SLY1 SAM _C structure calculation | 94 |
| 4.6.1 | Secondary structure determination from chemical shift data using TALOS+ and hydrogen bond couplings | 94 |
| 4.6.2 | χ_1 -angle restraints determined from scalar couplings | 96 |
| 4.6.3 | Distance restraints determined from NOE data | 98 |
| 4.6.4 | Characterisation of the dimer interface by isotope-filtered/edited NOESY spectra and solvent PREs | 99 |
| 4.7 | NMR structure calculation | 102 |
| 4.7.1 | Analysis of the NMR derived structure ensemble | 102 |
| 4.7.2 | Solution structure of the SLY1 SAM _C cross-linked dimer | 107 |
| 4.7.3 | Interface of the cross-linked SAM _C dimer | 108 |
| 4.8 | SAM _{wt} structure determination by X-ray crystallography | 111 |
| 4.8.1 | Screening of crystallisation conditions for SLY1 SAM _{wt} | 111 |
| 4.8.2 | Phasing, model building and refinement | 113 |
| 4.8.3 | SAM _{wt} structure | 113 |
| 4.8.4 | Crystallographic SAM _{wt} dimer | 116 |
| 5 | Discussion | 121 |
| 5.1 | SLY1 SAM _{wt} forms symmetric homodimers | 121 |
| 5.2 | Cross-linking of the wild-type SAM domain does not affect the overall protein fold | 121 |
| 5.3 | Quality of the NMR-derived SAM _C dimer structure ensemble | 129 |
| 5.4 | Crystal and solution structure of the SAM domain are very similar | 132 |
| 5.5 | SAM _{wt} domain forms a novel type of symmetric dimer for SAM domains | 134 |

| | |
|---|--------------|
| 5.6 Importance of the extended N-terminus of the SAM domain in dimer formation..... | 139 |
| 5.7 Biological implications of the SLY1 SAM dimer formation..... | 140 |
| 6 Conclusion and Outlook | 143 |
| 7 References | 145 |
| 8 Appendix | 161 |
| 8.1 Xml script for NOE resonance assignment and structure calculation by ARIA and CNS..... | 161 |
| 8.2 ^{15}N , ^{13}C - ω_1 -filtered NOESY- ^1H , ^{15}N -HSQC NOE cross-peak assignments | 164 |
| 8.3 ^{15}N , ^{13}C - ω_1 -filtered NOESY- ^1H , ^{13}C -HSQC NOE cross-peak assignments | 166 |
| Publications and presentations | XV |
| Publications | XV |
| Oral Presentations..... | XV |
| Danksagung | XVI |
| Eidesstattliche Erklärung..... | XVIII |

Tables

| | |
|--|-----|
| Table 1: List of the instruments used in this work | 27 |
| Table 2: List of specialised chemicals used | 28 |
| Table 3: List of enzymes used | 29 |
| Table 4: List of columns and resins used for preparative and analytical column chromatography | 29 |
| Table 5: List of commercially available kits used | 29 |
| Table 6: List of all commercially available kits for sparse matrix screening of crystallisation conditions used for SLY1 SAM _{wt} crystallisation | 30 |
| Table 7: List of bacterial strains used for molecular cloning and recombinant gene expression | 30 |
| Table 8: Plasmids used for recombinant gene expression..... | 31 |
| Table 9: Primers used for standard PCR reactions and sequencing | 32 |
| Table 10: Primers used for site-directed mutagenesis by "QuikChange"-PCR..... | 32 |
| Table 11: Software and databases used for various applications | 32 |
| Table 12: Composition of bacterial growth media for recombinant gene expression | 35 |
| Table 13: Composition of TS2 solution and vitamin solution used in M9-medium preparation | 36 |
| Table 14: Compositions of TAE-buffer and 5× concentrated loading dye | 39 |
| Table 15: PCR thermocycling program for nucleic acid amplification | 40 |
| Table 16: Composition of the 25 μL reaction mix for Site-specific mutagenesis by QuikChange..... | 40 |
| Table 17: The PCR thermocycling program for site-specific mutagenesis..... | 41 |
| Table 18: Buffers used for the purification of the various SLY1 SAM constructs | 43 |
| Table 19: Buffers used as eluent for preparative scale size exclusion chromatography on SLY1 SAM ... | 45 |
| Table 20: Composition of the stacking and the separation gels in SDS-PAGE | 47 |
| Table 21: Compositions of the buffers required for an SDS-PAGE | 48 |
| Table 22: Selected acquisition parameters of the NMR experiments conducted for the SLY1 SAM domain resonance assignment and structure determination | 53 |
| Table 23: Composition of NMR-buffer used for NMR spectroscopy on the SLY1 SAM domain..... | 54 |
| Table 24: Empiric relationship between the rotameric state of the C ^α -C ^β bond and the strength of the ³ J _{HaHβ2/3- and the ³J_{NHβ2/3-}-couplings (Case <i>et al.</i>, 1994).....} | 59 |
| Table 25: Mixing times used during the acquisition of the five 3D NOESY-HSQC experiments | 60 |
| Table 26: Adjustable parameters in ARIA 2.3 for the assignment, conversion and use of the NOE data in the recorded NOESY spectra..... | 62 |
| Table 27: Symmetry restraints parameter setting for a homodimer with C ₂ symmetry | 63 |
| Table 28: Parameters of the CNS-based simulated annealing protocol used for the SAM _C dimer structure calculation | 64 |
| Table 29: Assignment and structural statistics for the SLY1 SAM _C structure ensemble..... | 106 |
| Table 30: Hydrogen bonds and salt bridges across the SAM _C dimer interface predicted by PISA in the NMR structure..... | 111 |
| Table 31: Crystallisation conditions that resulted in SAM _{wt} crystal formation..... | 112 |
| Table 32: Data collection and refinement statistics. | 115 |

| | |
|--|-----|
| Table 33: Hydrogen bonds and salt bridges across the SAM _{wt} dimer interface predicted by PISA based on x-ray data | 119 |
|--|-----|

Figures

| | |
|---|----|
| Fig. 1: Schematic depiction of the murine SLY1 domain organisation. | 2 |
| Fig. 2: Crystal structure of the SAM domain of the ephrin type-A receptor 4 (EphA4) in ribbon representation..... | 4 |
| Fig. 3: Mid-loop–end-helix interaction between PHC3 SAM domains. | 6 |
| Fig. 4: Interfaces described for SAM-SAM domain interactions..... | 7 |
| Fig. 5: Fourier transformation of the FID of a single spin into a function of frequency (ν)..... | 12 |
| Fig. 6: Schematic depiction of standard 2D and 3D NMR experiments | 14 |
| Fig. 7: 1J - and 2J -coupling constants in peptides and proteins (Sattler et al., 1999). | 16 |
| Fig. 8: Backbone and side chain torsion angles of a glutamate-leucine dipeptide. | 17 |
| Fig. 9: Schematic depiction of the NOE transfers during 3D ^{15}N , ^{13}C - ω_1 filtered, ^{13}C -edited and ^{15}N -edited NOESY-HSQC experiments. | 19 |
| Fig. 10: Simulated 1D NMR spectra of nuclei undergoing chemical exchange between two states (A and B) at different exchange rates (k_{ex}) (reproduced with permission from Keeler, 2010). | 20 |
| Fig. 11: Schematic depiction of cross-peaks representing nuclei undergoing fast exchange. | 21 |
| Fig. 12: Flowchart of the structure determination process of biomolecules by NMR spectroscopy. | 23 |
| Fig. 13: Schematic representation of the SLY1 SAM domain constructs. | 72 |
| Fig. 14: Agarose gel of wt and mutant plasmid DNA digested with <i>BtsI</i> | 73 |
| Fig. 15: Sequencing results of pGEX-6p-2_SLY1SAM_Ig mutants. | 73 |
| Fig. 16: SDS-PAGE (15 %) gel slices containing protein samples taken during the SLY1 SAM purification. | 75 |
| Fig. 17: Superposition of 2D ^1H , ^{15}N -HSQC spectra of SAM _{Ig} at three concentrations. | 77 |
| Fig. 18: Weighted average ^1H and ^{15}N chemical shift differences, $\Delta\delta_{\text{ave}}$, between resonances recorded in 2D ^1H , ^{15}N -HSQC spectra of SAM _{Ig} at 500 and 5 μM | 78 |
| Fig. 19: Fluorescence measured in each capillary without the application of any laser power..... | 80 |
| Fig. 20: Binding isotherms from MST experiments conducted with SLY1 SAM _{Ig} | 81 |
| Fig. 21: 2D ^1H , ^{15}N -HSQC spectra of cross-linked SAM _{Ig_C} and non-cross-linked dimer..... | 83 |
| Fig. 22: Weighted average chemical shift perturbation analysis of the backbone amide resonances in 2D ^1H , ^{15}N -HSQC spectra of SAM _{Ig} and SAM _{Ig_C} under oxidising and reducing conditions. | 84 |
| Fig. 23: Sedimentation equilibrium experiments with SLY1 SAM _{wt} to specify oligomeric state and oligomerisation affinity. | 85 |
| Fig. 24: Fluorescence of samples in the absence of heating (capillary scan) | 86 |
| Fig. 25: MST-based isotherm reflecting the monomer-dimer equilibrium | 87 |
| Fig. 26: 2D ^1H , ^{15}N -HSQC spectra of SAM _{wt} and SAM _C at varying concentrations demonstrate the effect of disulfide cross-linking on SAM domain self-association. | 88 |

| | |
|--|-----|
| Fig. 27: Sequential assignment of resonances arising from backbone nuclei of residues K285-E294 of cross-linked SAM _C | 89 |
| Fig. 28: Fully assigned 2D ¹ H, ¹⁵ N-HSQC spectra of cross-linked and non-cross-linked SAM _C | 91 |
| Fig. 29: Comparison of the 2D ¹ H, ¹⁵ N-HSQC of cross-linked and non-cross-linked SAM _C with the 2D ¹ H, ¹⁵ N-HSQC of SAM _{wt} | 93 |
| Fig. 30: Weighted average chemical shift ($\Delta\delta_{ave}$) perturbation analysis between cross-linked SAM _C and SAM _{wt} (red), and non-cross-linked SAM _C and SAM _{wt} (blue)..... | 94 |
| Fig. 31: Secondary structure restraints derived from TALOS+ torsion angle prediction and ^{h3} J _{NC} -couplings. | 95 |
| Fig. 32: ³ J _{HαHβ2/3} - and ³ J _{NHβ2/3} -couplings for the determination of <i>trans</i> , <i>gauche</i> (+) and <i>gauche</i> (-) rotamer states..... | 97 |
| Fig. 33: Observed intra-residue (grey), sequential (red), medium range (green) and long range (blue) NOEs per SAM _C residue. | 98 |
| Fig. 34: Characterisation of the SLY1 SAM _C dimer interface by 3D ¹⁵ N, ¹³ C- ω_1 -filtered, ¹³ C-edited and ¹⁵ N-edited NOESY-HSQC spectra and solvent PRE analysis..... | 100 |
| Fig. 35: Superposition of the 15 lowest energy structures calculated from NMR data..... | 102 |
| Fig. 36: RMSD from the mean atom position of the backbone (A) and all (B) heavy atoms per residue. | 103 |
| Fig. 37: Ramachandran plot of the backbone torsion angles of the SAM _C 15 lowest energy structures. | 105 |
| Fig. 38: Structure of SLY1 SAM _C cross-linked dimer as determined from the NMR derived restraints and calculated by ARIA. | 107 |
| Fig. 39: Residues at the interface of the SLY SAM _C dimer structure..... | 108 |
| Fig. 40: Dimer interface of the SLY1 SAM _C cross-linked dimer in ribbon and surface representation... .. | 109 |
| Fig. 41: Non-covalent interactions stabilising the dimer interface of SAM _C cross-linked dimer..... | 110 |
| Fig. 42: SAM _{wt} crystal and the respective 1° diffraction image..... | 112 |
| Fig. 43: Structure of SLY1 SAM _{wt} determined from X-ray crystallography data. | 114 |
| Fig. 44: Dimer of SLY1 SAM _{wt} predicted by PISA..... | 116 |
| Fig. 45: Dimer interface of the SLY1 SAM _{wt} dimer in ribbon and surface representation. | 117 |
| Fig. 46: Non-covalent interactions stabilizing the dimer interface of SAM _{wt} | 118 |
| Fig. 47: Chemical shift variations of amide resonances of selected residues in a SAM _{wt} concentration series..... | 124 |
| Fig. 48: Flexibility of SAM _C and SAM _{wt} per residue as indicated by the RCI and B-factors. | 126 |
| Fig. 49: 2D Strips of ¹⁵ N, ¹³ C-filtered, ¹³ C-edited NOESY spectra of SAM _{Ig} and SAM _C | 128 |
| Fig. 50: Superposition of the X-ray crystallography and NMR derived SLY1 SAM structures..... | 134 |
| Fig. 51: ML-EH interfaces observed in homooligomerisation. | 136 |
| Fig. 52: Ribbon representation of different structures of symmetric SAM domain homodimers. | 138 |

Abbreviations

| | |
|------------------------|--|
| 1D | One-dimensional |
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| Amp | Ampicillin |
| ANKS | Ankyrin repeat and SAM domain containing protein |
| APS | Ammonium peroxodisulfate |
| ARIA | Ambiguous restraints for iterative assignments |
| AUC | analytical ultracentrifugation |
| BMRB | Biological magnetic resonance data bank |
| BLAST | Basic local alignment search tool |
| BSA | Buried surface area |
| CDS | Coding sequence |
| CNK | Connector enhancer of KSR |
| CNS | Crystallography & NMR System |
| COSY | Correlation spectroscopy |
| CCP4 | Collaborative computational project number 4 |
| CcpNmr | Collaborative computational project for NMR |
| CT | constant time |
| CV | Column volume |
| DGK δ 1 | Diacylglycerol kinase δ 1 |
| DNA | Deoxyribonucleic acid |
| DNAse | Deoxyribonuclease |
| dNTP | Deoxyribonucleotide triphosphate |
| DOL | Degree of labelling |
| <i>D. melanogaster</i> | <i>Drosophila melanogaster</i> |
| DSS | 2,2-dimethyl-2-silapentane-5-sulfonic acid |
| DSSP | Define secondary structure of proteins |
| DTT | Dithiothreitol |
| ϵ | Molar extinction coefficient |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | Ethylenediaminetetraacetic acid |
| EH | End-helix |
| EphA | Ephrin type-A receptor |
| EphB | Ephrin type-B receptor |
| <i>et al.</i> | <i>et alii</i> |
| <i>g</i> | gravitational acceleration |
| GSH | Glutathione |
| GST | Glutathione-S-transferase |

| | |
|-------------|--|
| HLH | Helix-loop-helix |
| HSQC | heteronuclear single quantum coherence |
| HYP | Hyphen |
| Hz | Hertz |
| IPTG | Isopropyl β -D-1-thiogalactopyranoside |
| K_D | equilibrium dissociation constant |
| k_{ex} | exchange rate |
| KSR-1 | Kinase suppressor of Ras-1 |
| L | litre |
| LB | Lysogeny broth |
| LHS | Left-handed staple |
| MAPKKK | Mitogen-activated protein kinase kinase kinase |
| ML | Mid-loop |
| MST | Microscale thermophoresis |
| MWCO | Molecular weight cut-off |
| NCR | N-terminal conserved regions |
| NK cell | Natural killer cell |
| NLS | Nuclear localisation signal |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NOESY | Nuclear Overhauser enhancement spectroscopy |
| OD | Optical density |
| PAGE | polyacrylamide gel electrophoresis |
| <i>p.a.</i> | <i>pro analysi</i> |
| pI | Isoelectric point |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| PHC3 | Polyhomeotic homologue 3 |
| RHS | Right-handed staple |
| RNA | Ribonucleic acid |
| rMD | Restrained molecular dynamics |
| RMSD | Root-mean-square-deviation |
| RPF | Recall, Precision, F-measure |
| rpm | Revolutions per minute |
| RT | Room temperature |
| RTK | Receptor tyrosine kinase |
| SA | Simulated annealing |
| SAM | Sterile alpha motif |

| | |
|----------|--|
| SAMSN1 | SAM domain, SH3 domain and nuclear localization signals protein 1 |
| SASH1 | SAM and SH3 domain containing protein 1 |
| SDS | Sodium dodecyl sulfate |
| SEC | size exclusion chromatography |
| SEP | Yeast sterility, Ets-related, PcG proteins |
| SLY1 | SH3 domain containing protein expressed in lymphocytes |
| SPM | Scm, Ph, lethal-3 malignant brain tumour |
| SSM | Secondary structure matching |
| STIM1 | Stromal interaction molecule-1 |
| σ | Standard deviation |
| TAE | Tris-Acetate-EDTA |
| TALOS | Torsion angle likelihood obtained from shift and sequence similarity |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TEMED | Tetramethylethylenediamine |
| TNKS | Tankyrase |
| TOCSY | Total correlation spectroscopy |
| Tris | tris(hydroxymethyl)aminomethane |
| T | Temperature |
| t | Time |
| UV | ultraviolet |
| wt | wild-type |

1 Introduction

1.1 Adaptor proteins

Signal transduction is crucial for many different cellular processes from proliferation to the programmed cell death. In many cases, interacting protein kinases and phosphatases are involved in signal transduction as well as the regulation and amplification of the signals. In the highly complex signalling networks, adaptor and scaffolding proteins play a major role in mediating the required protein-protein interactions.

Adaptor proteins are characterised by their lack of enzymatic activity and by the presence of often multiple protein interaction domains (Jordan *et al.*, 2003). Among the best-characterised types of interaction domains in cell signalling are the SH2 (Src homology 2) and SH3 (Src homology 3) domains. SH2 and SH3 domains bind to phosphorylated tyrosine residues (Liu & Nash, 2012) and proline-rich amino acid sequences (Teyra *et al.*, 2017; Verschueren *et al.*, 2015), respectively. The sterile alpha motif (SAM) domain is another common interaction module that plays a role in the formation of protein complexes through various forms of homo- and heterooligomerisation (Section 1.4) (Qiao & Bowie, 2005).

Adaptor proteins are capable of binding multiple proteins through their interaction domains, thus supporting the assembly of transient protein complexes (Jordan *et al.*, 2003; Nussinov & Jang, 2014), which are essential for the correct spatiotemporal localisation of the effector molecules, e.g., enzymes. Thus, studying how adapter proteins mediate the assembly of protein complexes at the molecular level aids our understanding of cellular signal transduction networks.

1.2 The adaptor protein SLY1

The murine protein SLY1 (SH3 domain containing protein expressed in lymphocytes-1), also known as SAM and SH3 domain containing protein 3 (SASH3) (UniProtKB - Q8K352), is a putative signalling adaptor protein that is preferentially expressed by lymphocytes. SLY1 consists of 380 residues, with an N-terminal bipartite nuclear localisation signal (NLS), and an SH3 (aa 176–232) and a SAM domain (aa 254–316) as protein interaction motifs (Beer *et al.*, 2001) (Fig. 1). The function of SLY1 as an adaptor protein has been deduced from its domain organisation and the absence of a catalytic domain. Sequence comparison using BLAST (Basic local alignment search

tool) revealed a homologue protein of 94% amino acid sequence identity in the human genome (UniProtKB - O75995) (Beer *et al.*, 2001).

SLY1 is a member of the SLY family of adapter proteins. In addition to SLY1, the family also contains the proteins SAMSN1 (SAM domain, SH3 domain and nuclear localization signals protein 1; also referred to as HSC1, SLY2 and NASH1) (Claudio *et al.*, 2001) and SASH1 (SAM and SH3 domain containing protein 1) (Zeller *et al.*, 2003). All three family members share a very similar protein domain organisation consisting of the bipartite NLS, and SH3 and SAM domains as interaction motifs. SASH1, in contrast to SLY1 and SLY2, comprises one additional SAM domain at its C-terminal end and a coiled-coil domain at the N-terminus, making it the largest member of the SLY family.



Fig. 1: Schematic depiction of the murine SLY1 domain organisation.

SLY1 contains an N-terminal bipartite nuclear localisation signal (NLS), a Src homology 3 (SH3) domain and a C-terminal sterile alpha motif (SAM) domain.

SLY1 was initially identified during an adhesion screen of a murine T-cell lymphoma cDNA library, although involvement of SLY1 in adhesion could not be verified (Beer *et al.*, 2001). Another group independently identified SLY1 as a novel substrate of serine kinases during antigen receptor signalling in T-cells (Astoul *et al.*, 2003).

Studies to understand SLY1 function were conducted using SLY1 deficient mice (SLY1^{-/-}) and mice carrying a truncated version of the *sly1* gene (SLY1^{Δ/Δ}). The truncated SLY1 lacks residues 20–100, thereby eliminating one half of the bipartite NLS. Both mouse strains showed a severe reduction in cell numbers of lymphatic organs and a prolonged allograft survival rate after transplantation of semi-allogenic organs. SLY1 appears to be involved in T-cell maturation, leading to a reduced number of thymocytes in SLY1 deficient mice and a reduced antigen receptor-mediated proliferation rate (Beer *et al.*, 2005; Reis *et al.*, 2009). Furthermore, the number of marginal zone B cells was reduced significantly in the absence of a functional SLY1 protein, resulting in defects in the humoral immune responses to both T-cell-dependent

and -independent antigens (Beer *et al.*, 2005; Scheikl *et al.*, 2009). Since SLY1^{-/-} mice display a highly similar phenotype to mice producing the truncated version of SLY1 (SLY1^{ΔΔ}), it was hypothesised that the SLY1 function in B- and T-lymphocytes is strongly dependent on its intact NLS, and therefore, its ability to be imported into the nucleus. SLY1 function in natural killer (NK) cells differs from its role in B- and T-cells. In NK cells, SLY1 is exclusively located in the cytoplasm and appears to contribute to ribosomal stability (Arefanian *et al.*, 2016). In oligodendrocytes, SAM domain deficient SLY1 has been implicated in the formation of oligodendrogliomas, although its role in that type of cell is unknown (Erdem-Eraslan *et al.*, 2015).

Considering its diverse functions and its domain architecture, it is likely that SLY1 is capable of interacting with multiple interaction partners. Currently, SLY1 has been shown to interact with 14-3-3 proteins upon phosphorylation of serine 27, which leads to the export of SLY1 from the nucleus and its retention in the cytoplasm (Schäll *et al.*, 2015). Furthermore, both, the SH3 and SAM domain, are capable of binding ribosomal proteins (Arefanian *et al.*, 2016). In addition, SLY1 appears to be capable of forming dimers through its SH3 domain (Brandt, 2010), although the biological implication of the dimerisation has yet to be determined. Dimerisation in general is often an essential means for the regulation of protein function (Marianayagam *et al.*, 2004).

The function of adapter proteins is inevitably connected to the biophysical properties of their interaction motifs. Knowledge of the structure, possible interaction surfaces as well as the oligomerisation state of domains such as SH3 and SAM could help shed light on the role of SLY1 in the complex protein interaction network required for cell signalling at the molecular level.

1.3 SAM domains

Sterile alpha motif (SAM) domains are protein interaction modules of ~70 amino acids in size (Kim & Bowie, 2003; Ponting, 1995; Qiao & Bowie, 2005). They were originally also referred to as pointed domains (Klämbt, 1993; Slupsky *et al.*, 1998), SPM (Scm, Ph, lethal-3 malignant brain tumour) domains (Bornemann *et al.*, 1996; Peterson *et al.*, 1997), SEP (yeast sterility, Ets-related, PcG proteins) domains (Alkema *et al.*, 1997), NCR (N-terminal conserved region) and HLH (helix-loop-helix) domains (DeCamillis *et al.*, 1992; Nomura *et al.*, 1994). Currently, SAM domains have been identified in most eukaryotes, including mammals, and bacteria (PFAM database; Finn

et al., 2016). Most recently, SAM domains have also been identified in plant proteins (Denay *et al.*, 2017), beginning with the characterisation of the SAM domain of transcription factor LEAFY (Sayou *et al.*, 2016).

Solved SAM domain structures show that these domains are structurally very similar. The majority adopts a canonical tertiary fold of four shorter helices and one long C-terminal helix, with the shorter helices 1 and 2 as well as 3 and 4 arranging themselves in an antiparallel manner perpendicular to helix 5 (Fig. 2). For the SAM domains of the proteins DLC1 (deleted in liver cancer 1) and DLC2 (deleted in liver cancer 2), four-helix folds have been reported (Li *et al.*, 2007; Qiao & Bowie, 2005; Zhong *et al.*, 2009).

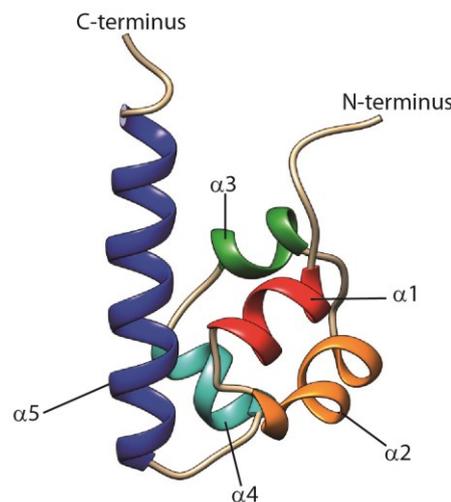


Fig. 2: Crystal structure of the SAM domain of the ephrin type-A receptor 4 (EphA4) in ribbon representation.

The SAM domain of EphA4 (PDB ID: 1BOX) adopts the canonical five-helix fold described for the majority of SAM domains with the shorter helices 1 (red) and 2 (orange) as well as 3 (green) and 4 (cyan) arranging themselves in an antiparallel manner perpendicular to helix 5 (blue).

SAM domains have been identified in various proteins with different functionalities like receptor tyrosine kinases (RTK), serine/threonine kinases, kinase interacting proteins, RNA binding proteins, scaffolding proteins and transcription factors (Denay *et al.*, 2017; Qiao & Bowie, 2005). Although SAM domains are structurally very similar, they have been shown to mediate interactions with a variety of different interaction partners, such as other proteins (Wei *et al.*, 2011; Wu *et al.*, 2015), as well as lipids (Li *et al.*, 2007; Sathyamurthy *et al.*, 2011) and RNA (Green *et al.*, 2003; Johnson & Donaldson, 2006). Some SAM domains function by interacting with other domains that belong to the same protein they are located on: The SAM domains of the proteins STIM1 (stromal

interaction molecule-1) and KSR-1 (kinase suppressor of Ras-1) have been found to interact closely with the EF-hand and the CC-domain, respectively, to which they are close in sequence and on which they exert a stabilising effect (Koveal *et al.*, 2012; Stathopoulos *et al.*, 2008). Nonetheless, most SAM domains mediate protein-protein interactions by engaging with other SAM domains. The isolated SAM domains of the human proteins ANKS3 (Ankyrin repeat and SAM domain containing protein 3), DGK δ 1 (diacylglycerol kinase δ 1), TNKS1 (Tankyrase 1), TNKS2 (Tankyrase 2) and PHC3 (polyhomeotic homologue 3) form large, helical homopolymers (Harada *et al.*, 2008; Leettola *et al.*, 2014; Mariotti *et al.*, 2016; Riccio *et al.*, 2016; Robinson, Leal, Nanyes *et al.*, 2012). The polymerisation is achieved by consecutive head-to-tail interactions of multiple SAM domains. This head-to-tail interaction involves predominantly the N-terminal end of helix 5 and residues in loop α 1/ α 2 interacting with residues of the helices α 3 and loops α 2/ α 3 and α 3/ α 4 of the following SAM domain (Fig. 3). Based on the residues involved in that type of interaction, the interface is referred to as mid-loop–end-helix (ML-EH) interface.

The ML-EH interaction is the most common form of the published SAM-SAM interactions and has also been described in the formation of distinct heterodimers, e.g., the interactions of the isolated SAM domains of ANKS3 and ANKS6 (Leettola *et al.*, 2014), of CNK (connector enhancer of KSR) and HYP (Hyphen) (Rajakulendran *et al.*, 2008), and of the transcription factor Yan and its regulator Mae (modulator of the activity of ETS) (Baker *et al.*, 2001; Qiao *et al.*, 2004). The two interacting SAM domains can also be located on the same protein, as is the case for the tandem SAM domains of the scaffolding protein ANKS1B (also referred to as Amyloid-beta protein intracellular domain-associated protein 1; AIDA-1) (Kurabi *et al.*, 2009). The structures of a dimeric subunit of the PHC3 SAM homopolymer and the ANKS3-ANKS6 SAM domain heterodimer are shown in Fig. 4.

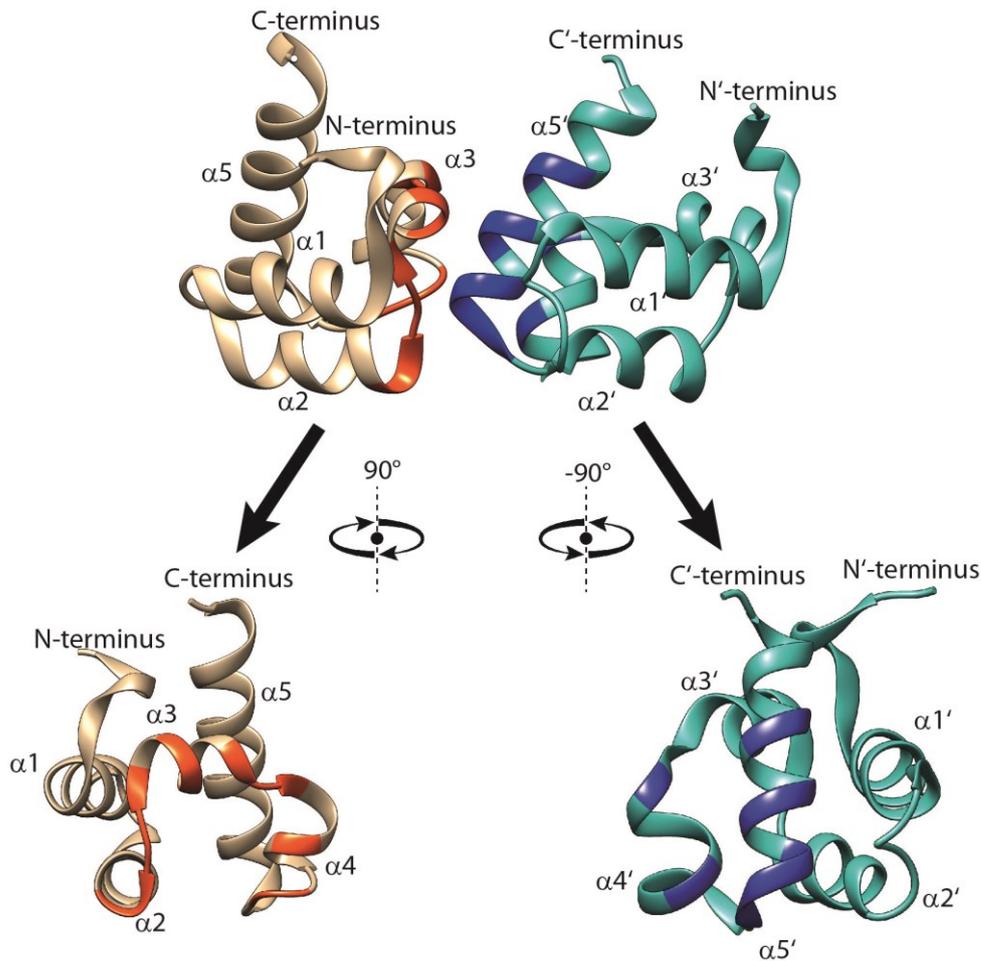


Fig. 3: Mid-loop–end-helix interaction between PHC3 SAM domains.

Shown above are two subunits A (tan) and A' (teal) of the PHC3 SAM domain homopolymer (PDB ID: 4PZO). The SAM domains of PHC3 oligomerise via the asymmetric ML-EH interface, which allows the SAM to form extended helical polymers. The N-terminal end of helix $\alpha 5$ and residues in loop $\alpha 1/\alpha 2$ of subunit A'(EH) interact with residues of the helices $\alpha 2$, $\alpha 3$ and $\alpha 4$ and loops $\alpha 2/\alpha 3$ and $\alpha 3/\alpha 4$ of subunit A (ML). The ML and EH interfaces are coloured on the depicted SAM domains in orange and blue, respectively.

Although SAM domain mediated interactions and SAM domain oligomerisation in particular has mostly been studied using isolated SAM domains, these interactions have been shown to be relevant for the function of multiple proteins, e.g., the SAM domain mediated formation of human TNKS1 and TNKS2 polymers, which is required for the induction of tankyrase dependent Wnt-signalling (Mariotti *et al.*, 2016; Riccio *et al.*, 2016). In the case of the *Drosophila melanogaster* transcription factor Yan, SAM domain mediated polymerisation is required for successful gene repression (Qiao *et al.*, 2004; Zhang *et al.*, 2010), whereas the SAM mediated interaction with its regulator Mae leads to depolymerisation and inactivation of Yan (Baker *et al.*, 2001; Qiao *et al.*, 2004). A functional role of SAM domain polymerisation has also been proposed for Polyhomeotic (Ph), a gene silencer of the Polycomb Group in *D. melanogaster*, and its

human homologue PHC2 (polyhomeotic homologue 2) (Isono *et al.*, 2013; Robinson, Leal, Chadwell *et al.*, 2012). In these cases, SAM domain polymerisation has been linked to the clustering of the polycomb-group repressive complex-1 (PRC1) and gene silencing.

Malfunctions of SAM domain-mediated interactions have been implicated in multiple diseases, like ankyloblepharon–ectodermal dysplasia–clefting syndrome (AEC) (Sathyamurthy *et al.*, 2011), polycystic kidney disease (PKD) (Leettola *et al.*, 2014), atherosclerosis (Meruelo & Bowie, 2009) and various types of cancer (Erdem-Eraslan *et al.*, 2015; Kim *et al.*, 2001; Mariotti *et al.*, 2016; Singh *et al.*, 2015). Furthermore, the SAM domain of the human restriction factor SAMHD1 (SAM and histidine/aspartate containing protein 1) is involved in viral protein x-mediated SAMHD1 degradation during human immunodeficiency virus (HIV) infection (Wu *et al.*, 2015).

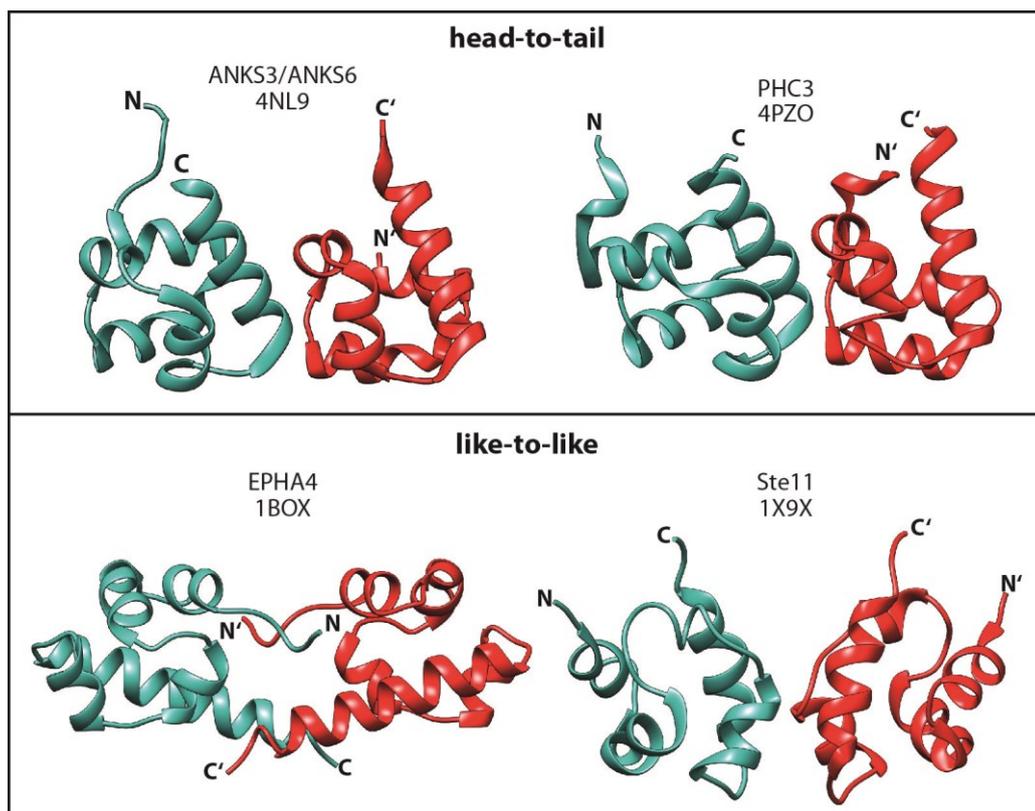


Fig. 4: Interfaces described for SAM-SAM domain interactions.

The different types of interfaces that have been reported for SAM-SAM interactions can be categorised into asymmetric head-to-tail and symmetric like-to-like interactions. The prevalent form of SAM domain homo- and heterooligomerisation is through the head-to-tail ML-EH interface, as shown for the ANKS3-ANKS6 SAM heterodimer and the PHC3 SAM homopolymer. Like-to-like interactions have been observed for the SAM domain homodimers of the human EphA4 and yeast protein Ste11.

Aside from the ML-EH interface, two further modes of interactions have been observed for two SAM domains that are capable of forming symmetric homodimers (Fig. 4). The SAM domain of the yeast mitogen-activated protein kinase kinase kinase (MAPKKK) Ste11 dimerises through a tail-to-tail type of interface, which consists of hydrophobic residues positioned in helix $\alpha 4$ and at the beginning of helix $\alpha 5$ (Bhattacharjya *et al.*, 2004). The homodimer of the SAM domain of the receptor tyrosine kinase EphA4 (ephrin type-A receptor 4) is formed by the two SAM domains facing each other with their N-termini and the C-terminal part of helix $\alpha 5$ extending from the domain core towards the dimer interface. The interface is further stabilised by residues in helices $\alpha 1$ and $\alpha 3$ (Stapleton *et al.*, 1999). However, the strength of the homodimerisation in both cases is very weak, with equilibrium dissociation constants (K_D) of 0.5 mM and 0.5–5 mM for the self-association of the SAM domains of Ste11 (Grimshaw *et al.*, 2004) and EphA4 (Stapleton *et al.*, 1999), respectively. As a consequence, the biological relevance of the SAM dimerisation for Ste11 and EphA4 protein function remains uncertain.

The variety of interaction partners and interaction interfaces in addition to low overall sequence similarity between SAM domains have made it challenging to predict SAM domain functions and modes of interaction based on primary sequences. Some effort to predict the oligomerisation state and the tendency of polymerisation of SAM domains based on the SAM domains primary sequences has been made with limited success (Meruelo & Bowie, 2009). Even in closely related proteins, the role of the respective SAM domains can differ quite dramatically, as can be seen for the SAM domains of EphA2 and EphA3, where the SAM domain has been shown to have a stabilizing effect on the dimer of unliganded EphA3 and a destabilizing effect on the dimerisation of unliganded EphA2 (Singh *et al.*, 2015, 2017). As a consequence, understanding the role of a SAM domain in the context of a specific protein requires thorough examination *in vivo* and *in vitro*. Biophysical and biochemical methods can be employed to determine the structure, oligomerisation state and interaction interfaces of SAM domains, which in turn may help to define the interactions mediated by the SAM domain and further the understanding of the functions of SAM domain containing proteins.

1.4 SAM domain of SLY1

The SLY1 SAM domain is a 62 amino acids (aa 254-316) long protein interaction domain located in the C-terminal region of SLY1 (Fig. 1). The primary sequence of the SAM domain of the murine SLY1 is identical to the SAM domain of its human homologue. Previous work shows that the SLY1 SAM domain self-associates and most likely forms homodimers (Thiagarajan, 2011).

1.5 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a method that relies on the physical property of nuclei to absorb and emit electromagnetic radiation when held under the influence of a constant magnetic field. The effect was first described by Rabi *et al.* in 1938 when measuring the effect of radio waves on a molecular beam of LiCl in a magnetic field. The first NMR experiments on liquids and solids were performed by Felix Bloch (Bloch *et al.*, 1946) and Edward Purcell (Purcell *et al.*, 1946) in the 1940s. Over the years, NMR spectroscopy as a technique has developed to allow more detailed studies on increasingly complex systems. Most notable in that context were the introduction of pulsed Fourier Transform NMR (FT-NMR) spectroscopy by Ernst & Anderson in 1966 and the conception of multidimensional NMR spectroscopy by Jean Jeener in 1971 (Jeener & Alewaeters, 2016).

Currently, NMR spectroscopy is commonly used to examine structures and dynamic processes of biological macromolecules. As with X-ray crystallography and more recently (with limitations) cryo-electron microscopy, NMR spectroscopy is capable of providing structural information of proteins at atomic resolution. In addition, and in contrast to the other methods, solution NMR spectroscopy can also provide data on molecular dynamics over timescales ranging from picoseconds to seconds. The applicability of solution NMR spectroscopy to the study of biological macromolecules was historically limited to proteins of ≤ 25 kDa in size. Larger biological macromolecules have unfavourable relaxation properties that result in line broadening and consequently in a noticeable loss of sensitivity, and increased spectral complexity (i.e., more signals as proteins size increases), which complicates the analysis of spectra. However, progress over the last two decades in methods development, sample preparation and hardware development have facilitated NMR studies of biological macromolecules of up to 900 kDa in size (Fiaux *et al.*, 2002).

1.5.1 Nuclear spins and nuclear spin angular momentum

The nuclear spin S is a quantum mechanical property of a nucleus that is the prerequisite for any NMR experiment. Many nuclei possess a spin. In the classical model, the spin describes the rotation of the nucleus around its axis. However, like any spin angular momenta of physical systems, the nuclear spin is quantised,

$$S = \sqrt{I(I + 1)} \hbar \quad [1]$$

where \hbar is the reduced Planck constant ($\hbar = h/2\pi$) and I the nuclear spin angular momentum quantum number. The nuclear spin angular momentum quantum number characterises the nuclear spin S and can adopt non-negative integer and half-integer numbers ($I = 0, 1/2, 1, 3/2, 2, \dots$). For protein NMR spectroscopy, the commonly used NMR-active nuclei are the spin- $1/2$ (i.e., $I = 1/2$) isotopes ^1H , ^{13}C , and ^{15}N . Incorporation of the non-abundant ^{13}C (1.07%) and ^{15}N (0.4%) isotopes into the target protein involves isotope labelling strategies in combination with biotechnological means for the production of biological macromolecules (Muchmore *et al.*, 1989; Studier & Moffatt, 1986). The nuclear spin S gives rise to a magnetic momentum μ . Spin and magnetic momentum are proportional to each other with an isotope specific constant of proportionality γ , which is termed the gyromagnetic ratio.

$$\mu = S\gamma \quad [2]$$

In the presence of an external static magnetic field (B_0), the quantisation of the spin allows two possible energy states: a high energy and a low energy state, or α and β states. The magnetic momentum of a nucleus in the field direction (μ_z) is defined by the magnetic quantum number m ($m = -I, -I+1, \dots, I-1, I$), the gyromagnetic ratio γ and \hbar .

$$\mu_z = m\gamma\hbar \quad [3]$$

For spin- $1/2$ nuclei, m can adopt the values of $1/2$ and $-1/2$. Therefore, two energy levels exist for spin- $1/2$ nuclei in a static magnetic field (B_0):

$$E = -m\gamma B_0\hbar \quad [4]$$

The energy difference between the two states $\Delta E_{\alpha\beta}$ can be calculated as:

$$\begin{aligned}\Delta E_{\alpha\beta} &= +\frac{1}{2}\gamma\hbar B_0 - \left(-\frac{1}{2}\hbar\gamma B_0\right) \\ \Delta E_{\alpha\beta} &= \hbar\gamma B_0\end{aligned}\quad [5]$$

The $\Delta E_{\alpha\beta}$ changes proportionally to the strength of B_0 . The overall energy difference for nuclear spins is relatively small. Consequently, populations in the α - and β -state at equilibrium are very nearly equal according to the Boltzmann distribution,

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E_{\alpha\beta}/k_B T} \quad [6]$$

where N_α and N_β are the number of nuclei adopting the α - and β -state, respectively, k_B is the Boltzmann constant and T the absolute temperature. It is the small excess of nuclei with spins in the low energy state versus in the high energy state that gives rise to the net magnetisation vector in the direction of B_0 (z -direction; M_z). As the energy difference $\Delta E_{\alpha\beta}$ increases with increasing B_0 , the population difference is greater at higher B_0 , and as a consequence, M_z is larger.

1.5.2 Basic principles of NMR experiments

Transitions of the nuclear spins between the upper and lower energy states can be stimulated by adding energy to the system in the form of electromagnetic radiation. This energy transfer is only possible if the energy of the electromagnetic radiation corresponds to the energy difference $\Delta E_{\alpha\beta}$. According to the Planck-Einstein relation, the frequency of the required electromagnetic radiation (ν) is given by:

$$\begin{aligned}\nu &= \frac{\Delta E_{\alpha\beta}}{h} \\ \nu &= \frac{\gamma B_0}{2\pi}\end{aligned}\quad [7]$$

This frequency is called the Larmor frequency (ν_0) for nuclear spins and is in the radio frequency (rf) (MHz) range. It describes the precession of the nuclear magnetic moment

about an external magnetic field. During an NMR experiment, irradiation at ν_0 disturbs the spin population equilibrium and changes the orientation of the net magnetisation vector of the spins. The change in orientation of the net magnetisation and its relaxation back to equilibrium is detected as a signal in an NMR experiment. The signal is called the free induction decay (FID) and has the form of a damped sine wave (Fig. 5). The FID, which is a function of time, can be transformed into a function of frequency by Fourier transformation, resulting in a spectrum which shows peaks at the Larmor frequencies of the excited nuclei (Fig. 5). If multiple nuclei with differing Larmor frequencies are excited during an NMR experiment, the spectrum contains multiple peaks. The different Larmor frequencies of nuclei are often given in the dimensionless unit δ and are referred to as chemical shift (Section 1.5.4).

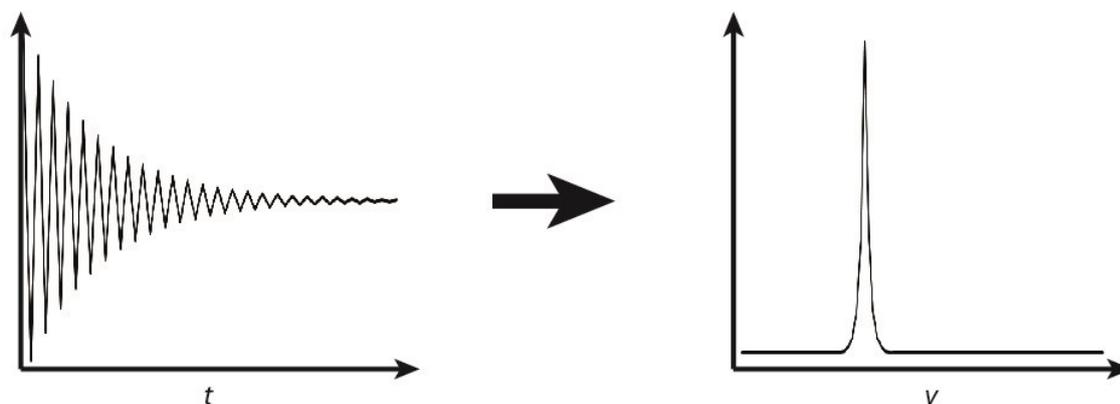


Fig. 5: Fourier transformation of the FID of a single spin into a function of frequency (ν).

In one-dimensional (1D) NMR experiments, as depicted in Fig. 5, the NMR spectrum is composed of only one frequency domain on the x-axis and the signal intensity on the y-axis. 1D NMR experiments often provide insufficient resolution when studying proteins because of a large number of ^1H signals with similar Larmor frequencies. Thus, the majority of NMR experiments used in protein solution NMR spectroscopy are multidimensional heteronuclear NMR experiments (Section 1.5.3) that correlate the frequencies of multiple spins with each other.

1.5.3 Multidimensional NMR spectroscopy

The principle of multidimensional NMR spectroscopy was first introduced in 1971 by Jean Jeener (Jeener & Alewaeters, 2016). Multidimensional NMR experiments correlate resonances of nuclei that are either connected through bonds or in spatial proximity to each other. The signals observed for correlated nuclei in an N-dimensional spectrum are referred to as cross-peaks. Their frequency coordinates correspond to the Larmor frequencies of the correlated nuclei. These correlations and the additional frequency dimensions increase resonance dispersion in NMR spectra and therefore significantly reduce spectral overlap of detected signals often observed in 1D and two-dimensional (2D) NMR experiments. The correlation of proton resonances with those of ^{15}N and or ^{13}C nuclei are especially helpful because the chemical shift range (Section 1.5.4) of the heteronuclei is much broader ($^{15}\text{N} = 100\text{--}135$ ppm; $^{13}\text{C}_{\text{aliphatic}} = 5\text{--}75$ ppm) than that of proton resonances (0–11 ppm). The most common forms of multidimensional NMR experiments are three-dimensional (3D) heteronuclear experiments, although 4D (Kay *et al.*, 1992) and even 5D (Motáčková *et al.*, 2010) experiments have been established.

The correlation of nuclear spins that are connected through covalent bonds can provide information on networks of covalently bound spins (e.g., total correlation spectroscopy (TOCSY) and correlation spectroscopy (COSY) type of experiments). The correlation of selected covalently-bound nuclei in experiments like 3D (H)CC(CO)NH, 3D H(CCO)NH and the 3D TOCSY- ^1H , ^{15}N -HSQC in protein NMR allow the identification of the amino acid type to which the observed cross-peaks belong. Sequence-specific assignments of cross-peaks belonging to nuclei in consecutive amino acids can be derived from experiments such as the 3D HNCA and 3D HNCACB. Schematic depictions of standard 2D and 3D correlation experiments used for resonance assignment of proteins are shown in Fig. 6. Through-bond correlations amongst nuclear spins rely on the use of scalar couplings (Section 1.5.5). Another type of NMR experiment, namely nuclear Overhauser effect spectroscopy (NOESY) (Section 1.5.6), allows the correlation of nuclear spins that are spatially close to each other and therefore provide internuclear distance information.

NMR parameters of ^1H , ^{15}N and ^{13}C spins in proteins that are measured in multidimensional NMR experiments and that provide essential information required for protein structure determination are the chemical shifts (δ), scalar couplings (J), the

nuclear Overhauser effect (NOE) and residual dipolar couplings (RDC). Chemical shift, scalar couplings and the NOE are described in more detail the following sections.

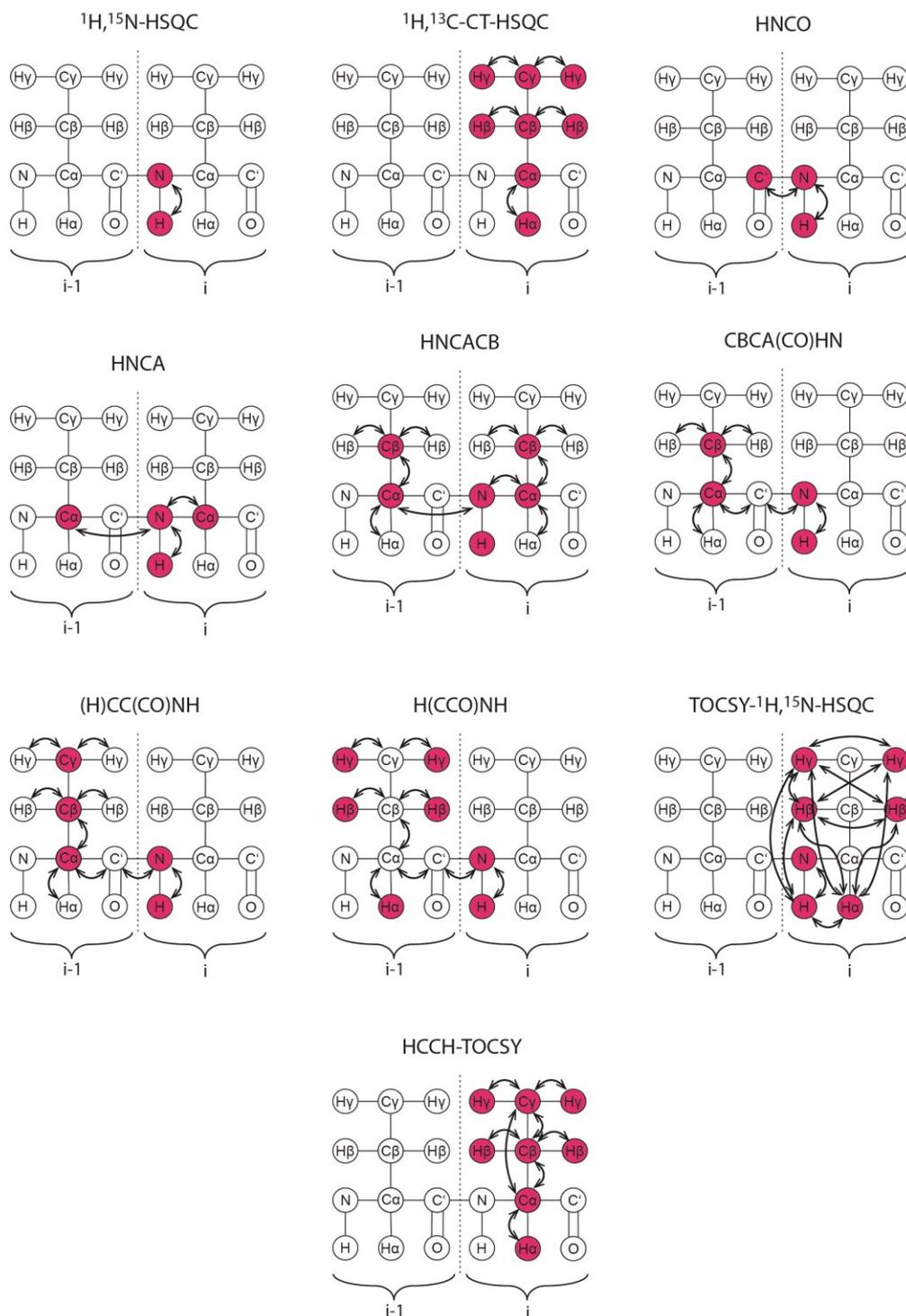


Fig. 6: Schematic depiction of standard 2D and 3D NMR experiments

The diagrams show the nuclei involved in each NMR experiment. The arrows indicate through which nuclei magnetisation is transferred. The chemical shifts of the nuclei coloured in red are measured during the experiment, whereas open circles indicate that magnetisation is passed through this nucleus, but that the chemical shift is not recorded.

1.5.4 Chemical shift

The resonance frequency of a nuclear spin is dependent on the magnetic field strength at the position of the nucleus (Equation [7]). The effective magnetic field usually differs slightly from the applied field B_0 because of the electron density surrounding the nucleus. The motions of the electrons generate secondary magnetic fields. The effect of the secondary fields is termed nuclear shielding and can either augment or diminish the effect of B_0 . Therefore, the resonance frequency of a nuclear spin is also dependent on the chemical environment of the nuclear spin. This results in physically identical nuclei resonating at slightly different frequencies. The deviation of the resonance frequencies in a sample from that of an isotope specific reference substance is referred to as the chemical shift. The chemical shift is given as the dimensionless unit δ , which is independent of B_0 and is defined as:

$$\delta_{sample} = \frac{\nu_{sample} - \nu_{reference}}{\nu_{reference}} \quad [8]$$

where $\nu_{reference}$ is the resonance frequency of a standard reference compound, with $\nu_{reference} = 0$ ppm. Since $\Delta\nu$ is very small (in the range of a few Hz) in comparison to the resonance frequencies (MHz), δ is given in ppm.

Multiple effects contribute to the chemical shifts of nuclei. Among the most important are the magnetic anisotropies of neighbouring functional groups, aromatic ring-current effects, electric field effects (charged functional groups) and dipole-dipole effects.

In the case of proteins and peptides, the chemical shift is subdivided into two components, which represent different contributions to the overall chemical shift: the random coil chemical shift and the secondary chemical shift. The random coil chemical shift is the ensemble averaged chemical shift that a nucleus adopts in an unfolded state. This chemical shift is dominated by the solvent of the sample, the covalently bound atoms and the neighbouring amino acids adjacent to the observed nucleus (nearest neighbour effect) (Wishart, 2011). The secondary chemical shift is the difference between the tabulated random coil chemical shift of a nucleus and the observed chemical shift of a signal (Schwarzinger *et al.*, 2001; Wishart, 2011). This secondary chemical shift reflects contributions from the secondary and tertiary structure of a protein, e.g., the nucleus relative orientation to aromatic rings, existing hydrogen bonds and the orientation of backbone and side chain torsion angles (Neal *et al.*, 2003; Shen &

Bax, 2010; Wishart, 2011). The structural information reflected in the chemical shift can be employed for the prediction of secondary structure elements of a biological macromolecule (Shen *et al.*, 2009; Wishart, 2011).

1.5.5 Scalar couplings

Scalar couplings or J -coupling occur between two nuclei that are connected by covalent bonds. Scalar couplings are only observed between nuclei that are one to three chemical bonds apart in proteins. The nomenclature used for naming scalar couplings is:

$${}^nJ_{AX}$$

where n is the number of covalent bonds between nuclei A and X, and J is the scalar coupling constant (Hz).

Scalar couplings arise for an AX spin system when the spin polarisation of nucleus A perturbs the polarisation of the electrons shared with the bonded nucleus X. The electrons transmit the magnetic polarisation to nucleus X. This gives rise to a splitting of the resonance arising from X and the separation between the two resonances is equal to the scalar coupling constant. The size of the coupling constant depends on the conformational strain of the bond network, torsion angles, orbital hybridisation, bond length, the presence of π -electrons and the bound substituents. Typical coupling constant values found in proteins are provided in Fig. 7.

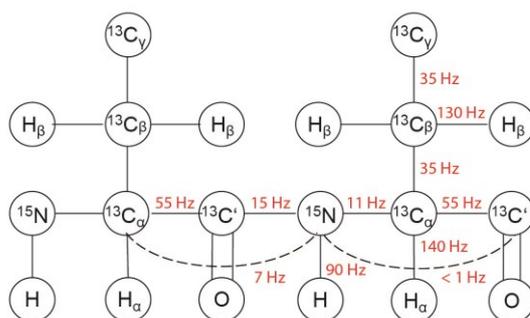


Fig. 7: 1J - and 2J -coupling constants in peptides and proteins (Sattler *et al.*, 1999).

Coupling constants (red) for typical 1J - (solid lines) and 2J -couplings (dotted lines) used for the transfer of magnetisation between ${}^1\text{H}$, ${}^{15}\text{N}$ and ${}^{13}\text{C}$ nuclei in protein NMR spectroscopy.

Scalar couplings can be used to correlate NMR signals of atoms that are chemically bonded to each other and are the basis for many multidimensional NMR experiments (Section 1.5.3). In protein NMR, manipulation of the magnetisation by exploiting specific coupling constants allows for the correlation of resonances of a specific subset

of bonded nuclei in the protein (e.g., $^1\text{H}^{\text{N}}$, ^{15}N and $^{13}\text{C}^{\alpha}$ in the 3D HNCA experiment). These correlation experiments are used in combination to make sequential resonance assignments by mapping the nuclei of the studied protein to the resonances observed. 3J -coupling constants also provide structural information because the size of the coupling can be related to the dihedral angle of the bond network which connects the two nuclei. The relationship between dihedral angles and the size of the coupling constant is described by the empirically derived Karplus equation (Karplus, 1959, 1963),

$$J(\phi) = A \cos^2 \phi + B \cos \phi + C \quad [9]$$

where $J(\phi)$ is the coupling constant of the 3J -coupling and ϕ the dihedral angle. The parameters A , B and C in Equation [9] are dependent on the atoms and substituents attached to the bond of interest and are semi-empirically derived by correlating observed 3J -coupling constants with corresponding dihedral angles measured in high-resolution structures. Based on the Karplus relation, the $^3J_{\text{HNH}\alpha}$, $^3J_{\text{NH}\beta 2/3}$ and the $^3J_{\text{H}\alpha\text{H}\beta 2/3}$ couplings can be used to determine the ϕ -angle of the protein backbone and the χ_1 torsion angle of side chains (Fig. 8), respectively (Case *et al.*, 1994; Pérez *et al.*, 2001). The determined torsion angles can be included as torsion angle restraints in structure calculations.

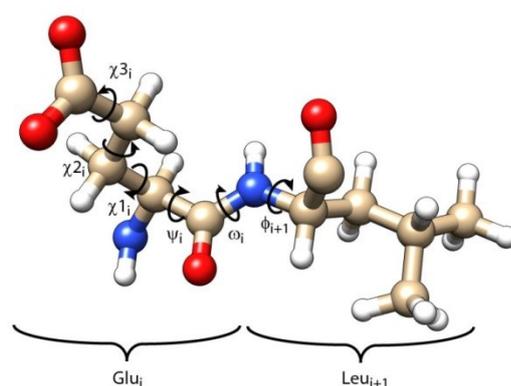


Fig. 8: Backbone and side chain torsion angles of a glutamate-leucine dipeptide.

The backbone of peptides comprises three torsion angles, i.e. the angles around the $\text{N}-\text{C}^{\alpha}-\text{C}'-\text{N}$ bond (ψ), the peptide bond ω and the $\text{C}'-\text{N}-\text{C}^{\alpha}-\text{C}'$ bond (ϕ). Since the peptide bond possesses partial double-bond character, the bond is planar, and ω adopts either 0° (cis) or 180° (trans) conformations, with trans isomers being the more common conformation in folded proteins. The $\text{C}^{\alpha}-\text{C}'$ and $\text{N}-\text{C}^{\alpha}$ bonds can rotate more freely, only hindered by possible steric clashes. Energetically favourable conformations of the angles ψ and ϕ in folded proteins are summarised in the Ramachandran plot (Ramachandran *et al.*, 1963). In addition to the backbone torsion angles, amino acids also possess side chain torsion angles, which define the side chain rotamer of the amino acids and are named $\chi_1-\chi_n$, e.g. $\chi_1-\chi_3$ for glutamate.

1.5.6 Nuclear Overhauser effect

The nuclear Overhauser effect (NOE) is a cross-relaxation process between two spins (AX) and is a through space interaction. The perturbation of the spin population levels at one nucleus (A) influences the population distribution of a nearby nucleus (X) as it relaxes back to equilibrium. This cross-relaxation between A and X results in either a decrease or an increase in the spin population difference between the lower and higher energy states of nucleus X and therefore an enhancement or reduction of the net magnetisation M_z of X. The change in M_z is called NOE. Whether M_z is enhanced or reduced depends on the rotational correlation time (τ_c) of the molecule. For large molecules like proteins, which have a comparatively large τ_c , the influence of the NOE on M_z is predominantly subtractive. In multidimensional NOESY experiments, cross-relaxation between two nuclei and the resulting NOEs are observed as cross-peaks correlating the chemical shifts of the two nuclei involved.

Cross-relaxation is distance-dependent and can only occur when two nuclei are in close proximity to each other. The rate of cross-relaxation is proportional to the inverse sixth power of the internuclear distance, with the maximum distance to observe an NOE correlation being $\sim 6 \text{ \AA}$ in the case of two protons. In the initial rate approximation, NOE cross-peak intensities that are observed in NOESY experiments are proportional to the cross-relaxation rate. As a result, the NOE cross-peak intensity can be used as a measure of the internuclear distance between nuclei in a molecule. Under the assumption that τ_c is the same for all nuclei in a protein, intra- and intermolecular interproton distances (r_{ij}) can be estimated from NOE data if a proton pair of known distance, r_{ref} , (e.g., from covalent geometry) serves as a reference [10]:

$$r_{ij} = r_{ref} \sqrt[6]{\frac{I_{ref}}{I_{ij}}} \quad [10]$$

in which I_{ref} and I_{ij} are the NOE cross-peak intensities.

Thus, the recorded NOE intensities can be translated into interproton distance restraints, which usually make up the majority of restraints when calculating protein structures from solution NMR data.

Special cases of NOESY experiments are isotope-filtered/edited NOESY experiments (Breeze, 2000). Isotope-filtered/edited NOESY spectra comprise NOEs from protons

attached to $^{12}\text{C}/^{14}\text{N}$ atoms to protons directly bound to $^{13}\text{C}/^{15}\text{N}$ heteronuclei (Fig. 9). In isotope-filtered/edited NOESY-HSQC spectra, the initial magnetisation arising from protons attached to ^{15}N or ^{13}C nuclei is quenched (filtering step), so that NOE cross-relaxation can only originate from protons bound to $^{12}\text{C}/^{14}\text{N}$ nuclei. As a second step, the build-up of the NOE is allowed among protons of the non-isotope-labelled and isotope-labelled protein. The last step involves the editing of the NOE-signals by selecting only signals arising from protons bound to either ^{15}N or ^{13}C nuclei (editing step).

These experiments are used for the characterisation of interaction interfaces of protein complexes, where one component is completely isotope-labelled, and the other is not. For the characterisation of the protein-protein interfaces of symmetric dimers, isotope-filtered/edited NOESY experiments are recorded on samples that contain an equimolar mixture of $^{13}\text{C}/^{15}\text{N}$ -labelled and unlabelled proteins. These experiments yield exclusively NOEs between protons located at the interface because the interface is the only region that enables interatomic distances $< 6 \text{ \AA}$ between protons attached to ^{15}N or ^{13}C nuclei and protons attached to ^{14}N or ^{12}C nuclei.

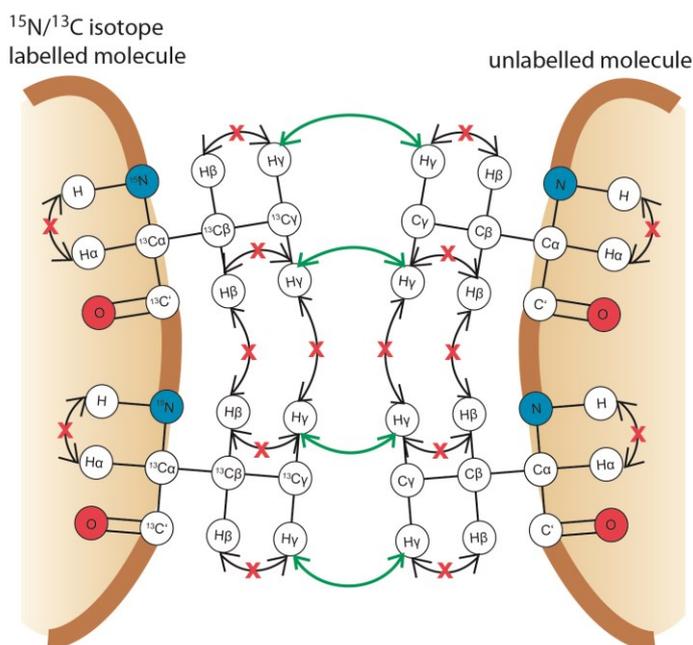


Fig. 9: Schematic depiction of the NOE transfers during 3D ^{15}N , ^{13}C - ω_1 filtered, ^{13}C -edited and ^{15}N -edited NOESY-HSQC experiments.

In isotope-filtered/edited NOESY spectra, NOEs from protons attached to $^{12}\text{C}/^{14}\text{N}$ atoms (unlabelled protein) to protons attached to isotope-labelled heteronuclei ($^{15}\text{N}/^{13}\text{C}$ labelled molecule) are detected, which are represented by the green arrows. In a sample containing isotope-labelled and unlabelled proteins, these NOEs can only be observed between nuclei associated with residues at the interface. Only at the interface are protons attached to unlabelled heteronuclei (unlabelled protein) in sufficiently close proximity (i.e., $< 6 \text{ \AA}$) to protons attached to isotope-labelled heteronuclei (labelled molecule).

1.5.7 Dynamic processes and chemical exchange

In solution NMR spectroscopy, the proteins move freely in solution and can therefore undergo a variety of dynamic processes, such as conformational changes or ligand binding. If these processes change the NMR parameters of the observed nuclei, the process is termed chemical exchange. Depending on the exchange rate (k_{ex}) with which the examined molecule switches between the states, the effect on the recorded NMR spectra varies. These effects will be described for the chemical shift in the following example for a two-state equilibrium (i.e., $A \leftrightarrow B$) where the populations of the two states are equal (Fig. 10).

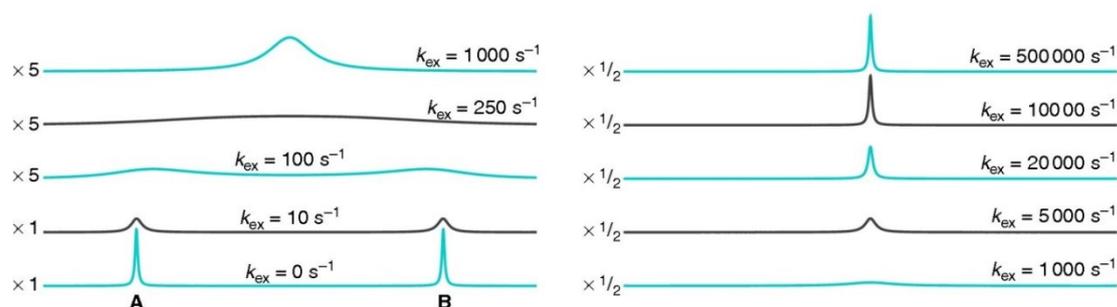


Fig. 10: Simulated 1D NMR spectra of nuclei undergoing chemical exchange between two states (A and B) at different exchange rates (k_{ex}) (reproduced with permission from Keeler, 2010).

In the absence of exchange ($k_{\text{ex}} = 0 \text{ s}^{-1}$) the two states are easily distinguishable with resonances appearing at two different chemical shifts ($\Delta\nu = 160 \text{ Hz}$). Since population A and B are equal, both peaks have the same intensity. The increase in the exchange rate first causes a decrease in intensity due to line broadening until the two peaks coalesce when $k_{\text{ex}} \approx \Delta\nu$. As k_{ex} increases further, the merged lines narrow and one signal at the average chemical shift is observed. If $k_{\text{ex}} \gg \Delta\nu$, the exchange process does not exhibit any influence on the line-shape of the resonance any longer.

If k_{ex} is much slower than the change in chemical shift ($\Delta\delta$) and thus the difference in the resonance frequencies between the two states ($\Delta\nu$), the process is termed slow exchange.

$$k_{\text{ex}} \ll \Delta\nu$$

If a nucleus undergoes slow exchange, peaks for both states are observed in the NMR spectrum. The intensity ratio of the peaks represents the fraction of nuclei which populate each of the two states.

If k_{ex} is much faster than $\Delta\nu$, the nucleus undergoes fast exchange.

$$k_{ex} \gg \Delta\nu$$

For nuclei undergoing fast chemical exchange (Fig. 10), only one peak is visible, which represents the population-weighted averaged chemical shift of the two states. A change in the population distribution, for example by addition of a ligand during a titration experiment, leads to a change in the chemical shift of the resonance (Fig. 11). This ligand-induced chemical shift perturbation can be used to identify amino acids that are involved in the interaction. This approach is called chemical shift mapping.

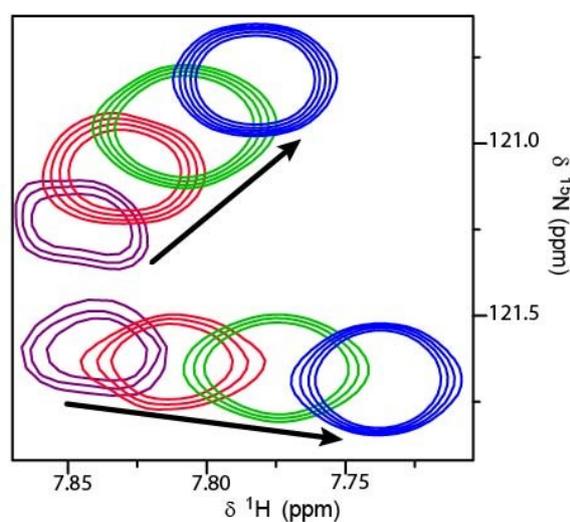


Fig. 11: Schematic depiction of cross-peaks representing nuclei undergoing fast exchange.

The spectrum shows a superposition of cross-peaks in 2D ^1H , ^{15}N -HSQC experiments of nuclei that undergo fast exchange as a result of ligand binding. The different colours represent different ligand concentrations at which the spectra were recorded. The arrows indicate an increase in ligand concentration. The observed signals display the population-averaged chemical shift of the states the nucleus can adopt. Changes in the population ratio of the two states due to the addition of ligand cause chemical shifts to approach the chemical shift of the bound state.

If the k_{ex} is close to $\Delta\nu$, the nucleus is undergoing intermediate exchange.

$$k_{ex} \approx \Delta\nu$$

The intermediate exchange regime is characterised by a change in the signal line-shape of the recorded signal. As long as one of the two states A and B is much higher populated than the other, one resonance is visible at the chemical shift representing the respective state. If the population of the two states becomes equal, the line-shape of the resonance broadens and the signal intensity is diminished (exchange broadening) (Fig. 10). The signals of the two species coalesce and cannot be

detected any longer in an NMR spectrum. This effect can be problematic for NMR-experiments of systems undergoing intermediate exchange since it can cause significant reduction of resonance intensities independent of concentration.

Seeing that the exchange regime is dependent on $\Delta\nu_{AB}$, it is possible to observe two or even all three forms of chemical exchange in one spectrum, because nuclei that experience vastly different chemical environments in the two states will show a larger chemical shift perturbation than those that are only marginally involved in the process.

1.6 Structure calculation from NMR-derived experimental restraints

NMR spectroscopy provides fundamental insights into biological macromolecules such as proteins. However, NMR spectroscopy does not directly observe 3D structures but measures NMR observable parameters like chemical shift (Section 1.5.4), J -couplings (Section 1.5.5) and dipolar cross-relaxation (NOE) (Section 1.5.6), which can be translated into angular and distance information between NMR-active nuclei. A schematic overview of the steps required to solve a protein structure by NMR spectroscopy is shown in Fig. 12.

If a protein sample is amenable to NMR spectroscopy, the signals observed in NMR experiments are assigned to the corresponding nuclei in a process called resonance assignment. Ideally, all NMR observable nuclei in a protein are assigned their corresponding resonances. With chemical shift assignments at hand, additional NMR experiments are conducted to define structural restraints, such as inter-proton distances and backbone and sidechain torsion angles. These restraints are then implemented in a computational structure calculation process. Therefore, when solving structures of proteins by NMR spectroscopy, the structures are determined indirectly by computational methods from the acquired NMR data.

The recorded experimental data can suffer from uncertainties and inconsistencies because of experimental errors or hardware and software limitations, such as erroneous signal intensities due to spectral overlap, spin diffusion, chemical exchange processes or spectral artefacts. Hence, restraints are implemented into structure calculations with a specified error margin. Therefore, NMR-based structure determination always entails the examination of multiple structures (structural ensembles) that sample the conformational space of the protein as defined by the restraints. Standard covalent

geometry restraints, e.g., bond lengths and bond angles, are also included in the structure calculation protocol (Braun & Gö, 1985; Nilges *et al.*, 1988; Stein *et al.*, 1997), because NMR restraints alone would not suffice to determine an accurate structure. The most common strategy for structure determination from NMR data is restrained molecular dynamics (rMD) simulations in combination with simulated annealing (SA). This approach is employed by structure generation software programs such as XPLOR-NIH (Schwieters *et al.*, 2003), ARIA/CNS (Ambiguous Restraints for Iterative Assignments/Crystallography & NMR System) (Bardiaux *et al.*, 2012; Brünger *et al.*, 1998) and CYANA (Güntert *et al.*, 1997; Güntert & Buchner, 2015).

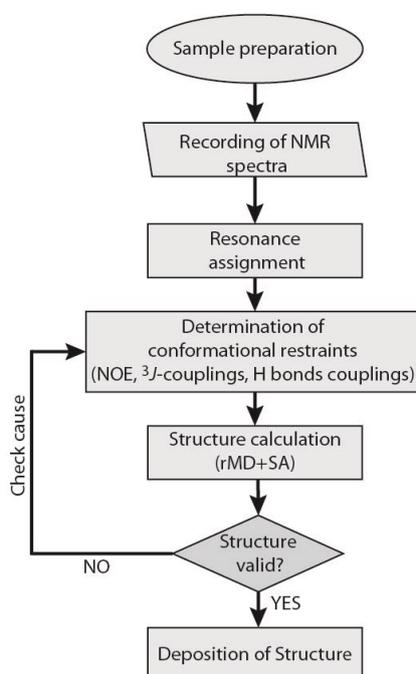


Fig. 12: Flowchart of the structure determination process of biomolecules by NMR spectroscopy.

In MD, a force field defines covalent bond parameters (bond length and angles) and non-bonded interactions (electrostatic and van der Waals forces) in the form of energy functions. In rMD, the force field is supplemented by experimental data, which, if violated, increases the energy of the system (Clare *et al.*, 1985, 1987). Simulated annealing helps to overcome local minima in the energy landscape by heating the system to a very high temperature followed by a stepwise cooling process (Nilges *et al.*, 1988; Nilges & O'Donoghue, 1998). Under ideal conditions, this combination of restraints yields an ensemble of structures with low overall energy because of good

accordance with the experimental data and with a high convergence. The quality of 10 to 20 structures with the lowest overall energy is subsequently validated. Validation tools assess the quality of a structural ensemble based on multiple criteria. Among those are the number and quality of restraint violations, overall convergence of the models by the root-mean-square-deviation (RMSD) of atom positions, the number of side chain and backbone torsion angle outliers (Laskowski *et al.*, 1996), and near atom clashes (Chen *et al.*, 2010; Davis *et al.*, 2007). Favourable assessment leads to deposition of the structure in the protein data bank (PDB) (Berman *et al.*, 2003).

1.7 Structure determination of symmetric homodimers by NMR spectroscopy

The main obstacle associated with structure determination of symmetric oligomers by NMR is the degeneracy of the recorded NMR spectra. In symmetric oligomers, all subunits give rise to the same set of resonances. While the redundancy in chemical shifts often facilitates the resonance assignment because it reduces spectral overlap, it results in highly ambiguous inter- and intramolecular NOE information. Intermolecular NOEs arise from the transfer of magnetisation through space between protons of one monomer to protons of the other monomer and thus provide essential distance restraints for the definition of the oligomer interface. However, distinguishing *a priori* between intra- and intermolecular restraints is impossible in standard 3D NOESY experiments (Breeze, 2000; Neuhaus & Williamson, 2000). Additional NMR experiments such as isotope-filtered/edited NOESY, RDC, hydrogen exchange and paramagnetic relaxation enhancement (PRE) measurements can be recorded to gain additional restraints that complement the NOE data (Göbl *et al.*, 2014; Wang *et al.*, 2008). Different orthogonal techniques including small angle X-ray scattering (SAXS) can also be used to aid structure determination of symmetric oligomers (Göbl *et al.*, 2014).

1.8 Aims

The aim of this work was to elucidate the oligomeric state of the SAM domain of SLY1 and to determine the structure of the SLY1 SAM domain by multidimensional heteronuclear NMR spectroscopy. The SLY1 SAM domain has been shown to self-associate (Thiagarajan, 2011). Moreover, previous efforts to solve the structure of a

SLY1 SAM domain construct (G249–E321) failed because of significant line-broadening of NMR signals due to the self-association equilibrium, which interfered with full resonance assignments and collection of distance restraint information.

In this thesis, an intermolecular cross-linking strategy was used to quench the chemical exchange process between monomer and dimer state and thus enable structure determination of the SLY1 SAM dimer by NMR spectroscopy. The SAM domain self-association was characterised by biophysical methods and verified to be a dimerisation process of moderate affinity without any higher order oligomerisation observed. A SLY1 SAM domain mutant containing a single cysteine outside of the SAM core domain was cloned and recombinantly produced in *E. coli*. The SAM domain mutant enabled site-specific cross-linking of two SAM domains via the introduced cysteine residues, thereby giving rise to a population of only SAM dimer. The effect of the mutation and the cross-linking on the SLY1 SAM dimer were assessed by NMR spectroscopy. Multidimensional heteronuclear NMR spectra were acquired and analysed for structure determination. NMR restraints derived from NMR data analysis were used to calculate the structure. The structure of the wild-type (wt) SAM domain was also determined by X-ray crystallography.

2 Materials

2.1 Equipment

Table 1: List of the instruments used in this work

| Equipment | Manufacturer/Distributor |
|--|--|
| Äkta purifier | GE Healthcare (Chicago, IL, USA) |
| Äkta pure | GE Healthcare (Chicago, IL, USA) |
| Agarose gel electrophoresis chamber | GE Healthcare (Chicago, IL, USA) |
| Amicon stirred cell concentrators (3 mL, 10 mL, 50 mL) | Merck Millipore (Burlington, MA, USA) |
| Analytical ultracentrifuge ProteomLab X-LA | Beckman-Coulter (Brea, CA, USA) |
| Branson sonifier 250 | Branson Ultrasonics (Danbury, CT, USA) |
| Centrifuge 5417R | Eppendorf (Hamburg, Germany) |
| Centrifuge 5702R | Eppendorf (Hamburg, Germany) |
| Centrifuge 5804R | Eppendorf (Hamburg, Germany) |
| Centrifuge Avanti J-20 XP | Beckman Coulter (Brea, CA, USA) |
| FPLC system | Pharmacia (Stockholm, Sweden) |
| Gel Doc XR+ gel documentation system | Biorad (Hercules, CA, USA) |
| Benchtop homogeniser M-110P | Microfluidics Cord (Westwood, MA, USA) |
| Hoefer Mighty Small SE 260 vertical electrophoresis unit | Hoefer (Holliston, MA, USA) |
| Monolith NT.115 blue/red | Nanotemper technologies (München, Germany) |
| MilliQ-Biocell-system | Merck Millipore (Burlington, MA, USA) |
| Multitron Pro incubation shaker | Infors (Einsbach, Germany) |
| Bruker Avance III HD NMR (600 MHz) | Billerica, MA, USA |
| Bruker Avance III HD NMR (700 MHz) | Billerica, MA, USA |
| POLARstar OPTIMA plate reader | BMG Labtech (Ortenburg, Germany) |
| Rotor JA 25.50 | Beckmann Coulter (Brea, CA, USA) |
| Rotor JLA 8.100 | Beckmann Coulter (Brea, CA, USA) |
| Thermocycler iCycler | Biorad (Hercules, CA, USA) |
| Rotor An-50 Ti Analytical 8-Place | Beckmann Coulter (Brea, CA, USA) |
| Thermomixer compact | Eppendorf (Hamburg, Germany) |

| | |
|------------------------------------|-------------------------------------|
| Spectrophotometer UV 1800 | Shimadzu (Kyoto, Japan) |
| Lambda 25 UV/Vis spectrophotometer | PerkinElmer (Waltham, MA, USA) |
| Spectrophotometer (Cells) | Schott Instruments (Mainz, Germany) |
| NanoPhotometer 300 | Implen (München, Germany) |

2.2 Chemicals, enzymes and media components

The purity of the chemicals, enzymes and media components was dependent on the requirements of the experiments. Overall, the grade of chemicals corresponded to *pro analysi* (p.a.) or higher purity.

Chemicals, enzymes and media components not mentioned in Table 2 and Table 3 were usually purchased from one of the following distributors: AppliChem GmbH (Darmstadt, Germany), Roth (Karlsruhe, Germany), Merck KGaA (Darmstadt, Germany), Serva (Heidelberg, Germany) and Sigma-Aldrich (St. Louis, MI, USA).

Table 2: List of specialised chemicals used

| Chemical | Manufacturer/Distributor |
|---------------------------------------|---|
| Agarose | Serva (Heidelberg, Germany) |
| Alexa Fluor 488 | Thermo Fisher Scientific (Waltham, MA, USA) |
| Ampicillin sodium salt | AppliChem (Darmstadt, Germany) |
| cOmplete protease inhibitor cocktail | Roche (Basel, Germany) |
| D ₂ O | Cambridge isotope laboratories (Tewksbury, MA, USA) |
| DNA loading dye (6×) | Thermo Fisher Scientific (Waltham, MA, USA) |
| DSS | Sigma-Aldrich (St. Louis, MI, USA) |
| DTT | AppliChem (Darmstadt, Germany) |
| EDTA-disodium salt | AppliChem (Darmstadt, Germany) |
| Glutathione (reduced) | AppliChem (Darmstadt, Germany) |
| IPTG | UBPbio (Aurora, USA) |
| OMNISCAN | GE Healthcare (Chicago, IL, USA) |
| Protein Assay Dye Reagent Concentrate | Biorad (Hercules, CA, USA) |
| TCEP | Sigma-Aldrich (St. Louis, MI, USA) |
| Unstained protein marker | Thermo Fisher Scientific (Waltham, MA, USA) |

Table 3: List of enzymes used

| Enzyme | Manufacturer/Distributor |
|-----------------------|---|
| Lysozyme | AppliChem (Darmstadt, Germany) |
| PreScission-Protease | GE Healthcare (Chicago, IL, USA) |
| DNase | AppliChem (Darmstadt, Germany) |
| <i>Bsp120I/PspOMI</i> | Thermo Fisher Scientific (Waltham, MA, USA) |
| <i>BtsI</i> | New England Biolabs (Ipswich, MA, USA) |
| <i>XhoI</i> | Thermo Fisher Scientific (Waltham, MA, USA) |
| SAP | New England Biolabs (Ipswich, MA, USA) |
| <i>T4</i> DNA Ligase | Thermo Fisher Scientific (Waltham, MA, USA) |
| Vent DNA Polymerase | New England Biolabs (Ipswich, MA, USA) |

2.3 Columns and resins

Table 4: List of columns and resins used for preparative and analytical column chromatography

| Column/Resin | Manufacturer |
|-----------------------------|----------------------------------|
| Glutathione Sepharose 4B | GE Healthcare (Ipswich, MA, USA) |
| Hiload 26/60 Superdex 75 pg | GE Healthcare (Ipswich, MA, USA) |
| Superdex 75 10/30 HR | GE Healthcare (Ipswich, MA, USA) |
| Superdex 75 5/150 GL | GE Healthcare (Ipswich, MA, USA) |

2.4 Kits

Table 5: List of commercially available kits used

| Kit | Use | Manufacturer/Distributor |
|--|--|---|
| Bradford Protein Assay | Determination of protein concentration | Biorad (Hercules, CA, USA) |
| QuikChange II XL Site-Directed Mutagenesis Kit | Site-directed mutagenesis of nucleic acids | Agilent Technologies (Santa Clara, CA, USA) |
| QIAGEN Plasmid Midi Kit | Plasmid isolation from <i>E. coli</i> | Qiagen (Hilden, Germany) |
| InnuPREP Plasmid Mini Kit | Plasmid isolation from <i>E. coli</i> | AnalytikJena (Jena, Germany) |
| Gel Filtration Calibration Kit (LMW) | Calibration of size exclusion chromatography columns | GE Healthcare (Ipswich, MA, USA) |

2. Materials

| | | |
|--|--|--|
| Monolith Protein Labeling Kit BLUE-NHS | Fluorescent labelling of SAM _{wt} by NHS-amine-coupling | Nanotemper technologies (München, Germany) |
| NucleoSpin Gel and PCR Clean-Up kit | Elution of DNA fragments from agarose gel and subsequent clean-up of DNA after enzymatic reactions | Macherey-Nagel, (Düren, Germany) |

2.5 Commercially available crystallisation screens

Table 6: List of all commercially available kits for sparse matrix screening of crystallisation conditions used for SLY1 SAM_{wt} crystallisation

| Screen | Manufacturer/Distributor |
|--------------------------|---|
| AmSO ₄ Suite | Qiagen (Hilden, Germany) |
| Crystal Screen I/II | Hampton Research (Aliso Viejo, CA, USA) |
| JCSG Core Suite I/II/III | Qiagen (Hilden, Germany) |
| PACT Suite | Qiagen (Hilden, Germany) |
| PEGs Suite I/II | Qiagen (Hilden, Germany) |
| PEG/Low Ion | Qiagen (Hilden, Germany) |
| ProPlex Eco | Molecular Dimensions (Suffolk, UK) |
| Wizard I/II | Rigaku (Ettingen, Germany) |

2.6 Bacterial strains

Table 7: List of bacterial strains used for molecular cloning and recombinant gene expression

| Strain | Genotype | Reference/Source |
|---|---|---|
| <i>E. coli</i> Mach1-T1 | <i>W ΔrecA1398 endA1 tonA Φ80ΔlacM15 ΔlacX74 hsdR(r_K-m_K+)</i> | Thermo Fisher Scientific (Waltham, MA, USA) |
| <i>E. coli</i> OmniMAX 2 T1R | <i>F' {proAB lacIq lacZΔM15 Tn10(Tet^R) Δ(ccdAB)} mcrA Δ(mrr hsdRMS-mcrBC) Φ80(lacZ)ΔM15 Δ(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD</i> | Thermo Fisher Scientific (Waltham, MA, USA) |
| <i>E. coli</i> XL10-Gold ultracompetent cells | <i>endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 Tet^R F'[proAB lacI^qZΔM15 Tn10(Tet^R Amy Cm^R)]</i> | Agilent Technologies (Santa Clara, CA, USA) |

| | | |
|---------------------------|---|---|
| <i>E. coli</i> BL21 (DE3) | F ⁻ <i>ompT gal dcm lon hsdSB(rB⁻ mB⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>]) [<i>malB⁺</i>] _{K-12} (λ ^S) | Studier & Moffatt, 1986 Jeong <i>et al.</i> , 2009 |
|---------------------------|---|---|

2.7 Plasmids

Table 8: Plasmids used for recombinant gene expression

| Plasmids for gene expression | Genotype | Reference/Source |
|------------------------------|--|--|
| pGEX-6p-2 | Amp ^R , <i>lacI^q</i> , P _{tac} , pBR322 origin, MCS, GST-tag, PresScission protease cleavage site | GE Healthcare (Ipswich, USA) |
| pGEX-6p-2_kombiP | pGEX-6p-2 with a deletion of the <i>PspOMI</i> recognition site by a point mutation in <i>laqI^q</i> | ICS-6 Forschungszentrum Jülich (Jülich, Germany) |
| pGEX-6p-2_SLY1SAM_lg | pGEX-6p-2_kombiP containing integrated <i>SlyIc</i> DNA-fragment, coding for SLY1 ₂₄₉₋₃₂₁ (SAM _{lg}) (<i>Mus musculus</i>) | Thiagarajan, 2011 |
| pGEX-6p-2_SLY1SAM_lg_C | pGEX-6p-2_SLY1SAM_lg coding for SLY1 ₂₄₉₋₃₂₁ containing a S320C substitution (SAM _{lg_C}) | this thesis |
| pGEX-6p-2_SLY1SAM_wt | pGEX-6p-2_kombiP containing an integrated <i>SlyIc</i> DNA-fragment, coding for SLY1 ₂₅₄₋₃₂₁ (SAM _{wt}) (<i>Mus musculus</i>) | this thesis |
| pGEX-6p-2_SLY1SAM_C | pGEX-6p-2_SLY1SAM_wt coding for SLY1 ₂₅₄₋₃₂₁ containing a S320C substitution (SAM _C) | this thesis |

2.8 Oligonucleotides

All RP-HPLC purity-grade oligonucleotides have been purchased from Biotex (Berlin, Germany). Endonuclease recognition sites (ERS) present in the oligonucleotides purchased are underlined in the respective sequences. In primers purchased for site-directed mutagenesis by "QuikChange"-PCR (Table 10), the three-nucleotide codon carrying the point mutation is also underlined.

2. Materials

Table 9: Primers used for standard PCR reactions and sequencing

| Oligonucleotide | Nucleotide sequence (5'→3') | Endonuclease | Use |
|-----------------|---|-----------------------------------|---|
| pGEX-5' | GGG CTG GCA AGC CAC GTT TGG | | Sequencing of all pGEX constructs |
| F-Sly-short | CCA <u>GGG</u> <u>CCC</u> CAA GAC TCT GCA TGA ACT GCT G | <i>Bsp120I</i> / <i>PspOMI</i> | pGEX-6p- 2_SLY1SAM_wt pGEX-6p-2_SLY1SAM _C |
| R-Sly-short | CCG CTC GAG TTA TTC ACT GC | <i>XhoI</i> | pGEX-6p- 2_SLY1SAM_wt |
| R-Sly-mut-short | CCG <u>CTC</u> <u>GAG</u> TTA TTC ACA GC | <i>XhoI</i> | pGEX-6p-2_SLY1SAM _C |

Table 10: Primers used for site-directed mutagenesis by "QuikChange"-PCR

| Oligonucleotide | Nucleotide sequence (5'→3') | Use |
|--------------------|---|----------------------------|
| SLY1SAM_QC_S74C_fw | G CTA CTG GAC TAT GAC ACT GGC <u>TGT</u> GAA TAA CTC | pGEX-6p- 2_SLY1SAM_lg_C |
| SLY1SAM_QC_S74C_rv | CCG CTC CAG TTA TTC <u>ACA</u> GCC AGT GTC ATA GTC | pGEX-6p- 2_SLY1SAM_lg_C |

2.9 Software and databases

Table 11: Software and databases used for various applications

| Software/Database | Use | Reference/distributor |
|--|---|--|
| ARIA version 2.3.2 (in conjunction with CNS) | NOE resonance assignment and calculation of 3D protein structures based on NMR data | Rieping <i>et al.</i> , 2007 |
| AIMLESS | Processing of diffraction data | Evans & Murshudov, 2013 |
| BEST | Calculation of recording parameters for X-ray diffraction data | Bourenkov & Popov, 2010 |
| BMRB Databank | Open-access repository for experimentally derived NMR chemical shifts | Seavey <i>et al.</i> , 1991 http://www.bmrwisc.edu/ |

| | | |
|---|---|---|
| Clonemanager Professional Suite 9 | <i>In silico</i> planning of molecular cloning strategies, primer-design and analysis of nucleotide sequences | Scientific & Educational Software (Denver, CO, USA) |
| CCP4 | Diffraction data processing, molecular replacement | Winn <i>et al.</i> , 2011 |
| CcpNmr Analysis Software version 2.4.2 | Visualisation and analysis of processed NMR spectra | Vranken <i>et al.</i> , 2005 |
| Chimera | Visualisation and analysis of protein structures | Pettersen <i>et al.</i> , 2004 |
| CNS 1.2 path level 1 | Calculation of protein structures based on NMR data | Brünger <i>et al.</i> , 1998 |
| Coot | Visualisation of 3D protein structures and electron density maps for X-ray structure determination | Emsley & Cowtan, 2004 |
| Disulfide Bond Dihedral Angle Energy Server | Calculation of disulfide dihedral angles and the dihedral energy | https://services.mbi.ucla.edu/disulfide/ Weiner <i>et al.</i> , 1984 Katz & Kossiakoff, 1986 |
| DSSP | Assignment of SSEs from atomic coordinates | W Kabsch & Sander, 1983 Touw <i>et al.</i> , 2015 |
| Expasy-ProtParam | Determination of biophysical properties of proteins based on their primary sequence | Gasteiger <i>et al.</i> , 2005 http://web.expasy.org/protparam/ |
| Expasy SWISS-Model | Model building for molecular replacement | Arnold <i>et al.</i> , 2006 Biasini <i>et al.</i> , 2014 https://swissmodel.expasy.org/ |
| LSQMAN | Superposition of 3D dimer structures | Kleywegt, 1996 |
| MATTHEWS_COEF | Estimation of the number of molecules in the asymmetric unit | Kantardjieff & Rupp, 2003 |
| Molprobrity | All-atom contact validation | Davis <i>et al.</i> , 2007 Chen <i>et al.</i> , 2010 |
| MOLREP | Molecular replacement | Vagin & Teplyakov, 1997 |

2. Materials

| | | |
|--|---|---|
| MOSFLM | Determination of possible point group and orientation of the crystal lattice | Battye <i>et al.</i> , 2011 |
| NCBI BLASTP | Sequence alignment | Altschul <i>et al.</i> , 1997 |
| NT analysis software version 1.5.41 | Evaluation and fitting of MST data | Nanotemper technologies |
| nmrPipe | Processing of NMR spectra | Delaglio <i>et al.</i> , 1995 |
| nmrDraw | Visualisation of processed NMR spectra | Delaglio <i>et al.</i> , 1995 |
| PHENIX | Refinement of X-ray data | Adams <i>et al.</i> , 2010 |
| PISA web server | Dimer modelling and interface analysis | Krissinel & Henrick, 2007 www.ebi.ac.uk/pdbe/prot_int/pistart.html |
| POINTLESS | Identification of the space group in protein crystals | Evans, 2011 |
| POLYGON | Comparative structure validation | Urzhumtseva <i>et al.</i> , 2009 |
| PROCHECK-NMR | Stereochemical quality analysis and validation of NMR-derived protein structures | Laskowski <i>et al.</i> , 1996 |
| RCI server | Determination of the <i>random coil index</i> (RCI) of a structure based on chemical shifts | Berjanskii & Wishart, 2007 http://randomcoilindex.com/ |
| RPF server | Statistical validation of the agreement between NOE data and the calculated structure | Huang <i>et al.</i> , 2012 |
| SSM | Superposition of 3D structures | Krissinel & Henrick, 2004 |
| TALOS+ version 3.80F1 Rev 2012.080.14.41 | Determination of protein backbone torsion angle restraints based on chemical shifts | Shen <i>et al.</i> , 2009 |
| XDS | Analysis and processing of X-ray diffraction data | Kabsch, 2010 |

3 Methods

3.1 Microbiological methods

3.1.1 Bacterial growth media

Bacterial growth media and liquid media components (Table 12 and Table 13) were prepared with ultrapure water (resistivity typically 18.2 MΩ cm at 25 °C). These media were sterilised by either moist heat sterilisation in an autoclave or, in the case of heat-labile components like vitamins, antibiotics and the TS2 solution, by filtration with a membrane filter (average pore diameter 0.22 μm). Antibiotics were added directly before the medium was used.

For solid growth media, liquid medium was supplemented with 2% (w/v) agar-agar. Heat-labile components like ampicillin were added to the medium-agar-mixture when the temperature of the agar solution was below 60 °C.

Table 12: Composition of bacterial growth media for recombinant gene expression

| LB-Medium (Sambrook & Russell, 2001) | | M9-Medium^a | |
|---|--------|--|---------------|
| Tryptone | 10 g/L | CaCl ₂ ·2H ₂ O | 14.7 mg/L |
| NaCl | 10 g/L | Mg ₂ SO ₄ | 0.49 g/L |
| Yeast extract | 5 g/L | Na ₂ HPO ₄ ·2 H ₂ O | 8.3 g/L |
| pH 7.2 | | KH ₂ PO ₄ | 3 g/L |
| | | NaCl | 0.5 g/L |
| | | NH ₄ Cl/ ¹⁵ NH ₄ Cl | 1 g/L |
| | | Glucose/ ¹³ C-glucose | 4 g/L / 2 g/L |
| | | Thiamine hydrochloride | 5 mg/L |
| | | TS2-solution | 0.2% (v/v) |
| | | Vitamin stock solution | 0.1% (v/v) |
| | | Iron (III)-citrate | 2.5 mg/L |
| | | pH 7.4 | |

^amodified from Sambrook & Russell, 2001

Table 13: Composition of TS2 solution and vitamin solution used in M9-medium preparation

| TS2 trace element solution (mg/L) | | Vitamin solution (g/L) | |
|---|-----|---|-----|
| MnCl ₂ ·4H ₂ O | 30 | D-biotin | 1 |
| ZnSO ₄ ·7H ₂ O | 100 | Choline chloride | 1 |
| H ₃ BO ₃ | 300 | Folic acid | 1 |
| CoCl ₂ ·6H ₂ O | 200 | Nicotinamide | 1 |
| NiCl ₂ ·6H ₂ O | 20 | Sodium-D-pantothenate | 1 |
| CuCl ₂ ·2H ₂ O | 10 | Pyridoxal hydrochloride | 1 |
| Na ₂ MoO ₄ ·2H ₂ O | 900 | Riboflavin | 0.1 |
| Na ₂ SeO ₃ | 20 | dissolved in 20 mM phosphate buffer, pH 7 | |
| dissolved in ultrapure water | | | |

3.1.2 Cultivation of bacteria

Cultivation of *E. coli* bacteria was performed at 37 °C either on LB-agar plates or in liquid media in baffled Erlenmeyer flasks. Unless otherwise described, the flasks contained media to a maximum of ¼ of the total volume of the flask. Cultures were incubated in incubation shakers (Table 1) with shaking at 120–130 rpm to ensure sufficient aeration. Small volume cultures with a maximum of 5 mL were incubated in 15 mL glass test tubes. Ampicillin (100 µg/mL) was added to the media as selection marker.

The cell density of liquid cultures was determined by measuring the optical density of the cultures at $\lambda = 600$ nm (OD₆₀₀) using a spectrophotometer (Table 1) with sterile medium used as the reference. Cuvettes with an optical depth of 1 cm were used.

3.1.2.1 Overnight precultures

Overnight cultures were inoculated using single colonies from agar plates and grown for a minimum of 16 h in LB-medium. The addition of 2% (w/v) glucose helped to increase the cell density and to suppress basal gene expression of the plasmids encoding the heterologous genes.

3.1.2.2 Overexpression of *Sly1SAM* variants

For the production of the recombinant SLY1 SAM domain by overexpression of the *Sly1SAM* coding sequences, *E. coli* BL21 (DE3) (Table 7) were transformed with one of the respective plasmids listed in Table 8. Five hundred mL culture medium containing ampicillin (100 µg/mL) in a 2 L-Erlenmeyer flask was inoculated with 5 mL of an overnight grown preculture. The bacteria were grown until the culture reached an OD₆₀₀ of ~0.6. Heterologous gene expression was then induced by adding IPTG at a final concentration of 0.2 mM. After an incubation period of 5 h post-induction, the cells were harvested by centrifugation (5000 × g, 30 min, 4 °C) and the cell pellets stored at -20 °C until used for protein purification.

For the production of non-isotope-labelled SLY1 SAM domain, the cells were grown in LB medium. Uniformly ¹⁵N-single labelled or uniformly ¹⁵N,¹³C-double-labelled SLY1 SAM was produced in M9-medium containing either isotope labelled ¹⁵NH₄Cl or ¹⁵NH₄Cl and ¹³C-glucose.

3.1.3 Preparation of chemically competent *E. coli*

For the preparation of chemically competent *E. coli*, the target strain was cultivated without the addition of an antibiotic. A culture of 100 mL LB-medium in a 500 mL Erlenmeyer flask was inoculated by adding 1 mL of an overnight preculture. The main culture was incubated until the culture density reached an OD₆₀₀ of 0.6. The bacteria were harvested by centrifugation at 3000 × g for 10 min at 4 °C. All the following steps were carried out on ice or with pre-chilled solutions.

The cell pellets were resuspended in 10 mL CaCl₂-solution and spun at 3000 × g for 10 min at 4 °C. The centrifugation step was followed by the resuspension of the cells in 20 mL CaCl₂-solution and an incubation period of 30 min on ice. The bacteria were then harvested by centrifugation and resuspended in 5 mL CaCl₂-solution. After another short incubation period on ice for 15 min, the cell suspension was supplemented with 11% (v/v) of sterile glycerol and gently mixed by pipetting. As a final step, the cell suspension was split into aliquots of 200 µL in 1.5 mL reaction tubes, frozen in liquid nitrogen and stored at -80 °C.

3.1.4 Transformation of chemically competent *E. coli* with plasmid DNA

Chemically competent *E. coli* were transformed with a maximum of 100 ng plasmid DNA by using the heat shock method. After the DNA had been added to the competent cells, the cells were warmed at 42 °C for 1 min. Cells were then incubated on ice for another 2 min before 700 µL of LB-medium was added. An incubation period of 45–60 min at 37 °C with gentle agitation enabled the transformed bacteria to express the acquired genes coding for ampicillin resistance and therefore survive an antibiotic selective pressure. Subsequently, the bacteria were plated on LB-Amp-agar plates and grown overnight at 37 °C.

3.1.5 Storage of bacterial cells

Storage of bacteria was dependent on how they were cultivated and prepared, and their further use. Agar plates containing bacteria after transformation or bacteria streaked out from a cryo-stock were stored short-term at 4 °C for a maximum of one week. Cell pellets prepared for either protein or plasmid purification were stored in 50 mL conical centrifuge tubes at –20 °C for no more than one year. Chemically competent bacteria (Section 3.1.3) were stored at –80 °C as glycerol stocks containing 11% (v/v) glycerol and stored for up to a year. Recombinant strains were stored at –80 °C indefinitely as cryo-stocks prepared by the addition of 11% (v/v) glycerol.

3.2 Genetic methods

3.2.1 Preparation of plasmid-DNA

The preparation of plasmid DNA from *E. coli* strains Mach1 T1 or OmniMAX 2 T1R (Table 7) was carried out on a small scale (0.5–1.5 µg DNA) using the InnuPREP Plasmid Mini Kit or on a medium scale (10–50 µg DNA) using the QIAGEN Plasmid Midi Kit (Table 5). In both cases, the plasmids were extracted and purified according to the manufacturer's instructions. The extraction method employed by both kits is based on the alkaline lysis approach first described by Birnboim & Doly, 1979.

3.2.2 Agarose gel electrophoresis

Agarose gel electrophoreses (Sambrook & Russell, 2001) were performed for the preparative separation of DNA fragments, the analysis of restriction patterns of cleaved plasmid DNA and the analysis of PCR products. Gels were prepared with an agarose concentration of 1 %, which ensured sufficient separation capacity for all the before mentioned applications. Five μL ethidium bromide was added per 100 mL of agarose gel during gel preparation to enable visualisation and documentation of the separated DNA fragments using a Gel Doc XR+ gel documentation system (Table 1).

Before being loaded onto the gel, the samples were mixed with either 5 \times concentrated loading dye (Table 14) or commercially available DNA loading dye (6 \times) (Table 2).

Agarose gel electrophoreses were performed in TAE-buffer at a voltage of 120 V for 20–30 min until adequate separation of DNA fragments was achieved.

Table 14: Compositions of TAE-buffer and 5 \times concentrated loading dye

| TAE-buffer | | DNA loading dye (10 mL) (5 \times) | |
|---------------------------|-----------|---------------------------------------|-------------|
| Tris base | 4.84 g/L | Glycerol | 34% (v/v) |
| Acetic acid | 1.15 mL/L | Bromphenol blue | 0.05% (w/v) |
| 0.5 M EDTA solution, pH 8 | 2 mL/L | 0.5 M EDTA solution, pH 8 | 2 mL |

3.2.3 DNA extraction from agarose gels

Cleaved plasmids and PCR products were extracted from agarose gels using the NucleoSpin Gel and PCR Clean-up kit (Table 5). The extraction was performed according to the manufacturer's instructions. Samples containing extracted DNA were either stored at 4 °C for continued use or frozen at –20 °C for long-term storage.

3.3 Polymerase Chain Reaction

The polymerase chain reaction (PCR) (modified from Saiki *et al.*, 1988) was employed for the fast and specific amplification of the DNA fragments coding for both SAM_{wt} and SAM_C. The pGEX_6p_2_kombiP plasmids carrying the inserts coding for the respective longer SAM variants (SLY1_{249–321}) were used as templates. Both reactions used 100 ng of template DNA, 30 pmol of each primer, 20 nmol dNTPs and 2 U Vent DNA polymerase (Table 3) in a total volume of 50 μL . The PCR was performed in an iCycler

thermocycler (Table 1), according to the program described in Table 15. The amplified DNA was analysed by agarose gel electrophoresis (Section 3.2.2). In the case of minor unspecific amplification leading to more than one product, a preparative agarose gel was performed and only the desired product was extracted from the gel.

Table 15: PCR thermocycling program for nucleic acid amplification

| Step | T °C | Duration (min:s) | |
|----------------------|------|------------------|-------|
| Initial denaturation | 95 | 05:00 | |
| Denaturation | 95 | 00:30 | } ×30 |
| Annealing | 62 | 00:30 | |
| Elongation | 72 | 00:30 | |
| Final elongation | 72 | 10:00 | |

3.3.1 Site-specific mutagenesis by QuikChange-PCR

The exchange of S320 to a cysteine in SAM_{Ig} was performed by introducing a single nucleotide mutation into the wt *SlyISAM* coding sequence using the QuikChange II XL Site-Directed Mutagenesis Kit (Table 5) for circular mutagenesis. The plasmid pGEX-6p-2_SLY1SAM_{Ig} served as the template. The primer design (Table 10) followed the manufacturer's instructions. The components used in the 25 µL reaction mix are listed in Table 16:

Table 16: Composition of the 25 µL reaction mix for Site-specific mutagenesis by QuikChange

| | |
|--|--------|
| 10× reaction buffer | 2.5 µL |
| pGEX-6p-2_SLY1SAM _{Ig} | 10 ng |
| SLY1SAM_QC_S74C_fw | 6 pmol |
| SLY1SAM_QC_S74C_rv | 6 pmol |
| dNTP mix | 0.5 µL |
| Quik solution | 1.5 µL |
| ultrapure water to a final volume of 25 µL | |

The “Quik-Change”-PCR was performed in an iCycler thermocycler following the program described in Table 17.

Table 17: The PCR thermocycling program for site-specific mutagenesis.

| Step | T °C | Duration (min: s) | |
|----------------------|------|-------------------|-------|
| Initial denaturation | 95 | 01:00 | |
| Denaturation | 95 | 00:50 | } ×18 |
| Annealing | 60 | 00:50 | |
| Elongation | 68 | 05:30 | |
| Final elongation | 68 | 07:00 | |

Paternal DNA was digested post-amplification by adding 10 U of the methylation sensitive restriction endonuclease *DpnI* (Table 3) to the reaction mix and incubating this solution at 37 °C for 3 h. Subsequently, chemically competent *E. coli* XL-Gold cells (Table 7) were transformed with the amplified plasmid according to the manufacturer's protocol and plated on LB-Amp-agar plates (Section 3.1.1). Plates were incubated overnight at 37 °C and stored at 4 °C until further use.

3.3.2 Molecular cloning

PCR products and plasmids were cleaved at specific positions in their sequence using restriction endonucleases. Ten U of the restriction enzyme was used to cleave a maximum of 1 µg DNA. Cleavage of DNA was carried out at the recommended temperature for the enzymes used. The reaction was incubated for 1–2 h. The cleavage was followed by heat inactivation of the enzymes at 80 °C for 20 min. Cleaved DNA was further analysed by agarose gel electrophoresis, as described in Section 3.2.2 and, if required, purified by preparative gel electrophoresis and extracted from the gel (Section 3.2.3).

Linearised plasmids were dephosphorylated to increase the probability of successful insertion of the desired DNA fragments by preventing religation of the plasmids. Dephosphorylation was achieved by treating the cleaved plasmid with shrimp alkaline phosphatase (SAP) (Table 3) for 1.5 h at 37 °C, followed by a heat inactivation step at 60 °C. The plasmid and the desired DNA insert were ligated using T4 ligase following the standard protocol supplied by the manufacturer (Table 3). The ligation mix was subsequently used to transform chemically competent *E. coli* Mach1-T1 or OmniMAX 2 T1R cells (Table 7).

3.3.3 DNA sequencing

The cloning and mutagenesis results were checked by sequencing the samples at SeqLab, Göttingen, Germany. Samples were prepared to contain 0.5–1 µg of plasmid DNA in a total volume of 15 µL. For the sequencing of all pGEX_6p_2 derivatives, the sequencing primer pGEX-5' (Table 9) was used, which binds at the end of the coding sequence of GST.

3.4 Preparation of protein samples

3.4.1 Purification of SLY1 SAM variants from the cell extract

3.4.1.1 Buffers and solutions

All buffers and solutions used for the purification of the SLY1 SAM domain variants from cell extracts were prepared with ultrapure water, sterilised by passage of the solutions through a membrane filter (0.2 µm average pore diameter), and degassed in a vacuum desiccator for 30 min. DTT or TCEP was added after degassing for buffers that required these chemicals. The composition of buffers and solutions used in the SLY1 SAM domain purification procedure are listed in Table 18.

3.4.1.2 Preparation of the *E. coli* cell extract

Optimal lysis of the cells required thorough resuspension of the cells in Lysis-buffer (Table 18). Cell pellets which had been frozen and stored at –20 °C were thawed on ice for 1–2 h and resuspended in ice-cold phosphate buffered saline (PBS) (1×) containing cOmplete protease inhibitor. The cell suspension was subjected to three rounds of homogenisation using the benchtop homogeniser M-110P (Table 1) at a pressure of 20,000 psi to disrupt the bacterial cells. The cell suspension was incubated on ice between each homogenisation step for 1 min to prevent the sample from overheating. Afterwards, DNase and lysozyme were added to the homogenate at a concentration of 20 µg/mL and 100 µg/mL, respectively, and the solution was incubated at RT for 1 h on a rotary shaker. Following the incubation period, the lysate was cleared from cell debris and membrane fragments by centrifugation at 48,000 × g for 45 min at 11 °C, during which the soluble and insoluble fractions were separated. Immediately after centrifugation, the sediment was discarded and the cleared cell extract subjected to GSH-Sepharose affinity chromatography (Section 3.4.1.3).

Table 18: Buffers used for the purification of the various SLY1 SAM constructs

| 1× PBS (Sambrook & Russell, 2001) | |
|---|-----------|
| NaCl | 8 g/L |
| KCl | 0.2 g/L |
| KH ₂ PO ₄ | 0.24 g/L |
| Na ₂ HPO ₄ ·2H ₂ O | 1.805 g/L |
| pH 7.4 | |

| Lysis-buffer | |
|------------------------------------|----------------|
| 1× PBS | |
| DNase A | 20 µg/mL |
| Lysozyme | 100 µg/mL |
| cOmplete protease inhibitor tablet | 1 tablet/50 mL |

| Standard-buffer | |
|---------------------------------|----------|
| KH ₂ PO ₄ | 36.25 mM |
| K ₂ HPO ₄ | 13.75 mM |
| NaCl | 20 mM |
| EDTA | 0.2 mM |
| pH 6.4 | |

| Proteolysis-buffer | |
|---------------------------|--------|
| Tris | 50 mM |
| NaCl | 150 mM |
| DTT | 5 mM |
| pH 7 | |

| GST-binding-buffer | |
|---------------------------|------|
| 1× PBS | |
| DTT | 5 mM |

| Tris-buffer | |
|--------------------|--------|
| Tris | 50 mM |
| NaCl | 150 mM |
| pH 7 | |

3.4.1.3 Purification of GST-SAM fusion proteins from cells extracts by GSH-affinity chromatography

All SLY1 SAM domain variants were initially produced as fusion proteins containing an N-terminal glutathione-S-transferase-tag (GST-tag). The recombinant GST-SAM fusion proteins were purified from cell extracts (Section 3.4.1.2) by GSH-affinity chromatography. The stationary phase for this type of chromatography consists of reduced glutathione (GSH) coupled to a Sepharose matrix (Table 4). GSH binds GST with high affinity and specificity and therefore adsorbs GST-fusion proteins while other proteins and contaminants are removed by washing the resin material under mild conditions (e.g., neutral pH). All steps involved in the purification were performed on ice or with pre-chilled buffers.

The resin volume was chosen according to the size of overexpression cultures, where 3 mL of GST-resin was used for 1 L of expression culture. The resin was equilibrated in GST-binding-buffer (Table 18) and subsequently mixed with the cell extract in 50 mL conical centrifuge tubes. The slurry was incubated for 30–60 minutes at 10 °C to ensure that the majority of the GST-fusion protein was bound to the resin, followed by the separation of bound and unbound material by application to a gravity flow column. Contaminants were washed through the column matrix with 2×5 column volumes (CV) of GST-binding-buffer and 2×5 CV of proteolysis-buffer.

3.4.1.4 Cleavage of GST-fusion proteins

The GST-SAM protein bound to the resin was mixed with two CV of Proteolysis-buffer and transferred into a sterile 50 mL conical centrifuge tube. Two milligrams of PreScission protease per litre of expression culture was added to the slurry and gently mixed by pipetting up and down. The mixture was incubated overnight with gentle agitation at 10 °C. The slurry was then transferred back into the gravity flow column, and the buffer containing the isolated SAM domain was collected and kept on ice. The resin was washed with 5 CV of Proteolysis-buffer to remove any SAM domain that remained within the column matrix. Each wash fraction was collected separately and kept on ice. Cell lysis, purification and proteolysis were examined by SDS-PAGE (Section 3.5.1).

3.4.1.5 Preparative size exclusion chromatography

Preparative size exclusion chromatography was performed as a final purification and buffer exchange step in the purification of the SLY1 SAM domain constructs. After GSH-affinity chromatography, the fractions containing the SAM domain were pooled and concentrated to a final volume of 10 mL (Section 3.4.4). This sample was applied to a Hiload 26/60 Superdex 75 pg column (Table 4) connected to an FPLC system (Table 1) at 10 °C. The protein was eluted from the column using 1.5 CV of the appropriate buffer (Table 19) at a flow rate of 1.5 mL/min. The elution was fractionated in 5 mL fractions by a fraction collector. The elution of proteins was monitored by the absorbance of the eluent at $\lambda = 280$ nm. The fractions containing the SLY1 SAM domain were checked for purity by SDS-PAGE (Section 3.5.1), pooled and concentrated to the desired concentration. In the case of SAM_{Ig_C} and SAM_C, the

reducing agent present during the initial purification steps was removed after SEC by dialysis to enable the formation of the stabilised dimer. Finally, the samples were filtered and stored at 4 °C until further use.

Table 19: Buffers used as eluent for preparative scale size exclusion chromatography on SLY1 SAM

| Sample | Application | SEC-eluent |
|---------------------------------|-----------------------|------------------------------|
| SAM _{Ig} | NMR | Standard-buffer |
| SAM _{Ig} | MST | Tris-buffer |
| SAM _{Ig_C} | NMR | Standard-buffer (+5 mM TCEP) |
| SAM _{Ig_C} | MST | Tris-buffer (+5 mM DTT) |
| SAM _{Ig_C} (Alexa 488) | MST | Standard-buffer |
| SAM _{wt} | NMR | Standard-buffer |
| SAM _{wt} | X-Ray crystallography | Tris-buffer |
| SAM _C | NMR | Standard-buffer (+5 mM TCEP) |

3.4.2 Preparation of the fluorescent labelled SLY1 SAM domain

The SLY1 SAM domain constructs were labelled with fluorescent dyes either by maleimide-thiol-coupling or NHS-amine-coupling. SAM_{Ig_C} was labelled with the fluorescent dye Alexa Fluor 488 (Table 2) via maleimide-thiol-coupling. Alexa Fluor 488 in powder form was dissolved in DMSO to give a 10 mM fluorescent dye solution. The solution was split into aliquots of 10 µL each and lyophilised, which facilitated the handling of the dye and minimised the possibility of oxidation. Dissolving and aliquoting of the label was performed under hypoxic conditions in a glove box. Two hundred twenty-nine µL of a 43.6 µM SAM_{Ig_C} solution in Tris-buffer was treated for 1 h with a 10-fold excess of TCEP at RT to reduce possible disulfide bonds. The content of one aliquot of Alexa Fluor 488 was then redissolved in 10 µL DMSO and added to the protein solution. The protein-dye mixture was incubated in the dark for 2 days at RT. The removal of excess dye was performed by size exclusion chromatography. For the separation of Alexa Fluor 488-labelled SAM_{Ig_C} from the unbound fluorescent dye, the sample containing the labelled protein was loaded onto a Superdex 75 pg 10/30 HR column (Table 4) connected to an Äkta purifier chromatography system (Table 1). The sample was eluted with 1.5 CV of buffer at a flow rate of 0.5 mL/min. The elution of the fluorescent labelled protein and the unbound

Alexa Fluor 488 was monitored by the absorbance of the eluent at $\lambda = 280$ nm and at $\lambda = 495$ nm. Half-millilitre fractions were collected. The run was performed at RT. The fractions containing the protein were also examined by SDS-PAGE (Section 3.5.1).

SAM_{wt} was labelled using the Monolith Protein Labeling kit BLUE-NHS (Table 5) according to the manufacturer's instructions.

The degree of labelling (DOL) was calculated from the amount of total protein c_{prot} (mol L⁻¹) in the sample, as determined by the Bradford method (Section 3.5.4), and the amount of bound dye c_{dye} (mol L⁻¹) as determined by measuring the absorbance at $\lambda = 495$ nm (Section 3.5.3).

$$\text{DOL} = \frac{c_{\text{dye}}}{c_{\text{prot}}} \quad [11]$$

3.4.3 Buffer exchange by dialysis

For the exchange of buffers and the removal of reducing agents, the samples were dialysed against the desired buffer conditions using dialysis tubing with a molecular weight cut-off (MWCO) of 3 kDa. Dialysis was carried out at 10 °C or room temperature.

3.4.4 Raising the sample concentration by ultrafiltration

The protein concentration in a sample was increased by either pressure-based or centrifugal ultrafiltration. Volumes of 2–50 mL were concentrated in pressurised Amicon stirred cells (Table 1) using a regenerated cellulose ultrafiltration membrane with an MWCO of 3 kDa. Small volumes of 0.2–2 mL were concentrated in Amicon Ultra-0.5 Centrifugal Filter units with an MWCO of 3.5 kDa at 14,000 × g.

3.5 Analytical methods

3.5.1 SDS-polyacrylamide gel electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) is a method for the separation of proteins according to their molecular mass (Lämmli, 1970). SDS-PAGE was used for the analysis of protein samples regarding their composition, purity, and possible degradation.

Before loading samples onto an SDS-PAGE gel, samples were supplemented with SDS-PAGE loading buffer and heated to 95 °C for at least 5 min. For samples that contained proteins in aqueous solution, 30% (v/v) of 4× concentrated SDS-PAGE loading dye was added to the sample. Samples taken from bacterial cultures were harvested by centrifugation and resuspended in 1× concentrated SDS-PAGE loading dye. The number of cells in SDS-PAGE samples was kept constant to enable comparison between expression cultures. For the estimation of the molecular mass of the proteins, 5 µL of the unstained protein marker (Table 2) was loaded onto the gels.

The polyacrylamide gels consisted of a 5% (w/v) acrylamide stacking gel and a 15% (w/v) acrylamide separation gel (Table 20). SDS-PAGE gels were run in SDS-PAGE running buffer with an electric current of 40 mA per gel for ~40 min or until the dye front reached the end of the gel.

The gels were stained for 30 min in pre-warmed Coomassie staining solution and destained by boiling the gel in deionised water until the protein bands were clearly visible. Polyacrylamide gels were documented with the Gel Doc XR+ gel documentation system (Table 1). The composition of all buffers used in an SDS-PAGE is listed in Table 21.

Table 20: Composition of the stacking and the separation gels in SDS-PAGE

| Stacking gel (5% acrylamide) | | Separation gel (15% acrylamide) | |
|------------------------------|-------------|---------------------------------|--------------|
| Acrylamide | 4.85% (w/v) | Acrylamide | 14.55% (w/v) |
| N,N'-Methyldisacrylamid | 0.15% (w/v) | N,N'-Methyldisacrylamid | 0.45% (w/v) |
| Tris-HCl, pH 6.8 | 125 mM | Tris-HCl, pH 8.8. | 375 mM |
| SDS | 0.1% (w/v) | SDS | 0.1% (w/v) |
| APS | 0.1% (v/v) | APS | 0.1% (v/v) |
| TEMED | 0.1% (v/v) | TEMED | 0.1% (v/v) |

Table 21: Compositions of the buffers required for an SDS-PAGE

| SDS-PAGE loading dye (4×) | | Coomassie staining solution | |
|---------------------------|-----------|-------------------------------|-------------|
| Tris-HCl, pH 7.4 | 200 mM | 2-propanol | 25% (v/v) |
| Glycerin | 40% (v/v) | acetic acid | 10% (v/v) |
| SDS | 8% (w/v) | Coomassie Brilliant blue R250 | 0.05% (w/v) |
| β-mercaptoethanol | 8% (v/v) | | |

| SDS-PAGE running buffer | |
|-------------------------|------------|
| Tris-HCl, pH 8.3 | 50 mM |
| Glycine | 385 mM |
| SDS | 0.1% (w/v) |

3.5.2 Analytical ultracentrifugation

Molecular masses, as well as a binding constant for the monomer-dimer equilibrium, were determined by sedimentation equilibrium centrifugation in an analytical ultracentrifuge (Table 1) equipped with absorbance optics. Sample volumes of 120 μL were filled into standard aluminium double sector cells with 1.2 cm optical path length. Centrifugation was performed in an eight-hole-rotor at different speeds (28,000 rpm, 35,300 rpm, 42,700 rpm, 50,000 rpm) and a temperature of 30 °C. Radially dependent protein concentrations were recorded as absorbance values at 280 nm with a radial step size of 10 μm . For each radius position 20 single absorbance measurements were averaged. Spinning times for reaching equilibrium were estimated before and approved by checking scans recorded two hours apart after reaching equilibrium for identity.

Data evaluation was performed with the global fitting routine implemented in UltraScan II. The molecular mass of the monomer and the molar extinction coefficient at 280 nm were estimated with the ExPASy protparam tool (Gasteiger *et al.*, 2005) (Table 11). The following parameters were used: the partial specific volume of SAM_{wt} $\bar{v} = 0.7382 \text{ cm}^3/\text{g}$; solvent density $\delta = 1.0054 \text{ g}/\text{cm}^3$ and extinction coefficient at 280 nm $\epsilon_{280} = 2,980 \text{ M}^{-1} \text{ cm}^{-1}$. Data fits were attempted using a model of a monomer-dimer equilibrium of reversibly self-associating species, a model of one single ideal species (one-component model), and a model of two ideal, non-interacting species (two-component model). Starting from the simplest model, the data was fitted according to different models until a satisfactory fit was achieved.

3.5.3 Quantitation of analyte concentrations in solution by UV/Vis spectrophotometry

The concentration of DNA and protein in solution was determined by UV/Vis spectrophotometry. The method relies on the absorption of UV light by biomolecules at a characteristic wavelength (proteins: $\lambda = 280$ nm; DNA: $\lambda = 260$ nm). The concentration of the analyte was calculated from the absorbance of the sample according to the Lambert-Beer-Law:

$$\lg\left(\frac{I_0}{I_1}\right) = A_\lambda = \varepsilon_\lambda c l \quad [12]$$

Where I_0 is the radiant flux received by the sample, I_1 the radiant flux transmitted by the sample, A the absorbance at wavelength λ , c the concentration of the analyte in solution, ε_λ the molar extinction coefficient at the wavelength λ and l the path length of the cuvette. The molar extinction coefficient ε was estimated for each construct using the ExPASy protparam tool (Table 11). The extinction coefficient of double-stranded DNA is sequence-independent and given as an average mass extinction coefficient (μ/ρ_m) of 0.02 ($\mu\text{g}/\text{mL}$) $^{-1}$ cm^{-1} . The concentration of Alexa Fluor 488 (Table 2) was determined by measuring the absorbance at $\lambda = 495$ nm. Alexa Fluor 488 molar extinction coefficient at $\lambda = 495$ nm is $71,000$ M^{-1} cm^{-1} .

The protein concentration was routinely determined in a quartz cuvette using standard spectrophotometers (Table 1), whereas the DNA concentration was determined with a nanophotometer (Table 1)

3.5.4 Quantitation of the protein concentration using the Bradford method

Due to the comparatively high ε of Alexa Fluor 488 at $\lambda = 280$ nm of $\sim 9,500$ M^{-1} cm^{-1} , the concentration of SAM_{Ig_C} could not be determined by its absorbance at $\lambda = 280$ nm in the presence of the fluorescent dye. Thus, the protein concentration of fluorescent labelled SAM was determined using the Biorad Bradford protein assay (modified from Bradford, 1976; Table 5). The assay was performed according to the manufacturer's instructions for the standard protein assay in microtiter plates.

A protein standard was prepared from the SAM_{Ig} or SAM_{wt} stock solutions by diluting the sample with buffer to concentrations in the range of 90 to 10 μ M. The concentrations of the stock solutions were determined by UV/Vis spectrophotometry as described in Section 3.5.3. This concentration range corresponded to the likely concentration of the labelled protein in the sample and to the range of concentrations where the Bradford assay is most accurate (0.05–0.5 mg BSA/mL). Three times 10 μ L of each dilution were transferred into a microtiter plate. The sample buffer served as the reference. Three times 10 μ L of sample and also the 1:2 dilution of the sample were transferred into the microtiter plate. Six millilitres of protein assay reagent was mixed in a ratio of 1:4 with ultrapure water and filtered. Two hundred microliters of the diluted reagent were added to each well containing either the protein standard or the protein sample in the microtiter plate. The mixture was incubated for 5 min, and the absorbance at $\lambda = 595$ nm of the reaction mixtures was measured by a microtiter plate reader (Table 1).

3.5.5 Microscale thermophoresis

Microscale thermophoresis (MST) (Dühr & Braun, 2006; Jerabek-Willemsen *et al.*, 2011) was used to determine the K_D of the SLY1 SAM domain dimerisation. During a series of microscale thermophoresis experiments, the thermophoretic properties of a fluorescent labelled analyte are monitored as a function of the concentration of an unlabelled interacting species (ligand).

The thermophoretic properties of the analyte are reflected in the ratio F_{norm} of the fluorescence intensity measured after (F_{hot}) and before (F_{cold}) the induction of a temperature gradient at the boundary of a selectively heated spot in the sample. Without any ligand or at ligand concentrations significantly below the K_D , the fluorescent labelled analyte moves along the temperature gradient in a manner which is characteristic of its predominantly free, unbound state. A movement from warmer to colder areas is termed positive, the opposite movement negative thermophoretic behaviour. This thermophoretic effect leads to a characteristic increase or decrease of fluorescence in the heated sample area during the heating period. Complex formation with the ligand may change the thermophoretic behaviour of the fluorescent labelled analyte and may therefore result in a change of F_{hot} in the complexed versus F_{hot} in the free state. In this case, a change in the population of the free and the complexed state

upon addition of unlabelled ligand can be monitored by the change in $F_{\text{hot}}/F_{\text{cold}}$, i.e. F_{norm} .

The change in fluorescence intensity observed as a result of localised sample heating is not entirely due to thermophoresis, but also a result of the so-called temperature jump or T-Jump. The T-Jump reflects the temperature dependence of the fluorescence properties of the dye and occurs during the initial heating of the sample until a steady state temperature in the heated spot is established. This temperature equilibration process occurs rapidly (< 1 s) and may result in an increase or decrease of the fluorescence without any molecular motion. Thermophoresis in comparison is a slow process based on molecular diffusion that typically takes 20 to 50 s to establish a new quasi-equilibrium state. Binding events and complex formation can not only affect the thermophoretic properties of the analyte but also its response to the T-jump. By defining the intervals for measurement of F_{cold} and F_{hot} , the acquired data can either be dominated by T-Jump effects or thermophoresis or include both types of effects.

In the case of the experiments performed to characterise the SAM_{wt} and the SAM_{Ig} dimerisation, a low concentration of fluorescent labelled SAM_{wt} (70 nM) or SAM_{Ig_C} (50 nM) (Section 3.4.2) was added to samples containing increasing concentrations of unlabelled SAM_{wt} or SAM_{Ig} , respectively. The samples were prepared by a 1:2 serial dilution of the ligand stock solution (SAM_{wt} : 1.65 mM; SAM_{Ig} : 1.2 mM) using standard-buffer (Table 18). After an equilibration period in the dark of 30 minutes, the samples were transferred into specialised glass capillaries required for thermophoresis measurements in a Monolith NT.1.15 blue/red instrument (Table 1). For all experiments, the duration of the heating period ('Laser on' time) was 30 s and the duration of the re-equilibration period ('Laser off' time) was 5 s. Parameters including LED power and MST power were varied to optimise the results of an MST experiment. The fluorescence levels in the samples during the MST measurements were kept between 200 and 2000 units, as suggested by Nanotemper technologies (München, Germany). The optimal MST power was determined empirically by recording the experiment at different MST powers (i.e. 20%, 40%, 60%, 80%). The data recorded at the lowest level of MST power that provided a sufficient signal-to-noise ratio was used for the subsequent analysis and K_D fit. The analysis of the recorded data was performed using the manufacturer supplied NT Analysis software version 1.5.41 (Table 11) and Sigma Plot for the fitting of the data to a monomer-dimer equilibrium model

$$f(F_{\text{norm}}) = F_m + (F_d - F_m) * \left(1 + \frac{K_D}{4 * [A]_0} - \sqrt{\left(1 + \frac{K_D}{4 * [A]_0} \right)^2 - 1} \right) \quad [13]$$

where F_m is F_{norm} for the monomer state, F_d is F_{norm} for the dimer state and $[A]_0$ the total concentration of protein (labelled and unlabelled) in the sample. In the experiments conducted for the characterisation of the SLY1 SAM domain dimerisation, both thermophoresis and T-Jump effects were taken into account.

3.6 Nuclear magnetic resonance spectroscopy

All NMR experiments were recorded at 35 °C on Bruker spectrometers operating at ^1H frequencies of 600 (14.1 Tesla) and 700 MHz (16.4 Tesla) (Table 1). The spectrometers were equipped with cryogenically cooled 5 mm z-gradient $^1\text{H}/^{15}\text{N}/^{13}\text{C}$ probes to reduce thermal noise (Styles *et al.*, 1984). Calibration of the sample temperature was performed with a deuterated methanol sample (99.8% methanol- d_4), as described in Findeisen *et al.*, 2007. Proton chemical shifts were directly referenced to DSS. The ^{15}N and ^{13}C chemical shifts were indirectly referenced according to the $\gamma_{^{13}\text{C}}/\gamma_{^1\text{H}}$ and $\gamma_{^{15}\text{N}}/\gamma_{^1\text{H}}$ ratios given by Wishart, 2011.

Water suppression in the NMR experiments was achieved by water flip-back, watergate or gradient coherence selection methods. The detection modes used for quadrature detection in the indirect dimension were States (States *et al.*, 1982) or States-TPPI (Marion *et al.*, 1989).

The digital resolution of the recorded data was enhanced by zero filling to twice the number of acquired points and the application of a square-sine-bell apodisation function. All processing of the spectra was conducted with the program NMRPipe (Delaglio *et al.*, 1995) (Table 11). Visualisation and analysis of the spectra were carried out in NMRDraw (Delaglio *et al.*, 1995) and the program Analysis of the collaborative computational project for NMR (CcpNmr) (Vranken *et al.*, 2005) (Table 11), respectively. Selected acquisition parameters of the NMR experiments recorded are listed in Table 22.

Table 22: Selected acquisition parameters of the NMR experiments conducted for the SLY1 SAM domain resonance assignment and structure determination

| Experiment | sw1 (ppm) | t ₁ (complex points) | offset (ppm) | sw2 (ppm) | t ₂ (complex points) | offset 2 (ppm) | sw3 (ppm) | t ₃ (complex points) | offset3 (ppm) | Number of scans | Duration (h) | Reference |
|--|-----------|---------------------------------|------------------------|-----------|---------------------------------|---------------------------|-----------|---------------------------------|---------------------------|-----------------|--------------|---|
| 2D ¹ H, ¹⁵ N-HSQC | 16 | 1024 | 4.68 (¹ H) | 28 | 192 | 116.99 (¹⁵ N) | | | | 48 | 8.75 | Palmer <i>et al.</i> , 1991 Kay <i>et al.</i> , 1992 |
| 2D ¹ H, ¹³ C- CT-HSQC | 16 | 1024 | 4.68 (¹ H) | 74 | 320 | 42.64 (¹³ C) | | | | 32 | 9.5 | Palmer <i>et al.</i> , 1991 Vuister & Bax, 1992 |
| 3D HNCO | 16 | 1024 | 4.68 (¹ H) | 27 | 32 | 116.99 (¹⁵ N) | 13 | 64 | 177.11 (¹³ C) | 8 | 27 | Sattler <i>et al.</i> , 1999 |
| 3D HNCA | 16 | 1024 | 4.68 (¹ H) | 27 | 48 | 117 (¹⁵ N) | 29 | 64 | 57.64 (¹³ C) | 8 | 40.75 | |
| 3D HNCACB | 16 | 1024 | 4.68 (¹ H) | 27 | 36 | 116.99 (¹⁵ N) | 62 | 64 | 45.64 (¹³ C) | 16 | 60.5 | |
| 3D CBCA(CO)NH | 16 | 1024 | 4.68 (¹ H) | 27 | 36 | 117 (¹⁵ N) | 60 | 64 | 48.64 (¹³ C) | 8 | 118.5 | |
| 3D C(CO)NH | 16 | 1024 | 4.68 (¹ H) | 27 | 25 | 116.99 (¹⁵ N) | 67 | 64 | 42.64 (¹³ C) | 16 | 30.75 | |
| 3D H(CCO)NH | 16 | 1024 | 3.48 (¹ H) | 27 | 25 | 116.99 (¹⁵ N) | 6,5 | 64 | 1.88 (¹ H) | 16 | 43 | |
| 3D TOCSY- ¹ H, ¹⁵ N-HSQC | 16 | 1024 | 4.68 (¹ H) | 27 | 32 | 116.99 (¹⁵ N) | 14 | 64 | 4.68 (¹ H) | 16 | 54.5 | |
| 3D HCAN | 16 | 1024 | 4.68 (¹ H) | 30 | 64 | 55.64 (¹³ C) | 27 | 32 | 116.99 (¹⁵ N) | 16 | 53.5 | |
| 3D HCCH-TOCSY | 7 | 512 | 3.28 (¹ H) | 7 | 105 | 1.48 (¹ H) | 67 | 41 | 42.64 (¹³ C) | 16 | 111 | |
| 2D (HB)CB(CGCD)HD | 16 | 1024 | 4.68 (¹ H) | 40 | 55 | 41.64 (¹³ C) | | | | 8 | 0.25 | |
| 2D (HB)CB(CGCDCE)HE | 16 | 1024 | 4.68 (¹ H) | 40 | 55 | 41.64 (¹³ C) | | | | 8 | 0.25 | |
| 3D NOESY- ¹ H, ¹⁵ N-HSQC | 16 | 1024 | 4.68 (¹ H) | 27 | 32 | 116.99 (¹⁵ N) | 12 | 90 | 4.68 (¹ H) | 16 | 82.75 | |
| 3D NOESY- ¹ H, ¹³ C-HSQC (aliphatic) | 14 | 1024 | 4.68 (¹ H) | 67 | 55 | 42.64 (¹³ C) | 11 | 80 | 4.68 (¹ H) | 16 | 117 | |
| 3D NOESY- ¹ H, ¹³ C-HSQC (aromatic) | 16 | 1024 | 4.68 (¹ H) | 28 | 36 | 128.13 (¹³ C) | 8 | 66 | 4.68 (¹ H) | 16 | 63.75 | |
| 3D ¹⁵ N, ¹³ C- ω ₁ -filtered NOESY- ¹ H, ¹³ C -HSQC | 16 | 1024 | 4.68 (¹ H) | 65 | 25 | 41.64 (¹³ C) | 11 | 64 | 4.68 (¹ H) | 32 | 88.5 | |
| 3D ¹⁵ N, ¹³ C-ω ₁ -filtered NOESY- ¹ H, ¹⁵ N-HSQC | 16 | 1024 | 4.68 (¹ H) | 25,5 | 11 | 115.59 (¹⁵ N) | 12 | 64 | 4.68 (¹ H) | 48 | 60.75 | |
| 3D HAHB(CACO)NH | 16 | 1024 | 4.68 (¹ H) | 24 | 21 | 115.69 (¹⁵ N) | 8 | 64 | 3.2 (¹ H) | 40 | 69 | |
| 3D HNHB | 16 | 1024 | 4.68 (¹ H) | 27 | 32 | 116.99 (¹⁵ N) | 11 | 64 | 4.68 (¹ H) | 16 | 54 | Düx <i>et al.</i> , 1997 |
| 2D long range ¹³ J _{NC'} HNCO | 16 | 1024 | 4.68 | 11 | 65 | 177.1 | | | | 1920 | 136 | Cordier & Grzesiek, 1999 |

3.6.1 NMR sample preparation

After preparative size exclusion chromatography (Section 3.4.1.5) the sample containing the isotope-labelled SAM domain was concentrated to the desired concentration and dialysed against NMR-buffer. The NMR-buffer had the following composition (unless reported otherwise):

Table 23: Composition of NMR-buffer used for NMR spectroscopy on the SLY1 SAM domain

| NMR-buffer | |
|---------------------------------|-------------|
| KH ₂ PO ₄ | 36,25 mM |
| K ₂ HPO ₄ | 13.75 mM |
| NaCl | 20 mM |
| EDTA | 0.2 mM |
| NaN ₃ | 0.03% (w/v) |
| pH 6.4 | |

The reducing agent TCEP was added to the sample when required to disrupt and inhibit reformation of disulfide bridges. Seven per cent D₂O was added to the NMR samples during sample preparation.

3.6.2 Assignment of resonances

Sequential resonance assignment information was obtained from *J*-correlated 3D triple resonance experiments. Resonance assignment was performed manually using the program CcpNmr Analysis. Backbone nuclei (¹H^N, ¹⁵N; ¹H^α, ¹³C^α, ¹³C^γ) and ¹³C^β were assigned using 3D HNC(O), HNCA, CBCA(CO)NH, HNCACB experiments, whereas the assignment of aliphatic side chain (¹H, ¹³C) nuclei was conducted using C(CO)NH, H(CCO)NH, TOCSY-¹H,¹⁵N-HSQC and HCCH-TOCSY experiments (Table 22).

The sequence-specific assignment of resonances in heteronuclear NMR spectra of a protein usually starts with the assignment of the backbone nuclei. This procedure involves assigning resonances by linking backbone nuclei of residues in the protein sequence with their sequential neighbours. This assignment information is transferred to the 2D ¹H,¹⁵N-HSQC spectrum (¹H^N and ¹⁵N assignments) and partly to the 2D ¹H,¹³C-CT-HSQC spectrum (i.e., ¹³C^α and ¹³C^β nuclei and the directly bound ¹H^α and ¹H^β nuclei). The 3D triple resonance HNCA and HNCACB experiments correlate the amide proton and nitrogen chemical shifts with the chemical shifts of the intra (*i*) and inter

($i-1$) $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ nuclei (only the HNCACB). The resonances derived from $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ nuclei in the HNCACB experiment are distinguishable by the 180° phase difference, which results in one set of resonances being positive and the other negative. The signals arising from inter-residue correlations between the amide group and the carbons of the preceding residue can usually be identified by the weaker intensity when compared with the resonances arising from intramolecular correlations. Nonetheless, the unambiguous distinction between intra- and inter-correlations for each $^1\text{H}_i-^{15}\text{N}_i$ are achieved by recording the 3D CBCA(CO)NH which only transfers magnetisation between the $^1\text{H}_i$ and $^{15}\text{N}_i$ and the $^{13}\text{C}_{i-1}^\alpha$ and $^{13}\text{C}_{i-1}^\beta$ nuclei.

Particular amino acids have an easily recognisable carbon chemical shift pattern, which makes these amino acids distinguishable from other residues. For example, threonine $^{13}\text{C}^\beta$ exhibits a downfield chemical shift because of the strong deshielding effect of the hydroxyl side chain moiety (Fig. 27), whereas alanine $^{13}\text{C}^\beta$ resonances are shifted upfield because of the strong shielding effect of the attached methyl hydrogens. Glycine gives rise to only a $^{13}\text{C}^\alpha$ resonance in the HACBCA/CBCA(CO)NH spectra. These amino acid resonance patterns (and others) can be used for orientation and starting points during sequential assignment (also termed sequential walk) of the spectra. The backbone carbonyl carbon ($^{13}\text{C}'$) was assigned using the 3D HNCO experiment and the assigned 2D $^1\text{H},^{15}\text{N}$ -HSQC spectrum. This 3D experiment correlates the amide (i) chemical shift with the inter-residue $^{13}\text{C}'_{i-1}$. The chemical shift information on the nitrogen of P300 was obtained by recording an HCAN spectrum, which correlates $^1\text{H}^\alpha$ with $^{15}\text{N}_i$ and $^{15}\text{N}_{i+1}$ nuclei and therefore bypasses the use of the amide proton for identification of the amide nitrogen chemical shifts.

The 3D triple resonance TOCSY- $^1\text{H},^{15}\text{N}$ -HSQC, C(CO)NH and H(CCO)NH experiments correlate the side chain carbon and proton resonances to the amide group resonances. In the case of the 3D TOCSY- $^1\text{H},^{15}\text{N}$ -HSQC, the correlated side chain protons and the amide group belong to the same residue (i), whereas the 3D C(CO)NH and H(CCO)NH correlate the amide group with side chain carbons and protons of the preceding residue ($i-1$). Using the assignment information obtained for the backbone $^1\text{H}_i-^{15}\text{N}_i$ correlations, the side chain resonances were first assigned to a specific amino acid type. As a second step, the majority of the side chain resonances were assigned to the respective nuclei with the help of typical ranges of chemical shifts known for each nucleus available from the Biological magnetic resonance data bank (BMRB) (Seavey et al., 1991) (these chemical shift distributions are derived from protein structures

solved by NMR spectroscopy). In the case of ambiguous assignments, the 3D HCCH-TOCSY was used to distinguish resonances belonging to the same side chain with very similar chemical shifts, e.g. methyl protons of leucine side chains.

As the last step, all assignments were transferred to the 2D $^1\text{H}, ^{13}\text{C}$ -CT-HSQC spectra because these spectra provide a higher resolution than the 3D experiments in the indirect dimensions (^{13}C , ^{15}N), i.e., higher chemical shift precision (Table 22). The assigned 2D heteronuclear chemical shift correlation spectra served as a basis for the assignment of the 3D-NOESY experiments, the long range HNC0 experiment (Section 3.6.5) and the 3D HAHB(CACO)NH and HNHB (Section 3.6.4.2) recorded for the structure calculation of the SAM_C cross-linked dimer.

Aromatic ring protons, as well as proline amide nitrogens, were assigned using 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE, and 3D HCAN experiments, respectively.

3.6.3 Chemical shift perturbation analysis

The average weighted chemical shift perturbation between two resonances in the recorded 2D $^1\text{H}, ^{15}\text{N}$ -HSQC spectra was calculated according to Equation [14] (Grzesiek *et al.*, 1997):

$$\Delta\delta_{\text{ave}} = \frac{1}{2} \sqrt{(\Delta\delta_{\text{HN}})^2 + \left(\frac{\Delta\delta_{\text{N}}}{5}\right)^2} \quad [14]$$

With $\Delta\delta_{\text{ave}}$ being the normalised average chemical shift perturbation of a resonance in the $^1\text{H}^{\text{N}}$ and ^{15}N dimension, $\Delta\delta_{\text{H}}$ the chemical shift perturbation in the proton dimension and $\Delta\delta_{\text{N}}$ chemical shift perturbation in the nitrogen dimension.

3.6.4 Determination of torsion angle restraints

Torsion angle restraints used in the structure calculation were determined by the two strategies described in the following sections.

3.6.4.1 Chemical shift derived backbone torsion angle restraints

The relationship between secondary chemical shifts and backbone torsion angles is well established (1.5.4). Various software tools employ empirical relationships to predict the backbone torsion angles of proteins solely from chemical shift data (Wishart, 2011). The program TALOS+ (Torsion angle likelihood obtained from shift and sequence similarity +) (Shen *et al.*, 2009) predicts the protein backbone torsion angles φ and ψ from the ^{15}N , ^{13}C , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^1\text{H}^\text{N}$ and $^1\text{H}^\alpha$ nuclei chemical shifts. The prediction is based on TALOS+ screening short overlapping tri-peptides consisting of the SAM_C primary sequence against a curated database. The database contains protein entries comprised of nearly complete backbone assignments complimented by high-resolution X-ray structures, i.e. torsion angle information. The chemical shift and sequence of the query peptides are matched against those of the peptides available in the database. The averages of φ - and ψ -angles of the centre residue of the ten best matches found in the database are assigned as the backbone torsion angles of the centre residue of the query tri-peptide. As long as the standard deviation (σ) of the ten φ - and ψ -angles are below 45° , the prediction is considered to be reliable. In that case, TALOS+ assigns the attribute GOOD to the prediction. Otherwise, the prediction is considered to be AMBIGUOUS. Predictions that have a particularly large σ are given the attribute BAD, those with a high random coil chemical shift contribution (*random coil index* –RCI; Berjanskii & Wishart, 2005) are termed DYNAMIC. Only reliable (GOOD) dihedral angle predictions were used in the structure calculations.

3.6.4.2 χ_1 -angle restraints from scalar couplings

χ_1 side chain torsion angle restraints were extracted from the coupling constants of the ^{15}N - $^{13}\text{C}^\alpha$ - $^{13}\text{C}^\beta$ - $^1\text{H}^\beta$ and the $^1\text{H}^\alpha$ - $^{13}\text{C}^\alpha$ - $^{13}\text{C}^\beta$ - $^1\text{H}^\beta$ vicinal scalar couplings (1.5.5). With the help of the empirically derived Karplus relations between χ_1 -angles and the $^3J_{\text{HaH}\beta 2/3}$ - and $^3J_{\text{NH}\beta 2/3}$ -coupling constants (Case *et al.*, 1994; Pérez *et al.*, 2001), the experimentally determined 3J -coupling constants were translated into one of the three favourable rotamer states of the C_α - C_β bond. The magnitude of the 3J -couplings was obtained from modified versions of the quantitative 3D HAHB(CACO)NH and HNHB experiments (Table 22). In these experiments, the magnetisation is transferred from either the $^1\text{H}^\alpha$ (HAHB(CACO)NH) or the ^{15}N (HNHB) backbone nuclei to the $^1\text{H}^\beta$ protons. The magnetisation transfer leads to a reduction of the reference ($^1\text{H}^\alpha$ or $^1\text{H}^\text{N}$)

cross-peak intensity which in turn causes an increase in the intensity of just one or both H^β cross-peaks. The efficiency of this transfer is relying on the size of the 3J -coupling. The more efficient the magnetisation transfer, the stronger is the intensity of the ${}^1H^\beta$ cross-peak in comparison to the reference peaks, which corresponds to a larger 3J -constant.

The ${}^3J_{HaH\beta 2/3}$ - and ${}^3J_{NH\beta 2/3}$ -couplings for each H^β proton were calculated from the ratio of the peak intensities of the respective H^β proton cross-peak to the correlated reference peak according to the equation [15]

$$\frac{I_c}{I_d} = -\tan^2(T\pi J) \quad [15]$$

Where I_c and I_d are the intensities of the cross-peaks and diagonal peaks, respectively. J is the coupling constant in Hz and T a delay period which allows the evolution of the 3J -coupling during the NMR experiment. Table 24 provides the criteria for assigning a *trans* (180°), *gauche* ($-$) (60°) or *gauche* ($+$) (-60°) staggered conformation for the χ_1 -angle for a particular residue based on the determined 3J -coupling values. The most abundant conformation is *gauche* ($+$), which has the gamma side chain atoms opposite to the residue main chain carbonyl group when viewed along the $C^\alpha-C^\beta$ bond. It is therefore energetically favourable. The least favoured is *gauche* ($-$), which occurs when the side chain is opposite the hydrogen substituent on the C^α atom. This confirmation is unstable because the gamma atom is in close contact with the main chain carbonyl and amide groups. The tolerance given for each determined conformational restraint was $\pm 30^\circ$.

Table 24: Empiric relationship between the rotameric state of the $C^\alpha-C^\beta$ bond and the strength of the ${}^3J_{\text{HaH}\beta 2/3}$ - and the ${}^3J_{\text{NH}\beta 2/3}$ -couplings (Case *et al.*, 1994)

| Rotamer | | | |
|-----------------------------|-------------------|--------------|-------------------|
| | <i>gauche</i> (-) | <i>trans</i> | <i>gauche</i> (+) |
| | 60° | 180° | -60° |
| χ_1 | | | |
| ${}^3J_{\text{HaH}\beta 2}$ | ~3.4 Hz | ~3.4 Hz | ~12.9 Hz |
| ${}^3J_{\text{HaH}\beta 3}$ | ~3.4 Hz | ~12.9 Hz | ~3.4 Hz |
| ${}^3J_{\text{NH}\beta 2}$ | ~5 Hz | ~1 Hz | ~1 Hz |
| ${}^3J_{\text{NH}\beta 3}$ | ~1 Hz | ~1 Hz | ~5 Hz |

3.6.5 Hydrogen bond coupling experiments

Hydrogen bond restraints were obtained by directly measuring the hydrogen bond couplings (${}^{\text{h}3}J_{\text{NC}}$) between the amide nitrogen and the carbonyl carbon of the protein backbone. A 2D version of the long range HNCO experiment, as described in (Cordier & Grzesiek, 1999), measures a scalar coupling between the donor amide nitrogen and the acceptor carbonyl carbon across the hydrogen bond. Unambiguously detected and assigned ${}^{\text{h}3}J_{\text{NC}}$ couplings were included in the structure calculation as an additional set of distance restraints. Each experimentally confirmed hydrogen bond was accounted for by distance restraints of 1.8 Å between the H^{N} and the O atoms and 2.8 Å between the N and the O atoms with an upper distance limit of + 0.5 Å.

3.6.6 NOE-derived distance restraints

Interatomic distances in the protein were determined by measuring the magnetisation transfer from one proton spin to another through space due to the nuclear Overhauser effect (NOE) (Section 1.5.6). ${}^{15}\text{N}$ - and ${}^{13}\text{C}$ -edited NOESY-HSQC spectra were recorded to obtain both intra- and intermolecular NOEs. Exclusively intermolecular distance information was obtained by recording 3D ${}^{15}\text{N}$, ${}^{13}\text{C}$ - ω_1 -filtered, ${}^{13}\text{C}$ -edited and ${}^{15}\text{N}$ -edited NOESY-HSQC spectra. NOESY spectra were recorded with mixing times as listed in Table 25.

Table 25: Mixing times used during the acquisition of the five 3D NOESY-HSQC experiments

| Experiment | Mixing times (ms) |
|--|-------------------|
| 3D NOESY- ¹ H, ¹⁵ N -HSQC | 150 |
| 3D NOESY- ¹ H, ¹³ C-HSQC (aliphatic) | 120 |
| 3D NOESY- ¹ H, ¹³ C-HSQC (aromatic) | 140 |
| 3D ¹⁵ N, ¹³ C- ω_1 -filtered NOESY- ¹ H, ¹³ C -HSQC | 150 |
| 3D ¹⁵ N, ¹³ C- ω_1 -filtered NOESY- ¹ H, ¹⁵ N -HSQC | 200 |

The isotope-filtered experiments were recorded using a sample that contained an equimolar amount of ¹⁵N/¹³C double-labelled and unlabelled SAM_C protein with a total SAM_C monomer concentration of 0.8 mM.

3.6.7 Surface mapping using paramagnetic relaxation enhancers

Dipolar interactions with the unpaired electron(s) of a chelated lanthanide ion (e.g., Gd³⁺) dissolved in the buffer cause a concentration-dependent increase of the nuclear relaxation rates, which results in additional line broadening of the NMR signal. Solvent-exposed nuclei, such as those on the surface of a protein, should be more susceptible to paramagnetic relaxation enhancement (PRE) and thus show the strongest reduction in resonance intensity. Nuclei that are buried in the core of the protein or shielded by a second monomeric unit at the interface should be less affected by PRE (Göbl *et al.*, 2014).

A set of 2D ¹H,¹⁵N-HSQC of a sample containing 200 μ M cross-linked SAM_C were recorded with varying concentrations (0–6 mM) of Gadodiamide (Trade name: Omniscan; Table 2), to identify the SAM_C dimer interface. Omniscan contains the gadolinium cation (Gd³⁺) complexed with 2-[bis[2-(carboxylatomethyl)-(methylcarbamoylmethyl)amino]ethyl]amino]acetate.

3.6.8 Semiautomatic NOE resonance assignment and structure calculation

Resonance assignment of NOESY spectra was performed in a semi-automated fashion, in which the HSQC- (through-bond) dimensions of the cross-peaks in the NOESY spectra were manually assigned in CcpNmr Analysis using the assigned 2D HSQC spectra (3.6.2). Regions with strong spectral overlap were not assigned. The NOE

dimension in the NOESY spectra was initially left unassigned except for unambiguous resonance assignments in the isotope-filtered experiments.

The assignment of the NOE dimension was performed as a second step in conjunction with the structure calculation and was carried out using the program ARIA (Ambiguous Restraints for Iterative Assignment) (Rieping *et al.*, 2007) in combination with CNS (Crystallography and NMR system) (Brünger *et al.*, 1998) and CcpNmr Analysis (Table 11). Distance restraints were obtained by integrating the peak volume using the truncated box sum model. The structure determination by ARIA and CNS is an iterative process of NOE resonance assignments, peak integration and distance calibration, followed by the structure calculation using restrained molecular dynamics (rMD) and simulated annealing (SA). The assignments are subsequently re-evaluated based on the calculated structures. This process is repeated eight times with each iteration refining the data from the previous step. The iterative character of the structure calculation protocol allows ARIA the use of highly ambiguous distance restraints (ADRs) (Nilges & O'Donoghue, 1998). Unambiguous NOEs only have one assignment option. In contrast, ambiguously assigned NOE cross-peaks arise because of: (i) contributions from multiple NOEs involving several nuclei or (ii) NOEs with more than one assignment option that could not be unambiguously assigned. ARIA initially allows all assignment options for NOE cross-peaks based on matching chemical shifts. During the iterative process, ARIA reduces assignment ambiguity. This concept of ADRs is especially helpful for the structure determination of symmetric homo oligomers, where symmetric degeneracy of the NMR spectra prohibits an *ab initio* distinction between intra- and intermolecular NOEs. The final step is an energy minimisation in water as explicit solvent. Restraints that do not agree with the final calculated structure ensemble are flagged as violations. In the case of distance restraint violations, the corresponding cross-peaks were manually checked for alternative assignments of the HSQC-dimensions that would be in better agreement with the structural ensemble. Violations that clearly arose from peak picking spectral artefacts such as t_1 noise ridges and NOEs arising through spin diffusion were removed. ARIA allows the adjustment of several parameters to minimise the number of ambiguous assignments and restraint violations as well as to yield better structural convergence. These parameters will be described in more detail in the following sections.

3.6.8.1 Parameters settings for NOE cross-peak assignment

A list of the chemical shifts of all assigned nuclei and lists of picked NOE cross-peaks that had been HSQC-base assigned were used by ARIA to assign the NOEs. The following user adjustable parameters have been used to optimise the assignment process (Table 26). Since the unambiguous manual assignment of NOE cross-peaks has been limited to clearly unambiguous assignments, the *Use manual assignment* option was enabled in all five NOESY spectra. In this case, ARIA does not change the manual assignments. The *Filter diagonal peaks* option excludes all diagonal peaks from the structure calculation. Diagonal peaks do not provide any structural information. The *upper bound correction* was set to 6.5 Å. The activation of the *Upper bound correction* leads to a re-evaluation of previously rejected cross-peaks under the premise that they satisfy the more liberal 6.5 Å criterion. The nuclear frequency windows define the maximum allowed deviation of the resonance frequency of an NOE from the chemical shift provided in the chemical shift list. A too large frequency window increases the ambiguity of the NOE resonance assignments, whereas a too small value might prevent (correct) assignment. The *Ambiguity level* defines whether NOE resonances originate from intra- or intermolecular magnetisation transfer. Therefore, the *Ambiguity level* was set to intermolecular for the two ^{15}N , ^{13}C -filtered NOESY spectra and left undefined for the others.

Table 26: Adjustable parameters in ARIA 2.3 for the assignment, conversion and use of the NOE data in the recorded NOESY spectra

| | 3D NOESY- ^1H , ^{15}N - HSQC | 3D NOESY- ^1H , ^{13}C -HSQC (aliphatic) | 3D NOESY- ^1H , ^{13}C -HSQC (aromatic) | 3D ^{15}N , ^{13}C - ω_1 - filtered NOESY- ^1H , ^{13}C -HSQC | 3D ^{15}N , ^{13}C - ω_1 - filtered NOESY- ^1H , ^{15}N -HSQC |
|------------------------------------|---|--|---|--|--|
| Use manual assignment? | yes | yes | yes | yes | yes |
| Trust assigned peaks? | yes | yes | yes | yes | yes |
| Filter diagonal peaks | yes | yes | yes | yes | yes |
| Upper bound correction (Å) | yes (6.5) | yes (6.5) | yes (6.5) | yes (6.5) | yes (6.5) |
| Lower bound correction | no | no | no | no | no |
| Proton freq. window (direct) (ppm) | 0.04 | 0.04 | 0.04 | 0.06 | 0.05 |

| | | | | | |
|---|---------|---------|---------|---------------------|---------------------|
| Proton freq. window (indirect) (ppm) | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 |
| Hetero freq. window (indirect) (ppm) | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| Ambiguity level | Unknown | Unknown | Unknown | Intermolecular only | Intermolecular only |

3.6.8.2 Symmetry

If the option *Symmetry enabled* is activated, a number of symmetry-related parameters can be specified. For a symmetric homodimer, the symmetry type would be C2, which describes a 180° rotation about the symmetry axis. Enabling *packing restraints* keeps the two monomers in close proximity to each other during simulated annealing. *NCS restraints* minimise the RMSD differences between the two monomers.

Table 27: Symmetry restraints parameter setting for a homodimer with C₂ symmetry

| | |
|----------------------------|-----|
| Symmetry enabled | yes |
| Symmetry type | C2 |
| Number of monomers | 2 |
| NCS restrained enabled | yes |
| Packing restraints enabled | yes |

3.6.8.3 Disulfide bond

The intermolecular disulfide bond between C320 of the two monomers was implemented into the ARIA project as a predefined covalent bond.

3.6.8.4 Network anchoring

Network anchoring reduces the number of possible assignments of a given NOE cross-peak. Under the assumption that NOE cross-peak assignments have to be consistent with the assignments of neighbouring nuclei, the assignment options are ranked according to their conformity to the network of all distance restraints for the surrounding nuclei. Assignment options that are incompatible with the network are rejected. Network anchoring is usually enabled for the first four iterations (i.e., 0–3) of NOE assignment and structure calculation (Bardiaux *et al.*, 2009).

3.6.9 Structure calculation using restrained molecular torsion angle dynamics simulations

The program CNS employs an rMD strategy using either torsion angle dynamics (TAD) or Cartesian dynamics with a simplified MD forcefield complemented by experimentally determined restraints for structure calculation in conjunction with SA (Bardiaux *et al.*, 2012). In the case of the SAM_C cross-linked dimer structure calculation, the TAD protocol was chosen. TAD has become the standard in structure calculation due to reduced computational effort (i.e., reduced degrees of freedom when compared with that of Cartesian space dynamics) and a high success rate (Güntert *et al.*, 1997; Nilges & O’Donoghue, 1998; Stein *et al.*, 1997).

Experimental data such as NOEs or torsion angle values are converted into structural restraints with the help of energy functions, which penalise violations by increasing the energy term of the corresponding structure. The log-harmonic-potential implemented in ARIA 2.3 was used as the penalty function for distance restraints in structure calculations of SLY1 SAM in combination with automatic Bayesian weighting of the input restraints (Nilges *et al.*, 2008).

The parameters of the simulated annealing protocol had to be adjusted to allow convergence of the dimer structure. The high temperature time step was reduced from 3 to 2 fs, and the number of high steps, refine steps and cooling steps were increased. The parameters used in the final structure calculation are listed in Table 28.

Table 28: Parameters of the CNS-based simulated annealing protocol used for the SAM_C dimer structure calculation

| Type | Torsion angle |
|--------------------------------|---------------|
| Random seed | 89,764,443 |
| TAD high temperature | 10,000 K |
| TAD time-step factor | 9.0 |
| Time-step | 0.002 ps |
| Cool1 final temperature | 1,000 K |
| Cool2 final temperature | 50 K |
| High-temp steps | 15,000 |
| Refine steps | 8,000 |
| Cool1 steps | 20,000 |
| Cool2 steps | 20,000 |

3.6.10 Analysis and evaluation of the NMR derived structure ensembles

The overall quality of the calculated structures was validated using a variety of software tools listed in Table 11. First and foremost, a low number of restraint violations indicated a good agreement of the calculated structure ensemble with the provided structural data. Therefore, a list of distance restraint violations provided by ARIA post structure calculation was imported into CcpNmr Analysis and was checked manually for consistently violated distance restraints. Torsion angle restraint violations were calculated using CcpNmr Analysis based on the final 15 models refined in explicit solvent. Also, the agreement of the experimental data with the calculated structure was assessed by RPF (Recall, Precision and F-measure) score analysis (Huang *et al.*, 2005; Huang *et al.*, 2012). The RPF scores provide statistical measures of how well the NOE peak-list is explained by the resulting structure (Recall) and whether the number of observed NOEs in the peak-list is sufficient to explain the resulting structure (Precision). The F-score combines the Recall and Precision scores. The definition of the scores is described elsewhere in more detail (Huang *et al.*, 2005). The RPF algorithm is implemented as a subroutine in CcpNmr Analysis and makes use of the RPF web server. Isotope filtered/edited NOESY spectra were excluded from the RPF analysis because of their low sensitivity.

Further criteria examined during the assessment of the structure quality were the number of backbone torsion angle outliers, and the RMSD of atomic coordinates of the superimposed 15 models in the water refined structure ensemble. For the determination of the RMSD of the backbone and side chain heavy atoms, the 15 lowest energy structures were superimposed over the well-defined regions of the structure ensemble. The well-defined regions exclude all residues for which the $^{13}\text{C}^{\alpha}$ RMSD exceeds the mean $^{13}\text{C}^{\alpha}$ RMSD over all residues by more than 2σ (Linge & Nilges, 1999; Nilges *et al.*, 1987). Backbone and sidechain torsion angles were analysed using the programme PROCHECK_NMR (Laskowski *et al.*, 1996). Low scores in all three categories indicated convergence to the correct 3D structure. Disulfide dihedral angles and their dihedral energies were calculated using the atom coordinates and the Disulfide Bond Dihedral Angle Energy Server. Secondary structure elements were predicted based on the atom coordinates using the programme DSSP (Define Secondary Structure of Protein) (Kabsch & Sander, 1983; Touw *et al.*, 2015).

3.7 X-ray crystallography

In X-ray crystallography, the structure of a protein is determined in a crystallised, ordered state, which is a prerequisite for the recording of X-ray diffraction data. The derivation of structural information from the diffraction data requires several steps before and after conducting the X-ray diffraction experiment, which consists of sample preparation and protein crystallisation (Section 3.7.1), data recording, data processing (Section 3.7.3), phase determination and model building, and refinement (Section 3.7.4).

3.7.1 Sample preparation and crystallisation

For crystallisation screens, the purified SAM_{wt} was transferred into Tris-buffer (Table 18) either by dialysis (Section 3.4.3) or SEC (Section 3.4.1.5) and the sample sterilised by membrane filtration (pore diameter 0.2 µm). The protein concentration in the sample that was used for crystallisation was ~10 mg/mL.

The experimental setup for SAM_{wt} crystallisation was the ‘sitting drop vapour diffusion method’, which is based on the reduction of water associated with the increase of precipitation agent in the protein sample by vapour diffusion in a closed system. Commercially available crystallisation screens that were used in SAM_{wt} crystallisation trials are listed in Table 6.

The crystallisation setups were prepared in Greiner 96-well crystallisation plates for sitting drop experiments. The drops contained a 1:1 mixture of 0.7 µL protein solution and crystallisation buffer containing the precipitation agent. The reservoir was filled with 70 µL of the respective crystallisation buffer. Once the crystallisation setups were prepared, the plate was sealed with a transparent film and incubated at 19 °C.

Protein crystals of SAM_{wt} were obtained after 5–7 days of incubation. Selected crystals were placed in fibre loops, frozen in a $T = 100$ K cold gaseous nitrogen stream and stored in liquid nitrogen until required for X-ray diffraction experiments. Because of the high salt content in the buffers, in which SAM_{wt} crystallises, the addition of a cryoprotectant was not necessary in most cases. If necessary, 10% glycerol was added.

3.7.2 Data collection

The diffraction data of SAM_{wt} was recorded at the beamline ID30A-3 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using an Eiger X 4M detector (DECTRIS, Baden-Dättwil, Switzerland). The data sets were recorded at a temperature of $T = 100$ K to reduce radiation damage.

Initially, two single 1° rotation images with a separation of a $\varphi = 90^\circ$ rotation around the spindle axis were recorded, which allowed the determination of the possible point group and orientation of the crystal lattice relative to the X-ray beam using the program MOSFLM (Battye et al., 2011) (Table 11). Based on this information, a suitable starting point, rotation range and step size for the data acquisition was proposed to ensure as complete as possible coverage of the asymmetric unit. In addition, the program BEST (Bourenkov & Popov, 2010) (Table 11) provides optimised choices for the radiation dose, exposure time and crystal-to-detector distance, to ensure the highest possible resolution while taking radiation damage into account.

3.7.3 Data processing

After successful data acquisition, the recorded set of diffraction images were evaluated by the program XDS (Wolfgang Kabsch, 2010) (Table 11). XDS corrects for data distortion and determines the orientation of the crystal lattice relative to the X-ray beam, and the unit cell dimensions and space groups, followed by the integration of the intensities of all reflections in all images.

Based on the integrated intensities, the program POINTLESS (Evans, 2011) (Table 11) identifies the space group with the highest symmetry that could explain the experimental data. POINTLESS checks for systematic absences of reflections in the data set. Subsequently, the mean intensities of all images are scaled to match the mean intensity of the first image, and observations of symmetry-equivalent reflections and Friedel pairs were merged using the program AIMLESS (Evans & Murshudov, 2013) (Table 11). Furthermore, AIMLESS proposes a resolution cut-off based on the average ratio of reflection intensity over its estimated error ($I/\sigma(I)$), completeness of the data, and the correlation coefficient between random half data sets ($CC_{1/2}$) for the highest resolution. The number of SAM_{wt} molecules present in the asymmetric unit was determined using MATTHEWS_COEF (Kantardjieff & Rupp, 2003) (Table 11).

3.7.4 Model building and refinement

The initial phases for the structure determination of SAM_{wt} were obtained by molecular replacement (MR). The model of SAM_{wt} for MR was generated by homology modelling using the SWISS-MODEL web server (Arnold *et al.*, 2006; Biasini *et al.*, 2014) (Table 11). A suitable template was found by searching for proteins with sequence similarities to the SAM_{wt} primary sequence in the SWISS-MODEL template library (SMTL). Subsequently, MR was performed by the program MOLREP (Vagin & Teplyakov, 1997) (Table 11) using the generated model.

The phases for the solution found by MR were calculated in an initial refinement step by the program REFMAC (Table 11), which also creates the initial 2F_o-F_c and F_o-F_c electron density maps. Further refinement was conducted using the program Phenix (Adams *et al.*, 2010) (Table 11). Between subsequent refinement steps, the calculated model was inspected and manually modified with respect to the calculated weighted 2F_o-F_c and F_o-F_c density maps using the graphics software Coot (Emsley & Cowtan, 2004) (Table 11). As a measure of the agreement between the model and the recorded data, the *R*-values *R*_{work} and *R*_{free} were calculated, which represent the discrepancy between the observed and calculated structure factor amplitudes (Equation [16]). Before any modelling and refinement steps are performed, the recorded diffraction data is partitioned into a *working set* containing 95% of the observed reflections and a *free set* containing the remaining 5%, which are randomly chosen reflections from the data set. Only the *working set* is used in model building and refinement. *R*_{work} is calculated for all reflections in the *working set*, whereas *R*_{free} is calculated for the reflections in the *free set*. During the refinement process, both *R*_{work} and *R*_{free} should decrease with every refinement cycle. The relationship between *R*_{work} and *R*_{free} is used as an indicator of overfitting. Ideally, the difference between *R*_{work} and *R*_{free} for the final model should not exceed 5% (Brünger, 1992).

$$R = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)} \quad [16]$$

- F*_{obs} observed structure factors
*F*_{calc} structure factors calculated from the model
hkl Miller indices

3.7.5 Analysis and evaluation of the X-ray structure

The refined models were analysed using the structure validation software Molprobity (Chen *et al.*, 2010; Davis *et al.*, 2007) (Table 11). The refinement process was stopped when additional rounds of refinement did not further improve the quality criteria of the structure. These criteria included R-values below 30%, minimal bond length and angle deviations from ideal geometry, backbone torsion angles in the favoured regions of the Ramachandran plot and side chains in the commonly observed rotamer states (unless explained by other energy contributions). In addition, a comparison with other crystal structures of the same resolution was performed by using POLYGON analysis (Urzhumtseva *et al.*, 2009) (Table 11), which helped in assessing the overall quality of the final structure. Secondary structure elements were predicted based on the atom coordinates using the program DSSP.

The possible oligomerisation state of SAM_{wt} was analysed using the PISA web server (Krissinel & Henrick, 2007) (Table 11). Oligomerisation is a crystallisation inherent process, with molecules often having multiple contact points with other molecules present in the asymmetric unit or with symmetric equivalent molecules throughout the unit cell. The PISAweb server provides help for the distinction of possibly biologically relevant assemblies from crystallisation artefacts by estimating the free dissociation energy (ΔG^0_{diss}) of a possible complex. In that context, it provides information about stabilising non-covalent (e.g., hydrogen bonds, salt bridges) and covalent (disulfide bonds) interactions across the interface and the buried surface area (BSA) of the assembly. The criteria for the presence of hydrogen bonds and salt bridges by PISA are entirely dependent on the distance between donor and acceptor heteroatoms (hydrogen bonds: upper limit of 3.89 Å between the heavy atoms; salt bridge: upper limit of 4 Å between heavy atoms).

4 Results

4.1 Molecular cloning of long and short SAM constructs

All SLY1 SAM domain constructs were produced initially as fusion proteins with an N-terminal GST tag. Therefore, all constructs were incorporated into the backbone of the pGEX-6P-2 derivative pGEX-6p-2_kombiP plasmid (Table 8), which is an expression vector optimised for high-level protein production in *E. coli*. In contrast to pGEX-6P-2, the *Psp*OMI restriction endonuclease recognition site (5'-gggccc-3') located in the *laqI* gene of pGEX-6p-2_kombiP contains a mutation, so that the other *Psp*OMI site located in the multiple cloning sites is amenable for molecular cloning purposes (pers. comm. Dr Karen Hänel).

The sequences coding for a long (aa 249–321) and a short (aa 254–321) SLY1 fragment were inserted downstream of a linker region coding for a PreScission protease cleavage site, LEVLPQGP. The protease cleaves the cleavage site sequence between the glutamine and glycine residue. Therefore, protease cleavage of the fusion proteins containing the long SLY1 fragments yielded SLY1 constructs with two additional residues, i.e., GP, at the N-terminus, which are not part of the native SLY1 sequence. In the shorter construct, the N-terminal proline P254 in the SLY1 SAM primary sequence coincides with the C-terminal proline of the protease cleavage site, resulting in a SLY1 construct with only an additional glycine at the N-terminus. Depending on the size of the SLY1 fragment, the long and short constructs are referred to as SAM_{lg} and SAM_{wt}, respectively (Fig. 13). The long and short SLY1 SAM mutants, in which S320 was substituted by a cysteine, are named SAM_{lg_C} and SAM_C, respectively. The cloning of the original pGEX-6p-2_SLY1SAM_{lg} coding for SAM_{lg} was described previously (Thiagarajan, 2011). The molecular cloning strategies for the other constructs were based on pGEX-6p-2_SLY1SAM_{lg}.

4. Results

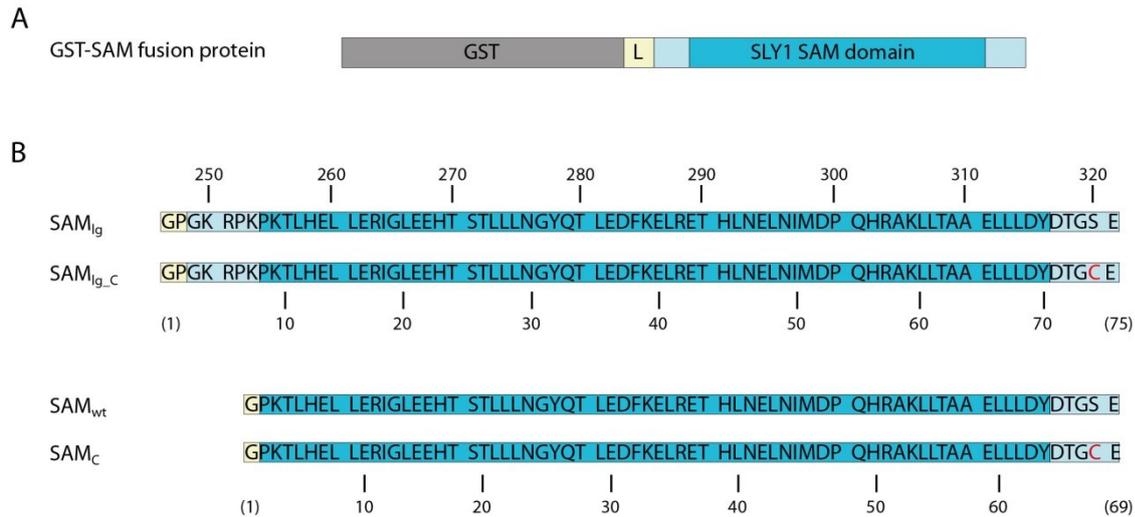


Fig. 13: Schematic representation of the SLY1 SAM domain constructs.

(A) The SAM domain constructs used in this study were produced fused to an N-terminal GST-tag. The GST (grey) and the SLY1 fragment (light blue) including the SAM domain (blue) are separated by a short linker of eight amino acids (yellow) that codes for a PreScission protease cleavage site. (B) Four different SLY1 SAM constructs were used in this study. The longer constructs SAM_{lg} and SAM_{lg_C} code for the entire folded core SAM domain (blue) and additional flanking residues of SLY1 (light blue). In the shorter constructs SAM_{wt} and SAM_C , the N-terminal flanking region was removed. All constructs contain an N-terminal GP-dipeptide, which is a remnant of the PreScission protease cleavage site (yellow). In the short constructs, the proline coincides sequentially with P254 in the SLY1 sequence. In SAM_{lg_C} and SAM_C , S320 was exchanged for a cysteine (red).

4.1.1 Molecular cloning of pGEX-6p-2_SLY1SAM_{lg_C}

Residue S320 in the SLY1 sequence was exchanged for a cysteine by site-directed mutagenesis to facilitate selective fluorescence labelling and cross-linking of the SAM domain monomers by disulfide bond formation. The plasmid coding for SAM_{lg} , pGEX-6p-2_SLY1SAM_{lg}, served as the template of the Quikchange-PCR (Section 3.3.1). A single nucleotide mutation from adenine to thymine in the coding sequence was sufficient for the S320C exchange and also resulted in the loss of the *BtsI* endonuclease recognition site, which gives rise to a change in the band pattern after *BtsI* treatment of the plasmid (Fig. 14). This difference in the restriction pattern was used to check clones for the successful incorporation of the mutation.

Four of the clones that showed a promising band pattern in the gel were further analysed by sequencing (Fig. 15) (Section 3.3.3). Clone number 5 carried the desired mutation while not showing any additional mutation in the SAM coding sequence and was chosen for protein production and purification of SAM_{lg_C} .

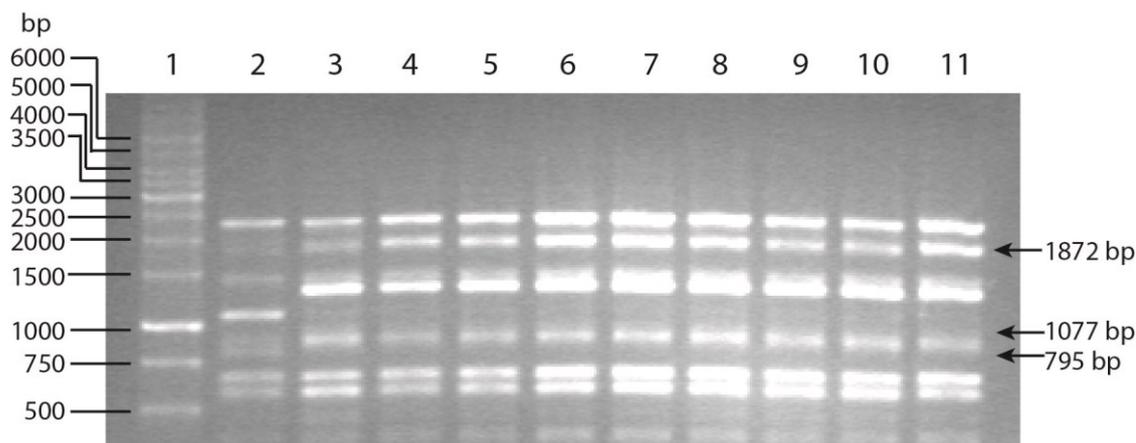


Fig. 14: Agarose gel of wt and mutant plasmid DNA digested with *BtsI*.

Then nucleotide exchange in the SAM_{lg} coding sequence to generate the S320C mutation eliminates the *BtsI* endonuclease recognition site. Thus, the DNA cannot be cleaved by *BtsI* at this site in the mutant DNA construct, which results in the presence of an 1872 kb fragment that would otherwise be digested to 1077 and 795 bp fragments. This change in band pattern can be observed in the agarose gel, where lane 1 contains the DNA ladder, lane 2 the wt plasmid and lanes 2–11 contain the mutant plasmid from nine different colonies.

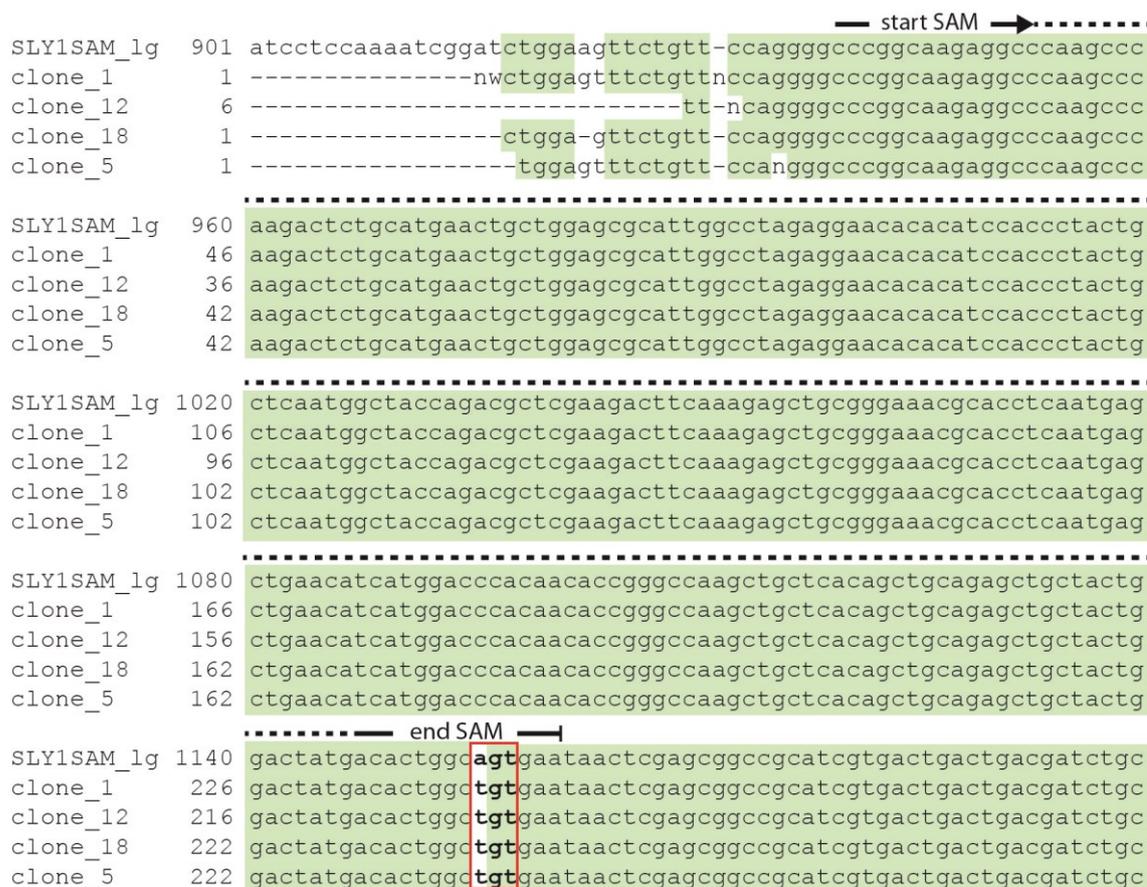


Fig. 15: Sequencing results of pGEX-6p-2_SLY1SAM_lg mutants.

The sequencing results above show that the selected four mutant clones (1,5,12,18) contain the desired mutation. The mutated codon leading to an S320C exchange at the C-terminus of the SAM domain is marked by a red box. As a comparison, the sequence of the wt SAM_{lg} sequence is also shown in the first line.

4.1.2 Molecular cloning of pGEX-6p-2_SLY1SAM_wt and pGEX-6p-2_SLY1SAM_C

The short SAM domain constructs, SAM_{wt} and SAM_C lack the first six residues of the SAM_{Ig} and SAM_{Ig_C} constructs after the N-terminal GP-dipeptide. The deletion of residue 249-254 was achieved by amplifying the desired sequence from the template plasmids by PCR (Section 3.3). The template was pGEX-6p-2_SLY1SAM_{Ig} and pGEX-6p-2_SLY1SAM_{Ig_C} for SAM_{wt} and SAM_C, respectively. Since the N-terminus and therefore the 5' end of both constructs is the same, the oligonucleotide F-Sly-short served as the forward directed primer in both cases (Table 9). The modification close to the 5' end containing the point mutation for the S320C substitution was accounted for by using two different reverse directed primers, named R-SLY-short and R-SLY-mut-short (Table 9). The forward and reversed primers additionally coded for a 5' *Psp*OMI and a 3' *Xho*I restriction endonuclease recognition site, respectively, thus allowing the insertion of the PCR products into the expression plasmid pGEX-6p-2_kombiP (Section 3.3.2). The newly cloned constructs were sequenced (3.3.3) to confirm successful insertion of the SAM domain coding sequences.

4.2 SAM domain production and purification

The production and purification protocol of the different SLY1 SAM domain variants was adapted from the protocol published by Thiagarajan, 2011. All SAM domain variants were expressed in *E. coli* BL21 (DE3) cells (Section 3.4.1). The bacteria were either cultivated in M9-medium for isotope-labelling or LB-medium.

The success of the protein production and subsequent protein purification was monitored by SDS-PAGE analysis. Figure 16 shows exemplary SDS-PAGE gel slices of samples taken during the expression and purification of each of the four SLY1 SAM constructs, i.e. SAM_{Ig} (A), SAM_{Ig_C} (B), SAM_{wt} (C) and SAM_C (D). The SDS-PAGE analysis of the samples of expression cultures of all constructs taken pre- and post-induction of recombinant gene expression showed the appearance of a strong band in the gel 4–5 hours post-induction (Fig. 16 A, B, C, D, lanes 3). The observed signal corresponds to a protein of a molecular mass of ~33 kDa, which agrees with the calculated molecular mass of the GST-SAM fusion-proteins of ~35 kDa, indicating the successful production of the GST- SAM fusion protein.

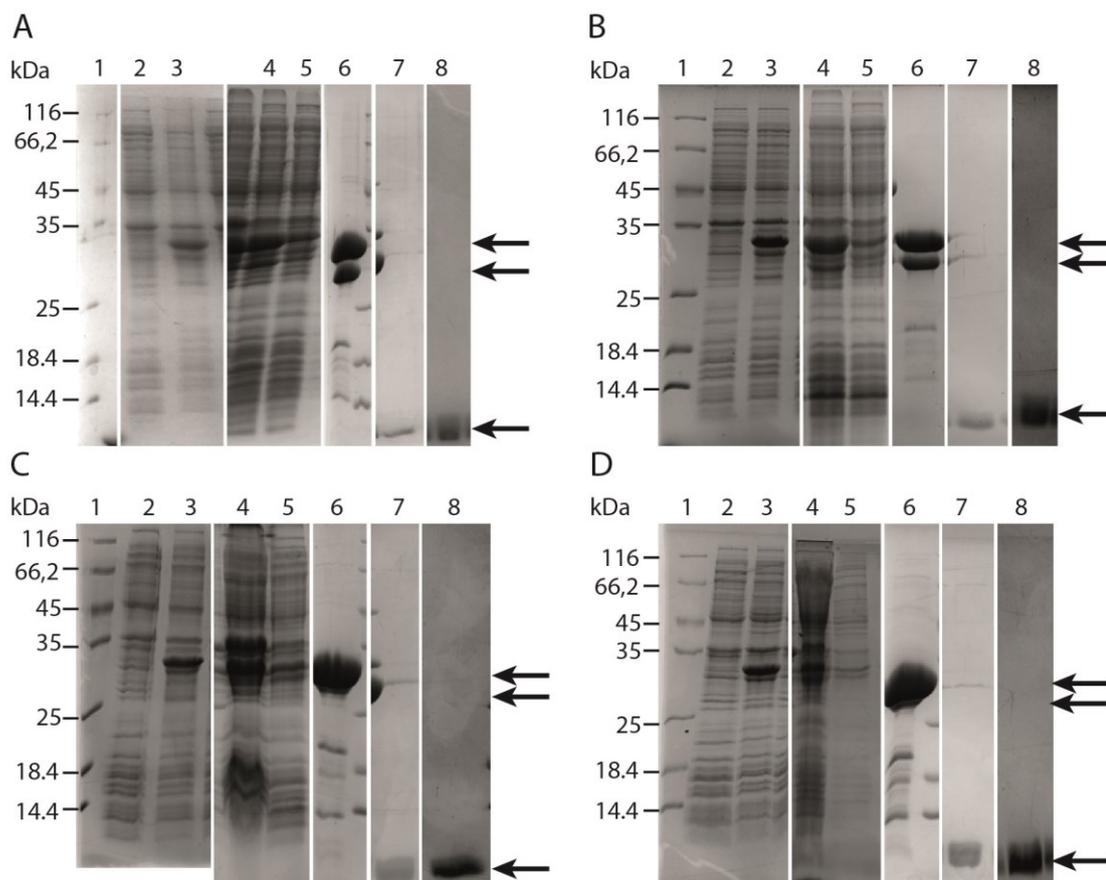


Fig. 16: SDS-PAGE (15 %) gel slices containing protein samples taken during the SLY1 SAM purification.

The figures A–D show samples that have been taken during the production and purification of SAM_{Ig} (A), SAM_{Ig_C} (B), SAM_{wt} (C) and SAM_C (D). The arrows indicate the position of GST-SAM (~35 kDa), GST (~26 kDa) and the SAM domain (8–9 kDa). The following fractions were loaded onto the gels. Lane 1: Unstained protein marker (Thermo Fisher Scientific); lane 2: expression culture before induction; lane 3: expression culture 5 h after induction; lane 4: cleared bacterial cell extract; lane 5: protein fraction not bound by the GST-Sepharose resin; lane 6: protein bound to GST-Sepharose resin; lane 7: cleaved SAM domain eluted from the GST-Sepharose resin; and lane 8: SAM domain after SEC.

The signal is also observed after the clearing of the cell extract by centrifugation (Section 3.4.1.2), showing that the majority of the produced fusion protein was soluble and not part of inclusion bodies (Fig. 16 A, B, C, D, lanes 4). The GST-SAM fusion proteins were separated from bacterial proteins present in the cell extract by GSH-affinity chromatography (Section 3.4.1), as can be observed in Lanes 5 and Lanes 6. Lanes 7 show the released SAM constructs after PreScission protease cleavage of the fusion proteins (Section 3.4.1.4). The significantly lower molecular mass of the isolated SAM construct (~8 kDa) is reflected by its faster migration through the gel, resulting in a band just below the protein marker signal corresponding to 14.4 kDa. Minor contaminations in the sample were observed on the gel in the form of weak signals corresponding to GST (26.5 kDa) and uncleaved GST-SAM fusion protein. These contaminants were removed from the samples by preparative SEC (Section 3.4.4).

Finally, the samples were concentrated to the desired volume, filtered and stored at 4 °C. The yield of purified SAM domain per L of culture varied in all cases between 5–10 mg.

4.3 Characterisation of the SAM_{Ig} dimerisation behaviour

Previous biophysical analysis of the SAM_{Ig} construct showed that the domain dimerised with an equilibrium dissociation constant (K_D) in the low micromolar range (Thiagarajan, 2011). Subsequent efforts to solve the structure of the SAM_{Ig} dimer by solution NMR spectroscopy were unsuccessful. NMR spectra suffered from severe line broadening or even disappearance of many resonances due to intermediate-to-fast chemical exchange as a result of the monomer-dimer equilibrium.

4.3.1 2D ¹H,¹⁵N-HSQC spectra of SAM_{Ig} show concentration-dependent line broadening due to chemical exchange

NMR analysis of the SAM_{Ig} construct was initially performed to reassess the effect of the chemical exchange on resonance line-shape and to enable a comparison of NMR data between SAM_{Ig} and the other SAM domain constructs. Two-dimensional ¹H,¹⁵N-HSQC spectra of SAM_{Ig} at three concentrations (500, 50 and 5 μM) were acquired and showed that the majority of the resonances experienced concentration-dependent continuous chemical shift changes. Furthermore, a number of resonances also showed extensive line broadening (e.g., K255, K305, L306 and L312) (Fig. 17A,B). These changes in resonance line-shape and position are indicative of intermediate-to-fast chemical exchange on the chemical shift time scale and in agreement with previous results on SLY1 SAM_{Ig} (Thiagarajan, 2011).

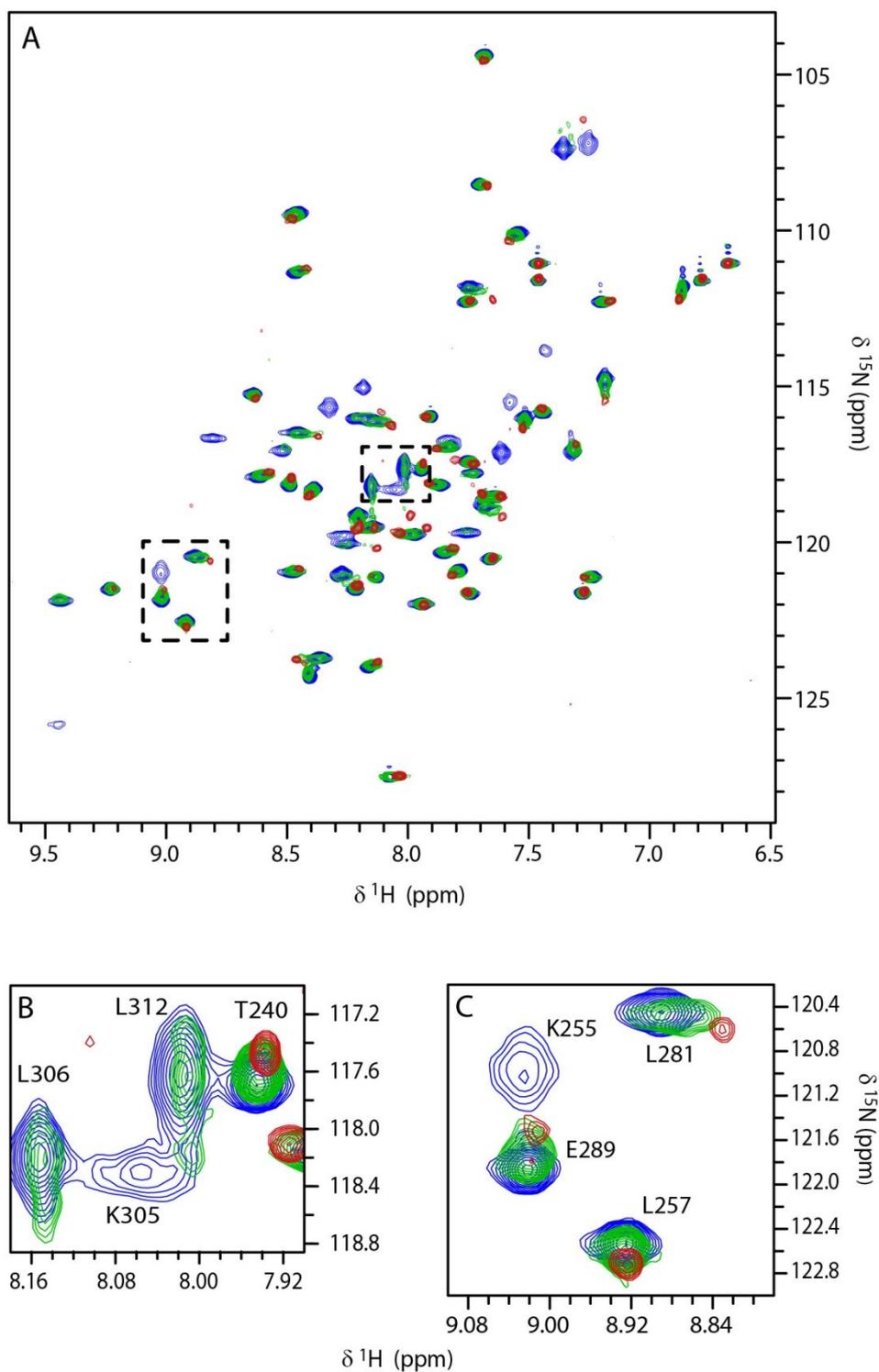


Fig. 17: Superposition of 2D ^1H , ^{15}N -HSQC spectra of SAM_{Ig} at three concentrations.

2D ^1H , ^{15}N -HSQC spectra of SAM_{Ig} recorded 35 °C at 600 MHz (^1H operating frequency) in NMR-buffer at three concentrations (blue: 500 μM ; green: 50 μM ; red: 5 μM). (A) Overlay of the 2D ^1H , ^{15}N -HSQC spectra. (B,C) Regions of the 2D ^1H , ^{15}N -HSQC spectra presented in (A). The spectra show resonances undergoing concentration-dependent broadening because of intermediate-to-fast chemical exchange on the NMR chemical shift time scale. In particular, resonances representing the amide groups of K305, L306 and L312 (B) and K255 (C) show extensive line broadening at the highest concentration and are broadened beyond detection at 50 and 5 μM . Resonances arising from amide groups of T240, L257, L281 and E289 show concentration-dependent changes in chemical shift because of fast chemical exchange

The available backbone amide assignments of SAM_{Ig} (Thiagarajan, 2011) were transferred to the recorded 2D ¹H,¹⁵N-HSQC spectra and the weighted average chemical shift differences ($\Delta\delta_{\text{ave}}$) between the 500 and the 5 μM samples were calculated as described in Section 3.6.3 (Fig. 18). Overall, $\Delta\delta_{\text{ave}}$ between the two spectra is very small, with the highest change measured as 70 parts per billion (ppb) for the cross-peaks arising from D286. Moreover, approximately half of the resonances observed in the spectrum recorded at 500 μM underwent severe line broadening at 50 μM or 5 μM protein concentration, which prevented a complete chemical shift perturbation analysis. The resonances that showed line broadening represent amide groups of residues found mostly among the first 20 and last 17 amino acids of SAM_{Ig}. The intermediate exchange those nuclei undergo indicates a stronger susceptibility to the dimerisation process, because their $\Delta\nu$ must be larger than that of nuclei undergoing fast exchange. Structure determination of the SAM dimer revealed that these N- and C-terminal residues of the SAM domain are positioned at the dimer interface.

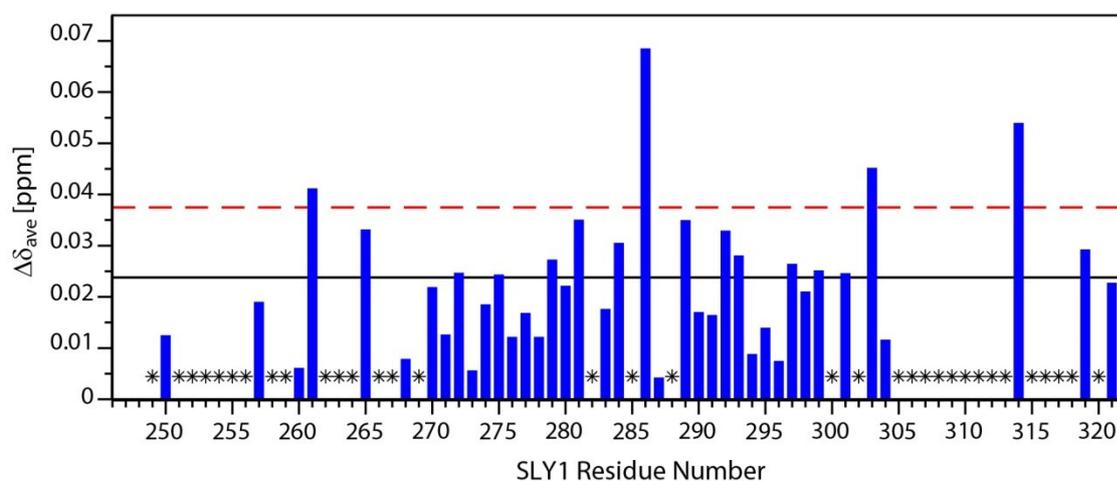


Fig. 18: Weighted average ¹H and ¹⁵N chemical shift differences, $\Delta\delta_{\text{ave}}$, between resonances recorded in 2D ¹H,¹⁵N-HSQC spectra of SAM_{Ig} at 500 and 5 μM .

The average chemical shift perturbations ($\Delta\delta_{\text{ave}}$) between amide resonances at 5 and 500 μM of SAM_{Ig} were calculated as described in Section 3.6.3. The backbone amide proton assignments were taken from Thiagarajan, 2011. The black line represents the average $\Delta\delta_{\text{ave}}$ and the red dashed line corresponds to 1σ above the mean. $\Delta\delta_{\text{ave}}$ values could not be determined for residues marked with an asterisk because of the absence of the corresponding cross-peaks at lower SAM domain concentrations.

4.3.2 Determination of the equilibrium dissociation constant of the SAM_{Ig} dimerisation by MST

Previous MST measurements using non-specific fluorescence labelling of lysine residues of SAM_{Ig} was conducted and yielded a K_D of approximately 5 μM (Thiagarajan, 2011). In this study, the SAM_{Ig_C} variant was labelled with the fluorescent dye Alexa Fluor 488 (Section 3.4.2) by maleimide-thiol-coupling of the dye with the reduced thiol group of the cysteine. Alexa Fluor 488 has an excitation maximum at ~ 490 nm and emits light at ~ 520 nm. Therefore, the Alexa Fluor 488 dye is compatible with the Monolith NT.115 blue/red system used for MST measurements. A series of SAM_{Ig} samples with decreasing concentrations was prepared in standard-buffer (Table 18) by 1:2 dilution steps. Fluorescent labelled SAM_{Ig_C} was added to the samples resulting in a SAM_{Ig} 1:2 dilution series ranging from 620–0.05 μM total SAM domain concentration with each sample containing 50 nM labelled SAM_{Ig_C}. Samples were transferred into either Premium coated or hydrophilic capillaries. The measurements were performed at 35 °C using 40–50% LED power. The thermophoretic behaviour was measured at each concentration three consecutive times per run using 40, 60 and 80% MST power. Each run was performed twice in Premium coated and hydrophilic capillaries. Only the data recorded at 80% MST power were used for further analysis of the SAM domain dimerisation because these data sets showed the strongest effect of the increased SAM domain concentration on the thermophoretic behaviour. It should be mentioned, however, that the same trend was observed at lower MST powers, albeit with lower signal-to-noise ratio.

The overall fluorescence of each capillary in the absence of a temperature gradient (capillary scan) was found to increase as the SAM_{Ig} concentration increased (Fig. 19). This effect most probably arose from non-specific interactions of the fluorescence dye with the reaction tube surface, which was reduced by the addition of unlabelled protein. Preparation of the samples in special low-binding tubes minimised the unspecific interaction, but could not quench it completely. Therefore, to reduce the variation of the fluorescence signal, the samples exhibiting a fluorescence intensity which deviated more than 1σ from the mean intensity were excluded from the analysis. Fluorescence intensities measured during the capillary scan presented in Fig. 19 showed that data for the samples containing 620 and 310 μM protein were above the defined threshold, whereas the data for samples containing 0.05 and 0.088 μM protein were below the defined threshold. Thus these data were excluded from K_D data fitting (shaded grey).

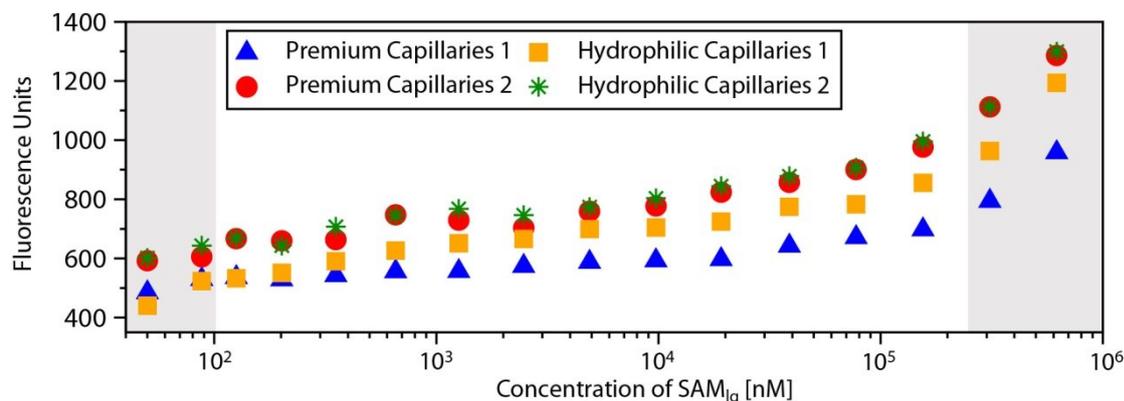


Fig. 19: Fluorescence measured in each capillary without the application of any laser power.

Fluorescence data recorded during the MST measurements without a temperature gradient. MST measurements were performed in duplicate using hydrophilic and Premium coated capillaries. The LED power was 40–50%. The increase in fluorescence as the concentration of SAM_{I_g} increased follows a similar trend in all data sets. For K_D determination, the concentrations shaded grey were excluded from data analysis because their fluorescence deviated by more than 1σ from the mean fluorescence intensity of the respective data set.

The F_{norm} values were calculated from F_{cold} and F_{hot} for all runs using the default settings in the NTanalysis software (Table 11), which takes both thermophoretic and T-Jump contributions into account. The results in Fig. 20 show that a positive thermophoretic behaviour increases as a function of SAM_{I_g} concentration, indicating that both, SAM domain monomers and dimers, show positive thermophoretic behaviour. However, positive thermophoresis is stronger for dimers than monomers. The K_D was extracted using a fitting routine for a monomer-dimer equilibrium model (Equation [13]). The calculated K_D values were $(19 \pm 5) \mu\text{M}$ and $(19 \pm 7) \mu\text{M}$ for samples in Premium coated and hydrophilic capillaries, respectively. While the K_D values determined in this study are still in the lower micromolar range as expected, they are significantly larger than the K_D of $5 \mu\text{M}$ reported in Thiagarajan, 2011. This divergence between the determined dissociation constants could have resulted from multiple factors: The SAM_{I_g_C} construct, which was used in this study, was labelled site-specifically at position 320, while the SAM_{I_g} in previous studies was labelled using the NHS-coupling, which reacts with amino groups of lysines and at the N-terminal end of the SAM domain. The different positions of the fluorescent dye could have affected the interaction. Moreover, the measurements for the data presented here were performed at 35°C , while the previous measurements were performed at room temperature. Furthermore, the previous data was fitted using a 1:1 binding model instead of the more appropriate monomer-dimer equilibrium model. Although the results presented here do not precisely match

the K_D reported previously, both data sets support the hypothesis of dimerisation of the SLY1 SAM domain with a K_D in the lower micromolar range.

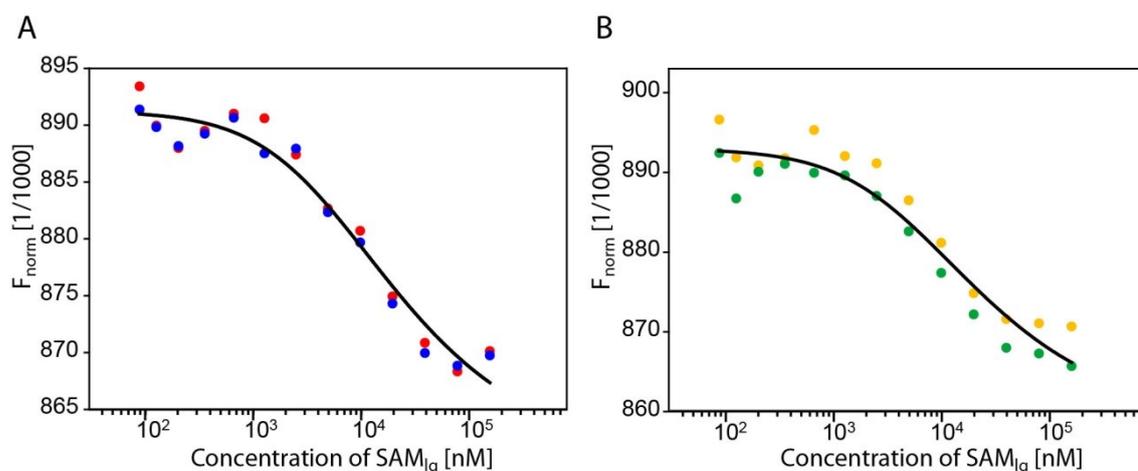


Fig. 20: Binding isotherms from MST experiments conducted with SLY1 SAM_{Ig}.

The thermophoretic behaviour of a 1:2 dilution series of SAM_{Ig} ranging from 155–0.076 μ M total SAM concentration, each containing 50 nM Alexa Fluor 488 labelled SAM_{Ig_C}, was examined using a Monolith NT.115 blue/red system at 35 °C. The samples were prepared twice in Premium capillaries and twice in hydrophilic capillaries. The F_{norm} data includes thermophoretic and T-Jump effects. **(A)** Samples analysed in Premium capillaries used 40% (blue) and 50% (red) LED power and 80% MST power. The black line represents the fitted data to yield a K_D of (19 ± 5) μ M. **(B)** Samples analysed in hydrophilic capillaries using 40% LED power and 80% MST power. The black line represents the fitted data to yield a K_D of (19 ± 7) μ M.

4.4 Characterisation of the SAM_{Ig_C} single-cysteine mutant

The SLY1 SAM_{Ig_C} construct was designed to enable site-specific C-terminal labelling with a fluorescent dye rather than non-specific labelling of lysine residues using the N-hydroxysuccinimide (NHS) coupling reaction. However, the labelling reaction was observed to be inefficient with more than 50% of SAM_{Ig_C} remaining unlabelled after an incubation period of 48 h. Acquisition of 2D ¹H,¹⁵N-HSQC spectra of samples containing SAM_{Ig_C} at 5, 50 and 500 μ M, which had been prepared after the removal of DTT by dialysis and without the addition of any fluorescent dye, showed well-dispersed resonances of uniform intensity and linewidth. The intensity and line-shape of the signals suggested that the chemical exchange due to the monomer-dimer equilibrium had been quenched by the formation of a stable dimer species via the formation of a disulfide bond (Fig. 21A,B,D). Consequently, the observed low efficiency of the maleimide-thiol-coupling reaction was most likely the result of the spontaneous formation of a disulfide-cross-linked SAM_{Ig_C} dimer. This hypothesis was supported by the observation that addition of the reducing agent TCEP to the cross-linked SAM_{Ig_C}

dimer sample gave rise to line broadened resonances in 2D ^1H , ^{15}N -HSQC spectra (Fig. 21C,E), as observed in spectra recorded on the SAM_{Ig} construct (Fig. 17).

Using the available backbone amide assignments of the SAM_{Ig} domain (Thiagarajan, 2011), weighted chemical shift perturbation analysis was used to examine the influence of the S320C mutation and the influence of disulfide-cross-linked dimer formation on the overall structure of the SAM domain (Fig. 22). Overlaying the 2D ^1H , ^{15}N -HSQC spectra of the wt SAM_{Ig} and reduced SAM_{Ig_C} domains enabled successful mapping of the assignment information of SAM_{Ig} to the backbone amide resonances of the reduced SAM_{Ig_C} domain. Of the expected 70 backbone amide resonances, 64 resonances could be assigned. The same procedure was performed for the resonances of the cross-linked SAM_{Ig_C} mutant. However, only those resonances could be assigned which allowed the unambiguous transfer of the assignments, therefore only 61 out of the expected 70 resonances were assigned.

The weighted chemical shift differences of the backbone amide resonances of SAM_{Ig} and non-cross-linked SAM_{Ig_C} are predominantly below 10 ppb with an average $\Delta\delta_{\text{ave}}$ of 4 ppb (Fig. 22). Overall low $\Delta\delta_{\text{ave}}$ indicates that the mutation did not affect the SAM domain fold, nor its dimerisation properties. The largest $\Delta\delta_{\text{ave}}$ values, as expected, exist for resonances representing residues neighbouring the mutation site under reducing conditions. Similar size $\Delta\delta_{\text{ave}}$ values are observed for resonances arising from these neighbouring residues in the disulfide cross-linked SAM_{Ig_C} dimer. However, some resonances of other backbone amide groups show a considerably larger $\Delta\delta_{\text{ave}}$ in the cross-linked SAM_C than the corresponding resonances in the reduced state. In the cross-linked SAM_{Ig_C}, resonances of residues K250–L260, T280–E286 and P300–D315 show larger than average values for $\Delta\delta_{\text{ave}}$.

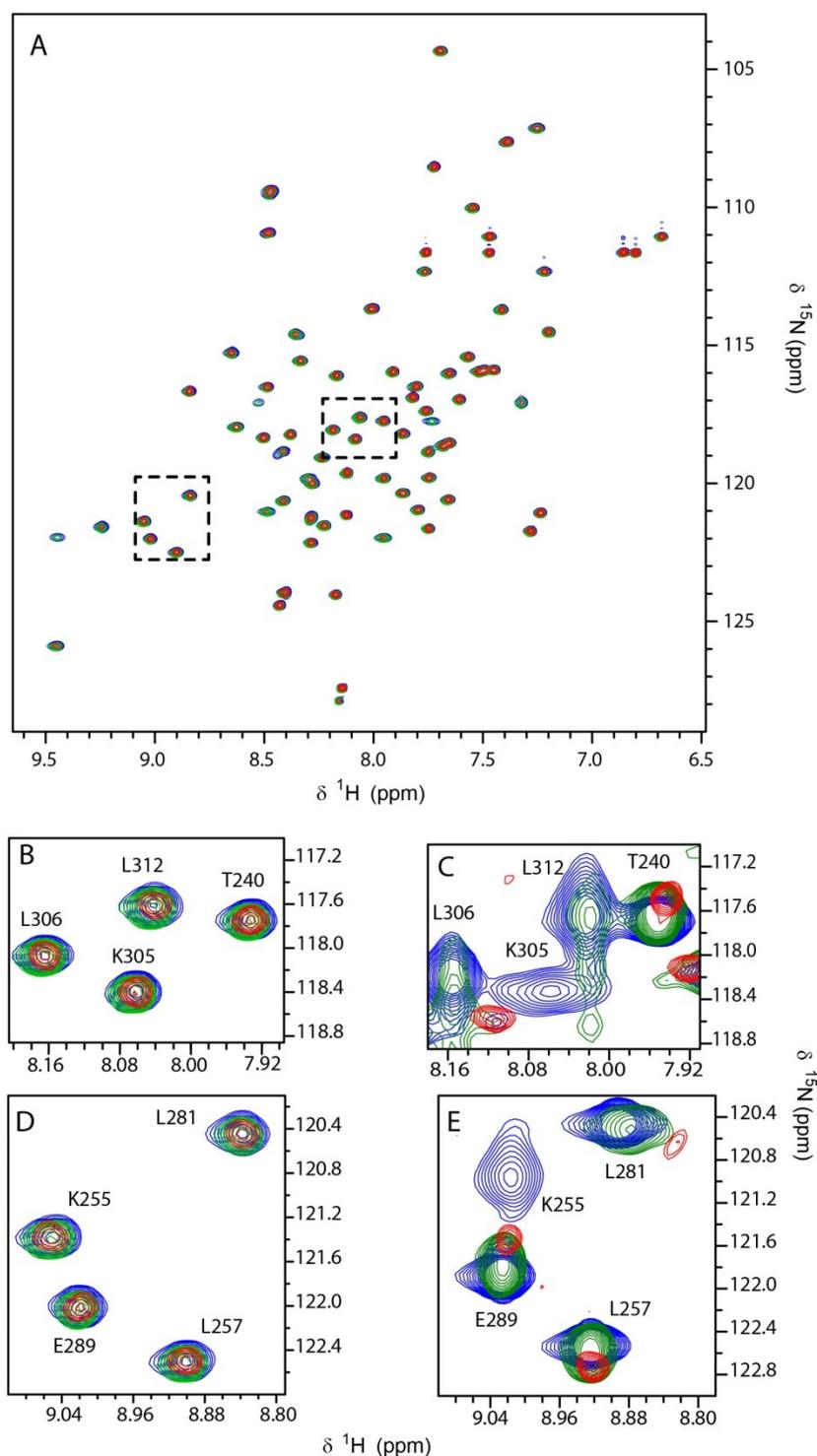


Fig. 21: 2D ^1H , ^{15}N -HSQC spectra of cross-linked SAM_{lg_C} and non-cross-linked dimer.

2D ^1H , ^{15}N -HSQC spectra of SAM_{lg_C} recorded 35 °C at 600 MHz (^1H carrier frequency) in NMR-buffer using three different concentrations (blue: 500 μM ; green: 50 μM ; red: 5 μM) under oxidising conditions and reducing conditions. (A) Overlay of the 2D ^1H , ^{15}N -HSQC spectra of SAM_{lg_C} under oxidising conditions. (B,D) Regions of the overlaid 2D ^1H , ^{15}N -HSQC spectra. Since the line broadening is quenched, all resonances in each spectrum are of uniform shape and intensity. (C,E) Addition of 10 mM TCEP caused the re-appearance of concentration-dependent line broadening in the 2D ^1H , ^{15}N -HSQC spectra of SAM_{lg_C} .

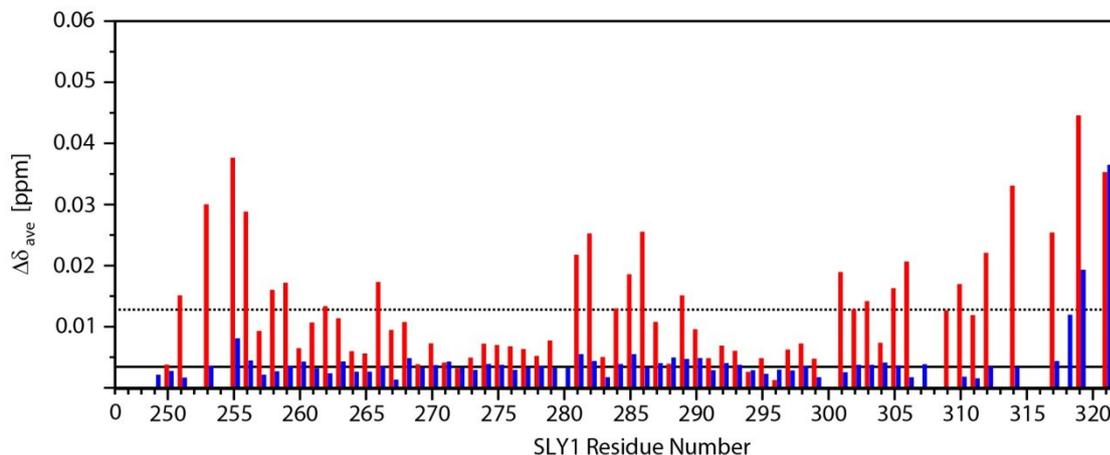


Fig. 22: Weighted average chemical shift perturbation analysis of the backbone amide resonances in 2D ^1H , ^{15}N -HSQC spectra of SAM_{ig} and SAM_{ig_C} under oxidising and reducing conditions.

The $\Delta\delta_{\text{ave}}$ between backbone amide resonances of SAM_{ig} and SAM_{ig_C} in the cross-linked (red) and non-cross-linked (blue) state were calculated as described in Section 3.6.3. The spectra of the non-cross-linked SAM_C were recorded after the addition of 10 mM TCEP to the cross-linked sample. The mean $\Delta\delta_{\text{ave}}$ is drawn as a solid line for the non-cross-linked SAM_C and as a dotted line for the cross-linked SAM_C . For the cross-linked SAM_C , unambiguous assignment information could not be obtained for G249, T280, L307, T308, T318 and C320. Under reducing conditions, the assignments for T308, A309 and C320 could not be obtained because resonances for these residues were missing, and in the case of the S320C mutation, unambiguous assignment was not possible. For L313, D315 and Y316 assignment information was missing from the previous study (Thiagarajan, 2011). The N-terminal G247 and the four proline residues (248, 252, 254, 300) are not observable.

4.5 SLY1 SAM_{wt} and SAM_C

The shorter variant of the SLY1 protein, SAM_{wt} , is five residues shorter than the longer version, lacking amino acids G249–K253 (GKRPK) (Fig. 13), which are not part of the folded SAM domain. Residues P₂₅₂ and P₂₅₄, in combination with the proline in the PreScission protease cleavage site in the longer construct, gave rise to multiple sets of signals for neighbouring residues because of cis/trans-isomerisation. These factors complicated the structure determination of SAM_{ig} by NMR. Therefore, the shorter construct was cloned (Section 4.1.2) and prepared (Section 4.2) to enable the determination of the SLY1 SAM domain structure by NMR. A corresponding single-cysteine mutant construct, SAM_C , was prepared to enable the formation of a stable cross-linked dimer.

4.5.1 Analytical ultracentrifugation verifies SAM_{wt} dimerisation

Analytical ultracentrifugation experiments were performed to characterise SAM_{wt} dimerisation. Sedimentation equilibrium experiments were performed using protein concentrations of 60, 120 and 300 μM (Section 3.5.2) to determine the K_D of the

dimerisation. Acceptable global fits of the data were achieved only with the monomer-dimer equilibrium model with low residuals throughout the radial concentration profile (Fig. 23). A global fit of the data to the monomer-dimer equilibrium model provided an equilibrium association constant (K_A) of $8.5 (2.4; 30) \times 10^3 \text{ M}^{-1}$, which corresponds to a K_D of $117 (33; 423) \mu\text{M}$, and a molecular mass of $(8,040 \pm 532) \text{ Da}$ for the monomer. Both, the one-component and the two-component model, provided inferior fits with significantly larger residuals than the monomer-dimer equilibrium model. The numbers in parenthesis specify the 95% confidence interval.

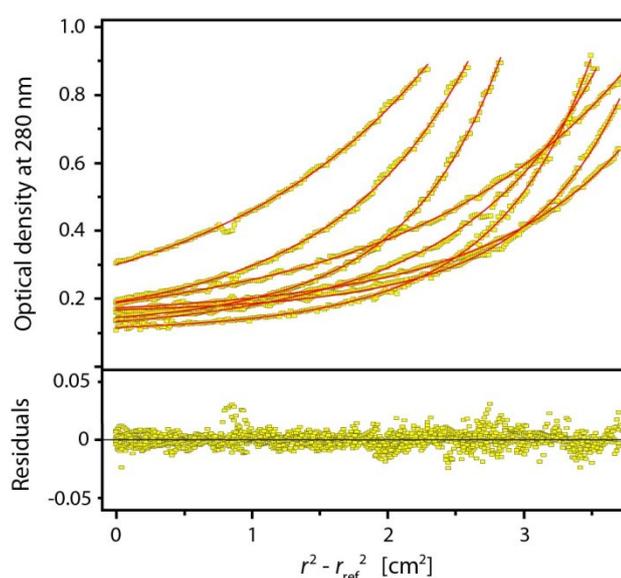


Fig. 23: Sedimentation equilibrium experiments with SLY1 SAM_{wt} to specify oligomeric state and oligomerisation affinity.

Three different concentrations of SAM_{wt} (60, 120, 300 μM) were studied at four different speeds. The data set was fitted globally assuming a monomer-dimer model. The upper panels display overlays of the concentration profiles recorded after the establishment of equilibrium between sedimentation and back diffusion at each selected speed and the calculated concentration distributions (red lines) based on a monomer-dimer model. The graph below shows the residuals of the fit.

4.5.2 Determination of the equilibrium dissociation constant of SAM_{wt} dimerisation by MST

The K_D of the SAM_{wt} dimerisation reaction was additionally determined by MST (Section 3.5.5). SAM_{wt} was labelled with a fluorescent probe using NHS-coupling (Section 3.4.2) with a labelling efficiency of $\sim 20\%$. A series of SAM_{wt} samples with decreasing concentrations was prepared in standard-buffer (Table 18) by 1:2 dilution steps. Fluorescent labelled SAM_{wt} was added to each dilution to give a final concentration of 70 nM labelled SAM_{wt} in a dilution series ranging from 0.4–824 μM of

total SAM concentration. Samples were transferred into Premium coated capillaries. The measurements were performed at room temperature using 35% LED power. The thermophoretic behaviour was measured twice at each concentration two consecutive times per run using 40% and 60% MST power. Data were acquired in triplicate by repeating the preparation of the dilution series from the same stock solution and running the MST experiment. As suggested by the manufacturer, the data recorded at the lowest MST power (i.e., 40%) which provided sufficient thermophoretic motion in the sample and yielded a good signal-to-noise ratio was used for further analysis of the SAM domain.

During the capillary scan, the signal intensity was found to be very similar for the majority of samples. However, the four lowest concentrations clearly showed a significant reduction in fluorescent signal (Fig. 24), which also lead to a poor signal-to-noise ratio during the thermophoresis measurements, effectively making these data points unreliable. These differences in fluorescence intensities are hypothesised to arise from non-specific interactions of the protein or the fluorescence dye with the reaction tube surface, which is reduced by the addition of unlabelled protein. All data from samples that varied by more than 1σ from the mean fluorescence intensity in the capillary scan were excluded from the K_D fitting to ensure similar conditions for the thermophoretic experiments for all concentrations examined.

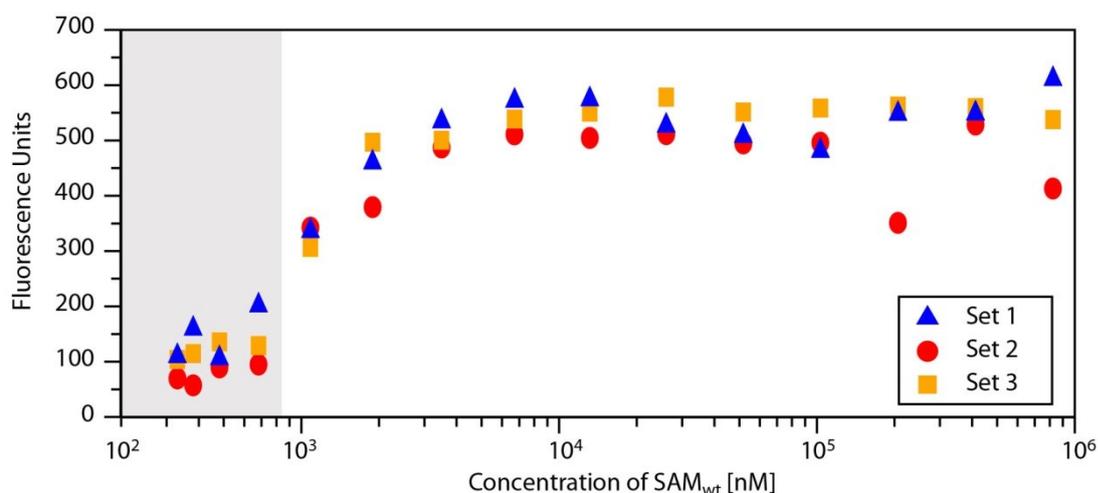


Fig. 24: Fluorescence of samples in the absence of heating (capillary scan)

Fluorescence data was recorded without the generation of a temperature gradient. Three data sets were recorded. The total concentration of unlabelled SAM_{wt} in the measured sample ranged from 0.4–824 μM . The LED power was kept constant at 35%. The measurements were performed in Premium capillaries. In all three recorded data sets a significant loss of fluorescence intensity was observed for the five lowest concentrations, For the K_D determination, the samples shaded in grey were excluded from the data analysis because their fluorescence intensity deviated by more than 1σ from the mean fluorescence intensity of the respective data set.

The F_{norm} values were calculated from F_{cold} and F_{hot} for all runs using the default settings in the NTanalysis software (Table 11), which takes both thermophoretic and T-Jump effects into account. The results in Fig. 25 show that the positive thermophoretic behaviour decreases as a function of SAM_{wt} concentration, reflecting SAM_{wt} dimerisation. The K_D was calculated by fitting the recorded data to a monomer-dimer equilibrium model (Equation [13]). Interestingly, the K_D value for SAM_{wt} was determined to be $153 \pm 25 \mu\text{M}$. This K_D corresponds to an 8-fold increase in concentration when compared with the K_D determined for SAM_{lg} (K_D of $19 \pm 5 \mu\text{M}$ and $19 \pm 7 \mu\text{M}$ for samples in Premium coated and hydrophilic capillaries, respectively; Section 4.5.2) and a 31-fold increase in comparison to the K_D of $5 \mu\text{M}$ reported by Thiagarajan, 2011.

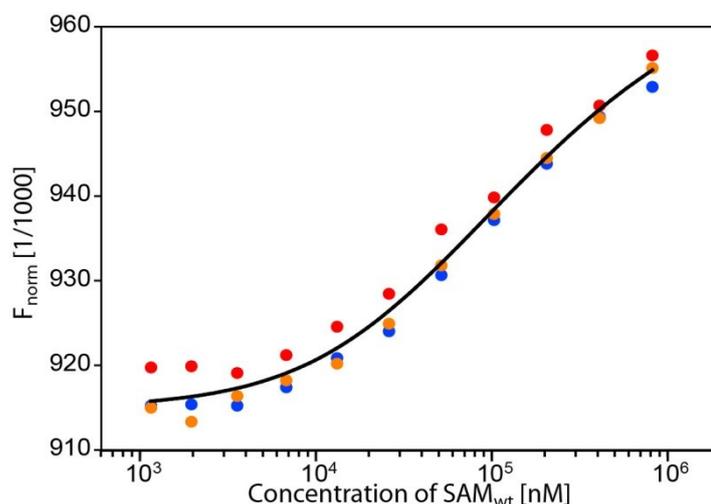


Fig. 25: MST-based isotherm reflecting the monomer-dimer equilibrium

The thermophoretic behaviour of SAM_{wt} was studied over a concentration range of $0.4\text{--}824 \mu\text{M}$ total SAM domain concentration. Each sample contained 70 nM fluorescently labelled SAM_{wt} . Experiments were run at room temperature. The LED power was maintained at 35% during the experiments, and the MST power was 40%. The normalised fluorescence data (F_{norm}) included thermophoretic and T-Jump effects. In all data sets, a concentration-dependent decrease in positive thermophoresis was observed. The fitted data yielded a K_D of $153 \pm 25 \mu\text{M}$.

4.5.3 2D $^1\text{H}, ^{15}\text{N}$ -HSQC spectra of SAM_{wt} and SAM_{C}

As a first step in the characterisation of SAM_{wt} and SAM_{C} by NMR, 2D $^1\text{H}, ^{15}\text{N}$ -HSQC spectra were recorded at varying concentrations to verify that both proteins adopt a stable three-dimensional fold (Fig. 26A,B). The 2D $^1\text{H}, ^{15}\text{N}$ -HSQC spectra showed well-dispersed resonances which indicates that both constructs adopt a stable tertiary fold. SAM_{wt} shows concentration-dependent changes in chemical shift and intensity of

resonances because of chemical exchange, i.e., monomer-dimer equilibrium (Fig. 26 A,C), which are absent in the spectrum of the cross-linked SAM_C (Fig. 26 B,C). This observation correlates with the biophysical characterisation of the short construct, which showed that SAM_{wt} is capable of self-association and adopts a monomer-dimer equilibrium. The self-associated SAM_{wt} homodimer can be stabilised successfully by disulfide cross-linking utilising the S320C mutation.

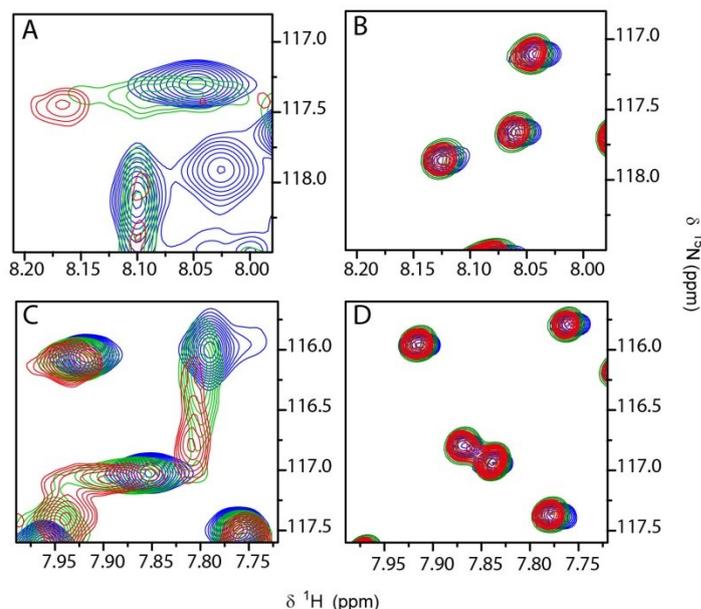


Fig. 26: 2D ^1H , ^{15}N -HSQC spectra of SAM_{wt} and SAM_C at varying concentrations demonstrate the effect of disulfide cross-linking on SAM domain self-association.

(A)+(C) Overlay of selected regions of 2D ^1H , ^{15}N -HSQC spectra of the SAM_{wt} recorded at SAM_{wt} concentrations of 14 (red), 50 (green) and 500 μM (blue). The dimerisation induced chemical exchange on the intermediate-to-fast timescale causes changes in the chemical shift and intensity of resonances. (B)+(D) Overlay of selected regions of 2D ^1H , ^{15}N -HSQC spectra of cross-linked SAM_C recorded at SAM_C concentrations of 14 (red), 140 (green) and 1400 μM (blue). In samples containing fully cross-linked SAM_C, the chemical exchange process observed by NMR for the wt domain successfully quenched the chemical exchange process observed by NMR for the wt domain.

4.5.4 Sequential assignment of backbone resonances of cross-linked SAM_C

The resonances of the backbone nuclei (i.e., $^1\text{H}^{\text{N}}$, $^1\text{H}^{\alpha}$, ^{15}N , $^{13}\text{C}^{\beta}$, $^{13}\text{C}^{\alpha}$) and $^{13}\text{C}^{\beta}$ of cross-linked SAM_C were assigned in a sequence-specific manner using the 2D and 3D triple resonances experiments as described in Section 3.6.2. Fig. 27 depicts a part of the sequential assignment of cross-linked SAM_C (K285–E294) using an HNCACB spectrum. The observable intra- and interresidual $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$ resonances are

connected to those of the preceding and succeeding resonances by matching chemical shifts at particular ${}^1\text{H}_i^{\text{N}}-{}^{15}\text{N}_i$ correlations.

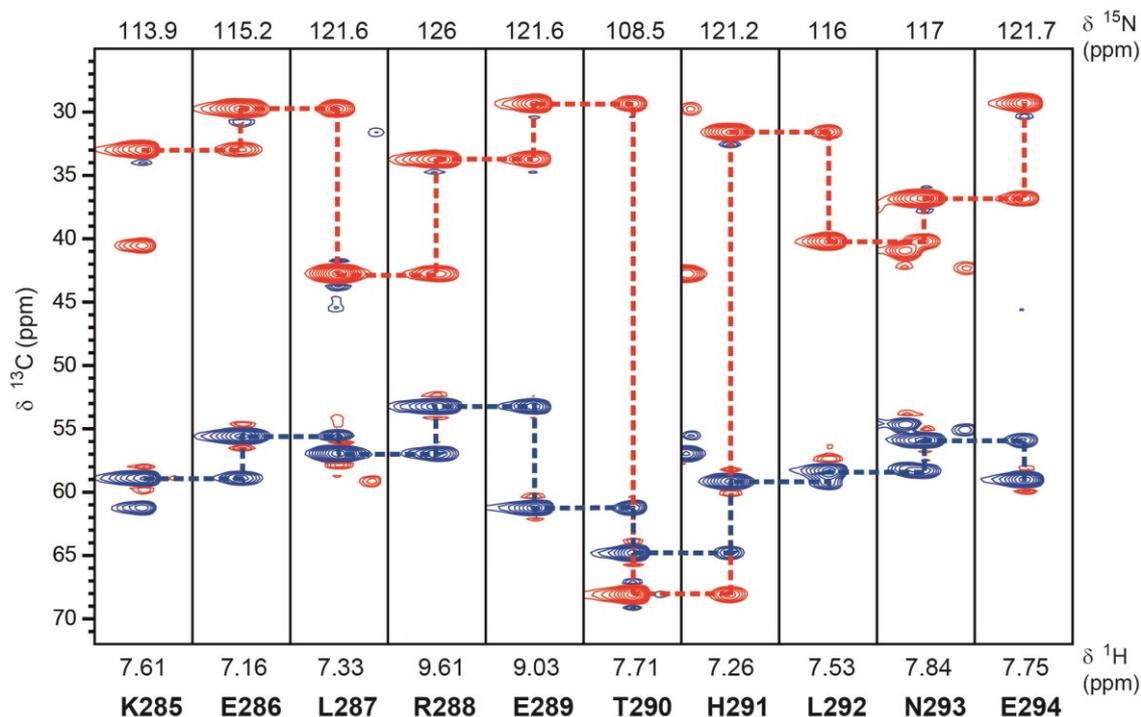


Fig. 27: Sequential assignment of resonances arising from backbone nuclei of residues K285-E294 of cross-linked SAM_C.

The ${}^1\text{H}_i^{\text{N}}-{}^{15}\text{N}_i$ strips taken from the 3D HNCACB experiment show resonances arising from ${}^{13}\text{C}^\alpha$ (blue) and ${}^{13}\text{C}^\beta$ (red) nuclei. In general, the chemical shift of the ${}^{13}\text{C}^\alpha$ resonances is downfield of the ${}^{13}\text{C}^\beta$ resonances, except T290 ${}^{13}\text{C}^\beta$, which is shifted to ~ 70 ppm because of the electron-withdrawing intra-residue side chain hydroxyl oxygen. The intra- and inter-residue connections between the ${}^{13}\text{C}^\alpha_i$ and ${}^{13}\text{C}^\alpha_{i-1}$ and ${}^{13}\text{C}^\beta_i$ and ${}^{13}\text{C}^\beta_{i-1}$ cross-peaks of the preceding and succeeding residues are indicated by the dotted lines.

This way, the 2D ${}^1\text{H}, {}^{15}\text{N}$ -HSQC spectrum of the cross-linked SAM_C was completely assigned (66 backbone amide cross-peaks) (Fig. 28). The N-terminal glycine and the two proline residues at positions 255 and 300 were excluded from the analysis because they do not give rise to ${}^1\text{H}_i^{\text{N}}-{}^{15}\text{N}_i$ correlations in the 2D ${}^1\text{H}, {}^{15}\text{N}$ -HSQC spectrum. The chemical shift information on the nitrogen of P300 was obtained by recording an HCAN spectrum. The presence of 66 backbone ${}^1\text{H}_i^{\text{N}}-{}^{15}\text{N}_i$ correlations corresponds to the expected number of backbone amide cross-peaks for one SAM_C molecule. A second conformation is not observable. Therefore, SAM_C must have been cross-linked in the form of a symmetric homodimer.

The side chain amide groups of all asparagine and glutamine residues were also sequence-specifically assigned. The assignment information for the side chain amides was obtained by comparing the chemical shift pattern of the side chain correlated

carbon resonances – $^{13}\text{C}^\beta$ and $^{13}\text{C}^\alpha$ for asparagine, $^{13}\text{C}^\gamma$ and $^{13}\text{C}^\beta$ for glutamine – in the HNCACB to the resonances assigned during the sequential assignment process.

Chemical shift information for $^1\text{H}^\alpha$ was derived from the TOCSY- ^1H , ^{15}N -HSQC and the H(CCO)NH experiments. The backbone carbonyl carbon ($^{13}\text{C}'$) was assigned using the 3D HNCO experiment using the assigned 2D ^1H , ^{15}N -HSQC spectrum. The carbonyl carbon resonances of 66 of 69 residues were assigned. The three missing assignments belonged to the N-terminal glycine and D299, which are both followed by prolines in the SLY1 SAM primary sequence (i.e., no amide proton at the i position), and the C-terminal D321. Overall, 337 of 346 (97.4%) protein backbone resonances were sequentially assigned (Table 29) and 64 of 65 possible C^β resonances.

4.5.5 Sequential assignment of backbone resonances of SAM_{wt} and non-cross-linked SAM_C

The sequential assignment of the backbone nuclei of reduced SAM_C was performed as described for the cross-linked SAM_C (Section 4.5.4) by correlating H^{N}_i and N_i nuclei with intra- (i) and inter-residue ($i-1$) $^1\text{H}^\alpha$, $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ using the same set of experiments as described above. Three-dimensional NMR data sets of reduced SAM_C were recorded at a concentration of 1.4 mM. At this concentration, which is more than one order of magnitude larger than the K_D of 117 μM determined by AUC, the monomer-dimer equilibrium is shifted significantly to the dimer state with a dimer population of ~82% (1.15 mM SAM domain in dimer state). Therefore, the resonances of SAM_C were less influenced by chemical exchange at this concentration than at lower concentrations and backbone assignments could in most cases successfully be made and mapped to the corresponding 2D ^1H , ^{15}N -HSQC (Fig. 28). Nonetheless, the signal belonging to A309 was still broadened beyond detection at this concentration. Its position was later identified using a sample with a SAM_{wt} concentration of 2.7 mM. Resonances belonging to the proline nitrogens were not assigned, because the HCAN spectrum was not recorded for the non-cross-linked SAM_C. All in all, 328 of 346 protein backbone resonances were assigned and 64 of possible 65 $^{13}\text{C}^\beta$ resonances.

Overlay of the 2D ^1H , ^{15}N -HSQC spectra of reduced and oxidised SAM_C samples revealed that the chemical shifts of the majority of the resonances are very similar in both states (cf. Fig. 28).

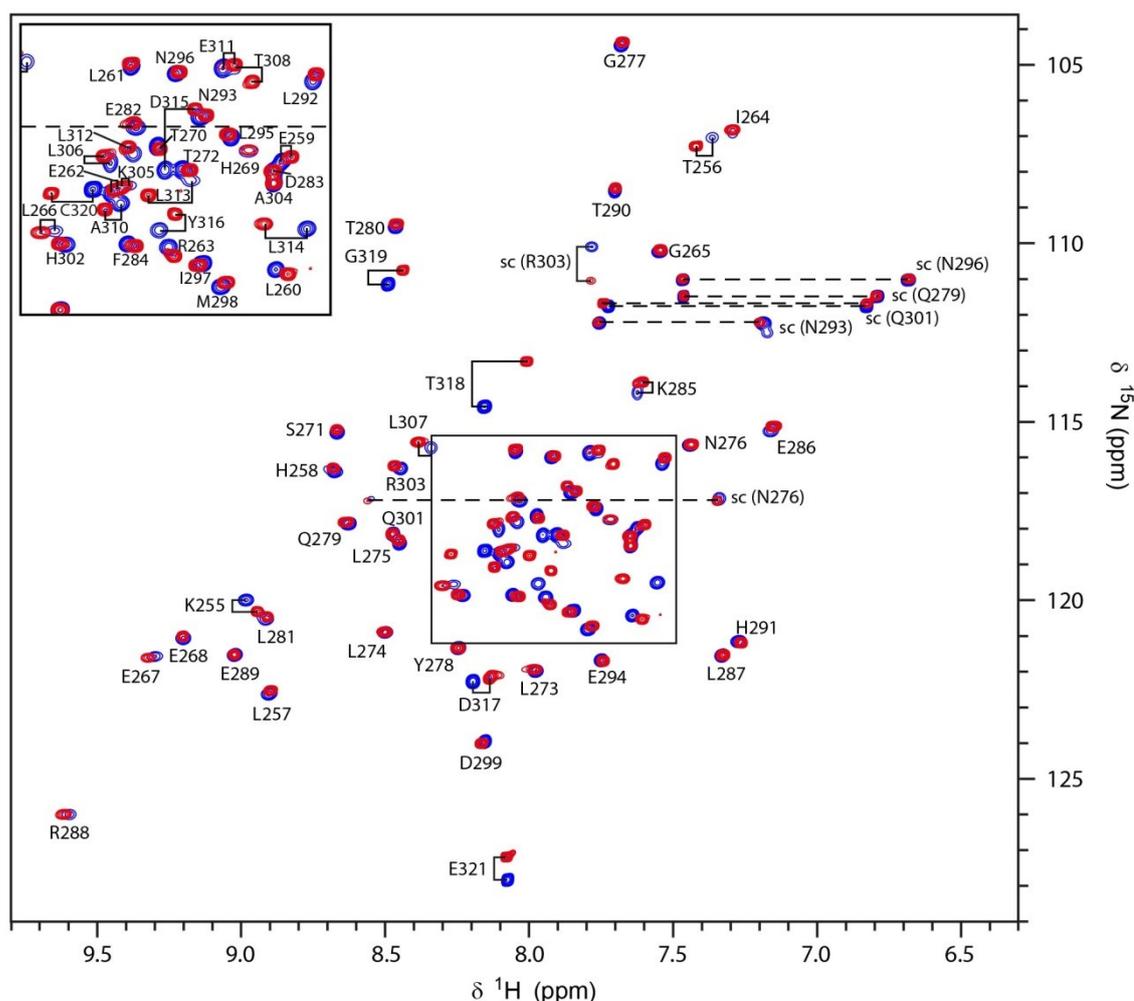


Fig. 28: Fully assigned 2D ^1H , ^{15}N -HSQC spectra of cross-linked and non-cross-linked SAM_C.

2D ^1H , ^{15}N -HSQC spectra were recorded for SAM_C at a concentration of 1.4 mM in the cross-linked (red) and non-cross-linked state (blue). The assignment of the backbone as well as asparagine and glutamine side chain amide resonances was performed, as described in Sections 4.5.4 and 4.5.5. Both 2D ^1H , ^{15}N -HSQC spectra show a very similar set of well-dispersed resonances, indicative of very high similarity between the cross-linked and the non-cross-linked dimer.

The assignments of the resonances arising from $^1\text{H}^{\text{N}}_i - ^{15}\text{N}_i$ correlations for the reduced SAM_C were mapped to the 2D ^1H , ^{15}N -HSQC spectrum of SAM_{wt} because of the very similar chemical shifts of the resonances (Fig. 29 and Fig. 30). The resonance of S320, which is not present in SAM_C, was identified by the elimination of all other possible assignment options. Sixty-six of the expected 66 backbone amid correlations could be assigned this way for SAM_{wt}.

4.5.6 Differences in the $^1\text{H}^{\text{N}}$ and ^{15}N chemical shifts between SAM_{wt} and SAM_{C}

As with the long construct, the $^1\text{H}^{\text{N}}$ and ^{15}N chemical shifts of the wt and the cross-linked and non-cross-linked short construct are similar with average $\Delta\delta_{\text{ave}}$ of 35 ppb and 14 ppb, respectively (Fig. 29). As observed with the longer construct, larger than average $\Delta\delta_{\text{ave}}$ values are observed for resonances corresponding to residues in close sequence and/or structural proximity to the mutation site. In addition to that, the cross-linked SAM_{C} shows larger than average chemical shift perturbations for resonances belonging to residues located at the N- and C-terminus, to residues K285 and E286 and to residues from K305–D315 when compared with that of the non-cross-linked SAM_{C} .

4.5.7 Assignment of side chain resonances of cross-linked SAM_{C}

The assignment information for the side chains of cross-linked SAM_{C} was obtained from *J*-correlated 3D NMR experiments as described in Section 3.6.2. The aromatic ring protons of phenylalanine and tyrosine residues were assigned using 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE spectra. N-terminal and lysine amino groups, guanidino groups of arginine, side chain hydroxyl protons of serine, threonine, tyrosine, thiol protons of cysteine, carboxyl resonances of aspartate and glutamate, and the quaternary aromatic carbons were excluded from the assignment process. ^1H nuclei belonging to the same methyl group, and phenylalanine/tyrosine $^1\text{H}^{\delta}$ and $^1\text{H}^{\epsilon}$ nuclei were counted as one signal. A total of 96.7% of all side chain nuclei were sequence-specifically assigned for cross-linked SAM_{C} using the available NMR data. A complete list of all assignments has been deposited in the BMRB under the deposition number 27432.

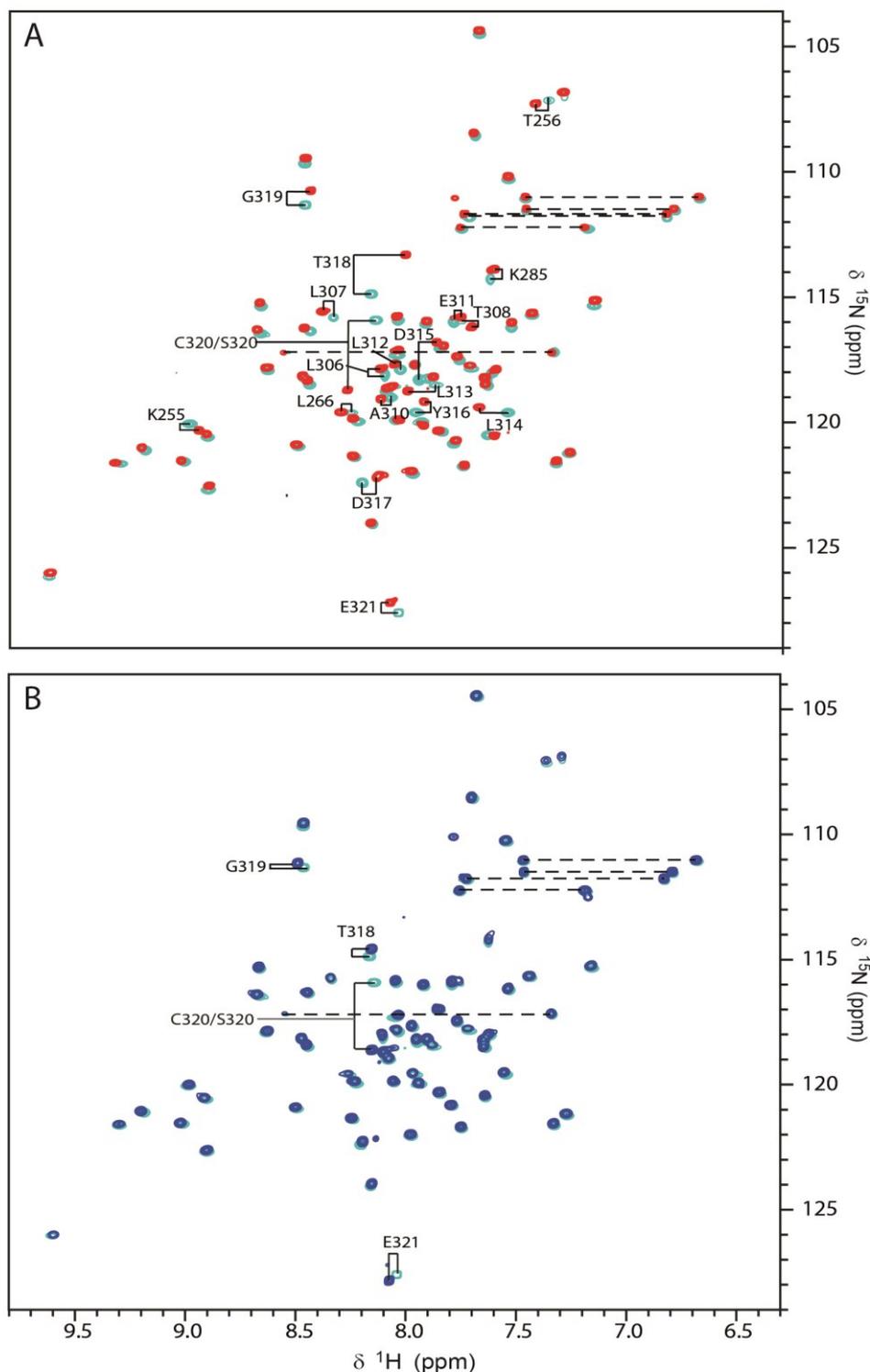


Fig. 29: Comparison of the 2D ^1H , ^{15}N -HSQC spectra of cross-linked and non-cross-linked SAM_C with the 2D ^1H , ^{15}N -HSQC spectrum of SAM_wt .

(A) Overlay of 2D ^1H , ^{15}N -HSQC spectra recorded on the cross-linked SAM_C (red) and SAM_wt (grey) at 1.4 mM and 2.7 mM, respectively. Only resonances showing a $\Delta\delta_\text{ave}$ value above the mean are annotated in the spectra (Fig. 30). (B) Overlay of 2D ^1H , ^{15}N -HSQC spectra recorded on the non-cross-linked SAM_C (blue) and SAM_wt (grey) at 1.4 mM and 2.7 mM, respectively. Only resonances showing a $\Delta\delta_\text{ave}$ value above the mean are annotated in the spectra (Fig. 30).

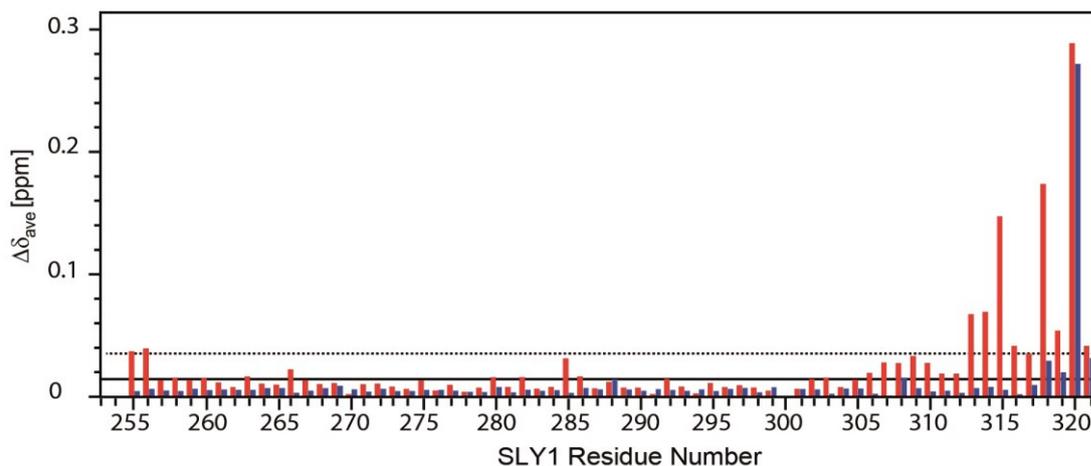


Fig. 30: Weighted average chemical shift ($\Delta\delta_{\text{ave}}$) perturbation analysis between cross-linked SAM_C and SAM_{wt} (red), and non-cross-linked SAM_C and SAM_{wt} (blue).

The $\Delta\delta_{\text{ave}}$ values per residue between wt and mutant SAM domains are small with a mean value of 35 ppb (dashed line) and 14 ppb (solid line) for the cross-linked and non-cross-linked mutant, respectively. Primarily residues in the vicinity of the mutation site (S320C) show $^1\text{H}^N$ and ^{15}N chemical shift changes that are above the mean values.

4.6 Determination of structural restraints from NMR data for SLY1 SAM_C structure calculation

4.6.1 Secondary structure determination from chemical shift data using TALOS+ and hydrogen bond couplings

Using the near complete resonance assignments of the cross-linked SAM_C , TALOS+ was used to predict the secondary structure elements and torsion angles φ and ψ of the protein (Section 3.6.4.1). Reliable (GOOD) torsion angle predictions could be made for 59 out of the first 64 residues, whereas C-terminal residues were mostly considered to be dynamic because of the high values of the random coil index (RCI), which is a measure of the random coil chemical shift contribution (Section 1.5.4) to the observed chemical shift (Fig. 31A).

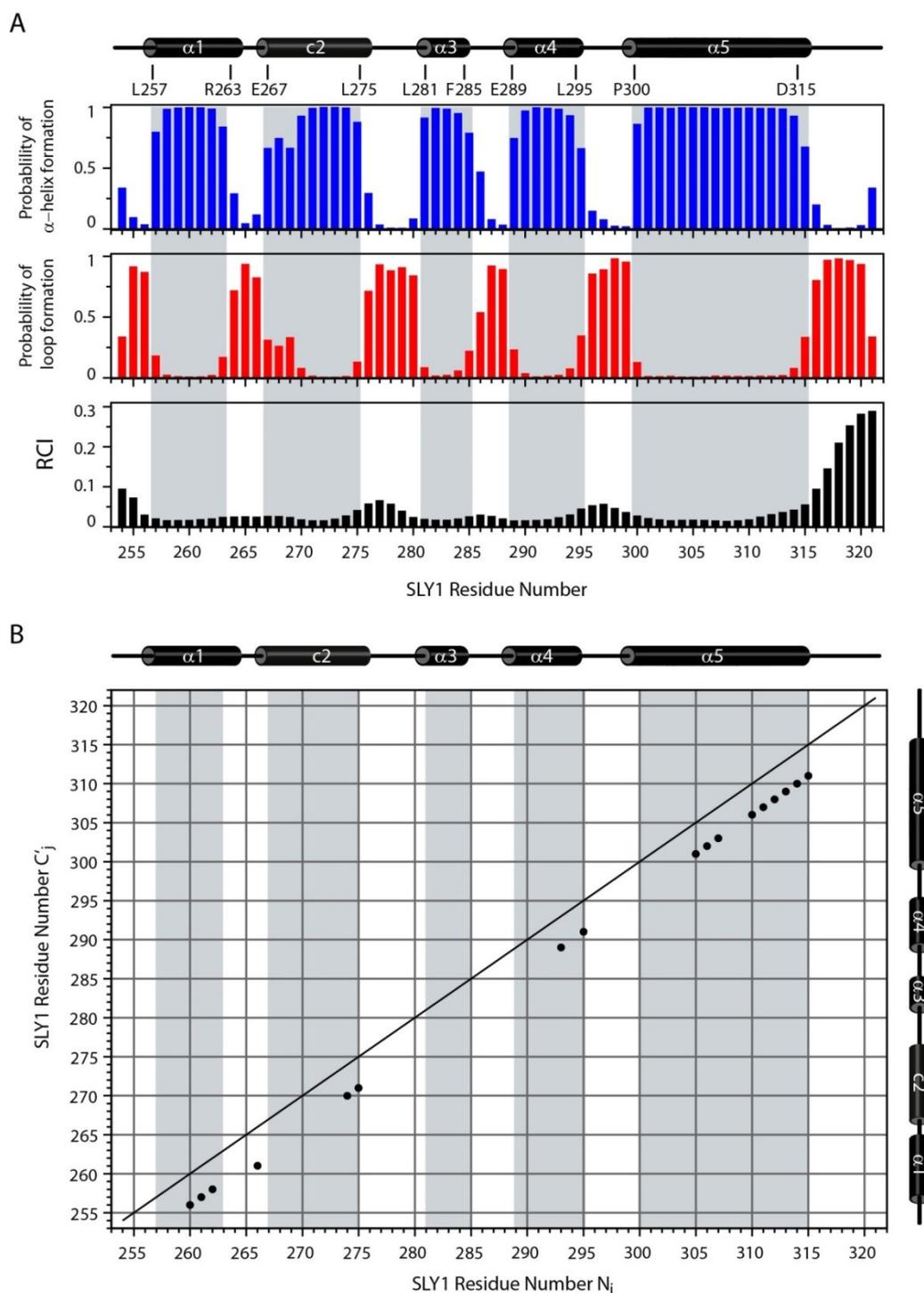


Fig. 31: Secondary structure restraints derived from TALOS+ torsion angle prediction and $^{\text{h}3}J_{\text{NC}}$ -couplings. (A) The probability of α -helix and loop formation calculated from chemical shifts by the TALOS+ artificial neuronal network (ANN) and the RCI of each residue. According to the TALOS+ prediction, SAM_C adopts the canonical five-helix fold of SAM domains. No secondary structure predictions could be made for the C-terminus. (B) Detection of hydrogen bonds by measuring $^{\text{h}3}J_{\text{NC}}$ -couplings between $^{15}\text{N}_i$ and $^{13}\text{C}_{i-4}$ in helices. The depiction of hydrogen coupled residue pairs by dots gives a line parallel to the diagonal of the diagram, which is characteristic for α -helices. A schematic representation of the secondary structure elements based on TALOS+ prediction is shown at the top and to the right.

The high helical content and the absence of β -sheet structures is typical for SAM domains and could largely be underpinned by the detection of α -helical hydrogen bond patterns using the long range HNCO experiment to measure the weak (< 1 Hz) $^{\text{h}3}J_{\text{NC}'}$ -couplings (Section 3.6.5). Fifteen $^{\text{h}3}J_{\text{NC}'}$ -couplings between $^{15}\text{N}_i$ and $^{13}\text{C}'_{i-4}$ could be unambiguously assigned to residues involved in helix formation (Fig. 31B). Only helix $\alpha 3$ did not give rise to any measurable couplings. However, helix $\alpha 3$ is the shortest of the predicted five helices consisting of only the minimum of required amino acids to form an α -helix, i.e., 4. Likewise, not all of the couplings expected from the secondary structure prediction were detected in other helices by the long range HNCO experiment. Interestingly, the first three residues of helix two (E267–L275) appeared to have a lower probability of α -helix formation than the average for residues involved in helix formation. This correlated with the calculated structure ensemble that revealed that the second helix is a composite helix (Pal *et al.*, 2005), which contains a 3_{10} -helix element (E267-H269) at the N-terminus of the α -helix. The presence of the 3_{10} -helix is corroborated by the absence of the typical α -helix hydrogen bond ($^{15}\text{N}_i - ^{13}\text{C}'_{i-4}$) pattern for these three residues. Therefore, the second helix will be referred to as composite helix c2, whereas the other four helices will be numbered as $\alpha 1$, $\alpha 3$, $\alpha 4$ and $\alpha 5$.

4.6.2 χ_1 -angle restraints determined from scalar couplings

The $^3J_{\text{H}\alpha\text{H}\beta 2/3}$ - and the $^3J_{\text{NH}\beta 2/3}$ -couplings were used to determine the conformation of the χ_1 torsion angles of amino acid side chains containing a methylene group ($-\text{CH}_2-$) at the C^β position. Measured coupling constants were translated into χ_1 rotameric states with boundaries of $\pm 30^\circ$ (Section 3.6.4.2).

The χ_1 -angles of 24 out of possible 51 methylene groups containing residues were determined (Fig. 32). Coupling constants of $6.5 \leq ^3J_{\text{H}\alpha\text{H}\beta 2/3} \leq 8.8$ Hz and $2.1 \leq ^3J_{\text{NH}\beta 2/3} \leq 3.4$ Hz correspond to χ_1 -angles that either adopt the energetically unfavourable eclipsed state (unlikely) or arise from the more likely dynamic averaging of the side chain rotating between the three preferred rotamer states. Such averaging is often observed in regions of a protein where the side chain has limited-to-no steric hindrance. Consequently, rather than the side chain adopting a single preferred state, it adopts all three states with differing populations. These 3J -coupling regions are shaded grey in Fig. 32.

The rotamer distribution of the 24 determined rotamer states is in agreement with the previously described (Section 3.6.4.2) preference for *gauche* (+) and *trans* rotamers. According to the measured coupling constants, the predominant rotamer state in the SAM_C cross-linked dimer is *gauche* (+) (83.3%). The rotamer state *trans* is adopted by 12.5%, and only E268 adopts the *gauche* (−) rotamer state, which may be due to the possible electrostatic interactions between E268, H269 and H302. The remaining 27 side chains could not be defined because they either exhibited rotameric averaging or because of chemical equivalence of the H^β protons. The determined rotamers were included in the structure calculation as conformational restraints.

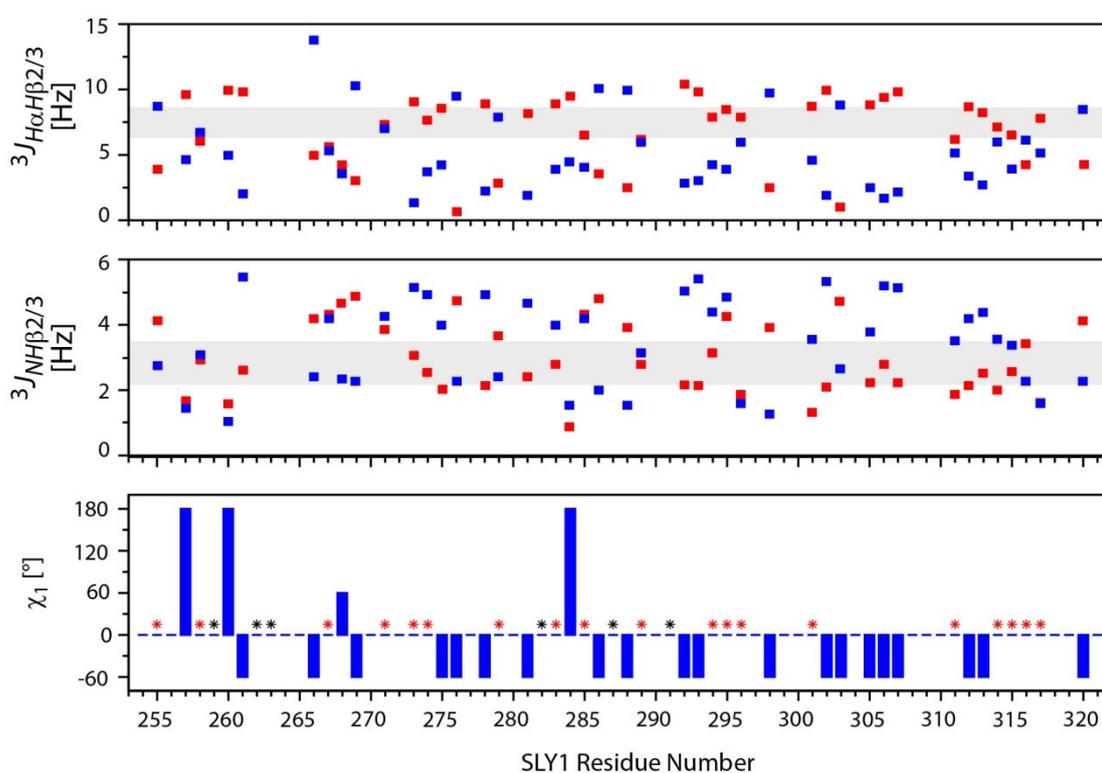


Fig. 32: ${}^3J_{H\alpha H\beta 2/3}$ - and ${}^3J_{NH\beta 2/3}$ -couplings for the determination of *trans*, *gauche* (+) and *gauche* (−) rotamer states.

χ_1 rotamer states were determined from scalar couplings using the χ_1 -related consensus Karplus coefficients (Case *et al.*, 1994; Pérez *et al.*, 2001) inferred from self-consistent J -coupling analysis using an amplitude of χ_1 -angle fluctuations of $\pm 30^\circ$ for each rotamer state. Thus, ${}^3J_{\alpha\beta}$ -couplings values of $6.5 \leq {}^3J_{H\alpha H\beta 2/3} \leq 8.8$ Hz and ${}^3J_{N\beta}$ -coupling values of $2.1 \leq {}^3J_{NH\beta 2/3} \leq 3.4$ Hz were considered to indicate dynamic averaging between χ_1 rotameric states. These regions are shaded grey in the figure. Rotamer states that could not be determined for residues due to rotameric averaging are marked with a red star. The rotamer state of residues that are marked with a black star could not be defined because of chemically equivalent H^β protons. Distinct rotamer states used in structure calculations of the SAM_C cross-linked dimer are indicated by the blue bars

4.6.3 Distance restraints determined from NOE data

Intra- and intermolecular distances between backbone, aromatic and aliphatic side chain protons were determined by NOE spectroscopy (Section 1.5.6). The resonance assignment of the 3D NOESY- ^1H , ^{15}N -HSQC and the aliphatic and aromatic 3D NOESY- ^1H , ^{13}C -HSQC was performed as described in Section 3.6.6 in a semi-automatic fashion. A total number of 5,026 NOEs were assigned. The removal of redundant NOEs, which are assigned in multiple spectra, reduced the number to 4,059 unique distance restraints which were used during the structure calculation. The observed NOEs are categorised by ARIA as ambiguously and unambiguously assigned NOEs. Unambiguous NOEs were further divided into intra- ($|i - j| = 0$) and inter-residue NOEs. Inter-residue NOEs were subcategorised into sequential ($|i - j| = 1$), medium range ($1 < |i - j| < 5$) and long range ($|i - j| \geq 5$) NOEs. The number of unambiguous NOEs per residues is depicted in Fig. 33, and the overall number of ambiguous, unambiguous, intra-residue, medium and long NOEs is presented in Table 29. There is a noticeable reduction in the number of NOEs for residues in helix $\alpha 5$ (Fig. 33; A304–D315). The reason for this reduction is the location of these residues at the domain interface. Therefore, many of the observed NOEs are intermonomer NOEs, which are not included (Fig. 34 B+C). The number of NOEs for the C-terminal residues D317-E321 is much lower than for other regions because this region is unstructured and flexible, as supported by the high RCI determined for these residues (Fig. 31 A).

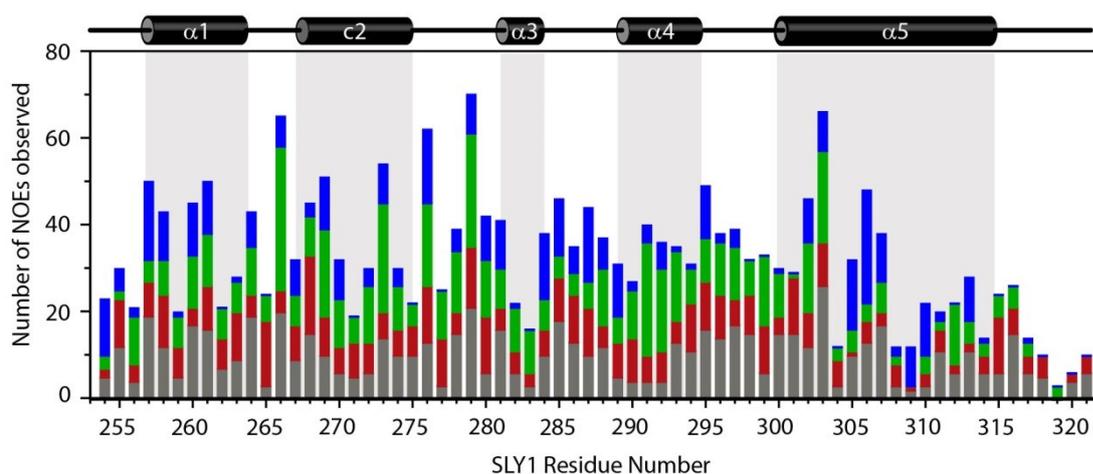


Fig. 33: Observed intra-residue (grey), sequential (red), medium range (green) and long range (blue) NOEs per SAM_C residue.

No inter-monomer NOEs are included here. A schematic representation of the secondary structure elements based on DSSP analysis of the final calculated structure ensemble is shown at the top.

4.6.4 Characterisation of the dimer interface by isotope-filtered/edited NOESY spectra and solvent PREs

Identifying and defining the dimer interface and the orientation of the two monomers relative to each other required the collection of isotope-filtered/edited NOESY spectra that provide spatial information about the residues positioned at the dimer interface (Section 1.5.6). ^{15}N , ^{13}C - ω_1 -filtered, ^{13}C -edited and ^{15}N -edited NOESY-HSQC spectra (Table 22) were recorded on a sample containing 50% $^{13}\text{C}/^{15}\text{N}$ -labelled SAM_C and 50% unlabelled SAM_C. Strips of the NOE dimension for residue I264 H^{72*} in the standard 3D NOESY- ^1H , ^{13}C -HSQC and 3D ^{15}N , ^{13}C - ω_1 -filtered NOESY- ^1H , ^{13}C -HSQC spectra are shown in Fig. 34A. Comparing the two strips elucidates the effect of the filtering step in the isotope-filtered/edited NOESY experiments. While the strip belonging to the standard 3D NOESY- ^1H , ^{13}C -HSQC shows intra- and intermolecular signals, only the intermolecular signals are present in the isotope-filtered/edited spectrum.

Two hundred and forty-nine intermolecular cross-peaks were identified in the two recorded isotope-filtered/edited NOESY-HSQC spectra. The majority of the intermolecular NOEs identified involved protons of residues located in helix α_1 , the loop region between helices α_1 and c_2 , and helix α_5 . Unambiguous NOEs observed in the filtered NOESY spectra were manually assigned, the remaining assignments were performed by ARIA (Section 3.6.6). Isotope-filtered/edited spectra can contain residual signals arising from intramolecular NOEs as a result of inefficient filtering. Therefore, the intermolecular NOE data were supported by information from experiments conducted in the presence of a paramagnetic relaxation enhancing agent. A set of 2D ^1H , ^{15}N -HSQCs of a sample containing 200 μM cross-linked SAM_C were recorded with varying concentrations (0–6 mM) of Gadodiamide (Section 3.6.7). The ratios of the intensities (I_{PRE}/I_0) of the amide group cross-peaks in the presence (I_{PRE}) and absence (I_0) of the solvent PREs were determined and analysed as a function of solvent PRE concentration. The strongest disparity between signal intensities of exposed and shielded nuclei was observed at a concentration of 4 mM Gadodiamide.

4. Results

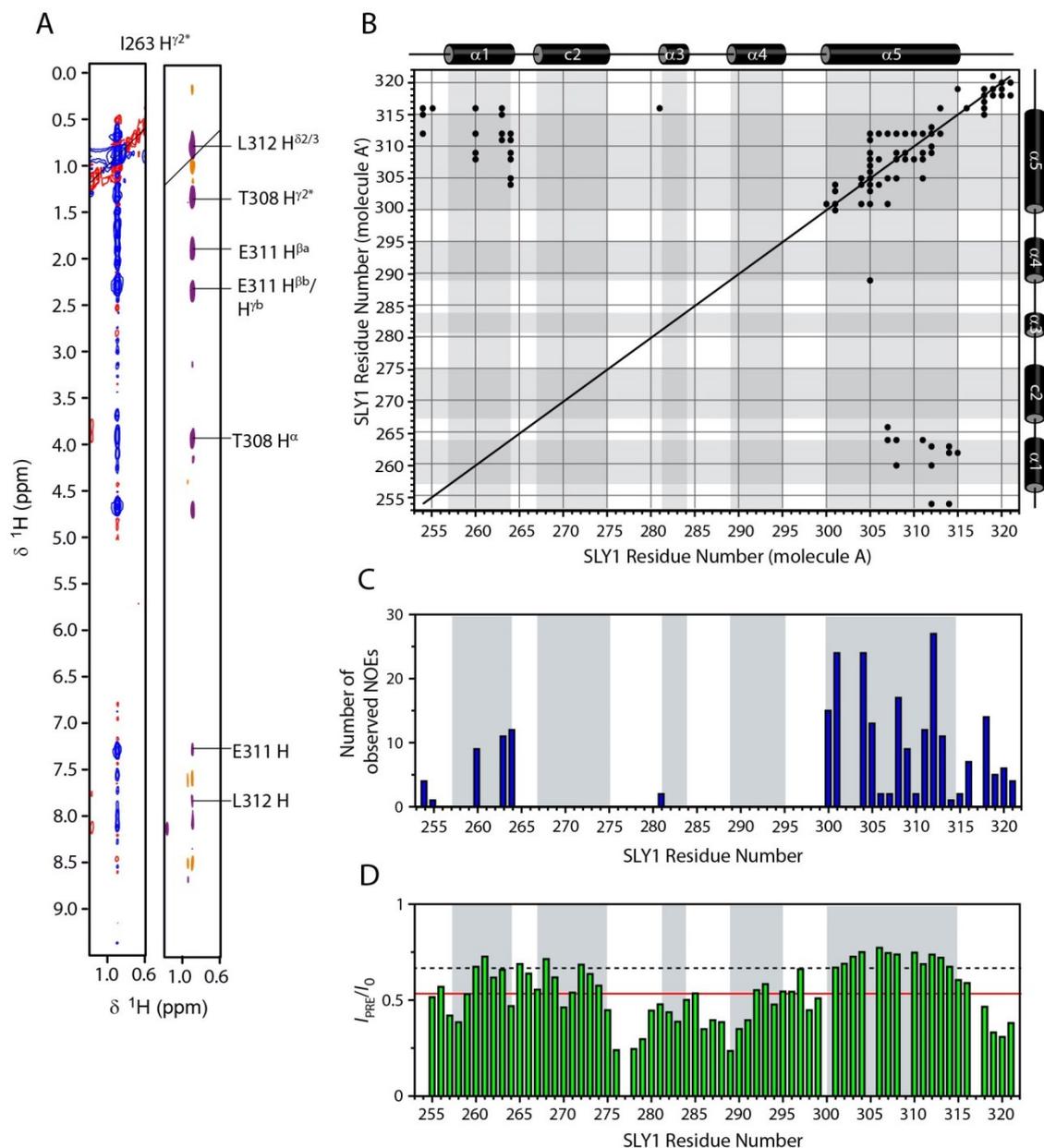


Fig. 34: Characterisation of the SLY1 SAM_C dimer interface by 3D ^{15}N , ^{13}C - ω ₁-filtered, ^{13}C -edited and ^{15}N -edited NOESY-HSQC spectra and solvent PRE analysis.

(A) Strips of the NOE dimension for I264 H γ 2* in the 3D NOESY- ^1H , ^{13}C -HSQC (left) and 3D ^{15}N , ^{13}C - ω ₁-filtered NOESY- ^1H , ^{13}C -HSQC (right) spectra. Assignments of intermolecular NOE resonances are listed on the right.

(B) Plot showing the unambiguous intermolecular NOEs between residues of molecule A and molecule A' of the SLY1 SAM_C cross-linked dimer. A schematic representation of the secondary structure elements based on the DSSP analysis of the final calculated structure ensemble is shown at the top and on the right.

(C) The number of observed unambiguously assigned intermolecular NOEs per residue. **(D)** The ratio of the amide group cross-peak intensities in the absence (I) and the presence (I_{PRE}) of 4 mM Gadodiamide. P254 and P300 do not give rise to a signal in 2D ^1H , ^{15}N -HSQC spectra, G277, K305, A309 and Y316 were excluded from the analysis because of signal overlap. The mean intensity and 1 σ above the mean intensity are indicated by the red and dotted line, respectively.

I_{PRE}/I_0 ratios of the cross-peak intensities at 4 mM Gadodiamide concentration are plotted against the SLY1 SAM residue number in Fig. 34D. All residues were influenced by the presence of 4 mM Gadodiamide, showing a reduction in signal intensity when compared with the corresponding intensities in the absence of Gadodiamide. This observation is in agreement with the distance dependency of the PRE effect, which affects the relaxation behaviour of nuclei in up to 20–30 Å distance from the paramagnetic molecule.

Nonetheless, there are clear differences in the intensity reductions of the cross-peaks (Fig. 34D). For example, the intensity of the signals corresponding to residues L275–N293 show a greater reduction in intensity compared with signals associated with other regions of the protein, thus indicating that these amide groups are affected more by the soluble paramagnetic agent due to greater solvent exposure. Least affected by the presence of the Gadodiamide were cross-peaks arising from backbone amide groups of residues L260, L261, G265, E268 and Q301–L312. These cross-peak ratios I_{PRE}/I_0 are more than 1σ above the mean intensity ratio of all cross-peaks. Notably, the cross-peaks belonging to Q301–L312 have uniform intensity ratios that are among the highest measured in the entire spectrum. These residues account for the majority of the predicted helix 5. For the amide groups of the residues belonging to the other four helical elements of SAM_C, periodic fluctuations in intensity are observed. These fluctuations probably depend on whether residues are buried in the core of the domain or are surface exposed, and therefore show a stronger susceptibility to PRE induced line broadening. These fluctuations are absent in helix $\alpha 5$, which indicates shielding of the 15 residue long helix $\alpha 5$ from the soluble paramagnetic agent, most likely by a second SAM_C molecule. This observation is in agreement with the high number of intermolecular NOEs identified for the last 20 residues, providing substantial evidence that helix $\alpha 5$ is located at the dimer interface. Therefore, in the cases of multiple HSQC-base assignment options of the NOE cross-peaks, assignment options for residues in helix $\alpha 5$ were favoured.

4.7 NMR structure calculation

The structure of the SLY1 SAM_C cross-linked dimer was determined from NMR data using the ARIA software, which includes CNS as the structure calculation engine (Section 3.6.7). Backbone and side chain torsion angle restraints from TALOS+ and 3J -couplings, respectively, and hydrogen bond restraints from $^{\text{h}^3}J_{\text{NC}'}$ -couplings were included in the structure calculation along with the 4,059 NOE derived distance restraints. The quality of the ensemble of the 15 lowest energy structures was evaluated regarding the number and quality of the violations of the experimentally derived restraints. In addition, the covalent geometry (allowed backbone and side chain angles), the agreement of the structures with the input data (RPF score) and the coordinate precision (RMSD) of the determined family of lowest energy structures were examined as described in Section 3.6.10.

4.7.1 Analysis of the NMR derived structure ensemble

During the process of structure determination, the input data was manually refined between cycles of automated NOE assignment and structure calculation by ARIA. The refinement predominantly included the correction of conflicting assignments and the identification of spectral artefacts. Refinement was stopped when no further improvement in the quality of the structure ensemble could be achieved.

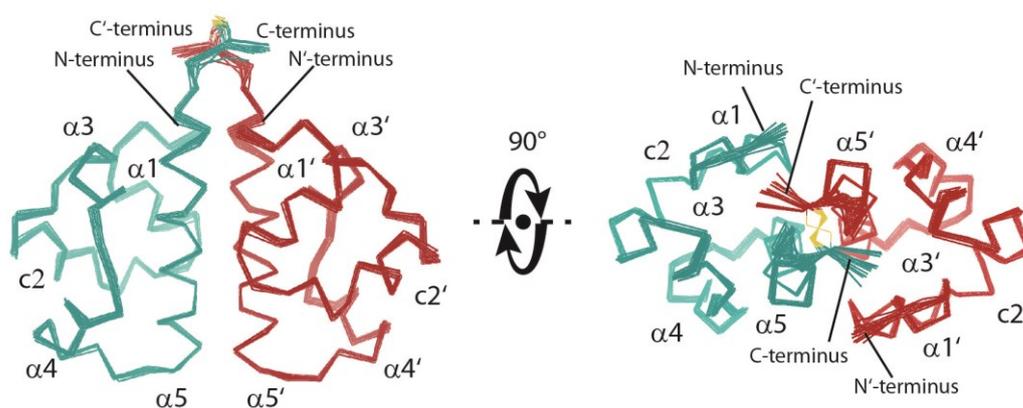


Fig. 35: Superposition of the 15 lowest energy structures calculated from NMR data.

The superposition of the 15 lowest energy models in the structural ensemble shows minimal deviation of the backbone heavy atoms from their average atomic position (RMSD: 0.31 Å). The strongest variation can be seen at the N- and C-termini. The stabilising cysteine at position 320 adopts a right-handed staple conformation in 14 of the 15 structures, while in one model the disulfide bond adopts a left-handed staple conformation. The molecules are drawn in line representation, with subunit A coloured in teal and subunit A' coloured in red. The S^γ atoms of the cysteine residues and the disulfide bond are yellow.

The final calculated structure ensemble of the SLY1 SAM_C cross-linked dimer (Fig. 35) fulfilled the vast majority of the input structural restraints, while at the same time exhibiting a high coordinate precision and proper covalent geometry. According to the ARIA derived violation analysis, of the 4,059 NOE derived distance restraints, none were violated by more than 0.1 Å. The RMSD from the 59 TALOS+ derived torsion angle restraints and the 24 experimentally determined χ_1 -angles was $(2.5 \pm 0.26)^\circ$ and $(1.27 \pm 0.62)^\circ$, respectively. All statistics regarding the fulfilment of experimental restraint data is listed in Table 29. The goodness-of-fit of the structure to the provided NOE data was also assessed using the RPF server (Huang *et al.*, 2012). Structures are considered to be correct solutions of the recorded NOE data when Recall and Precision scores and the F-measure adopt values above 0.8, and the DP score is above 0.7 for the calculated structural ensemble (Huang *et al.*, 2005). For the SAM_C dimer structures, Recall and Precision scores and F-measure values are 0.936, 0.946 and 0.941 respectively.

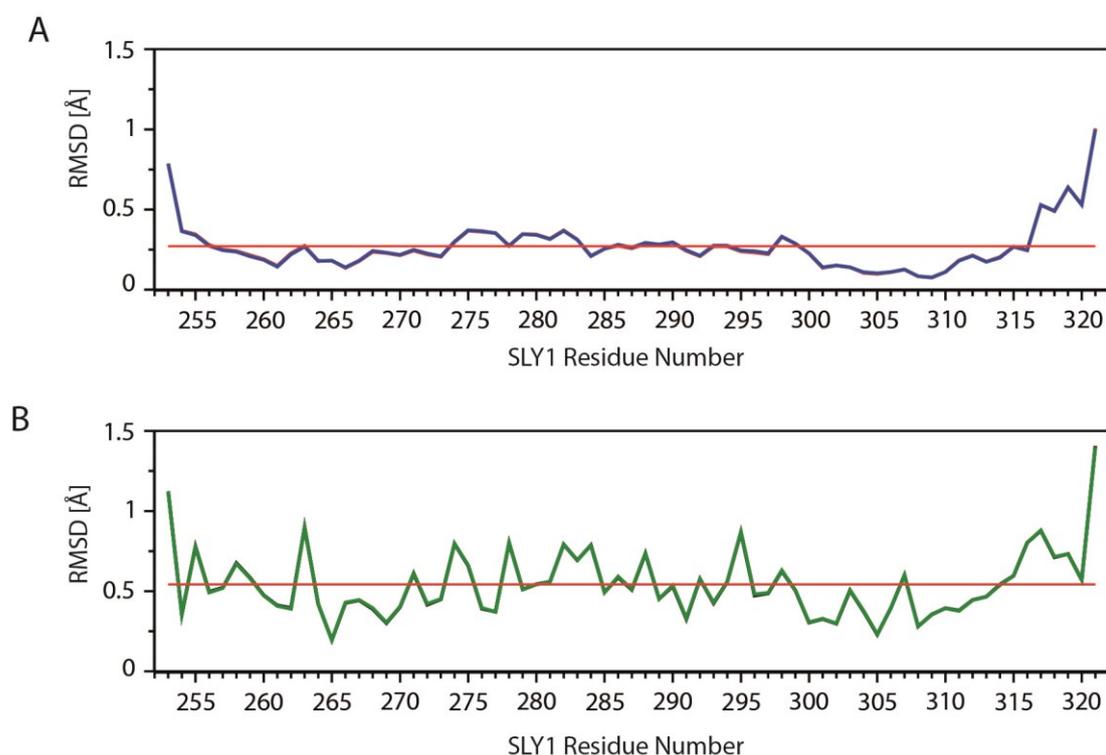


Fig. 36: RMSD from the mean atom position of the backbone (A) and all (B) heavy atoms per residue. The average RMSD of all residues is indicated by the red line and calculated to be 0.31 Å for the backbone atoms and 0.57 Å for all heavy atoms.

The RMSD of the structure ensemble for the SAM_C structure was calculated using the CcpNmr Analysis software and was provided on a per residue basis for the backbone (Fig. 36A) and all heavy atoms (Fig. 36B). The backbone RMSD values show that the deviation of the atom positions between P254 and Y316 are very small ($\sigma = 0.23 \text{ \AA}$) because this part of the protein is very well defined by the experimentally determined structural restraints. There are no restraints describing the structural position of the N-terminal glycine because the N-terminal residue was not observed in the recorded NMR experiments. The C-terminal part from Y316 onwards shows an increase in RMSD, which correlates with the decreasing number of NOE distance restraints (Fig. 33) describing the spatial position of these amino acid atoms. This, in addition to high RCI values (Fig. 31A) for the C-terminus, clearly indicates that the C-terminus is more flexible than the rest of the domain and samples a number of discreet conformations even though a disulfide bond was introduced. The average RMSD of the backbone atoms of all residues is 0.31 \AA . The average RMSD of all heavy atoms of all residues is 0.57 \AA . The region with the lowest variability in the structure ensemble for both, the backbone and side chains, is between P300 and D315. This region corresponds to helix $\alpha 5$. Helix $\alpha 5$ is positioned at the dimer interface and well defined by secondary structure restraints as well as intra- and inter-molecular NOE derived distance restraints.

Covalent geometry of the structure ensemble was assessed using the program PROCHECK_NMR. PROCHECK_NMR sorts the backbone torsion angles in energetically favoured, allowed and disallowed regions as proposed by Ramachandran *et al.*, 1963 and Morris *et al.*, 1992. More than 85% of the backbone torsion angles of the SAM_C cross-linked dimer are located in the most favoured regions of the Ramachandran plot, 13% in the additionally allowed regions and 0.8% in both, the generously allowed and disallowed regions. The only residue located in the generously allowed and disallowed regions of the plot is C320.

The stabilising disulfide bond at position 320 adopts a negative right-handed staple (–RHStaple) conformation (side chain dihedrals of the two involved cysteines negative except for χ_3 , which is positive) (Schmid *et al.*, 2006) in 14 of the 15 structures. The dihedral angles for C320 in subunit A and C320 in subunit A' with average χ_{1A} , χ_{2A} , χ_3 , $\chi_{1A'}$ and $\chi_{2A'}$ values of -62.9° , -66.6° , 124.8° , -62.8° and -66.6° , respectively. In one structure it adopts a negative left-handed staple (–LHStaple) conformation (χ_{1A} , $\chi_{1A'}$ and χ_3 are negative and χ_{2A} and $\chi_{2A'}$ are positive). The χ_{1a} , χ_{2a} , χ_3 , χ_{1b} and χ_{2b} values for this conformation were -83.1° , 83.6° , -125.5° , -83.5° and 84.1° .

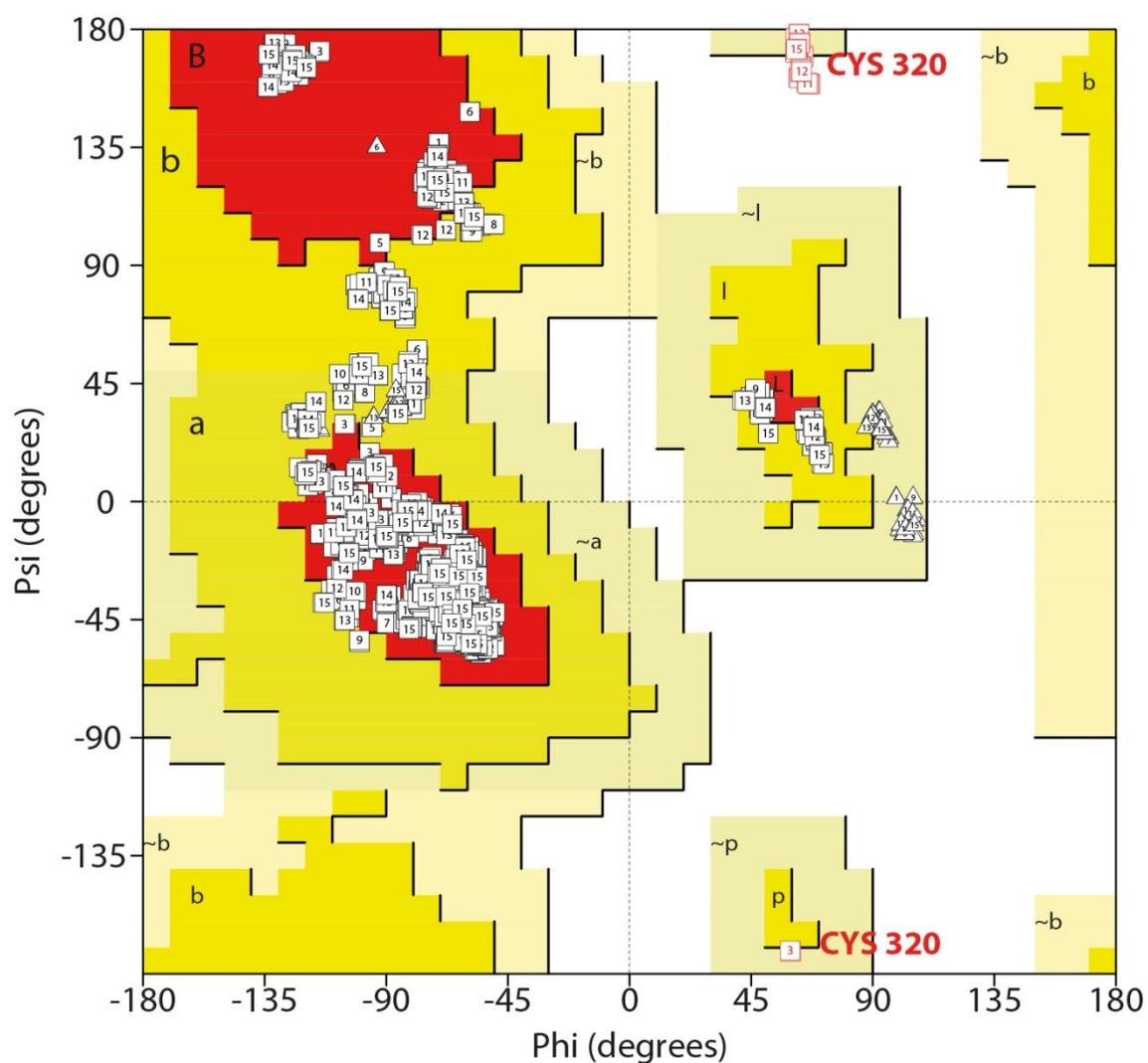


Fig. 37: Ramachandran plot of the backbone torsion angles of the SAM_C 15 lowest energy structures.

The plot shows the correlation of the ϕ, ψ -angles for each residue of all 15 lowest energy structures present in the calculated structural ensemble. All residue types are depicted as white squares, except for glycine residues, which are depicted as white triangles. The number inside the squares and triangles are identifiers for the particular model that the residue represents. The red areas of the plot are the favoured regions, the yellow areas are the additionally allowed regions, the light yellow regions are the generously allowed regions and the white areas are disallowed regions. (Morris *et al.*, 1992). The only residue that is repeatedly found in the generously allowed regions and disallowed regions is C320, probably because of the high-energy contribution of the disulfide bond to that residue.

4. Results

Table 29: Assignment and structural statistics for the SLY1 SAM_C structure ensemble.

| Assignment statistics | | | |
|---|-----------|--|----------|
| Completeness of resonance assignments (%) | | | |
| Protein, backbone ^a (337/345) | | 97.4 | |
| Protein, all atoms ^b (788/827) | | 97.0 | |
| Structural statistics | | | |
| NOE distance restraints | | | |
| Assigned | | 2466 | |
| Intraresidue ($i = j$) | | 694 | |
| Interresidue | | | |
| Sequential ($ i - j = 1$) | | 493 | |
| Medium range ($1 < i - j < 5$) | | 637 | |
| Long range ($ i - j \geq 5$) | | 406 | |
| Intermolecular | | 236 | |
| Ambiguous | | 1593 | |
| Hydrogen bonds | | 16 | |
| Dihedral angle restraints | | φ : 59; ψ : 59; χ_1 : 24 | |
| Experimental distance restraints violations $> 0.1 \text{ \AA}$ | | 0 | |
| RMSD value from TALOS+ torsion angle restraints ($^\circ$) (59) | | 2.5 ± 0.26 | |
| RMSD value from χ_1 -angle restraints ($^\circ$) (24) | | 1.27 ± 0.62 | |
| Deviations from idealised covalent geometry | | | |
| Bonds (\AA) | | 0.0026 ± 0.00007 | |
| Angles ($^\circ$) | | 0.5 ± 0.007 | |
| Improper ($^\circ$) | | 1.27 ± 0.077 | |
| Coordinate precision (\AA) ^c | | | |
| Backbone | | 0.31 ± 0.08 | |
| All heavy atoms | | 0.57 ± 0.06 | |
| RPF scores ^d | | | |
| Recall | Precision | F-measure | DP-score |
| 0.936 | 0.946 | 0.941 | 0.872 |
| Ramachandran statistics (%) ^e | | | |
| Residues in most favoured regions | | 85.5 ± 1.4 | |
| Residues in additionally allowed regions | | 13 ± 1.3 | |
| Residues in generously allowed regions | | 0.8 ± 0.8 | |
| Residues in disallowed regions ^f | | 0.8 ± 0.8 | |

^a Obtained from the $^1\text{H}^N$, $^{13}\text{C}^\alpha$, ^{15}N , $^{13}\text{C}^\gamma$ and $^1\text{H}^\alpha$ resonances.

^b Routinely assigned ^1H , ^{15}N and ^{13}C resonances are taken into account, excluding the N-terminal and lysine amino groups, guanidino groups of arginine, side chain hydroxyl protons of serine, threonine, tyrosine, thiol protons of cysteine, carboxyl resonances of aspartate and glutamate, and the quarternary aromatic carbons. ^1H belonging to the same methyl group and phenylalanine, tyrosine $^1\text{H}^\delta$, $^1\text{H}^\epsilon$ are counted as one signal.

^c Coordinate precision is the average root-mean-square-deviation of all the structures from the average structure coordinates.

^d Calculated RPF scores according to Huang *et al.*, 2012

^e Calculated by the program PROCHECK-NMR

^f Residues in the disallowed region are outside the core region of the domain

4.7.2 Solution structure of the SLY1 SAM_C cross-linked dimer

The solution structure of the SLY1 SAM cross-linked dimer consisted of two SAM_C subunits interlinked by an intermolecular disulfide bond at the C-terminus. The two SAM_C molecules adopt a five-helix fold, which is typical for SAM domains (Fig. 38A). Helices $\alpha 1$, $\alpha 3$, $\alpha 4$ and $\alpha 5$ are uniform α -helices, whereas helix c2 is a composite helix consisting of an N-terminal 3_{10} -turn (E267, E268, H269) followed by an α -helix. The secondary structure elements were assigned by DSSP. The two SAM_C subunits form a dimer with a 180° rotation symmetry, where the two $\alpha 5$ helices face each other, whereas the N-terminus of helix $\alpha 1$, the C-termini of helices c2 and $\alpha 4$ face outward and are not located at the interface.

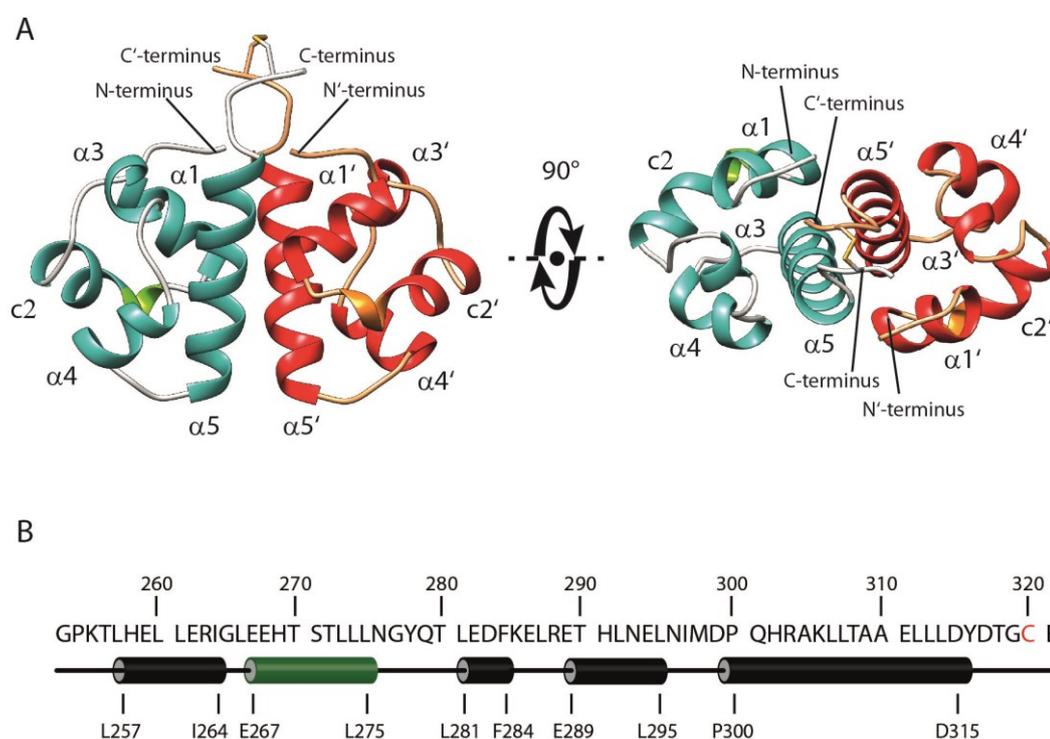


Fig. 38: Structure of SLY1 SAM_C cross-linked dimer as determined from the NMR derived restraints and calculated by ARIA.

(A) The closest to the average structure of the calculated structural ensemble consisting of the 15 lowest energy structures is shown. The two subunits A and A' adopt a five-helix folds. The five helical secondary structure elements are coloured in teal, with the 3_{10} -helix turn of helix c2 additionally highlighted in green for subunit A, and in red and orange for subunit A'. The SAM_C cross-linked dimer forms a symmetric homodimer with a 180° rotation symmetry. The dimer interface is formed primarily by residues in helix $\alpha 5$. (B) Secondary structure elements as determined by DSSP from the coordinates of the closest to average NMR structure of the NMR structure ensemble.

4.7.3 Interface of the cross-linked SAM_C dimer

One hundred and sixty-one of the 236 (68%) unambiguously assigned intermolecular NOEs arose from residues positioned in helix $\alpha 5$, whereas 36 intermolecular NOEs (15%) are accounted for by the adjacent C-terminal residues Y316–E321, which are held in spatial proximity by the disulfide bond. The remaining 36 (17%) intermolecular NOEs were assigned to residues P254, K255, L260, R263 and I264, localised at the N-terminus and in helix $\alpha 1$ (Fig. 39). The calculated dimer structure shows that the SAM_C monomers face each other with helices $\alpha 1$ and $\alpha 5$, and the N- and C-termini. As the NOE and PRE data indicated, the majority of the interface is formed by the $\alpha 5$ helices facing each other with an inter-axial angle Ω as defined by (Chothia *et al.*, 1977) between the packed helices of $\sim -50^\circ$. The N-terminus and helix $\alpha 1$ are near perpendicular to helix $\alpha 5$ in each monomer, and the side chains of residues in these structural elements extend across the $\alpha 5$ - $\alpha 5$ interface and interact with side chains of the opposing helix $\alpha 5$.

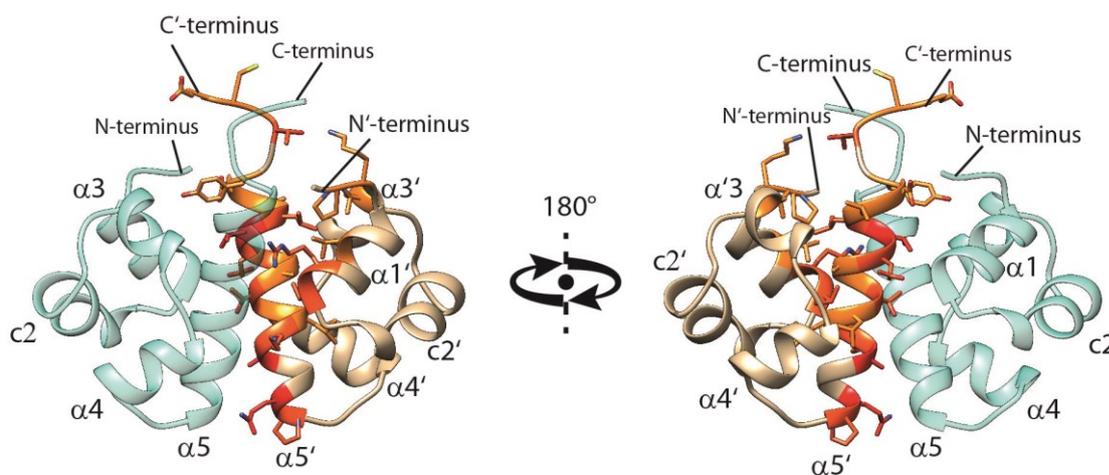


Fig. 39: Residues at the interface of the SLY SAM_C dimer structure

Structure of the SLY SAM_C homodimer with both monomers (cyan and tan) presented in ribbon representation. Side chains of residues that gave rise to intermolecular NOEs are drawn in stick representation for molecule A' (tan). These residues are localised in helix $\alpha 5$ and the unstructured N- and C-termini and helix $\alpha 1$. The colour-coded scheme for side chains in molecule A' illustrates the number of intermolecular NOEs detected, where orange indicates 1–10 NOEs; orange-red: 10–20 NOEs; and red: >20 NOEs.

In the surface representation of the SAM_C subunit A in Fig. 40, a hydrophobic groove is visible, which lies inbetween helix α_5 and helices α_1 and c_2 , and is sandwiched between negatively (red) and positively (blue) charged surfaces. In the dimer state, helix α_5 of subunit A' (ribbon representation) interacts along that hydrophobic patch with its hydrophobic residues A304, A309, L312 and L313, and polar residues T308 and Y316 reaching into the hydrophobic core of the pocket (Fig. 41A).

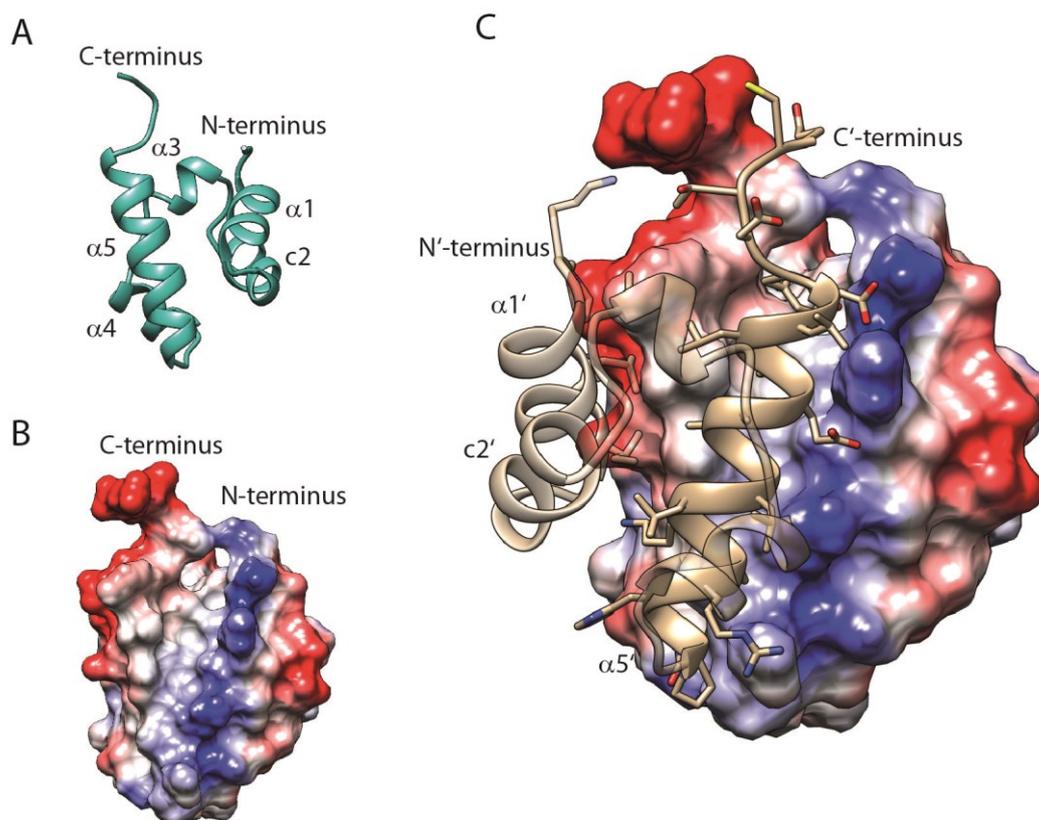


Fig. 40: Dimer interface of the SLY1 SAM_C cross-linked dimer in ribbon and surface representation. (A) SLY1 SAM_C subunit A in ribbon representation. (B) SLY1 SAM_C subunit A in a surface representation. The surface colouring is based on the electrostatic potential at pH 6.4 with negative charges in red and positive charges in blue. The visualisation of the surface charge reveals a hydrophobic stretch encompassed by helices α_5 , α_1 and c_2 . (C) Helix α_5 of molecule A' (in ribbon representation) aligns with the hydrophobic stretch of molecule A.

Additional analysis of the dimer interface of the calculated structures was performed using the PISA web server. PISA revealed that in addition to the disulfide bond, the dimer is stabilised by 12 hydrogen bonds and salt bridges, which are listed in Table 30. The negatively charged side chains of E311 and D315 (subunit A') interlock with the positively charged area formed by the N-terminal amino group of the N-terminal glycine and the guanidino group of R262 (subunit A) and form a network of salt bridges and hydrogen bonds that stabilise the dimer at the C-terminal end of helix α_5

(Fig. 41B). At the N-terminal end of the helices α_5 , the subunits A and A' are held together by the formation of hydrogen bonds across the interface formed by the side chain amide group of Q301 and the carbonyl group of P300 of the opposite subunit (Fig. 41C). The BSA was determined to be 1200 \AA^2 for the average NMR derived structure and ranging between $1168\text{--}1230 \text{ \AA}^2$ among the calculated structural ensemble.

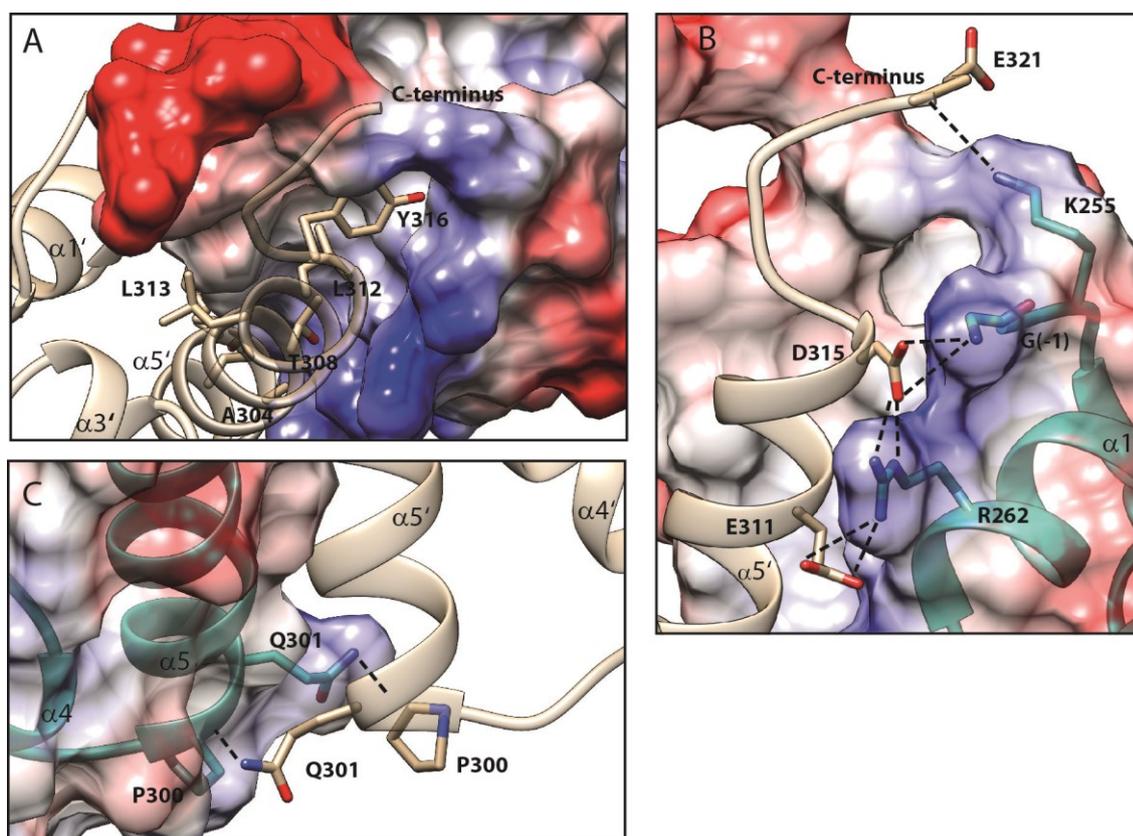


Fig. 41: Non-covalent interactions stabilising the dimer interface of SAM_C cross-linked dimer.

(A) Amino acids along helix α_5 of molecule A' reach into the hydrophobic pocket encompassed by helices α_1 , α_2 and α_5 of molecule A. (B) Network of hydrogen bonds and salt bridges between the guanidino group of R262, the side chain amino group of K256 and the N-terminal amino group of the N-terminal glycine (molecule A) and the side chain oxygens of E311 and D315 as well as the C-terminal backbone carboxyl group of E321 (subunit A'). (C) Hydrogen bonds predicted by PISA between the side chain amide nitrogen of N301 and the carbonyl oxygen of P300 of the respective opposite chain.

Table 30: Hydrogen bonds and salt bridges across the SAM_C dimer interface predicted by PISA in the NMR structure

| Molecule A | | | Molecule A': | | | Distance (Å) |
|------------|---------|--------------------|--------------------|---------|--------|-------------------|
| SSE | Residue | Atom | Atom | Residue | SSE | |
| N-term | P254 | [O] | [O ^η] | Y316 | C-term | 2.97 |
| N-term | K255 | [N ^ζ] | [OXT] | E321 | C-term | 2.70 |
| α1 | R262 | [N ^ε] | [O ^{δ1}] | D315 | α5 | 2.62 |
| α1 | R262 | [N ^{η1}] | [O ^{ε2}] | E311 | α5 | 3.44 ^a |
| α1 | R262 | [N ^{η1}] | [O ^{ε1}] | E311 | α5 | 2.99 |
| α5 | P300 | [O] | [N ^{ε2}] | Q301 | α5 | 2.91 |
| α5 | Q301 | [N ^{ε2}] | [O] | P300 | α5 | 2.91 |
| α5 | E311 | [O ^{ε1}] | [N ^{η1}] | R262 | α5 | 2.99 |
| α5 | E311 | [O ^{ε2}] | [N ^{η1}] | R262 | α5 | 3.44 ^a |
| α5 | D315 | [O ^{δ1}] | [N ^ε] | R262 | α1 | 2.62 |
| C-term | Y316 | [O ^η] | [O] | P254 | N-term | 2.97 |
| C-term | E321 | [OXT] | [N ^ζ] | K255 | N-term | 2.70 |

^aAccording to Jeffrey, 1997, hydrogen bonds beyond 3.4 Å are considered to be weak, electrostatic interactions. Distances larger 3.6 Å were excluded from the analysis.

4.8 SAM_{wt} structure determination by X-ray crystallography

4.8.1 Screening of crystallisation conditions for SLY1 SAM_{wt}

Different crystallisation conditions were screened for crystal formation of SLY1 SAM_{wt} (Section 3.7.1) using eight commercially available kits for sparse matrix screening (Table 6). The buffer conditions that resulted in crystal formation are listed in Table 31. An incubation period of 5–7 days in the listed conditions yielded tetragonal protein crystals. Fine-screen of the listed crystallisation conditions during which the components of the original conditions were varied yielded in most cases crystals of the same size, shape and stability. The incubation period was again 5–7 days. The crystal used for structure determination (Fig. 42A) was grown in 0.1 M K₂HPO₄, 2.2 M (NH₄)₂SO₄, 0.1 M NaCl and 0.1 M imidazole, which is based on composition 29 of the NH₄SO₄ suite by Qiagen. The same buffer supplemented with 10% glycerol was used for cryo-preservation of the crystal. The diffraction data were recorded at beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France (Section 3.7.2). Detailed information on data acquisition can be found in Table 32. The SAM_{wt}

4. Results

crystal belonged to the tetragonal space group $P 4_1 2_1 2$ and diffraction data could be recorded up to a resolution of 2.05 \AA (Fig. 42B). Based on the unit cell dimensions, the size of the domain and *Matthews coefficient* probabilities, one SAM_{wt} molecule is found in the asymmetric unit.

Table 31: Crystallisation conditions that resulted in SAM_{wt} crystal formation

| Screen | No. in composition list | Salt | Buffer | Precipitant |
|---|-------------------------|-------------------------------------|----------------------------|--|
| NH_4SO_4 suite (Qiagen) | 8 | 0.2 M $(\text{NH}_4)_2\text{HPO}_4$ | | 2.2 M $(\text{NH}_4)_2\text{SO}_4$ |
| | 29 | 0.2 M K_2HPO_4 | | 2.2 M $(\text{NH}_4)_2\text{SO}_4$ |
| | 42 | 0.2 M Na_2HPO_4 | | 2.2 M $(\text{NH}_4)_2\text{SO}_4$ |
| Wizard I/II (Rigaku) | 20 (Wizard I) | 0.2 M NaCl | 0.1 M Imidazole / HCl pH 8 | 0.4 M NaH_2PO_4 / 1.6 M K_2HPO_4 |

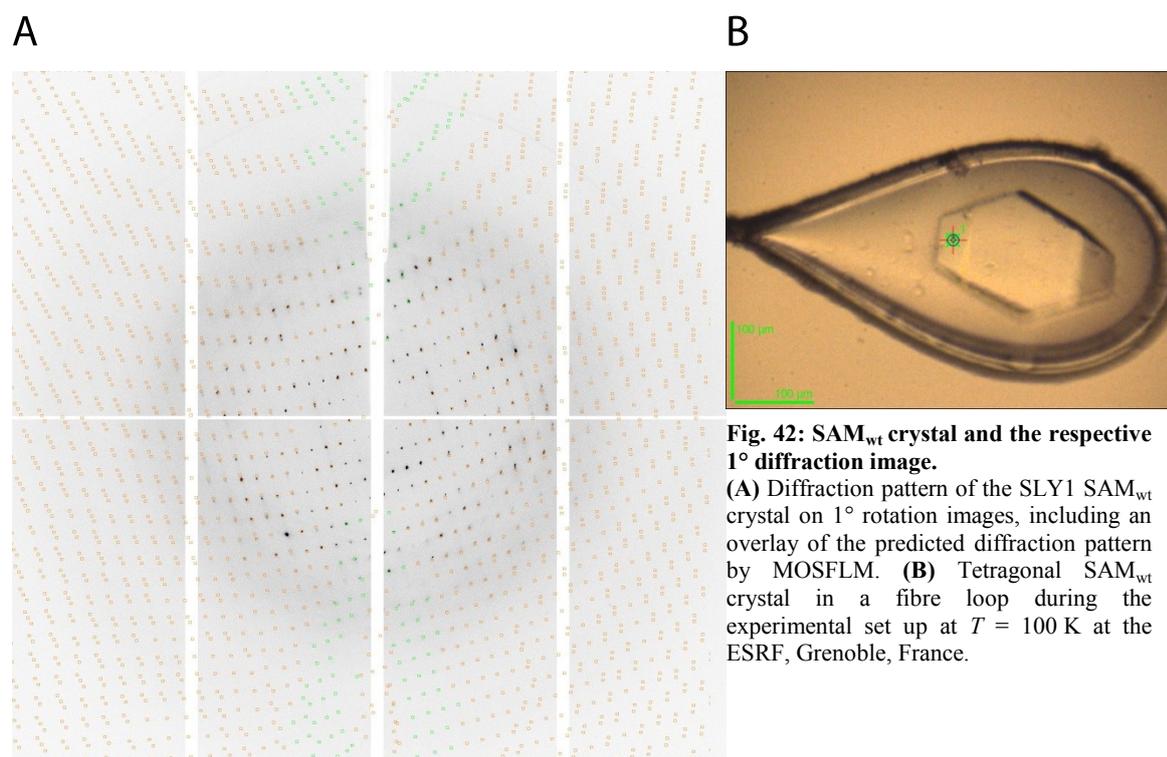


Fig. 42: SAM_{wt} crystal and the respective 1° diffraction image.
(A) Diffraction pattern of the SLY1 SAM_{wt} crystal on 1° rotation images, including an overlay of the predicted diffraction pattern by MOSFLM. **(B)** Tetragonal SAM_{wt} crystal in a fibre loop during the experimental set up at $T = 100 \text{ K}$ at the ESRF, Grenoble, France.

4.8.2 Phasing, model building and refinement

The initial phasing, model building and refinement were performed as described in Section 3.7.3 and 3.7.4. A starting model of SAM_{wt} was generated based on the structure of the SAM domain of the murine protein SAMSN1 (PDB ID: 1v38) by homology modelling using the SWISS-MODEL web server. The final structure comprised 65 of 69 residues (P254 to D317 plus the N-terminal glycine). For the remaining C-terminal residues T318–E321 of SAM_{wt} no electron density was visible in the weighted $2F_o - F_c$ map, probably due to the higher flexibility of the C-terminus. Therefore the position of the C-terminal amino acids could not be defined in the asymmetric unit. Detailed information on the refinement statistics is presented in Table 32.

4.8.3 SAM_{wt} structure

SAM_{wt} adopts the canonical five helix fold described previously for SAM domains (Chongwoo A Kim & Bowie, 2003), with the helices composed of residues L275–R263 ($\alpha 1$), residue E267–L274 (c2), L281–F284 ($\alpha 3$), E289–E294 ($\alpha 4$) and P300–Y316 ($\alpha 5$). Helix c2 is a composite helix that contains an N-terminal 3_{10} -turn in conjunction with a C-terminal α -helix. The helices are connected by short loop regions. The secondary structure elements were determined from atomic coordinates using DSSP.

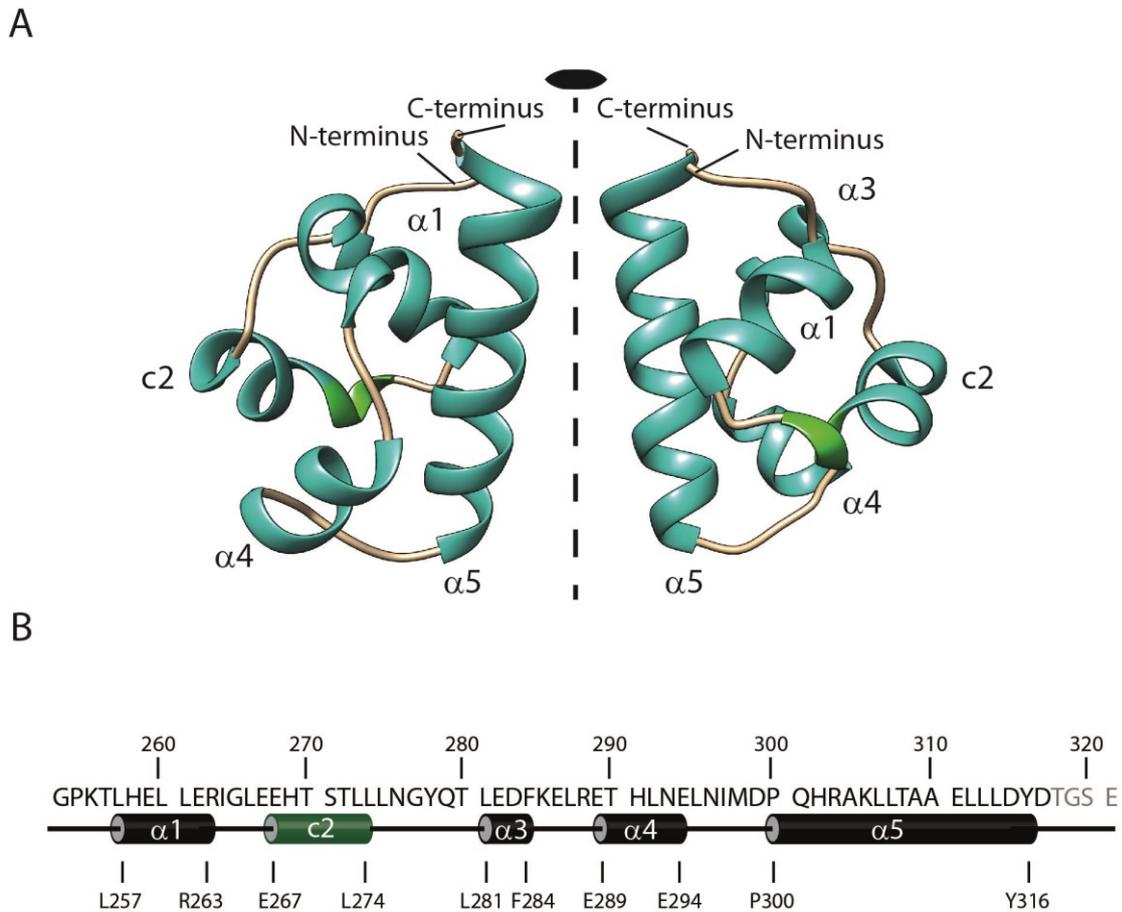


Fig. 43: Structure of SLY1 SAM_{wt} determined from X-ray crystallography data.

(A) Structure of the SLY1 SAM_{wt} molecule present in the asymmetric unit and its symmetric equivalent as determined by X-ray crystallography. The structure of SAM_{wt} exhibits the five-helix bundle conformation typical for SAM domains. The different secondary structure elements are coloured in teal (α -helices), tan (loops) and green (3_{10} -turn of helix c2) (B) Secondary structure elements as determined by DSSP. The SLY1 SAM domain consists of 5 helices numbered $\alpha 1$ – $\alpha 5$. Helices $\alpha 1$, $\alpha 3$, $\alpha 4$ and $\alpha 5$ are uniform α -helices, whereas helix c2 (marked in green) is a composite helix that contains an N-terminal 3_{10} -turn (E267, E268, H269) followed by an α -helix.

Table 32: Data collection and refinement statistics.

| Data acquisition | |
|--|-----------------------------------|
| Beamline/ Detector | ID30A/Eiger X 4M |
| Wavelength λ (Å) | 0.9677 |
| Resolution range (Å) | 47.34 - 2.05 (2.11 - 2.05) |
| Space group | P 41 21 2 |
| Unit cell a, b, c (Å) α, β, γ (°) | 44.09, 44.09, 94.67 90, 90, 90 |
| Total reflections | 80277 (6533) |
| Unique reflections | 6358 (481) |
| Multiplicity | 12.6 (13.6) |
| Completeness (%) | 99.9 (100.00) |
| Mean $I/\sigma(I)$ | 13.9 (2.1) |
| Wilson B-factor (Å ²) | 55.8 |
| R-merge | 0.092 (1.281) |
| R-meas | 0.096 (1.330) |
| R-pim | 0.027 (0.355) |
| CC (1/2) | 0.999 (0.381) |
| Refinement | |
| Resolution range in refinement (Å) | 39.97 - 2.05 (2.124 - 2.05) |
| Reflections used in refinement | 6318 (604) |
| Reflections used for R-free | 319 (29) |
| R-work | 0.2183 (0.3766) |
| R-free | 0.2421 (0.3764)0 |
| Number of non-hydrogen atoms | 547 |
| macromolecules | 537 |
| solvent | 10 |
| Protein residues | 65 |
| RMS(bonds) | 0.008 |
| RMS(angles) | 1.22 |
| Ramachandran favoured (%) | 96.83 |
| Ramachandran allowed (%) | 3.17 |
| Ramachandran outliers (%) | 0.00 |
| Rotamer outliers (%) | 1.67 |
| Clashscore | 7.63 |
| Average B-factor Å ² | 65.05 |
| macromolecules | 65.14 |
| solvent | 60.30 |
| Number of TLS groups | 4 |

4.8.4 Crystallographic SAM_{wt} dimer

Oligomerisation is a crystallisation inherent process, with molecules often having multiple contact points with other molecules present in the asymmetric unit or with symmetric equivalent molecules throughout the unit cell. In the case of SAM_{wt}, PISA identified one likely interface for dimerisation from the observed crystal contacts between symmetric equivalent molecules (molecule A and A') in the unit cell. The domains form a dimer interface via the N-terminal glycine of SAM_{wt} and residues in helices $\alpha 1$, $\alpha 5$ and the loop between $\alpha 1$ and c2. The majority of the interface is formed by the two helices $\alpha 5$ facing each other with an angle Ω between the two helix axes $\sim -50^\circ$.

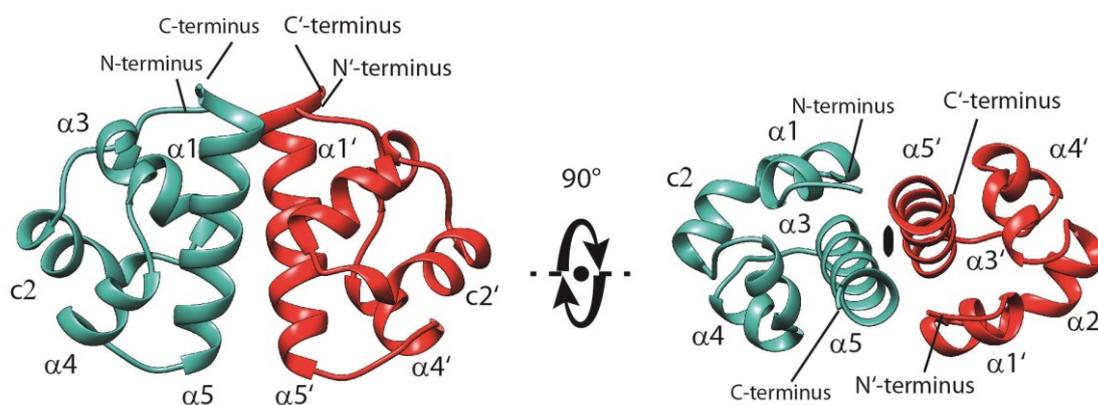


Fig. 44: Dimer of SLY1 SAM_{wt} predicted by PISA.

In the dimer interface determined by PISA the two molecules exhibit a 180° rotational symmetry, with the two $\alpha 5$ helices facing each other, and the N-terminal end of $\alpha 1$ and $\alpha 4$ and the C-terminal end of c2 near perpendicular to this interface.

The area between helices $\alpha 1$, c2, and $\alpha 5$ of molecule A forms a hydrophobic stretch, which is flanked by positively and negatively charged areas. Helix $\alpha 5$ of the molecule A' aligns with the hydrophobic stretch of molecule A upon dimer formation (Fig. 45). The hydrophobic amino acids A304, A309, L312, L313 as well as the polar T308 and Y316 are located at the interface and participate in hydrophobic interactions by reaching into the hydrophobic pocket (Fig. 46A).

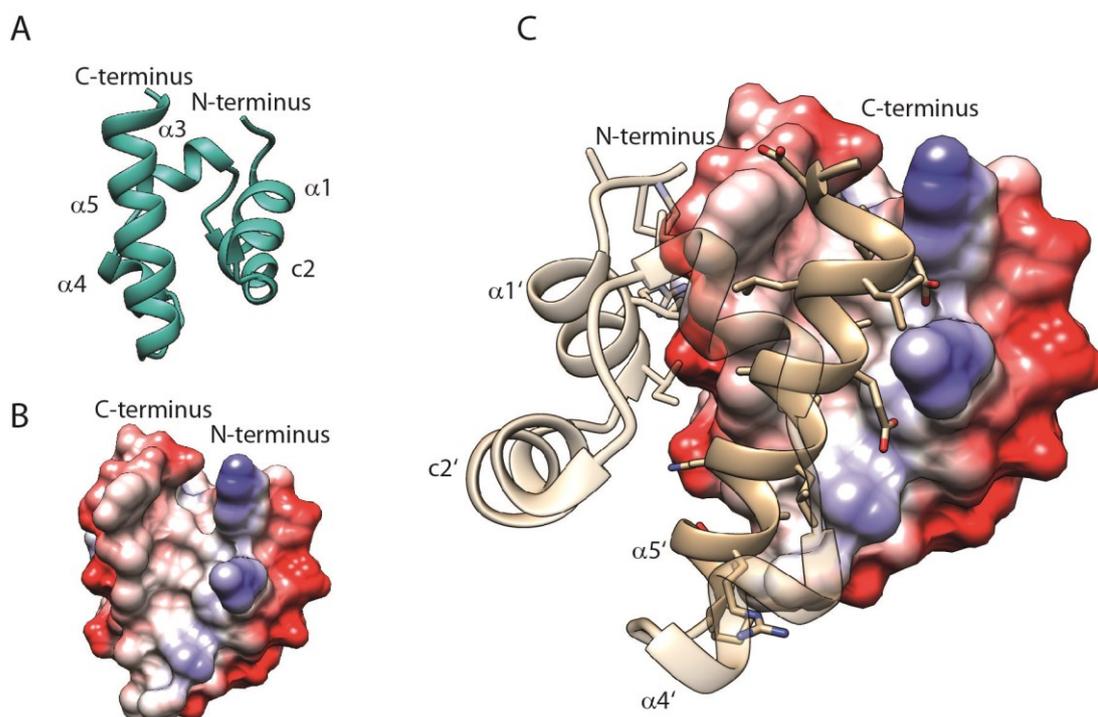


Fig. 45: Dimer interface of the SLY1 SAM_{wt} dimer in ribbon and surface representation.

(A) SLY1 SAM_{wt} monomeric subunit in ribbon representation. (B) SLY1 SAM_{wt} monomeric subunit in a surface representation. The surface colouring is based on electrostatic potential at pH 6.4 with negative charges in red and positive charges in blue. The visualisation of the surface charge reveals a hydrophobic stretch encompassed by helices $\alpha 5$, $\alpha 1$ and $c 2$. (C) Helix $\alpha 5$ of molecule A' (in ribbon representation) aligns with the hydrophobic stretch of molecule A.

The complex is further stabilised at the C-terminal end of $\alpha 5$ by the formation of salt bridges and hydrogen bonds between R262 and E310 and D315. At the N-terminus of $\alpha 5$, hydrogen bonds are formed across the two $\alpha 5$ helices by the sidechain amide group of N301 and the carbonyl carbon of the P300 of the opposite subunit (Fig. 46). A detailed list of salt bridges and hydrogen bonds across the dimer interface taken from the PISA analysis can be found in Table 33. Overall, the dimer interface has a BSA of $\sim 870 \text{ \AA}^2$, with a calculated energy gain of 4.7 kcal/mol upon dimer formation.

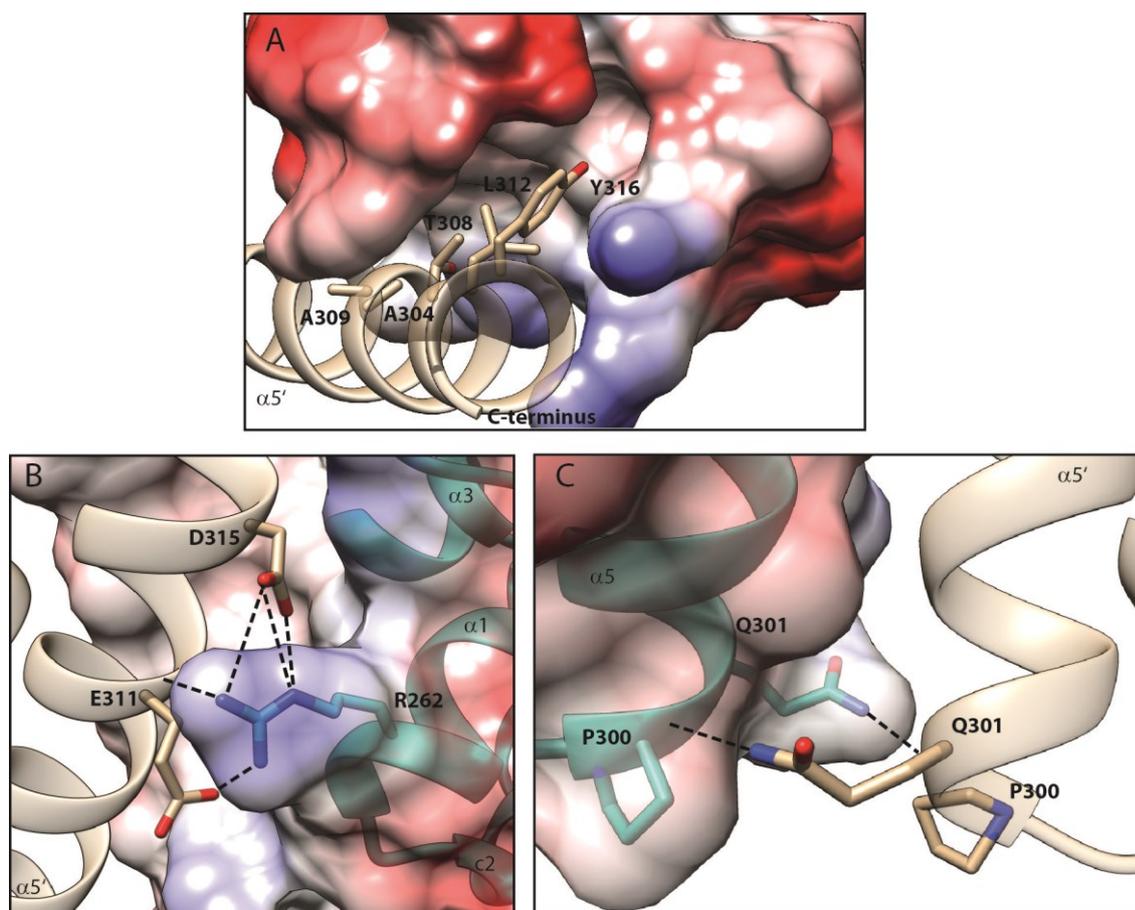


Fig. 46: Non-covalent interactions stabilizing the dimer interface of SAM_{wt}.

(A) Amino acids along helix $\alpha 5$ of molecule A' reaching into the hydrophobic pocket encompassed by helices $\alpha 1$, $\alpha 2$ and $\alpha 5$ of molecule A. (B) Network of hydrogen bonds and salt bridges between the guanidino group of R262 (molecule A'), the side chain and carbonyl oxygens of E311 and the side chain oxygens of D315 (subunit A). (C) Hydrogen bonds predicted by PISA between the side chain amide nitrogen of N301 and the carbonyl oxygen of P300 of the respective opposite chain.

Table 33: Hydrogen bonds and salt bridges across the SAM_{wt} dimer interface predicted by PISA based on x-ray data.

| Molecule A | | | Molecule A': | | | Distance (Å) |
|------------|---------|---|---|---------|------------|-------------------|
| SSE | Residue | Atom | Atom | Residue | SSE | |
| N-term | P254 | [O] | [OH] | Y316 | C-term | 3.28 |
| $\alpha 1$ | R262 | [N ^{ϵ}] | [O ^{$\delta 1$}] | D315 | $\alpha 5$ | 3.54 ^a |
| $\alpha 1$ | R262 | [N ^{ϵ}] | [O ^{$\delta 2$}] | D315 | $\alpha 5$ | 2.72 |
| $\alpha 1$ | R262 | [N ^{$\eta 2$}] | [O] | E311 | $\alpha 5$ | 3.55 ^a |
| $\alpha 1$ | R262 | [N ^{$\eta 2$}] | [O ^{$\delta 1$}] | D315 | $\alpha 5$ | 3.46 ^a |
| $\alpha 5$ | P300 | [O] | [N ^{$\epsilon 2$}] | E301 | $\alpha 5$ | 2.91 |
| $\alpha 5$ | N301 | [N ^{$\epsilon 2$}] | [O] | P300 | $\alpha 5$ | 2.91 |
| $\alpha 5$ | E311 | [O] | [N ^{$\eta 2$}] | R262 | $\alpha 1$ | 3.55 ^a |
| $\alpha 5$ | D315 | [O ^{$\delta 2$}] | [N ^{ϵ}] | R262 | $\alpha 1$ | 2.72 |
| $\alpha 5$ | D315 | [O ^{$\delta 1$}] | [N ^{ϵ}] | R262 | $\alpha 1$ | 3.54 ^a |
| $\alpha 5$ | D315 | [O ^{$\delta 1$}] | [N ^{$\eta 2$}] | R262 | $\alpha 1$ | 3.46 ^a |
| C-term | Y316 | [O] | [OH] | P254 | N-term | 3.28 |

^aAccording to Jeffrey, 1997, hydrogen bonds beyond 3.4 Å are considered to be weak, electrostatic interactions. Distances larger than 3.6 Å were excluded from the analysis.

5 Discussion

5.1 SLY1 SAM_{wt} forms symmetric homodimers

The concentration-dependent chemical exchange process observed in 2D ¹H, ¹⁵N-HSQC spectra recorded of SLY1 SAM_{wt} at different concentrations (Section 4.5.3) strongly indicated that SAM_{wt} is capable of reversible self-association. Since only one set of backbone amide signals was observable for each amide group present in SAM_{wt}, the self-association consequently must result in symmetric oligomers. Only in symmetric oligomers do equivalent nuclei of all subunits experience the same chemical environment. Comparable observations for the longer SAM domain construct of SLY1 have led to the hypothesis that SLY1 SAM forms symmetric homodimers (Thiagarajan, 2011). Therefore, the SAM_{wt} oligomerisation state was analysed using analytical AUC experiments.

AUC sedimentation equilibrium experiments (Section 4.5.1) found SAM_{wt} to exist in a reversible monomer-dimer equilibrium with a K_D of $\sim 117 \mu\text{M}$ and a calculated monomer mass of 8040 Da. The AUC-derived K_D agrees with the K_D of $153 \mu\text{M}$ determined by MST. One-component and the two-component models provided inferior fits with significantly larger residuals than the monomer-dimer equilibrium model. Additionally, these results of the AUC sedimentation equilibrium experiments show that the predominant species in the sample is monomers and dimers. A significant amount ($\geq 10\%$) of higher order species was not observed.

5.2 Cross-linking of the wild-type SAM domain does not affect the overall protein fold

Solving the structures of symmetric dimers by NMR spectroscopy remains a challenge. Distinguishing between restraints that define the interface and those that define the overall fold of the monomer unit can be difficult because of the degeneracy of the NMR spectra (Breeze, 2000; Neuhaus & Williamson, 2000). In addition to standard 3D NOESY experiments used for solving structures, intermolecular isotope-filtered/edited NOESY spectra can be utilised to define the dimer interfaces unambiguously. However, filtered spectra suffer from low sensitivity due to the reduced concentration of NMR observable nuclei (50%) in the sample and additional delay periods within the pulse

program. An additional problem for weaker protein complexes is chemical exchange, which can result in further reduction of signal intensity.

Superposition of 2D ^1H , ^{15}N -HSQC spectra recorded on samples with varying concentrations of the wt SLY1 SAM constructs SAM_{wt} and SAM_{lg} and following chemical shift mapping revealed that the SLY1 SAM domain dimerises with an exchange rate on the intermediate-to-fast chemical shift time scale. This exchange process prohibited acquisition of sufficient quality NMR data that could be used to define the dimer interface (Section 4.3.1; Section 4.5.3). As previously described, limited or incorrect distance restraint information can lead to incorrect orientation of the subunits in a protein oligomer (Breeze, 2000). The introduction of a cysteine at position 320 of the SAM domain by site-directed mutagenesis led to the formation of a disulfide bond between the two monomers, effectively locking the two subunits in the dimer state (Section 4.5.3). This cross-linking strategy ultimately allowed the structure determination of the SLY1 SAM domain by NMR. Similar strategies have also been used successfully in other NMR structural studies of homo- (Junius *et al.*, 1996) and heterodimers (Metcalf *et al.*, 2010) where disulfide bonds have been introduced to quench chemical exchange processes that have hampered NMR investigations.

The chemical shifts of the amide cross-peaks in the 2D ^1H , ^{15}N -HSQC spectra of SAM_{wt} and non-cross-linked SAM_{C} showed that the S320C mutation did not affect the SLY1 SAM domain structure, with above average chemical shift differences only occurring close to the mutation site (Section 4.5.6; Fig. 30). Comparison of the amide cross-peak chemical shifts of the cross-linked SAM_{C} with those of SAM_{wt} , however, showed above average chemical shift changes for cross-peaks belonging to residues in helix $\alpha 5$ (K305–D315) and at other C-terminal residues of SAM_{C} . Since helix $\alpha 5$ and the C-terminus are the main contributors to the dimer interface, this observation lead to the question of whether or not the cross-linking affects the dimer interface of the SAM_{C} dimer. The largest chemical shift differences clearly appear adjacent to the mutation site and are the result of the S320C exchange as well as the formation of the disulfide bond. Nonetheless, the mutation only affects the residues in spatial and sequential proximity (T318-E321), as can be seen for non-cross-linked SAM_{C} . The effect of the mutation would not translate throughout the whole of helix $\alpha 5$. Therefore, other aspects have to contribute to the observed changes in chemical shifts in SAM_{C} .

Aside from the mutation, the most likely explanation for these chemical shift changes around the dimer interface upon cross-linking is the difference in monomer

concentration between the cross-linked SAM_C and SAM_{wt} sample. The sample containing fully cross-linked SAM_C does not comprise any significant monomer population. However, based on the K_D of 117 μM for the SAM_{wt} monomer-dimer equilibrium as determined by AUC, the respective monomer fraction at 1.4 and 2.7 mM total concentration of SAM_{wt} would still be $\sim 258 \mu\text{M}$ (18.5%) and $\sim 447 \mu\text{M}$ (16.6%) respectively. Since the chemical shift of a resonance undergoing intermediate-to-fast chemical exchange represents the population-weighted average chemical shift, these monomer fractions would still contribute significantly to the observed chemical shifts. If these differences in the monomer populations were the sole reason for the existing chemical shift perturbations, the chemical shifts of the cross-peaks arising from the cross-linked SAM_C should follow the linear trend established by the corresponding amide cross-peaks of SAM_{wt} at different concentrations (0.003-1.7 mM) (Fig. 47). Consequently, non-linearity of the signal shifts would be indicative of the cross-linking having an additional effect on the structure, e.g. in the form of local influences on helix $\alpha 5$ covalent geometry. Examples of both linear and non-linear changes of the chemical shift perturbations are given in Fig. 47.

Many residues involved in the interface formation exhibit too strong line broadening in the 2D ^1H , ^{15}N -HSQC experiments recorded of SAM_{wt} to be amenable to chemical shift perturbation analysis. However, the chemical shift perturbation of the interface residues R303, L306, E311 and L312 can be analysed to a certain extent. The chemical shifts of the amide cross-peaks of R303, L306, E311 and L312 follow a linear trend in the SAM_{wt} concentration series. For the R303 and L312, this linearity is continued by the cross-peaks observed for cross-linked SAM_C. This observation supports the hypothesis that the chemical shift changes of cross-peaks between spectra of cross-linked SAM_C and SAM_{wt} are likely due to the remaining change in dimer population. In contrast, the amide cross-peaks of residues L306 and E311 of the cross-linked SAM_C slightly deviate from the linear trend observed in the SAM_{wt} concentration series, since the resonances are shifted slightly downfield (L306) and upfield (G311) in the proton dimension from the expected final position of the respective amide resonances.

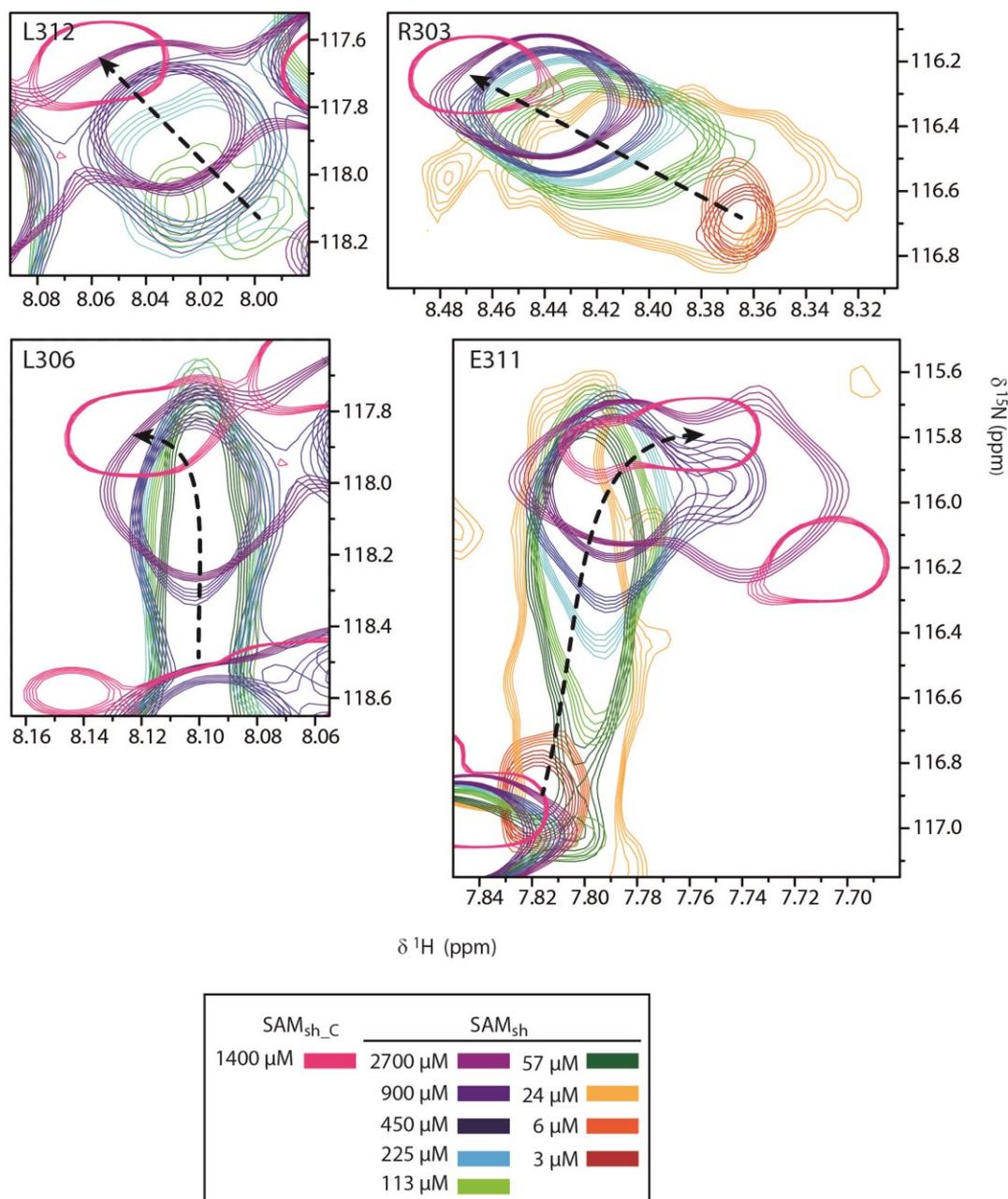


Fig. 47: Chemical shift variations of amide resonances of selected residues in a SAM_{wt} concentration series. 2D $^1\text{H},^{15}\text{N}$ -HSQC spectra were recorded of SAM_C at a concentration of 1400 μM and SAM_{wt} at varying concentrations using a spectrometer operating at a ^1H resonance frequency of 700 MHz. When superposing the spectra, the concentration-dependent chemical shift changes of the observed amide cross-peaks can be tracked. The amide cross-peaks of R303, L306, E311 and L312 were observable at various concentrations, although severe line broadening prevented the resonances of L306 and L312 to be tracked beyond 57 and 113 μM , respectively. The concentration-dependent chemical shift perturbation of R303 and L312 follow a linear trend which is further extended by the resonance in the cross-linked SAM_C spectrum. The amide resonances of residues L306 and E311 on the other hand in are shifted slightly more upfield or downfield, respectively, in the cross-linked state than an extension of the linearity of the resonances for SAM_{wt} would allow.

What type of process influences the chemical shift of these resonances can only be hypothesised. The overall increase of the perturbation towards the C-terminus could mean that the strain put on the protein backbone by the introduction of the disulfide bond is somehow translating along helix $\alpha 5$. It is known that the chemical shift of $^1\text{H}^{\text{N}}$ and ^{15}N nuclei of the protein backbone are not very susceptible to changes in the backbone torsion angles per se, but a change in hydrogen bond pattern and length can have a significant effect on the amide resonance chemical shifts. A change in hydrogen bond length of 0.1 Å can result in a perturbation of 1 ppm in ^{15}N chemical shift (Szilágyi, 1995; Wishart, 2011). It is conceivable that the cross-linking causes a minimal readjustment of the protein backbone in helix $\alpha 5$, resulting in an increase or decrease in hydrogen bond length. Furthermore, slight bending of solvent-exposed amphiphilic helices towards the hydrophobic core of proteins has been described as a result of shorter H-bonds on the “inner” side of the helix as opposed to the part facing the solvent (Cordier & Grzesiek, 2002; Szilágyi, 1995). Mechanical strain on the helices $\alpha 5$ and $\alpha 5'$ due to the cross-linking could straighten the helices slightly and therefore change the length of the stabilising hydrogen bond.

Another possibility could be the reduction of flexibility or fraying at the C-terminal end of the helix by cross-linking. Helix fraying has been described for α -helical peptides in solution that lack the stabilising effect of being part of an extended protein (Jaravine *et al.*, 2001). It could be argued that since helix $\alpha 5$ is very close to the C-terminal end of the SAM_{wt} construct, it exhibits some fraying or movement in the non-cross-linked state. This would agree with the increase in B-factor values from E311 onwards in the crystal structure of SAM_{wt} (Fig. 48).

Another contribution to the chemical shift perturbation upon cross-linking could be the ring current effects of Y316. Comparison between the average NMR structure and the crystal structure show differences in the backbone and sidechain torsion angles of Y316 which result in a $\sim 25^\circ$ rotation of the ring towards the N-terminus and a slight tilt of the ring towards the centre of helix $\alpha 5$ of the second SAM subunit. The chemical shift derived random coil index (RCI) (Berjanskii & Wishart, 2008) of the non-cross-linked SAM_C and the B-factors of the SAM_{wt} crystal structure show a significant increase starting from residue Y316 and L314, respectively, indicating an increase in the flexibility of the protein backbone for the C-terminal residues (Fig. 48). This flexibility is reduced upon cross-linking, as can be seen in the RCI of the cross-linked SAM_C. Thus, cross-linking may very well inhibit Y316 and other C-terminal residues to adopt

multiple conformations, which would de facto lead to changed ring current effects experienced by the surrounding nuclei.

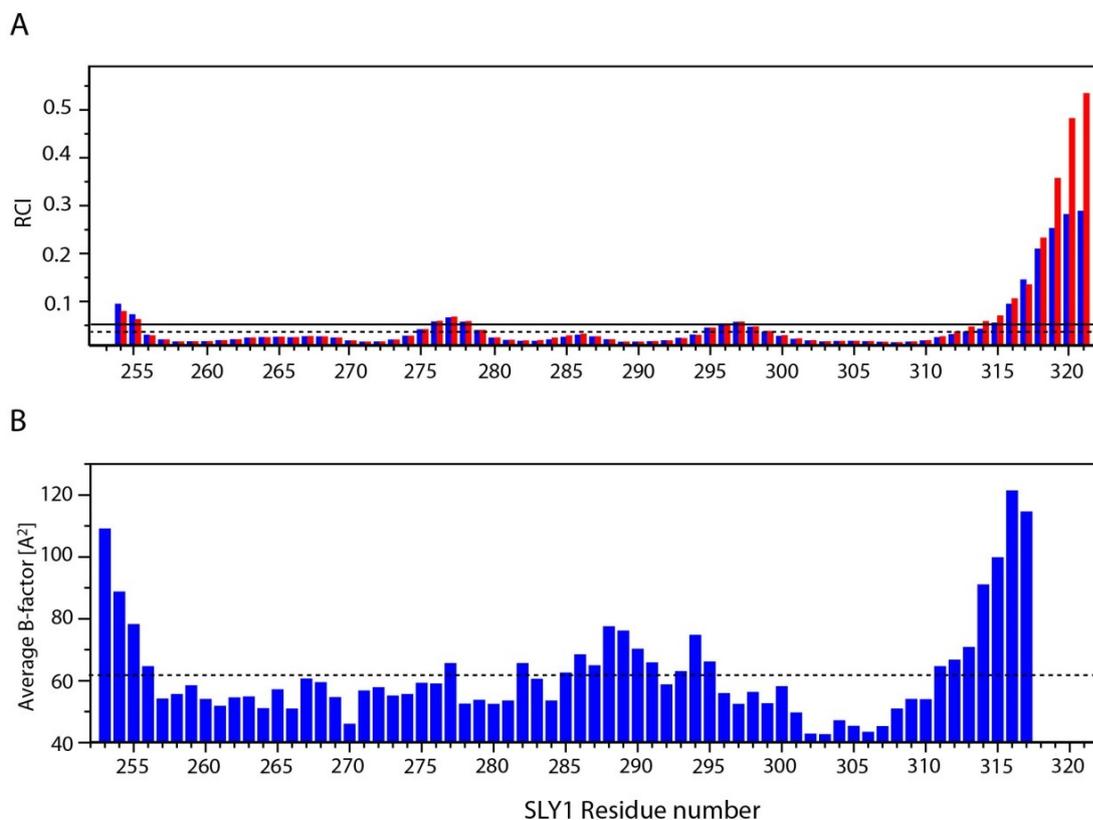


Fig. 48: Flexibility of SAM_C and SAM_{wt} per residue as indicated by the RCI and B-factors.

(A) RCI values for SAM_C in both, the cross-linked (blue) and non-cross-linked (red) state. Overall, the RCI values are almost identical throughout the whole protein. At the C-terminus however, the introduction of the disulfide bond resulted in a reduction in flexibility of the of the three C-terminal residue G319, C320, E321. The average RCI values for the cross-linked and the non-crosslinked state are 0.05 and 0.04, respectively, and are indicated by the solid and dotted black lines. (B) Average B-factors of the backbone atoms for the SLY1 SAM_{wt} crystal structure. Overall the positions of the backbone atoms from L256 to A310 are comparatively well defined with B-factors around the mean value (red line) of 61.5 Å², while the termini have much higher B-factors, which can be an indication of flexibility. Interestingly, the B-factors already start to noticeably increase from residue E311 onwards, although helix $\alpha 5$ continues until Y316 and α -helices are considered to be very stable secondary structure elements.

The strongest support that the interface is unaffected by the cross-linking was provided by comparison of the limited available intermolecular NOE information acquired for SAM_{Ig} (Thiagarajan, 2011) with the intermolecular ¹⁵N, ¹³C- ω_1 -filtered, NOESY-¹H, ¹³C-HSQC recorded for the cross-linked SAM_C.

Initial attempts to characterise the dimer interface by isotope-filtered/edited NOESY experiments conducted by Thiagarajan, 2011, yielded only a very limited number of NOE cross-peaks of weak intensity (due to chemical exchange). Only 28 intermolecular NOEs could be assigned for the interface of the SAM_{Ig} dimer. In the case of the cross-linked SAM_C dimer, 3D versions of the isotope-filtered/edited NOESY experiments

were recorded, which not only yielded a significantly higher number of intermolecular NOE cross-peaks but also offered increased signal dispersion through the introduction of the ^{13}C and ^{15}N dimensions to the experiments.

The few NOE cross-peaks found in the NOESY spectrum of SAM_{lg} could also be identified in the spectra recorded of SAM_{C} . Exemplary strips of the intermolecular NOESY spectra acquired for the wt and the cross-linked SAM domain are shown in Fig. 49. The strips show the cross-peaks of the intermolecular NOEs observed for T308 $\text{H}^{\gamma 2*}$, A304 $\text{H}^{\beta*}$ and A309 $\text{H}^{\beta*}$ nuclei. The comparison of the two spectra shows that signals observed in the spectrum recorded of the wt are also present in the cross-linked sample. This observation supports the assumption that the dimer interfaces of SAM_{lg} and cross-linked SAM_{C} are very similar. Hence, while the cross-linking of the two monomer subunits of SAM_{C} caused some reorientation and put a certain strain on the otherwise flexible C-terminal residues, which are not part of the canonical five-helix bundle of SAM domains, it did most probably not introduce significant changes into the dimer interface. Therefore, the cross-linking enabled the determination of the SLY1 SAM dimer interface in solution.

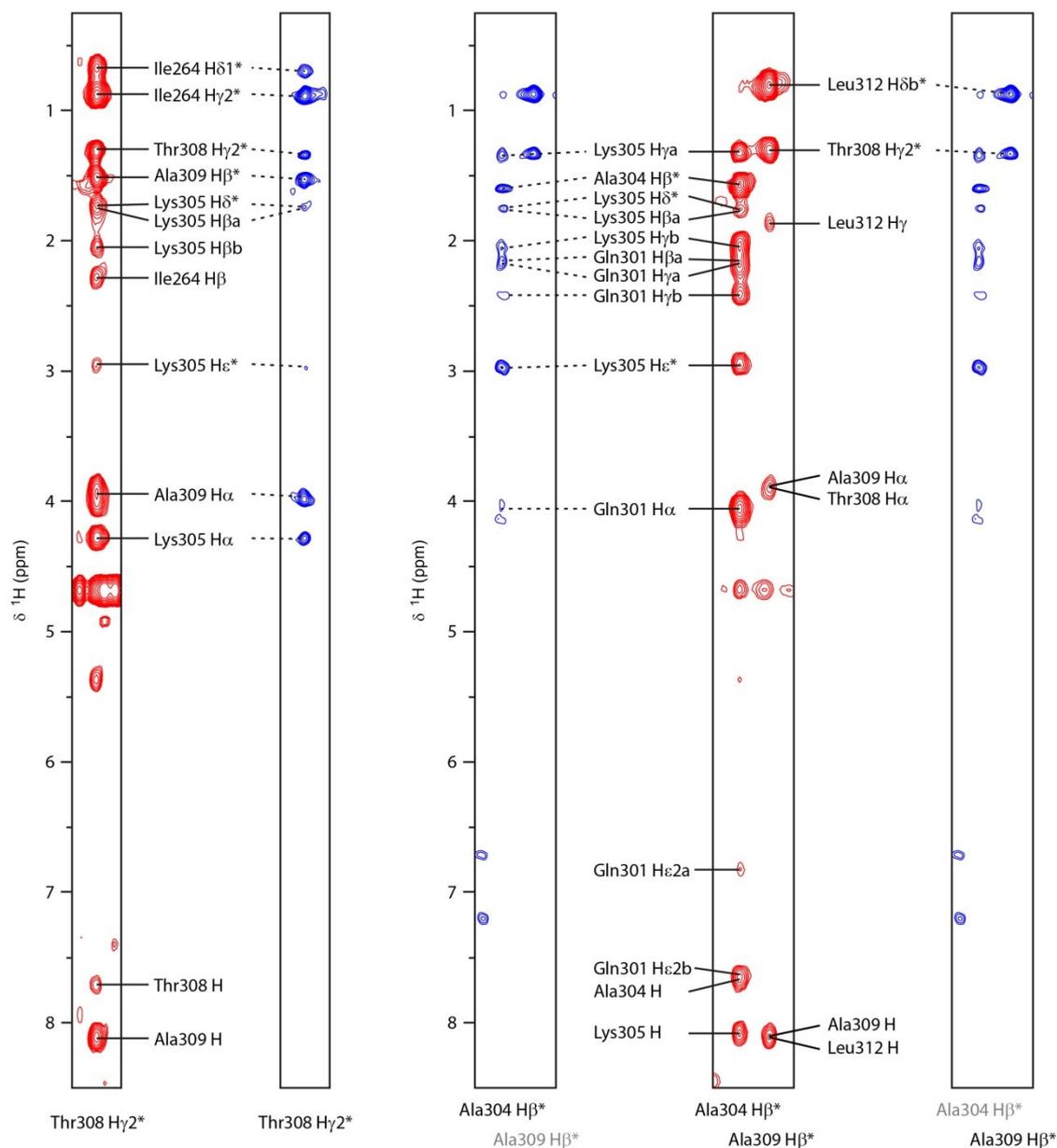


Fig. 49: 2D Strips of ^{15}N , ^{13}C -filtered, ^{13}C -edited NOESY spectra of SAM_{lg} and SAM_{c} .

Strips of the NOE dimension of the 3D and 2D ^{15}N , ^{13}C -filtered, ^{13}C edited NOESY spectra of SAM_{c} (red) and SAM_{lg} (blue), respectively. The strips show NOE cross-peaks for T308 $\text{H}\gamma^{2*}$, A304 $\text{H}\beta^*$ and A309 $\text{H}\beta^*$, which are located at the dimer interface of the SAM_{c} cross-linked dimer. All of the NOE cross-peaks observed in the strips for SAM_{lg} were also identified in the corresponding strips for the cross-linked SAM_{c} , thus corroborating that the dimer interface of the SLY1 SAM_{wt} domain dimer is unaltered by the cross-linking. Comparison of the signals between the two spectra shows an overall higher intensity of the NOE cross-peaks recorded for the cross-linked SAM_{c} because the monomer-dimer exchange process is suppressed. In addition, the absence of resonances in the spectrum recorded of SAM_{lg} result from the monomer-dimer equilibrium.

5.3 Quality of the NMR-derived SAM_c dimer structure ensemble

The validation of NMR-derived structural models is a critical aspect of the structure determination process by NMR spectroscopy. Experimental data, as in any other method, is never devoid of errors. In NMR spectroscopy, errors may be the result of software and hardware limitations, such as erroneous signal intensities due to spectral overlap, spin diffusion, chemical exchange processes or spectral artefacts. Furthermore, NMR data is very vulnerable to the experimentator's bias, especially in the treatment of violating restraints and unjustified cross-peak removal. NMR data is mostly comprised of relative restraints based on the 'perspective' of particular nuclei, which may or may not contradict each other in certain regions, especially those which may, in reality, adopt multiple structural conformations. NMR data does not provide absolute coordinates of atomic positions. Whether or not contradictory data represents the reality or is the result of artefacts and assignment error needs to be assessed very carefully. In addition, the information content provided by NMR data is comparatively low in comparison with X-ray crystallography. Therefore, it is highly reliant on the choice of forcefield and the therein provided restraints on covalent geometry and non-bonded interactions.

The first and foremost criterion on whether a calculated structural model is accurate is the fulfilment of the experimental restraints. Structures which are in good agreement with the data input and contain only a low number of violated restraints are desirable. Often distance restraint violation cut-offs of 0.3–0.5 Å are defined as 'serious' violations. Violating restraints can be corrected by correcting assignment errors, redefining the allowed error tolerance and, in the case of spectral artefacts, by complete removal of the restraint. However, especially unjustified removal of violating restraints can introduce strong bias into the data set and the resulting structure. Therefore, strategies and tools have been developed to assess the quality of the data set and the agreement of input data with the resulting structure ensemble to prevent overfitting of the data. In this study, the data were analysed using the RPF score algorithm (Section 3.6.10). The RPF scores provide statistical measures of how well the NOE peak-list is explained by the resulting structure ensemble and whether the number of observed NOE derived distance restraints utilised in the structure calculation is sufficient to explain the resulting structure (Huang *et al.*, 2005). Structures are considered to be correct solutions of the recorded NOE data when Recall and Precision scores and the F-measure adopt values above 0.8, and the DP score is above 0.7 for the calculated structural ensemble

(Huang *et al.*, 2005). For the SAM_C dimer structures, Recall and Precision scores and F-measure values are 0.936, 0.946 and 0.941 respectively. The DP score is 0.876. Therefore, the solved structure of the SAM_C dimer fulfils the criteria for a correct structure in all four categories and thus can be considered a reliable solution for the recorded structural restraints. The RPF scores furthermore indicate that the data set has not been subjected to excessive bias during the refinement process.

Another criterion for the overall quality of the structural model is the covalent geometry, especially the backbone conformation of the protein. Values for the backbone torsion angles φ and ψ are categorised into favoured, additionally allowed, generously allowed and disallowed regions as defined in the Ramachandran plot. The Ramachandran plot is based on the frequency of specific torsion angles conformations observed in high-resolution protein structures (Anderson *et al.*, 2005; Morris *et al.*, 1992; Ramachandran *et al.*, 1963). Ideally, the majority of the backbone torsion angles adopt values corresponding to the most favoured regions. The analysis of the backbone torsion angle conformation of the SAM_C dimer structure was performed using PROCHECK_NMR. More than 85% of the backbone torsion angles of the SAM_C cross-linked dimer are located in the most favoured regions of the Ramachandran plot, 13% in the additionally allowed regions and 0.8% in both the generously allowed and disallowed regions. These values correspond nicely to the average values of the four categories found in the PDB survey conducted by Morris *et al.*, 1992, on which the categorisation of the torsion angle space by PROCHECK_NMR is based. The low number of residues in the disallowed regions is an indicator of a high-quality structure. The only residue located in the generously allowed and disallowed regions of the plot is C320. Although torsion angles in the generously allowed and disallowed regions are considered to be undesirable and the result of incorrectly defined structural data, newer studies do not categorise the regions of the backbone torsion angle space in such a harsh way as done in Morris *et al.*, 1992. Anderson *et al.*, 2005 stated, that rare high energy states of backbone torsion angles could be genuine if they can be explained by other additional energy contributions. In the case of the disulfide-bonded SAM_C dimer, it is likely that the unusual orientation of the backbone torsion angles of C320 is strongly influenced by the presence of the introduced disulfide bond, and therefore could be genuine and not an indicator of poorly defined or incorrectly defined restraints. This hypothesis is corroborated by the finding that the intermonomer disulfide bond formed by C320 adopts the -LHS and -RHS staple conformations. These conformations on

average contain a higher dihedral strain energy than the average of all disulfides observed in deposited high-resolution structures (Schmidt *et al.*, 2006). The average strain energy of the disulfide bond in all 15 models of the SAM_C structure ensemble is $\sim 17 \text{ kJ mol}^{-1}$, whereas the average disulfide energy of all disulfides in the PDB is $\sim 7.5 \text{ kJ mol}^{-1}$ (Schmidt *et al.*, 2006). This additional strain on the disulfide bond likely translates to the backbone torsion angles of C320 and therefore forces them into an energetically unfavourable conformation.

The fact that the disulfide bond is present in two conformations in the calculated structure ensemble is not uncommon. The central region of the disulfide-bond is NMR silent and thus the χ_3 -angle of a disulfide cannot be determined readily by NMR-based experiments. Therefore, the definition of the disulfide conformation is reliant on the network of distance restraints of the surrounding residues, which in the case of an exposed disulfide bond in a relatively flexible region of the protein are comparatively few. In the case of the two different conformations observed in the SAM_C structural ensemble, the presence of both -LHS and -RHS staple conformations in NMR structural ensembles have been described before (Schmidt & Hogg, 2007).

The validation of the dimer interface was performed by searching for stabilising interactions, like hydrogen bonds and salt bridges. Excluding the artificial disulfide bond between the monomers, the SAM dimer is stabilised by the hydrophobic interactions between helix $\alpha 5$ of monomer A and a hydrophobic groove formed by helix $\alpha 5'$, helix $\alpha 1'$ and helix $c 2'$ of monomer A', and vice versa. The N- and C-terminal ends of helix $\alpha 5$ are involved in hydrogen bond and salt bridges formation across the dimer interface. The calculated buried surface area (BSA) is considered to be one of the critical parameters in identifying biologically relevant complex structures and is often correlated with the affinity of the interaction. For the SAM_C crosslinked dimer, the BSA, was determined to be 1200 \AA^2 for the average NMR derived structure and ranging between $1168\text{--}1230 \text{ \AA}^2$ among the calculated structural ensemble. The BSA for SAM_C fits the correlation found by Chen *et al.*, 2013, where a BSA of 500 and 1000 \AA^2 corresponds to an interaction of moderate affinity with a K_D in the μM range. Furthermore, the SAM_C domain dimer based on its affinity and surface area properties is comparable to other transient homodimers analysed by Nooren & Thornton, 2003.

5.4 Crystal and solution structure of the SAM domain are very similar

The asymmetric unit of the SAM_{wt} crystals contained one molecule. The structure of the molecule (P254-D317, including the N-terminal glycine) was solved at a resolution of 2.05 Å and corresponded to that of a SAM domain exhibiting the canonical five-helix fold described for SAM domains. The 3D structures of the SLY1 SAM monomer are highly similar in the crystal structure and the NMR structure. The superposition of subunit A of the closest to average NMR structure with the crystal structure at the C^α position using USCF chimera (Pettersen *et al.*, 2004) (Table 11) yielded an average RMSD of 0.84 Å. The superposition of the crystal structure with the average structure of the NMR structure ensemble is presented in Fig. 50A. The most pronounced difference between the crystal and NMR derived structures can be found at the termini. In both cases, the termini are the least defined regions. Furthermore, the crystal structure lacks the artificial disulfide bond at position 320, which influences the orientation of the C-termini of the NMR structure. Other small differences like the minimal reorientation of helix α 4 can be ascribed to crystal packing effects.

According to the analysis of the PISA web server, the molecule in the asymmetric unit is forming a crystallographic dimer with one of its symmetrically equivalent molecules which is likely to also form in solution. The closest to average NMR dimer structure and the crystallographic dimer was superposed using the program LSQMAN (Kleywegt, 1996) (Table 11), yielding a C^α RMSD of 1.13 Å. The superposition of the two dimer molecules (crystal and NMR derived) in Fig. 50B underline the similarity between the two structures. When aligning the helices α 5 of the NMR and crystal derived dimer, the backbone of both structures still follow a highly similar trend, although a slight translation of helix α 1- α 4 towards the interface can be observed in the NMR structure, making it slightly more compact than the crystal structure.

Garbuzynskiy *et al.*, 2005 and Sikic *et al.*, 2010 report an average RMSD between crystal structures and NMR derived structures of the same or homologous proteins of 1.7 and 1.5-2.5 Å, depending on which type of software was used for the superposition. The calculated average RMSD of 0.84 Å and 1.13 Å of the aligned monomer and dimer structures, respectively, are well below those values, confirming a very high level of conformity between both structures.

The comparison of the secondary structure elements by DSSP shows that the lengths of the four helices α 1, α 2, α 4 and α 5 differ slightly between the two structures: helices α 1, α 2,

and $\alpha 4$ are one residue shorter in the X-ray structure (I264, L275 and L295, respectively) while $\alpha 5$ is one residue longer (up to Y316). DSSP determines secondary structure elements from the hydrogen bond pattern of the structure, and the hydrogen bonds themselves are determined from the interatomic distances between potential donor and acceptor atoms. Slight differences in distance and orientation of the backbone, as often observed between structures derived from different methods (Garbuzynskiy *et al.*, 2005; Sikic *et al.*, 2010), can cause weak hydrogen bonds to be present in one structure but not the other, thereby influencing the secondary structure determination by DSSP. In fact, differences in hydrogen bond patterns in structures determined by crystallography and NMR seem to be the rule rather than the exception (Garbuzynskiy *et al.*, 2005). The reason for these differences could either be due to actual structural differences between proteins in a crystallised form versus in solution, or because of method inherent differences of the structural data and its implementation in the process of structure determination. A further observation is that helix c2 is a composite helix in both structures, comprising an N-terminal 3_{10} -helix turn and a C-terminal α -helix. This structural feature in both structures offers an argument that the composite helix is real and not an artefact of the structure calculations or crystal packing. The presence of the 3_{10} -helix is also corroborated by the comparatively low probability of helix formation for residues E267–L275 as determined by TALOS+ and the absence of α -helix typical hydrogen bonds (Section 4.6.1). Composite helices are not uncommon, with the 3_{10} -part of the helix contributing to a slight tilt at the junction of the two helices, which is hypothesised to reduce the surface area of proteins and thus make proteins more spherical (Pal *et al.*, 2005).

The dimer interface of the crystallographic dimer is the same as in the NMR derived homodimer structure. The interface consists predominantly of the helices $\alpha 5$ and $\alpha 5'$ of the SAM_{wt} subunits packing against each other with an interaxial angle of $\sim 50^\circ$. In addition to that, the residues L260, R263 and I264 in helix $\alpha 1$ and the loop region between helix $\alpha 1$ and c2 of subunit A further stabilise the dimer by contributing to hydrophobic interactions and salt bridge and hydrogen bond formation with residues in helix $\alpha 5'$ of subunit A'.

In summary, both X-ray crystallography and NMR strategies were successfully employed to determine the structure of the SLY1 SAM domain dimer.

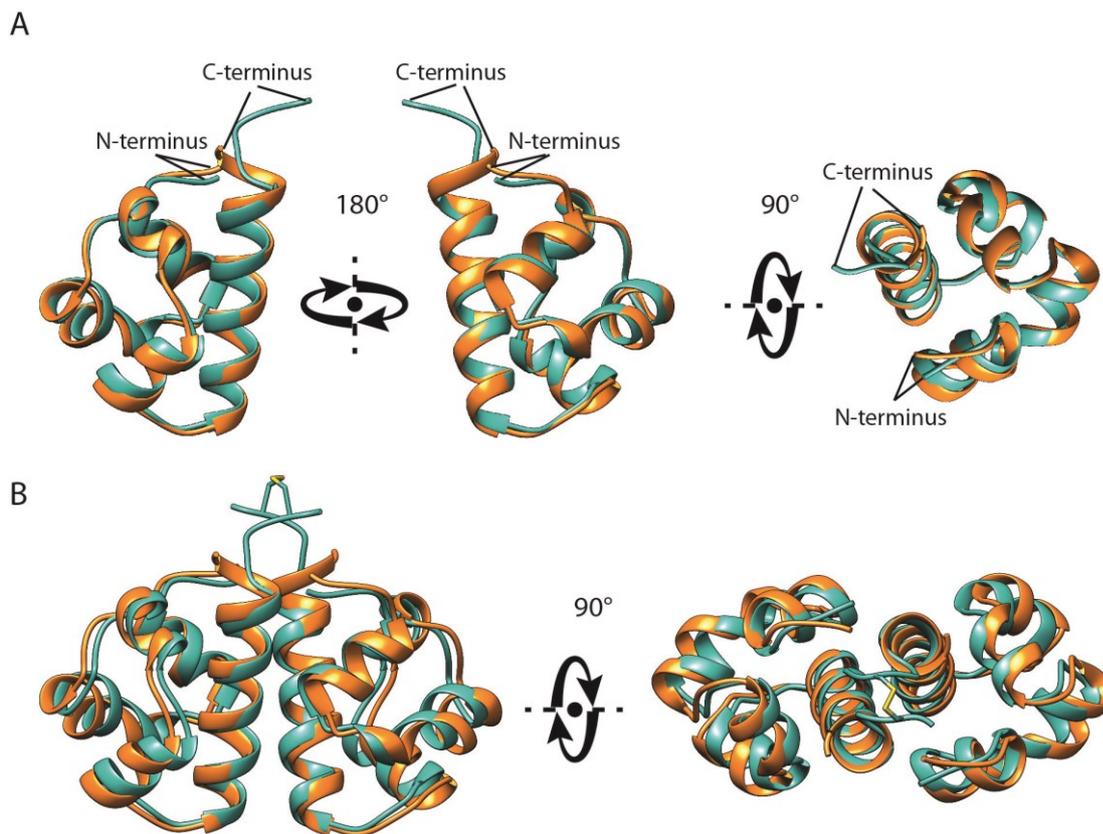


Fig. 50: Superposition of the X-ray crystallography and NMR derived SLY1 SAM structures.

The NMR structure closest to the average backbone structure of the NMR structure ensemble (teal) and the crystallographic structure (orange) is depicted. **(A)** Superposition of the monomer subunits with a C^α RMSD 0.84 Å. **(B)** Superposition of the crystallographic dimer and NMR-derived dimer shows a high similarity of the structures with a C^α RMSD of 1.13 Å. A slight translation of the backbone throughout helices $\alpha 1$ - $\alpha 4$ towards the interface can be observed in the NMR structure, which is most probably the result of diverging buffer conditions and crystal packing effects.

5.5 SAM_{wt} domain forms a novel type of symmetric dimer for SAM domains

The structure of the SLY1 SAM monomeric subunit determined by NMR spectroscopy and X-ray crystallography corresponds to the canonical five-helix bundle fold that has been described for the majority of SAM domain structures (Kim & Bowie, 2003; Qiao & Bowie, 2005). Although structural similarity between SAM domains is very high, the primary sequences of SAM domains have been described to vary significantly. The sequence of the SLY1 SAM domain was compared to amino acid sequences of other SAM domains, for which structures had been deposited in the PDB (December 2017), using the online tool NCBI BLASTP (Altschul *et al.*, 1997) (Table 11). The SLY1 SAM domain is closest in sequence to the SAM domain of the related murine signalling

protein SAMSN1 (UniProtKB-P57725) (67% sequence identity) and has a reasonable level of sequence similarity (46% sequence identity) with the SAM domain of human signalling protein ANKS1A (also referred to as Odin) (UniProtKB-Q92625). The remaining SAM domains found in the PDB share sequence identity with SLY1 SAM of below 40%, with most even below 30%.

Comparison of structural similarity using the Secondary Structure Matching algorithm (SSM) (Krissinel & Henrick, 2004) (Table 11) revealed that the subunits of the NMR dimer and the crystal structure share the highest structural identity with the SAM domains of the ephrin receptors. The SAM domain structures published for EphA1, EphA2, EphA4, EphA7, EphB2 and EphB4 make up 15 out of the 20 highest matches to the SLY1 SAM monomer structure. Interestingly, ephrin receptors dimerise (Singh, Ahmed *et al.*, 2015; Singh, Cao *et al.*, 2015) and the SAM domains of EphA4 is among the few SAM domains that form symmetric homodimers.

SAM domains for which structural data are available have been shown to exist in monomeric (Li *et al.*, 2007; Sathyamurthy *et al.*, 2011; Zhong *et al.*, 2009) and multiple oligomeric states, e.g., homo- and heterodimers (Kim *et al.*, 2005; Kwan *et al.*, 2006) and homo- and hetero-oligomers (Wei *et al.*, 2011), and can also form polymers and fibril-like structures (Kim *et al.*, 2005; Leettola *et al.*, 2014; Mariotti *et al.*, 2016; Robinson *et al.*, 2012; Sayou *et al.*, 2016). The most common mode of SAM domain homo- and heterotypic association is a head-to-tail interaction mediated by a ML-EH interface (Section 1.3). Of the 70 different SAM domains for which structural data is available in the PDB, 32 oligomerise through the ML-EH interface, with five of these structures exclusively forming extended, helical polymers, 12 forming heterodimers, and eight forming heterodimers as well as polymers.

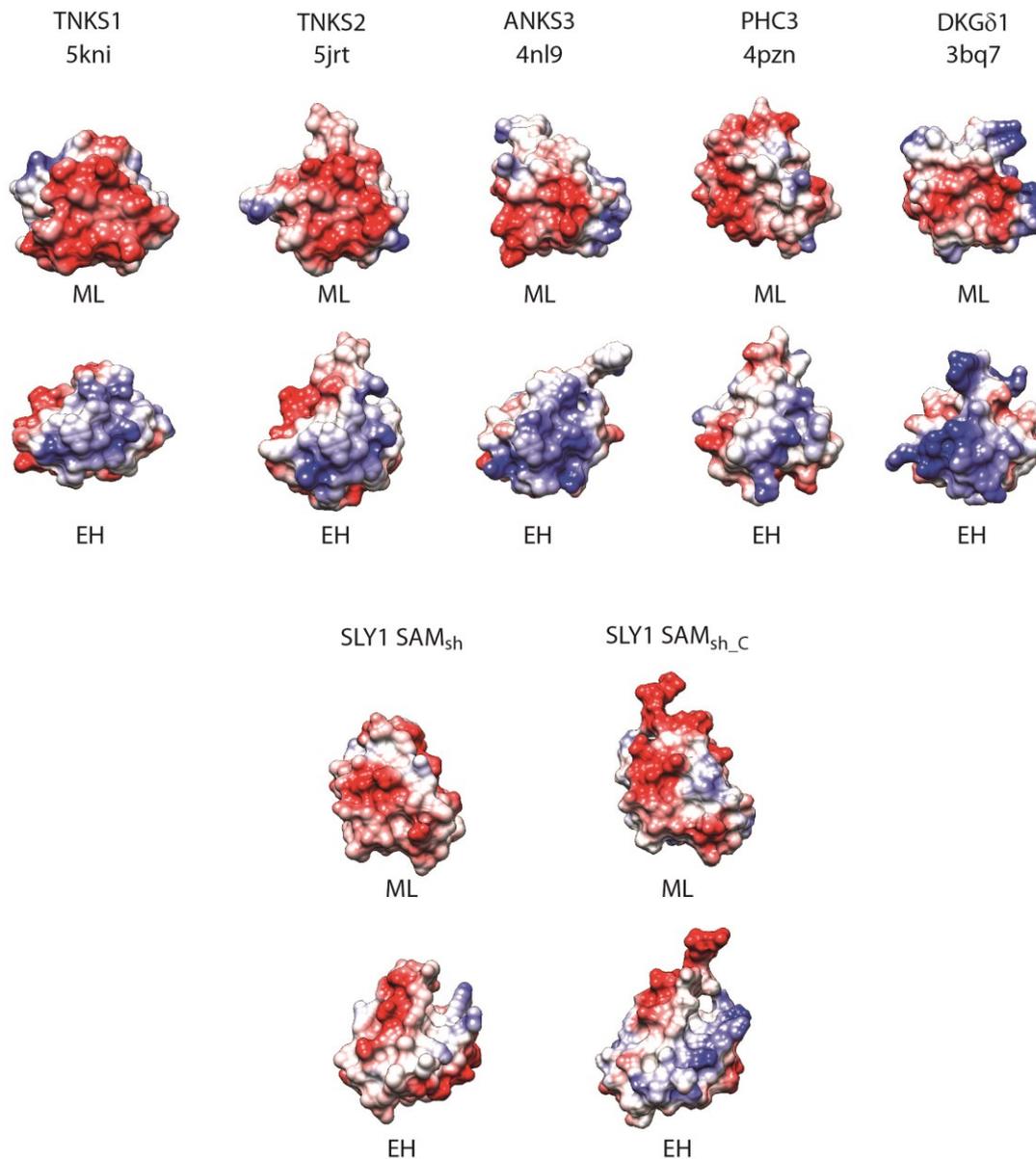


Fig. 51: ML-EH interfaces observed in homooligomerisation.

The SAM domains of the human proteins TNKS1, TNKS2, ANKS3, PHC3 and DKG δ 1 homooligomerise by employing the ML-EH interface. The surface representation with positive and negative charged residues coloured in blue and red, respectively, reveals complementing charges. Thus, the predominantly negatively charged ML interface readily interacts with the positively charged EH interface. SLY1 SAM also exhibits a negatively charged surface area around helices α 3 and α 4 and loops c 2/ α 3 and α 3/ α 4, which corresponds to the ML interface of other SAM domains. However, the N-terminal end of helix α 5 lacks the complementary positive charged residues on the surface, hence explaining why SLY1 SAM does not form large oligomers via the ML-EH interface.

The SAM domain of SLY1 does not show any indication to self-associate to larger oligomeric structures than the dimer described in this thesis. Analytical ultracentrifugation showed no evidence of SLY1 SAM polymers. In addition, no self-association or aggregation of NMR samples of SAM_{Ig} and SAM_{wt} were observed for extended periods of time (>1 year). Fig. 51 shows the ML and EH interfaces of the SAM domains of the human proteins ANKS3 (ankyrin repeat and SAM domain containing protein 3), DGK δ 1 (diacylglycerol kinase δ 1), TNKS1 (tankyrase 1), TNKS2 (tankyrase 2), PHC3 (polyhomeotic homologue 3) in surface representation (Harada *et al.*, 2008; Leettola *et al.*, 2014; Mariotti *et al.*, 2016; Riccio *et al.*, 2016; Robinson, Leal, Nanyes *et al.*, 2012). These five SAM domains are capable of forming homopolymers through ML-EH interactions. The ML interfaces of the polymerising SAM domains are predominantly negatively charged, whereas the EH interface consists of a cluster of positive charges. The complementary electrostatic charges of the ML and EH interfaces enable interaction between these SAM domains. The region in SLY1 SAM corresponding to the ML interface also contains predominantly negatively charged solvent-exposed residues (Fig. 51). However, the EH interface does not display a cluster of solvent-exposed positively charged residues but instead is dominated by hydrophobic residues on the surface of the N-terminal end of helix α 5. These hydrophobic residues contribute to the hydrophobic groove that forms the major part of the symmetric SLY1 SAM dimer interface. This lacking charge complementarity of the ML and EH surfaces of SLY1 SAM is very likely responsible for the lack of SLY1 SAM homopolymerisation. SLY1 SAM might form a discrete dimer by an ML-EH interface with another SAM domain. However, there is no evidence so far of interactions of SLY1 with other SAM domain containing proteins *in vivo*.

Currently, only SAM domains of the EphA4 (Stapleton *et al.*, 1999) and Ste11 (Bhattacharjya *et al.*, 2004) have been shown to form symmetric homodimers. In this context, the before mentioned structural similarity of SLY1 SAM to the SAM domains of the various ephrin receptor was of particular interest. The structural similarity of the monomeric subunits between SLY1 SAM and the ephrin receptor SAM domains suggested that there might also be similarities in the dimer structure. However, the SAM domain of EphA4 and the SAM domain of Ste11 self-associate by employing two types of interfaces (Fig. 52), which differ from each other and that of the SLY1 SAM homodimer.

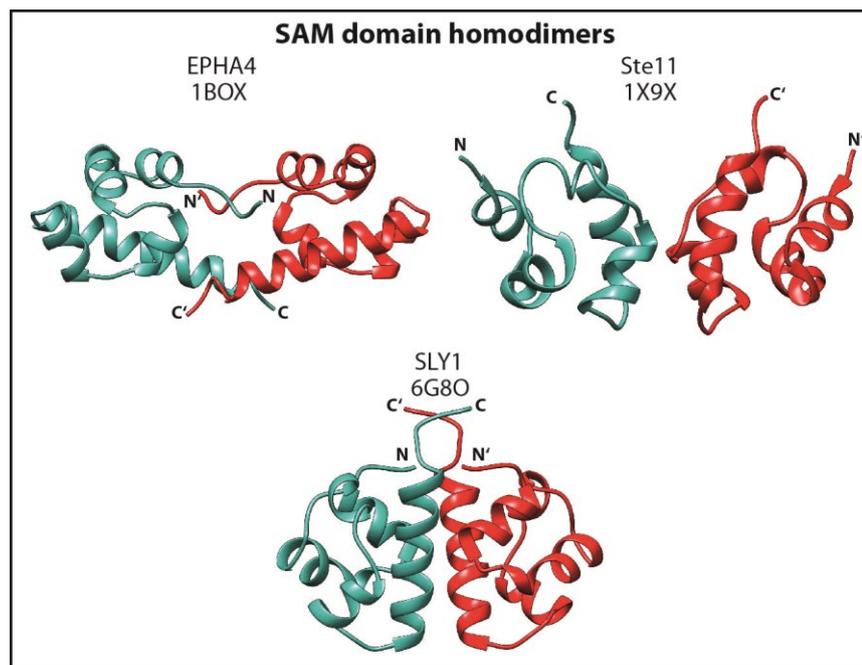


Fig. 52: Ribbon representation of different structures of symmetric SAM domain homodimers.

Out of the 70 SAM domains for which structural data is available in the PDB, only those of human EphA4 and yeast Ste11 form symmetric homodimers

Ste11 dimerises through interactions between hydrophobic residues positioned in helix $\alpha 4$ and at the beginning of the C-terminal helix $\alpha 5$ (Bhattacharjya *et al.*, 2004). EphA4 monomers interact in a pincer-like fashion, where the N-terminus and the C-terminal part of helix $\alpha 5$ extend from the subunit core and intertwine with the termini of the second SAM molecule. The interface is further stabilised by residues in helices $\alpha 1$ and $\alpha 3$ (Stapleton *et al.*, 1999). All of the homodimeric interactions described here are estimated to be of very low affinity with K_D values of 0.5 mM and 0.5-5 mM for Ste11 SAM (Grimshaw *et al.*, 2004) and EphA4 SAM (Stapleton *et al.*, 1999), respectively. Therefore, they are at least one order of magnitude weaker than the SLY1 SAM dimer. Inspection of the three SAM homodimer structures reveals that the buried surface area of the SLY1 SAM_{wt} homodimer (870 Å²) is larger than that of the Ste11 homodimer (504 Å²) but moderately lower than the EphA4 SAM homodimer (1009 Å²). Although the EphA4 SAM homodimer has a buried surface area of similar size to that of the SLY1 SAM homodimer, the number of stabilising hydrogen bonds and salt bridges is noticeably lower (2 versus 12), which likely explains the observed difference in homodimer stability. In contrast, for the Ste11 homodimer, a similar number of hydrogen bonds and salt bridges are present (8), but the buried surface area is considerably smaller and thus the homodimer is less stable.

However, the published data for the Ste11 SAM homodimer is contradictory, since other published studies are predicting asymmetric dimerisation of Ste11 SAM via a head-to-tail interface, as well as the formation of larger oligomeric structures, which are not explained by the published homodimer structures (Kwan *et al.*, 2004). Therefore, the biological relevance of the Ste11 symmetric homodimer is unclear. So far, neither of the two SAM homodimers described above have been assigned biological functions.

In conclusion, the typical head-to-tail interaction observed in multiple homotypic and heterotypic SAM domain complexes that employ ML and EH interfaces is unfavourable for SLY1 SAM dimerisation, probably because it lacks the necessary charge complementarity. Instead, the SLY1 SAM domain dimerises through a novel interface with a higher affinity than any other described symmetric SAM domain homodimer.

5.6 Importance of the extended N-terminus of the SAM domain in dimer formation

The flexibility of the N-terminal residues G(-2)–T256 (GPGKRPKPKT), and the multiple sets of signals arising from the three prolines (P(-1), P252, P254) hampered analysis of the NMR data of SAM_{Ig} and subsequent structure determination by NMR. These problems were addressed by shortening the SAM domain construct by the six residues G249–P254 to give SAM_{wt}. Unexpectedly, the five residues removed at the N-terminus appear to be involved in the SLY1 SAM domain dimerisation because the K_D determined for the SAM_{wt} dimerisation (i.e., 117 μ M by AUC; 153 μ M by MST; Sections 4.5.1; 4.5.2) is significantly higher than that determined for SAM_{Ig} (K_D = 19 μ M by MST; Section 4.3.2). Closer inspection of the SLY1 SAM dimer structure revealed that the deleted N-terminal residues, GKRPK, may be part of the dimer interface. The surface charge of the SAM monomer reveals a negatively charged surface patch formed by residues D315, D317 and E321 that are in close spatial proximity to the N-terminus of the opposing monomer. This negatively charged surface represents a possible binding site for the three positively charged residues K250, R251 and K253 that are part of SAM_{Ig}, but not of SAM_{wt} and SAM_C.

In support of the above-described hypothesis, an extended N-terminus or additional N-terminal residues beyond the boundaries of the SAM domains have been implicated to be involved in some SAM domain interactions. For example, the extended N-terminus of the protein SAMHD1 (SAM and histidine/aspartate domain containing protein 1) is

involved in the interaction with both the human protein DCAF (DDB1-and Cullin4-associated factor 1) and the HIV protein Vpx (Wu *et al.*, 2015). In the homodimer structures of EphA4, the unstructured N-terminus forms an integral part of the interface. Furthermore, the linker region N-terminally located from the SAM domain of the protein PHC3 appears to regulate the size of the SAM domain polymers (Robinson, Leal, Nanyes *et al.*, 2012). However, sequence analysis of the N-terminal regions of these proteins and SLY1 SAM show no sequence homology. The influence of the N-terminus on the dimerisation could be studied in more detail by structural characterisation of the SAM_{lg} construct. It would be interesting to study the SLY1 SAM N-terminus in the context of the SLY1 SH3 domain in a longer SLY1 construct comprising both SAM and SH3 domain. The SH3 domain of SLY1 is located 18 residues upstream of the SAM domain and has been identified to play a role in the dimerisation of SLY1 (Brandt, 2010). It would be interesting to see whether the SAM domain also dimerises in the presence of the SH3 domain and whether the two domains influence each other. The absence of the N-terminal residues clearly has a detrimental effect on the SAM-SAM interaction. Therefore structural reorientation of the linker region, e.g., upon binding of an SH3 binding partner, could modulate SAM domain dimerisation.

5.7 Biological implications of the SLY1 SAM dimer formation

Homodimerisation and oligomerisation is a standard mode of interaction for many SAM domains. Thus, homodimerisation of SLY1 SAM is in good agreement with previous observations describing SAM domain function. However, the formation of symmetric homodimers among SAM domains is uncommon. Only two other structures of symmetrical SAM domain homodimers have been published, all significantly distinct from each other and the determined SLY1 SAM dimer (Section 5.5). The protein SLY1 has been shown to dimerise (Brandt, 2010). However, the dimerisation appears to be reliant on the SH3 domain of the protein instead on the SAM domain, because deletion of the SH3 domain abolishes SLY1 dimerisation, while SLY1 still dimerises if SAM is removed (Brandt, 2010). It is plausible that the SAM domain modulates dimerisation of SLY1. The membrane-bound EphA2 and EphA3 receptors do not require the presence of their SAM domains for self-association, but it has been found that the absence of the SAM domain has a stabilizing effect on the EphA3 dimer (Singh *et al.*, 2015) and a

destabilizing effect on the dimer of EphA2 (Singh *et al.*, 2017). Also, the SH3 dependence and SAM independence of the SLY1 dimerisation has only been examined qualitatively by co-immunoprecipitation experiments and immunoblotting. A more quantitative approach would be to characterise the strength of the SLY1 dimerisation by determining the K_D in the absence and presence of its SAM domain.

Another interesting aspect about the SLY1 SAM dimerisation is the dimer interface itself. The type of interaction interface determined for the SLY1 SAM dimer has not been observed for any other SAM domain. The likely reason for the fact that SLY1 SAM does not self-associate via the ML-EH interfaces is that it could lead to the formation of higher order homopolymers, which could be detrimental to SLY1 function. It is plausible, however, that SLY1 SAM interacts with other SAM domains through the ML-EH interface to form heterodimers, as is the case for the SAM domains of Ste11 and its regulator Ste50 (Kwan *et al.*, 2006). The structure of the SLY1 SAM homodimer presents an outward facing ML surface on each monomer. Therefore it is plausible that other SAM domain containing proteins could interact with SLY1 SAM in its dimerised state via an ML-EH interaction, thereby enabling SLY1 to mediate interactions between other SAM domain containing proteins.

The affinities for the homodimerisations of the SAM domains of EphA4 and Ste11 are very low. It could be that the SLY1 SAM dimer differs from the other dimer structures because this interface allows a higher affinity interaction.

At the C-terminal end of SLY1 SAM, T318 and S320 have been identified as putative phosphorylation sites in mice and human SLY1 (Huttlin *et al.*, 2010; Mayya *et al.*, 2009). These residues are in proximity to the dimer interface. Phosphorylation of either of the two residues could likely affect the dimerisation kinetics of the SAM domain, as has been shown for interface proximal phosphorylation sites in other protein dimers (Graille *et al.*, 2005; Sluchanko & Gusev, 2012; Woodcock *et al.*, 2003). One could speculate that either the dimerisation of the SAM domain is preventing phosphorylation, or that the phosphorylation prevents dimerisation. However, understanding the connection of the phosphorylation site to the dimerisation would require additional experiments with phosphorylated SLY1 SAM domains.

What has been established is that the SLY1 SAM domain is capable of binding RNA binding proteins, which could be the result of its mostly negatively charged surface. Whether the dimerisation is of any importance in that context has not been assessed. Many nucleic acid binding proteins function as dimers. However, SLY1 SAM is most

probably not able to bind to RNA itself in the way other SAM domains do (Green *et al.*, 2003; Johnson & Donaldson, 2006) due to its lack of required positively charged residues in helix $\alpha 1$ and loop $\alpha 1/\alpha 2$ as pointed out by Arefanian *et al.*, 2016.

Another aspect which has yet to be addressed is the possible functional connection between the SAM domain and the SH3 located only 18 residues upstream of SAM. Aside from the possible influence of the SH3 domain on SAM domain dimerisation, SAM domains have also been found to stabilise other domains in proximity to them, resulting in the two domains functioning as one. Such composite domains have been described for SAM domains with coiled-coil and an EF-hand motifs (Koveal *et al.*, 2012; Stathopoulos *et al.*, 2008).

6 Conclusion and Outlook

The murine signalling adaptor protein SLY1 contains a SAM domain of unknown function. Therefore, determining the 3D structure of the SLY1 SAM domain and characterising its oligomerisation state should provide insight into the role of the SAM domain in the context of SLY1 biology.

For the structural studies conducted in this work, a SAM domain construct was created (SAM_{wt}) which comprised the complete 63 residues long folded SAM domain. As described for other SAM domains, SLY1 SAM_{wt} self-associates. However, in contrast to most other SAM homomers, SAM_{wt} forms distinct SAM homodimers as could be shown by AUC sedimentation equilibrium experiments. The affinity of the dimerisation has been determined to be moderate with a K_{D} of $\sim 153 \mu\text{M}$ by MST and $\sim 117 \mu\text{M}$ by AUC. For a construct with an extended N-terminus (SAM_{lg}), the affinity was increased by a factor of 6–8, indicating that the unstructured N-terminus, which is not part of the core SAM domain, strongly influences the dimerisation process.

The structure of the SLY1 SAM domain dimer was solved by solution NMR spectroscopy and X-ray crystallography. To make the SAM dimer amenable to NMR based structure determination, the single-cysteine mutant SAM_{C} was prepared. SAM_{C} allowed the cross-linking of the two dimer subunits via the introduced C-terminal cysteine, which inhibited dissociation and therefore chemical exchange induced line broadening. This way, standard 2D and 3D NMR experiments were successfully recorded to gain assignment and structural information of the SAM domain.

The NMR-derived structure of the cross-linked SLY1 SAM dimer represents a so far undescribed SAM-SAM interaction interface which involves the helices $\alpha 5$ of both subunits packing against each other, additionally supported by residues in helices $\alpha 1$ and the loop $\alpha 1/c2$. The validity of the cross-linked dimer structure and its interface is corroborated by intermolecular NOE data of the non-cross-linked SAM dimer. Additional credence is given by the SAM_{wt} crystal structure, which shows the same intersubunit contacts. The crystal structure was solved at a resolution of 2.05 Å and without any noteworthy Ramachandran or rotamer outliers. The superposition of the NMR derived structures with the SAM_{wt} crystal structure showed very high similarity for both, the monomeric subunits as well as the dimer structure.

The SLY1 SAM dimer is the first symmetric dimer structure described for SAM domains with a high enough affinity to suggest biological relevance, and the solved

structures now provide a solid foundation for new hypotheses of the SAM domain in SLY1. However, additional studies are required to assess the biological significance of the SLY1 SAM domain dimerisation. One future aspect of the SLY1 research could be the interaction of the SH3 and SAM domains of SLY1. Since the SH3 domain of SLY1 has been shown to be pivotal for the SLY1 dimerisation, the interplay of the two modules could help to understand whether the SAM dimerisation has a supportive or an obstructive influence on the overall SLY1 dimerisation. Additionally, it could help to assess, whether the SAM domain is required for SH3 domain stabilisation. Furthermore, the examination of the influence of interface proximal phosphorylations at T318 and S320 on the SAM domain dimerisation kinetics could help to elucidate the regulation of the SLY1 SAM domain dimerisation. At last, the search for potential interaction partners of the SLY1 SAM dimer as opposed to monomeric SAM could help to assess whether the SAM dimer functions as a supramolecular interaction surface for further interaction partners of SLY1.

7 References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213–221.
- Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van 't Veer, L. J., Berns, A. & van Lohuizen, M. (1997). Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes & Development*, *11*(2), 226–240.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.
- Anderson, R. J., Weng, Z., Campbell, R. K. & Jiang, X. (2005). Main-chain conformational tendencies of amino acids. *Proteins*, *60*(4), 679–689.
- Arefanian, S., Schäll, D., Chang, S., Ghasemi, R., Higashikubo, R., Zheleznyak, A., Guo, Y., Yu, J., Asgharian, H., Li, W., Gelman, A. E., Kreisel, D., French, A. R., Zaher, H., Plougastel-Douglas, B., Maggi, L., Yokoyama, W., ... Krupnick, A. S. (2016). Deficiency of the adaptor protein SLy1 results in a natural killer cell ribosomopathy affecting tumor clearance. *Oncoimmunology*, *5*(12), e1238543.
- Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, *22*(2), 195–201.
- Astoul, E., Laurence, A. D., Totty, N., Beer, S., Alexander, D. R. & Cantrell, D. A. (2003). Approaches to define antigen receptor-induced serine kinase signal transduction pathways. *The Journal of Biological Chemistry*, *278*(11), 9267–9275.
- Baker, D. A., Mille-Baker, B., Wainwright, S. M., Ish-Horowicz, D. & Dibb, N. J. (2001). Mae mediates MAP kinase phosphorylation of Ets transcription factors in *Drosophila*. *Nature*, *411*(6835), 330–334.
- Bardiaux, B., Bernard, A., Rieping, W., Habeck, M., Malliavin, T. E. & Nilges, M. (2009). Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA. *Proteins*, *75*(3), 569–585.
- Bardiaux, B., Malliavin, T. & Nilges, M. (2012). ARIA for Solution and Solid-State NMR. In *Protein NMR Techniques* (Vol. 831, pp. 453–483).

- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). iMOSFLM : a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D Biological Crystallography*, 67(4), 271–281.
- Beer, S., Scheikl, T., Reis, B., Hüser, N., Pfeffer, K. & Holzmann, B. (2005). Impaired immune responses and prolonged allograft survival in Sly1 mutant mice. *Molecular and Cellular Biology*, 25(21), 9646–9660.
- Beer, S., Simins, A. B., Schuster, A. & Holzmann, B. (2001). Molecular cloning and characterization of a novel SH3 protein (SLY) preferentially expressed in lymphoid cells. *Biochimica et Biophysica Acta*, 1520(1), 89–93.
- Berjanskii, M. V. & Wishart, D. S. (2005). A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *Journal of the American Chemical Society*, 127(43), 14970–14971.
- Berjanskii, M. V. & Wishart, D. S. (2007). The RCI server: Rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Research*, 35(SUPPL.2), 531–537.
- Berjanskii, M. V. & Wishart, D. S. (2008). Application of the random coil index to studying protein flexibility. *Journal of Biomolecular NMR*, 40(1), 31–48.
- Berman, H., Henrick, K. & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, 10(12), 980.
- Bhattacharjya, S., Xu, P., Gingras, R., Shaykhutdinov, R., Wu, C., Whiteway, M. & Ni, F. (2004). Solution structure of the dimeric SAM domain of MAPKKK Ste11 and its interactions with the adaptor protein Ste50 from the budding yeast: implications for Ste11 activation and signal transmission through the Ste50-Ste11 complex. *Journal of Molecular Biology*, 344(4), 1071–1087.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. & Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42(Web Server issue), W252-8.
- Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6), 1513–1523.
- Bloch, F., Hansen, W. W. & Packard, M. (1946). Nuclear Induction. *Physical Review*, 69(3–4), 127–127.
- Bornemann, D., Miller, E. & Simon, J. (1996). The Drosophila Polycomb group gene Sex comb on midleg (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development (Cambridge, England)*, 122(5), 1621–1630.
- Bourenkov, G. P. & Popov, A. N. (2010). Optimization of data collection taking radiation damage into account. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4), 409–419.

- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1–2), 248–254.
- Brandt, S. (2010). *Interaktionsanalyse der Adapterproteine SLy1 und SLy2*. Heinrich-Heine-Universität, Düsseldorf.
- Braun, W. & Gö, N. (1985). Calculation of protein conformations by proton-proton distance constraints. *Journal of Molecular Biology*, 186(3), 611–626.
- Breeze, A. L. (2000). Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 36(4), 323–372.
- Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359), 472–475.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica. Section D, Biological Crystallography*, 54(Pt 5), 905–921.
- Case, D. A., Dyson, H. J. & Wright, P. E. (1994). [13] Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins. In *Methods in Enzymology* (Vol. 239, pp. 392–416).
- Chen, J., Sawyer, N. & Regan, L. (2013). Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science: A Publication of the Protein Society*, 22(4), 510–515.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 1), 12–21.
- Chothia, C., Levitt, M. & Richardson, D. (1977). Structure of proteins: packing of alpha-helices and pleated sheets. *Proceedings of the National Academy of Sciences of the United States of America*, 74(10), 4130–4134.
- Claudio, J. O., Zhu, Y. X., Benn, S. J., Shukla, a H., McGlade, C. J., Falcioni, N. & Stewart, a K. (2001). HACS1 encodes a novel SH3-SAM adaptor protein differentially expressed in normal and malignant hematopoietic cells. *Oncogene*, 20(38), 5373–5377.
- Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1985). Solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*. *Journal of Molecular Biology*, 186(2), 435–455.
- Clore, G. M., Nilges, M., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1987).

- A comparison of the restrained molecular dynamics and distance geometry methods for determining three-dimensional structures of proteins on the basis of interproton distances. *FEBS Letters*, 213(2), 269–277.
- Cordier, F. & Grzesiek, S. (1999). Direct Observation of Hydrogen Bonds in Proteins by Interresidue $^3\text{H J NC}^{\prime}$ Scalar Couplings. *Journal of the American Chemical Society*, 121(7), 1601–1602.
- Cordier, F. & Grzesiek, S. (2002). Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. Edited by P. E. Wright. *Journal of Molecular Biology*, 317(5), 739–752.
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35(Web Server issue), W375–83.
- DeCamillis, M., Cheng, N. S., Pierre, D. & Brock, H. W. (1992). The polyhomeotic gene of *Drosophila* encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. *Genes & Development*, 6(2), 223–232.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), 277–293.
- Denay, G., Vachon, G., Dumas, R., Zubieta, C. & Parcy, F. (2017). Plant SAM-Domain Proteins Start to Reveal Their Roles. *Trends in Plant Science*, 22(8), 718–725.
- Duhr, S. & Braun, D. (2006). Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52), 19678–19682.
- Düx, P., Whitehead, B., Boelens, R., Kaptein, R. & Vuister, G. W. (1997). Measurement of ^{15}N - ^1H coupling constants in uniformly ^{15}N -labeled proteins: Application to the photoactive yellow protein. *Journal of Biomolecular NMR*, 10(3), 301–306.
- Emsley, P. & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), 2126–2132.
- Erdem-Eraslan, L., Heijnsman, D., de Wit, M., Kremer, A., Sacchetti, A., van der Spek, P. J., Sillevius Smitt, P. A. E. & French, P. J. (2015). Tumor-specific mutations in low-frequency genes affect their functional properties. *Journal of Neuro-Oncology*, 122(3), 461–470.
- Ernst, R. R. & Anderson, W. A. (1966). Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments*, 37(1), 93–102.
- Evans, P. R. (2011). An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 282–292.

- Evans, P. R. & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7), 1204–1214.
- Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wüthrich, K. (2002). NMR analysis of a 900K GroEL GroES complex. *Nature*, 418(6894), 207–211.
- Findeisen, M., Brand, T. & Berger, S. (2007). A ¹H-NMR thermometer suitable for cryoprobes. *Magnetic Resonance in Chemistry : MRC*, 45(2), 175–178.
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44(D1), D279–85.
- Garbuzynskiy, S. O., Melnik, B. S., Lobanov, M. Y., Finkelstein, A. V. & Galzitskaya, O. V. (2005). Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? *Proteins*, 60(1), 139–147.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607.
- Göbl, C., Madl, T., Simon, B. & Sattler, M. (2014). NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 80, 26–63.
- Graille, M., Zhou, C. Z., Receveur-Bréchet, V., Collinet, B., Declerck, N. & Van Tilbeurgh, H. (2005). Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. *Journal of Biological Chemistry*, 280(15), 14780–14789.
- Green, J. B., Gardner, C. D., Wharton, R. P. & Aggarwal, A. K. (2003). RNA recognition via the SAM domain of Smaug. *Molecular Cell*, 11(6), 1537–1548.
- Grimshaw, S. J., Mott, H. R., Stott, K. M., Nielsen, P. R., Evetts, K. A., Hopkins, L. J., Nietlispach, D. & Owen, D. (2004). Structure of the Sterile α Motif (SAM) Domain of the *Saccharomyces cerevisiae* Mitogen-activated Protein Kinase Pathway-modulating Protein STE50 and Analysis of Its Interaction with the STE11 SAM. *Journal of Biological Chemistry*, 279(3), 2192–2201.
- Grzesiek, S., Bax, A., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., Tjandra, N. & Wingfield, P. T. (1997). Refined solution structure and backbone dynamics of HIV-1 Nef. *Protein Science : A Publication of the Protein Society*, 6(6), 1248–1263.
- Güntert, P. & Buchner, L. (2015). Combined automated NOE assignment and structure calculation with CYANA. *Journal of Biomolecular NMR*, 62(4), 453–471.
- Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR

7. References

- structure calculation with the new program DYANA. *Journal of Molecular Biology*, 273(1), 283–298.
- Harada, B. T., Knight, M. J., Imai, S.-I., Qiao, F., Ramachander, R., Sawaya, M. R., Gingery, M., Sakane, F. & Bowie, J. U. (2008). Regulation of enzyme localization by polymerization: polymer formation by the SAM domain of diacylglycerol kinase delta1. *Structure (London, England : 1993)*, 16(3), 380–387.
- Huang, Y. J., Powers, R. & Montelione, G. T. (2005). Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *Journal of the American Chemical Society*, 127(6), 1665–1674.
- Huang, Y. J., Rosato, A., Singh, G. & Montelione, G. T. (2012). RPF: a quality assessment tool for protein NMR structures. *Nucleic Acids Research*, 40(Web Server issue), W542-6.
- Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., Villén, J., Haas, W., Sowa, M. E. & Gygi, S. P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*, 143(7), 1174–1189.
- Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E. & Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Developmental Cell*, 26(6), 565–577.
- Jaravine, V. A., Alexandrescu, A. T. & Grzesiek, S. (2001). Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide. *Protein Science*, 10(5), 943–950.
- Jeener, J. & Alewaeters, G. (2016). “Pulse pair technique in high resolution NMR” a reprint of the historical 1971 lecture notes on two-dimensional spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 94–95, 75–80.
- Jeffrey, G. A. (1997). *An Introduction to Hydrogen Bonding*. Oxford University Press.
- Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S.-H., Couloux, A., Lee, S.-W., Yoon, S. H., Cattolico, L., Hur, C.-G., Park, H.-S., Ségurens, B., Kim, S. C., Oh, T. K., Lenski, R. E., Studier, F. W., ... Kim, J. F. (2009). Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, 394(4), 644–652.
- Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. (2011). Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY and Drug Development Technologies*, 9(4), 342–353.
- Johnson, P. E. & Donaldson, L. W. (2006). RNA recognition by the Vts1p SAM domain. *Nature Structural & Molecular Biology*, 13(2), 177–178.
- Jordan, M. S., Singer, A. L. & Koretzky, G. A. (2003). Adaptors as central mediators of signal transduction in immune cells. *Nature Immunology*, 4(2), 110–116.
- Junius, F. K., O’Donoghue, S. I., Nilges, M.,

- Weiss, A. S. & King, G. F. (1996). High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *The Journal of Biological Chemistry*, 271(23), 13663–13667.
- Kabsch, W. (2010). Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 125–132.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, 22, 2577–2637.
- Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science: A Publication of the Protein Society*, 12(9), 1865–1871.
- Karplus, M. (1959). Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *The Journal of Chemical Physics*, 30(1), 11–15.
- Karplus, M. (1963). Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of the American Chemical Society*, 85(18), 2870–2871.
- Katz, B. A. & Kossiakoff, A. (1986). The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *The Journal of Biological Chemistry*, 261(33), 15480–15485.
- Kay, L. E., Wittekind, M., McCoy, M. A., Friedrichs, M. S. & Mueller, L. (1992). 4D NMR triple-resonance experiments for assignment of protein backbone nuclei using shared constant-time evolution periods. *Journal of Magnetic Resonance*, 98(2), 443–450.
- Kay, L., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society*, 114(26), 10663–10665.
- Keeler, J. (2010). *Understanding NMR Spectroscopy* (2nd Editio). Wiley.
- Kim, C. A. & Bowie, J. U. (2003). SAM domains: uniform structure, diversity of function. *Trends in Biochemical Sciences*, 28(12), 625–628.
- Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S. & Bowie, J. U. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *The EMBO Journal*, 20(15), 4173–4182.
- Klämbt, C. (1993). The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development (Cambridge, England)*, 117(1), 163–176.
- Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallographica. Section D, Biological Crystallography*, 52(Pt 4), 842–857.
- Koveal, D., Schuh-Nuhfer, N., Ritt, D., Page, R., Morrison, D. K. & Peti, W. (2012). A CC-SAM, for Coiled Coil-Sterile Motif, Domain Targets the Scaffold KSR-1 to Specific Sites in the Plasma Membrane.

- Science Signaling*, 5(255), ra94-ra94.
- Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), 2256–2268.
- Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3), 774–797.
- Kurabi, A., Brener, S., Mobli, M., Kwan, J. J. & Donaldson, L. W. (2009). A Nuclear Localization Signal at the SAM–SAM Domain Interface of AIDA-1 Suggests a Requirement for Domain Uncoupling Prior to Nuclear Import. *Journal of Molecular Biology*, 392(5), 1168–1177.
- Kwan, J. J., Warner, N., Maini, J., Chan Tung, K. W., Zakaria, H., Pawson, T. & Donaldson, L. W. (2006). Saccharomyces cerevisiae Ste50 binds the MAPKKK Ste11 through a head-to-tail SAM domain interaction. *Journal of Molecular Biology*, 356(1), 142–154.
- Kwan, J. J., Warner, N., Pawson, T. & Donaldson, L. W. (2004). The solution structure of the S.cerevisiae Ste11 MAPKKK SAM domain and its partnership with Ste50. *Journal of Molecular Biology*, 342(2), 681–693.
- Lämmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), 680–685.
- Laskowski, R. a, Rullmann, J. a, MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR*, 8(4), 477–486.
- Leettola, C. N., Knight, M., Cascio, D., Hoffman, S. & Bowie, J. U. (2014). Characterization of the SAM domain of the PKD-related protein ANKS6 and its interaction with ANKS3. *BMC Structural Biology*, 14(1), 17.
- Li, H., Fung, K.-L., Jin, D.-Y., Chung, S. S., Ching, Y.-P., Ng, I. O., Sze, K.-H., Ko, B. C. B. & Sun, H. (2007). Solution structures, dynamics, and lipid-binding of the sterile α -motif domain of the deleted in liver cancer 2. *Proteins: Structure, Function, and Bioinformatics*, 67(4), 1154–1166.
- Linge, J. P. & Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *Journal of Biomolecular NMR*, 13(1), 51–59.
- Liu, B. A. & Nash, P. D. (2012). Evolution of SH2 domains and phosphotyrosine signalling networks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367(1602), 2556–2573.
- Marianayagam, N. J., Sunde, M. & Matthews, J. M. (2004). The power of two: protein dimerization in biology. *Trends in Biochemical Sciences*, 29(11), 618–625.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application

- to the study of hydrogen exchange in proteins. *Journal of Magnetic Resonance* (1969), 85(2), 393–399.
- Mariotti, L., Templeton, C. M., Raney, M., Paracuellos, P., Cronin, N., Beuron, F., Morris, E. & Guettler, S. (2016). Tankyrase Requires SAM Domain-Dependent Polymerization to Support Wnt- β -Catenin Signaling. *Molecular Cell*, 63(3), 498–513.
- Mayya, V., Lundgren, D. H., Hwang, S.-I., Rezaul, K., Wu, L., Eng, J. K., Rodionov, V. & Han, D. K. (2009). Quantitative phosphoproteomic analysis of T cell receptor signaling reveals system-wide modulation of protein-protein interactions. *Science Signaling*, 2(84), ra46.
- Meruelo, A. D. & Bowie, J. U. (2009). Identifying polymer-forming SAM domains. *Proteins: Structure, Function, and Bioinformatics*, 74(1), 1–5.
- Metcalf, D. G., Moore, D. T., Wu, Y., Kielec, J. M., Molnar, K., Valentine, K. G., Wand, A. J., Bennett, J. S. & DeGrado, W. F. (2010). NMR analysis of the α IIb β 3 cytoplasmic interaction suggests a mechanism for integrin regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(52), 22481–22486.
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, 12(4), 345–364.
- Motáčková, V., Nováček, J., Zawadzka-Kazimierczuk, A., Kazimierczuk, K., Zidek, L., Sanderová, H., Krásný, L., Koźmiński, W. & Sklenář, V. (2010). Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *Journal of Biomolecular NMR*, 48(3), 169–177.
- Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. (1989). Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods in Enzymology*, 177, 44–73.
- Neal, S., Nip, A. M., Zhang, H. & Wishart, D. S. (2003). Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *Journal of Biomolecular NMR*, 26(3), 215–240.
- Neuhaus, D. & Williamson, M. P. (2000). *The nuclear Overhauser effect in structural and conformational analysis*. (M. P. Williamson, Ed.) (2nd ed.). New York NY: Wiley-VCH.
- Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T., Habeck, M. & Rieping, W. (2008). Accurate NMR Structures Through Minimization of an Extended Hybrid Energy. *Structure*, 16(9), 1305–1312.
- Nilges, M., Clore, G. M. & Gronenborn, A. M. (1987). A simple method for delineating well-defined and variable regions in protein structures determined from interproton distance data. *FEBS Letters*, 219(1), 11–16.
- Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). Determination of three-

7. References

- dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters*, 229(2), 317–324.
- Nilges, M. & O'Donoghue, S. I. (1998). Ambiguous NOEs and automated NOE assignment. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 32(2), 107–139.
- Nomura, M., Takihara, Y. & Shimada, K. (1994). Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: one of the early inducible clones encodes a novel protein sharing several highly homologous regions with a *Drosophila* polyhomeotic protein. *Differentiation; Research in Biological Diversity*, 57(1), 39–50.
- Nooren, I. M. A. & Thornton, J. M. (2003). Structural Characterisation and Functional Significance of Transient Protein–Protein Interactions. *Journal of Molecular Biology*, 325(5), 991–1018.
- Nussinov, R. & Jang, H. (2014). Dynamic multiprotein assemblies shape the spatial structure of cell signaling. *Progress in Biophysics and Molecular Biology*, 116(2–3), 158–164.
- Pal, L., Dasgupta, B. & Chakrabarti, P. (2005). 310-Helix adjoining α -helix and β -strand: Sequence and structural features and their conservation. *Biopolymers*, 78(3), 147–162.
- Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance (1969)*, 93(1), 151–170.
- Pérez, C., Löhr, F., Rüterjans, H. & Schmidt, J. M. (2001). Self-consistent Karplus parametrization of 3J couplings depending on the polypeptide side-chain torsion χ_1 . *Journal of the American Chemical Society*, 123(29), 7081–7093.
- Peterson, A. J., Kyba, M., Bornemann, D., Morgan, K., Brock, H. W. & Simon, J. (1997). A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Molecular and Cellular Biology*, 17(11), 6683–6692.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612.
- Ponting, C. P. (1995). SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Science: A Publication of the Protein Society*, 4(9), 1928–1930.
- Purcell, E. M., Torrey, H. C. & Pound, R. V. (1946). Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1–2), 37–38.
- Qiao, F. & Bowie, J. U. (2005). The many faces of SAM. *Science's STKE: Signal Transduction Knowledge Environment*, 2005(286), re7.
- Qiao, F., Song, H., Kim, C. A., Sawaya, M. R., Hunter, J. B., Gingery, M., Rebay, I.,

- Courey, A. J. & Bowie, J. U. (2004). Derepression by depolymerization; structural insights into the regulation of Yan by Mae. *Cell*, 118(2), 163–173.
- Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. (1938). A New Method of Measuring Nuclear Magnetic Moment. *Physical Review*, 53(4), 318–318.
- Rajakulendran, T., Sahmi, M., Kurinov, I., Tyers, M., Therrien, M. & Sicheri, F. (2008). CNK and HYP form a discrete dimer by their SAM domains to mediate RAF kinase signaling. *Proceedings of the National Academy of Sciences*, 105(8), 2836–2841.
- Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1), 95–99.
- Reis, B., Pfeffer, K. & Beer-Hammer, S. (2009). The orphan adapter protein SLY1 as a novel anti-apoptotic protein required for thymocyte development. *BMC Immunology*, 10, 38.
- Riccio, A. A., McCauley, M., Langelier, M.-F. & Pascal, J. M. (2016). Tankyrase Sterile α Motif Domain Polymerization Is Required for Its Role in Wnt Signaling. *Structure*, 24(9), 1573–1581.
- Rieping, W., Bardiaux, B., Bernard, A., Malliavin, T. E. & Nilges, M. (2007). ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3), 381–382.
- Robinson, A. K., Leal, B. Z., Chadwell, L. V., Wang, R., Ilangoan, U., Kaur, Y., Junco, S. E., Schirf, V., Osmulski, P. A., Gaczynska, M., Hinck, A. P., Demeler, B., McEwen, D. G. & Kim, C. A. (2012). The Growth-Suppressive Function of the Polycomb Group Protein Polyhomeotic Is Mediated by Polymerization of Its Sterile Alpha Motif (SAM) Domain. *Journal of Biological Chemistry*, 287(12), 8702–8713.
- Robinson, A. K., Leal, B. Z., Nanyes, D. R., Kaur, Y., Ilangoan, U., Schirf, V., Hinck, A. P., Demeler, B. & Kim, C. A. (2012). Human Polyhomeotic Homolog 3 (PHC3) Sterile Alpha Motif (SAM) Linker Allows Open-Ended Polymerization of PHC3 SAM. *Biochemistry*, 51(27), 5379–5386.
- Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. & Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839), 487–491.
- Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning A Laboratory Manual* (3rd ed.). Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.
- Sathyamurthy, A., Freund, S. M. V, Johnson, C. M., Allen, M. D. & Bycroft, M. (2011). Structural basis of p63 α SAM domain mutants involved in AEC syndrome. *The FEBS Journal*, 278(15), 2680–2688.
- Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance*

- Spectroscopy*, 34(2), 93–158.
- Sayou, C., Nanao, M. H., Jamin, M., Posé, D., Thévenon, E., Grégoire, L., Tichtinsky, G., Denay, G., Ott, F., Peirats Llobet, M., Schmid, M., Dumas, R. & Parcy, F. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nature Communications*, 7(7), 11222.
- Schäll, D., Schmitt, F., Reis, B., Brandt, S. & Beer-Hammer, S. (2015). SLy1 regulates T-cell proliferation during *Listeria monocytogenes* infection in a Foxo1-dependent manner. *European Journal of Immunology*, 45(11), 3087–3097.
- Scheikl, T., Reis, B., Pfeffer, K., Holzmann, B. & Beer, S. (2009). Reduced notch activity is associated with an impaired marginal zone B cell development and function in Sly1 mutant mice. *Molecular Immunology*, 46(5), 969–977.
- Schmidt, B., Ho, L. & Hogg, P. J. (2006). Allosteric Disulfide Bonds. *Biochemistry*, 45(24), 7429–7433.
- Schmidt, B. & Hogg, P. J. (2007). Search for allosteric disulfide bonds in NMR structures. *BMC Structural Biology*, 7(1), 49.
- Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society*, 123(13), 2970–2978.
- Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Marius Clore, G. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 160(1), 65–73.
- Seavey, B. R., Farr, E. A., Westler, W. M. & Markley, J. L. (1991). A relational database for sequence-specific protein NMR data. *Journal of Biomolecular NMR*, 1(3), 217–236.
- Shen, Y. & Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR*, 48(1), 13–22.
- Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, 44(4), 213–223.
- Sikic, K., Tomic, S. & Carugo, O. (2010). Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. *The Open Biochemistry Journal*, 4, 83–95.
- Singh, D. R., Ahmed, F., Paul, M. D., Gedam, M., Pasquale, E. B. & Hristova, K. (2017). The SAM domain inhibits EphA2 interactions in the plasma membrane. *Biochimica et Biophysica Acta*, 1864(1), 31–38.
- Singh, D. R., Cao, Q., King, C., Salotto, M., Ahmed, F., Zhou, X. Y., Pasquale, E. B. & Hristova, K. (2015). Unliganded EphA3 dimerization promoted by the SAM domain. *The Biochemical Journal*, 471(1), 101–109.
- Sluchanko, N. N. & Gusev, N. B. (2012). Oligomeric structure of 14-3-3 protein: what do we know about monomers? *FEBS*

- Letters*, 586(24), 4249–4256.
- Slupsky, C. M., Gentile, L. N., Donaldson, L. W., Mackereth, C. D., Seidel, J. J., Graves, B. J. & McIntosh, L. P. (1998). Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), 12129–12134.
- Stapleton, D., Balan, I., Pawson, T. & Sicheri, F. (1999). The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nature Structural Biology*, 6(1), 44–49.
- States, D. ., Haberkorn, R. . & Ruben, D. . (1982). A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. *Journal of Magnetic Resonance (1969)*, 48(2), 286–292.
- Stathopoulos, P. B., Zheng, L., Li, G.-Y., Plevin, M. J. & Ikura, M. (2008). Structural and mechanistic insights into STIM1-mediated initiation of store-operated calcium entry. *Cell*, 135(1), 110–122.
- Stein, E. G., Rice, L. M. & Brünger, a T. (1997). Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*, 124(1), 154–164.
- Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1), 113–130.
- Styles, P., Soffe, N. ., Scott, C. ., Crag, D. ., Row, F., White, D. . & White, P. C. . (1984). A high-resolution NMR probe in which the coil and preamplifier are cooled with liquid helium. *Journal of Magnetic Resonance (1969)*, 60(3), 397–404.
- Szilágyi, L. (1995). Chemical shifts in proteins come of age. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 27(4), 325–442.
- Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., Tempel, W., Min, J., Tong, Y., Kim, P. M., Bader, G. D. & Sidhu, S. S. (2017). Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure (London, England : 1993)*, 25(10), 1598–1610.e3.
- Thiagarajan, P. (2011). *NMR structure and biophysical characterization of SAM domain of SLY1 and the transmembrane and cytoplasmic domain of CD4*. Heinrich-Heine-Universität, Düsseldorf.
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P. & Vriend, G. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(Database issue), D364-8.
- Urzhumtseva, L., Afonine, P. V., Adams, P. D. & Urzhumtsev, A. (2009). Crystallographic model quality at a glance. *Acta Crystallographica. Section D, Biological Crystallography*, 65(Pt 3), 297–300.
- Vagin, A. & Teplyakov, A. (1997). MOLREP : an Automated Program for Molecular

- Replacement. *Journal of Applied Crystallography*, 30(6), 1022–1025.
- Vajpai, N., Gentner, M., Huang, J., Blackledge, M. & Grzesiek, S. (2010). Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from 3J Coupling Constants and Residual Dipolar Couplings. *Journal of the American Chemical Society*, 132(9), 3196–3203.
- Verschueren, E., Spiess, M., Gkourtsa, A., Avula, T., Landgraf, C., Mancilla, V. T., Huber, A., Volkmer, R., Winsor, B., Serrano, L., Hochstenbach, F. & Distel, B. (2015). Evolution of the SH3 Domain Specificity Landscape in Yeasts. *PLoS One*, 10(6), e0129229.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4), 687–696.
- Vuister, G. W. & Bax, A. (1992). Resolution enhancement and spectral editing of uniformly ^{13}C -enriched proteins by homonuclear broadband ^{13}C decoupling. *Journal of Magnetic Resonance (1969)*, 98(2), 428–435.
- Wang, X., Bansal, S., Jiang, M. & Prestegard, J. H. (2008). RDC-assisted modeling of symmetric protein homo-oligomers. *Protein Science: A Publication of the Protein Society*, 17(5), 899–907.
- Wei, Z., Zheng, S., Spangler, S. A., Yu, C., Hoogenraad, C. C. & Zhang, M. (2011). Liprin-mediated large signaling complex organization revealed by the liprin- α /CASK and liprin- α /liprin- β complex structures. *Molecular Cell*, 43(4), 586–598.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3), 765–784.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica, Section D, Biological Crystallography*, 67(Pt 4), 235–242.
- Wishart, D. S. (2011). Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(1–2), 62–87.
- Woodcock, J. M., Murphy, J., Stomski, F. C., Berndt, M. C. & Lopez, A. F. (2003). The dimeric versus monomeric status of 14-3-3 ζ is controlled by phosphorylation of Ser58 at the dimer interface. *Journal of Biological Chemistry*, 278(38), 36323–36327.
- Wu, Y., Koharudin, L. M. I., Mehrens, J., DeLucia, M., Byeon, C.-H., Byeon, I.-J. L., Calero, G., Ahn, J. & Gronenborn, A. M. (2015). Structural Basis of Clade-specific Engagement of SAMHD1 (Sterile

- α Motif and Histidine/Aspartate-containing Protein 1) Restriction Factors by Lentiviral Viral Protein X (Vpx) Virulence Factors. *Journal of Biological Chemistry*, 290(29), 17935–17945.
- Zeller, C., Hinzmann, B., Seitz, S., Prokoph, H., Burkhard-Goettges, E., Fischer, J., Jandrig, B., Schwarz, L.-E., Rosenthal, A. & Scherneck, S. (2003). SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is downregulated in breast cancer. *Oncogene*, 22(19), 2972–2983.
- Zhang, J., Graham, T. G. W., Vivekanand, P., Cote, L., Cetera, M. & Rebay, I. (2010). Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the Drosophila ETS family transcriptional repressor Yan. *Molecular and Cellular Biology*, 30(5), 1158–1170.
- Zhong, D., Zhang, J., Yang, S., Soh, U. J. K., Buschdorf, J. P., Zhou, Y. T., Yang, D. & Low, B. C. (2009). The SAM domain of the RhoGAP DLC1 binds EF1A1 to regulate cell migration. *Journal of Cell Science*, 122(3), 414–424.

8 Appendix

8.1 Xml script for NOE resonance assignment and structure calculation by ARIA and CNS

```
<!DOCTYPE project SYSTEM "project1.0.dtd">
<project name="SAM_Dimer" version="1.0" author="" date="Fri Aug 11 18:15:22 2017" description=""
comment="" references="" working_directory="/home/lkukuk/SAM/structurecalc/ARIA"
temp_root="/home/lkukuk/SAM/structurecalc/ARIA" run="SAM_Dimer_45" file_root="SAM" cache="yes"
cleanup="yes">
  <data>
    <ccpn_model filename="/home/lkukuk/SAM/structurecalc/analysis_201707245_run45_in"/>
    <molecule file="" format="ccpn" ccpn_id="MS1|A|B">
      <linkage_definition name="automatic" filename=""/>
      <parameter_definition name="automatic" filename=""/>
      <topology_definition name="automatic" filename=""/>
    </molecule>
    <spectrum enabled="yes" use_assignments="yes" trust_assigned_peaks="no" structural_rules="no"
filter_diagonal_peaks="yes" filter_unassigned_peaks="no">
      <shifts file="" format="ccpn" ccpn_id="defaultProject|1" default_shift_error="0.0"
use_shift_error="yes"/>
      <peaks file="" format="ccpn" ccpn_id="defaultProject|19|1|1" peak_size="volume"
freq_window_proton1="0.04" freq_window_hetero1="0.5" freq_window_proton2="0.04"
freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.5" enabled="yes"/>
      </peaks>
      <experiment_data molecule_correlation_time="0.0" spectrum_mixing_time="0.0"
spectrometer_frequency="0.0" ambiguity_type="all"/>
    </spectrum>
    <spectrum enabled="yes" use_assignments="yes" trust_assigned_peaks="no" structural_rules="no"
filter_diagonal_peaks="yes" filter_unassigned_peaks="no">
      <shifts file="" format="ccpn" ccpn_id="defaultProject|1" default_shift_error="0.0"
use_shift_error="yes"/>
      <peaks file="" format="ccpn" ccpn_id="defaultProject|20|1|1" peak_size="volume"
freq_window_proton1="0.04" freq_window_hetero1="0.5" freq_window_proton2="0.04"
freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.5" enabled="yes"/>
      </peaks>
      <experiment_data molecule_correlation_time="0.0" spectrum_mixing_time="0.0"
spectrometer_frequency="0.0" ambiguity_type="all"/>
    </spectrum>
    <spectrum enabled="yes" use_assignments="yes" trust_assigned_peaks="no" structural_rules="no"
filter_diagonal_peaks="yes" filter_unassigned_peaks="no">
      <shifts file="" format="ccpn" ccpn_id="defaultProject|1" default_shift_error="0.0"
use_shift_error="yes"/>
      <peaks file="" format="ccpn" ccpn_id="defaultProject|21|1|1" peak_size="volume"
freq_window_proton1="0.04" freq_window_hetero1="0.5" freq_window_proton2="0.04"
freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.5" enabled="yes"/>
      </peaks>
      <experiment_data molecule_correlation_time="0.0" spectrum_mixing_time="0.0"
spectrometer_frequency="0.0" ambiguity_type="all"/>
    </spectrum>
    <spectrum enabled="yes" use_assignments="yes" trust_assigned_peaks="no" structural_rules="no"
filter_diagonal_peaks="yes" filter_unassigned_peaks="no">
      <shifts file="" format="ccpn" ccpn_id="defaultProject|1" default_shift_error="0.0"
use_shift_error="yes"/>
      <peaks file="" format="ccpn" ccpn_id="defaultProject|26|1|1" peak_size="volume"
freq_window_proton1="0.06" freq_window_hetero1="0.5" freq_window_proton2="0.05"
freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.5" enabled="yes"/>
      </peaks>
      <experiment_data molecule_correlation_time="0.0" spectrum_mixing_time="0.0"
spectrometer_frequency="0.0" ambiguity_type="inter"/>
    </spectrum>
    <spectrum enabled="yes" use_assignments="yes" trust_assigned_peaks="no" structural_rules="no"
filter_diagonal_peaks="yes" filter_unassigned_peaks="no">
      <shifts file="" format="ccpn" ccpn_id="defaultProject|1" default_shift_error="0.0"
use_shift_error="yes"/>
      <peaks file="" format="ccpn" ccpn_id="defaultProject|27|1|1" peak_size="volume"
freq_window_proton1="0.05" freq_window_hetero1="0.5" freq_window_proton2="0.05"
freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.5" enabled="yes"/>
      </peaks>
      <experiment_data molecule_correlation_time="0.0" spectrum_mixing_time="0.0"
spectrometer_frequency="0.0" ambiguity_type="inter"/>
    </spectrum>
  </data>
</project>
```

8. Appendix

```
<hbonds file="/home/lkukuk/SAM/hbonds/hbonds_SAM.tbl" format="tbl" ccpn_id="" enabled="yes"
data_type="standard"/>
<dihedrals file="/home/lkukuk/SAM/TALOS/talos_test/Dihedrals_bb_20170221_A.tbl" format="tbl"
ccpn_id="" enabled="yes" data_type="talos"/>
<dihedrals file="/home/lkukuk/SAM/chil_angles/chil_AB.tbl" format="tbl" ccpn_id="" enabled="yes"
data_type="standard"/>
<dihedrals file="/home/lkukuk/SAM/TALOS/talos_test/Dihedrals_bb_20170221_B.tbl" format="tbl"
ccpn_id="" enabled="yes" data_type="talos"/>
<ssbridge residue1="68" segid1="A" residue2="68" segid2="B"/>
<symmetry enabled="yes" method="standard" n_monomers="2" symmetry_type="C2" ncs_enabled="yes"
packing_enabled="yes"/>
<initial_structure file="" format="iupac" ccpn_id="" enabled="no"/>
</data>
<structure_generation engine="cns">
  <cns local_executable="/opt/scisoft/usr/bin/cns-remote" keep_output="yes"
keep_restraint_files="yes" create_psf_file="yes" generate_template="yes"
nonbonded_parameters="PARALLHDG">
  <annealing_parameters>
    <unambiguous_restraints first_iteration="0" k_hot="10.0" k_cool1_initial="10.0"
k_cool1_final="25.0" k_cool2="25.0"/>
    <ambiguous_restraints first_iteration="0" k_hot="10.0" k_cool1_initial="10.0"
k_cool1_final="25.0" k_cool2="25.0"/>
    <hbond_restraints first_iteration="0" k_hot="10.0" k_cool1_initial="10.0"
k_cool1_final="50.0" k_cool2="50.0"/>
    <dihedral_restraints k_hot="20.0" k_cool1="20.0" k_cool2="20.0"/>
    <karplus_restraints parameter_class="1" a="6.98" b="-1.38" c="1.72" d="-60.0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0"/>
    <karplus_restraints parameter_class="2" a="6.98" b="-1.38" c="1.72" d="-60.0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0"/>
    <karplus_restraints parameter_class="3" a="6.98" b="-1.38" c="1.72" d="-60.0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0"/>
    <karplus_restraints parameter_class="4" a="6.98" b="-1.38" c="1.72" d="-60.0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0"/>
    <karplus_restraints parameter_class="5" a="6.98" b="-1.38" c="1.72" d="-60.0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0"/>
    <rdc_restraints parameter_class="1" method="SANI" first_iteration="0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0" r="0.4" d="8.0" border_hot_initial="0.1" border_hot_final="40.0"
border_cool1_initial="40.0" border_cool1_final="40.0" border_cool2_initial="40.0"
border_cool2_final="40.0" center_hot_initial="0.1" center_hot_final="0.1" center_cool1_initial="10.0"
center_cool1_final="10.0" center_cool2_initial="10.0" center_cool2_final="10.0"/>
    <rdc_restraints parameter_class="2" method="SANI" first_iteration="0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0" r="0.4" d="8.0" border_hot_initial="0.1" border_hot_final="40.0"
border_cool1_initial="40.0" border_cool1_final="40.0" border_cool2_initial="40.0"
border_cool2_final="40.0" center_hot_initial="0.1" center_hot_final="0.1" center_cool1_initial="10.0"
center_cool1_final="10.0" center_cool2_initial="10.0" center_cool2_final="10.0"/>
    <rdc_restraints parameter_class="3" method="SANI" first_iteration="0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0" r="0.4" d="8.0" border_hot_initial="0.1" border_hot_final="40.0"
border_cool1_initial="40.0" border_cool1_final="40.0" border_cool2_initial="40.0"
border_cool2_final="40.0" center_hot_initial="0.1" center_hot_final="0.1" center_cool1_initial="10.0"
center_cool1_final="10.0" center_cool2_initial="10.0" center_cool2_final="10.0"/>
    <rdc_restraints parameter_class="4" method="SANI" first_iteration="0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0" r="0.4" d="8.0" border_hot_initial="0.1" border_hot_final="40.0"
border_cool1_initial="40.0" border_cool1_final="40.0" border_cool2_initial="40.0"
border_cool2_final="40.0" center_hot_initial="0.1" center_hot_final="0.1" center_cool1_initial="10.0"
center_cool1_final="10.0" center_cool2_initial="10.0" center_cool2_final="10.0"/>
    <rdc_restraints parameter_class="5" method="SANI" first_iteration="0" k_hot="0.0"
k_cool1="0.2" k_cool2="1.0" r="0.4" d="8.0" border_hot_initial="0.1" border_hot_final="40.0"
border_cool1_initial="40.0" border_cool1_final="40.0" border_cool2_initial="40.0"
border_cool2_final="40.0" center_hot_initial="0.1" center_hot_final="0.1" center_cool1_initial="10.0"
center_cool1_final="10.0" center_cool2_initial="10.0" center_cool2_final="10.0"/>
    <flat_bottom_harmonic wall m_rswitch_hot="0.5" m_rswitch_cool1="0.5" m_rswitch_cool2="0.5"
rswitch_hot="0.5" rswitch_cool1="0.5" rswitch_cool2="0.5" m_asymptote_hot="-1.0" m_asymptote_cool1="-
1.0" m_asymptote_cool2="-0.1" asymptote_hot="1.0" asymptote_cool1="1.0" asymptote_cool2="0.1"/>
    <symmetry_restraints k_packing_hot="15.0" k_packing_cool1="10.0" k_packing_cool2="5.0"
last_iteration_packing="8" k_ncs="50.0"/>
    <logharmonic_potential enabled="yes" use_auto_weight="yes" weight_unambig="25.0"
weight_ambig="25.0" weight_hbond="25.0"/>
  </annealing_parameters>
  <md_parameters dynamics="torsion" random_seed="89764443" tad_temp_high="10000.0"
tad_timestep_factor="9.0" cartesian_temp_high="2000.0" cartesian_first_iteration="0" timestep="0.002"
temp_cool1_final="1000.0" temp_cool2_final="50.0" steps_high="15000" steps_refine="8000"
steps_cool1="20000" steps_cool2="20000"/>
</cns>
<job_manager default_command="csh -f">
  <host enabled="yes" command="qsub -S /opt/scisoft64/bin/csh"
executable="/opt/scisoft64/usr/bin/cns-remote" n_cpu="100" use_absolute_path="yes"/>
</job_manager>
</structure_generation>
<protocol floating_assignment="yes">
  <iteration number="0" n_structures="30" sort_criterion="total_energy" n_best_structures="10"
n_kept_structures="0">
    <assignment/>
    <merging method="standard"/>
    <calibration relaxation_matrix="yes" distance_cutoff="6.0" estimator="ratio_of_averages"
error_estimator="distance"/>
    <violation_analysis violation_tolerance="1000.0" violation_threshold="0.5"/>
    <partial_assignment weight_threshold="1.0" max_contributions="20"/>
    <network_anchoring high_residue_threshold="4.0" enabled="yes" min_residue_threshold="1.0"
min_atom_threshold="0.25"/>
  </iteration>
</protocol>
</structure_generation>
```



```

<water_refinement solvent="water" n_structures="15" enabled="yes" write_solvent_molecules="no"/>
</protocol>
<analysis>
  <structures_analysis enabled="yes"/>
  <procheck executable="/opt/scisoft64/usr/bin/procheck" enabled="yes"/>
  <prosa executable="/usr/bin/prosa" enabled="yes"/>
  <whatif executable="whatif" enabled="no"/>
  <clashlist executable="/opt/scisoft64/usr/bin/clashlist" enabled="yes"/>
</analysis>
<report>
  <ccpn export_assignments="yes" export_noe_restraint_list="last" export_structures="yes"/>
  <molmol enabled="yes"/>
  <noe_restraint_list pickle_output="no" text_output="yes" xml_output="no"/>
  <spectra write_assigned="no" write_assigned_force="no" iteration="last"
write_unambiguous_only="yes"/>
</report>
</project>

```

8.2 ^{15}N , ^{13}C - ω_1 -filtered NOESY- ^1H , ^{15}N -HSQC NOE cross-peak assignments

| Assign F1 | Assign F2 |
|---------------------------------|---|
| 11 Arg H | 60 Leu H $^{\delta a*}$ |
| 11 Arg H | 59 Glu (H $^{\gamma b}$ /H $^{\beta b}$) |
| 12 Ile H | 60 Leu H $^{\delta a*}$ |
| 12 Ile H | 59 Glu (H $^{\gamma b}$ /H $^{\beta b}$) |
| 12 Ile H | (59 Glu/56 Thr)H $^{\alpha}$ |
| 49 Gln H | 52 Ala H $^{\beta*}$ |
| 49 Gln H | 49 Gln H $^{\gamma a}$ |
| 49 Gln H | (48 Pro/49 Gln)(H $^{\beta b}$ /H $^{\gamma b}$) |
| 49 Gln H | 49 Gln H $^{\alpha}$ |
| 49GlnHe2a | 52AlaH $^{\beta*}$ |
| 49 Gln H $^{\epsilon 2a}$ | 48 Pro H $^{\beta a}$ |
| 49 Gln H $^{\epsilon 2a}$ | 48 Pro H $^{\gamma b}$ |
| 49 Gln H $^{\epsilon 2a}$ | 48 Pro H $^{\beta b}$ |
| 49 Gln H $^{\epsilon 2a}$ | 48 Pro H $^{\alpha}$ |
| 49 Gln H $^{\epsilon 2\beta}$ | 52 Ala H $^{\beta*}$ |
| 49 Gln H $^{\epsilon 2\beta}$ | (48 Pro/51 Arg) H $^{\beta*}$ |
| 49 Gln H $^{\epsilon 2\beta b}$ | 48 Pro H $^{\beta b}$ |
| 49 Gln H $^{\epsilon 2\beta}$ | 48 Pro H $^{\beta b}$ |
| 49 Gln H $^{\epsilon 2\beta}$ | (52Ala/48 Pro) H $^{\alpha}$ |
| 52 Ala H | 53 Lys H $^{\gamma a}$ |
| 52AlaH | 52 Ala H $^{\beta*}$ |
| 52AlaH | 53 Lys H $^{\gamma b}$ |
| 52AlaH | 49 Gln H $^{\gamma a}$ |
| 52AlaH | 49 Gln H $^{\gamma b}$ |
| 52AlaH | 53 Lys H $^{\epsilon 1}$ |
| 52AlaH | 49 Gln H $^{\alpha}$ |
| 52AlaH | 53 Lys H |
| 53LysH | 52 Ala H |

| | |
|--------|--|
| 53LysH | (49 Gln/53 Lys) H ^α |
| 53LysH | 52 Ala H ^{β*} |
| 53LysH | (53 Lys/56 Thr) (H ^{γ^a} /H ^{γ^{2*}}) |
| 53LysH | (54 Leu/56 Thr/57 Ala) H ^α |
| 53LysH | 53 Lys (H ^{β^b} /H ^{γ^b}) |
| 54LeuH | 60 Leu H ^{δ*} |
| 54LeuH | 56 Thr H ^{γ^{2*}} |
| 54LeuH | 52 Ala H ^{β*} |
| 55LeuH | 53 Lys (H ^{β^a} /H ^{δ*}) |
| 56ThrH | 60 Leu H ^{δ^{b*}} |
| 56ThrH | (53 Lys/56 Thr) (H ^{γ^a} /H ^{γ^{2*}}) |
| 56ThrH | 53 Lys (H ^{β^a} /H ^{δ*}) |
| 56ThrH | 53 Lys (H ^{γ^b} /H ^{β^b}) |
| 56ThrH | 53 Lys H ^{ε*} |
| 56ThrH | 53 Lys H ^α |
| 56ThrH | 56 Thr H ^β |
| 57AlaH | 60 Leu H ^{δ^{b*}} |
| 57AlaH | 56 Thr H ^{γ^{2*}} |
| 57AlaH | (56 Thr/57Ala) H ^α |
| 57AlaH | 56 Thr H ^β |
| 58AlaH | 60 Leu H ^{δ^{b*}} |
| 58AlaH | 56 Thr H ^{γ^{2*}} |
| 59GluH | 60 Leu H ^{δ^{b*}} |
| 59GluH | (56 Thr/ Ala57/53 Lys) H ^α |
| 59GluH | 12 Ile H ^{γ^{2*}} |
| 60LeuH | 60 Leu H ^{δ^{b*}} |
| 60LeuH | (57 Ala/60 Leu)H ^{β*} |
| 60LeuH | 60 Leu (H ^{β^b} /H ^γ) |
| 60LeuH | 57 Ala H ^α |
| 60LeuH | 8 Leu H ^{δ^{a*}} |
| 61LeuH | 60 Leu H ^{δ^{b*}} |
| 61LeuH | 60 Leu H ^{β^a} |
| 61LeuH | 60 Leu H ^{β^b} |
| 66ThrH | 66 Thr H ^{γ^{2*}} |
| 66ThrH | 66 Thr H _α |
| 66ThrH | 67 Gly H |
| 67GlyH | 66 Thr H ^{γ^{2*}} |
| 67GlyH | 67 Gly H ^{α*} |
| 67GlyH | 66 Thr H ^α |
| 67GlyH | 66 Thr H |
| 67GlyH | 69 Glu H |

| | |
|--------|------------------------|
| 68CysH | 68 Cys H ^{βa} |
| 68CysH | 68CysH ^{βb} |
| 68CysH | 67 Gly H ^{α*} |
| 68CysH | 66 Thr H [*] |
| 69GluH | 68 Cys H ^{βa} |
| 69GluH | 68 Cys H ^{βb} |

8.3 ¹⁵N,¹³C-ω₁-filtered NOESY-¹H,¹³C-HSQC NOE cross-peak assignments

| Assign F1 | Assign F2 |
|------------------------|--|
| 2 Pro H ^{δ*} | 60 Leu H ^{δa*} |
| 2 Pro H ^{δ*} | 63 Asp H ^{βa} |
| 2 Pro H ^{δ*} | 64 Tyr H ^{ε*} |
| 2 Pro H ^{γb} | 60 Leu H ^{δa*} |
| 8 Leu H ^{δa*} | 56 Thr H ^{γ2*} |
| 8 Leu H ^{δa*} | 60 Leu (H ^γ /H ^{βb}) |
| 8 Leu H ^{δa*} | 60 Leu H |
| 8 Leu H ^{δa*} | 60 Leu H ^{δ*} |
| 8 Leu H ^{δa*} | 64 Tyr H ^{ε*} |
| 8 Leu H ^{δb*} | 60 Leu H ^α |
| 8 Leu H ^{δb*} | 60 Leu H ^{βa} |
| 8 Leu H ^{δb*} | 60 Leu H ^γ |
| 8 Leu H ^{δb*} | 64 Tyr H ^{δ*} |
| 8 Leu H ^{δb*} | 64 Tyr H ^{ε*} |
| 8 Leu H ^{δb*} | (57 Ala/56 Thr) H ^α |
| 8 Leu H ^γ | 60 Leu H ^{δ*} |
| 11 Arg H ^{β*} | 59 Glu (H ^{γb} /H ^{βb}) |
| 11 Arg H ^{β*} | 59 Glu H ^α |
| 11 Arg H ^{β*} | 60 Leu H |
| 11 Arg H ^{β*} | 60 Leu H ^{δa*} |
| 11 Arg H ^{δa} | 60 Leu H ^{δa*} |
| 11 Arg H ^{δa} | 63 Asp H ^{βb} |
| 11 Arg H ^{δa} | (60 Leu/62 Leu) (H ^{βb} /H ^γ) |
| 11 Arg H ^{δb} | 60 Leu H ^{δa*} |
| 11 Arg H ^{γa} | 60 Leu H ^{δa*} |
| 11 Arg H ^{γa} | (63 Asp/64Tyr) H |
| 11 Arg H ^{γb} | 60 Leu H ^{δa*} |
| 11 Arg H ^{γb} | (64 Tyr/63 Asp) H |
| 12 Ile H ^β | 56 Thr H ^{γ2*} |
| 12 Ile H ^β | 60 Leu H ^{δ*} |

| | |
|-------------------------|--|
| 12 Ile H ^{δ1*} | 52 Ala H ^{β*} |
| 12 Ile H ^{δ1*} | 56 Thr H ^α |
| 12 Ile H ^{δ1*} | 56 Thr H ^{γ2*} |
| 12 Ile H ^{δ1*} | 59 Glu (H ^{γa} /H ^{βb}) |
| 12 Ile H ^{δ1*} | 60 Leu H ^{δ*} |
| 12 Ile H ^{δ1*} | 60 Leu (H ^{βa} /H ^{βb} /H ^γ) |
| 12 Ile H ^{δ1*} | (53 Lys/57 Ala/60 Leu) H |
| 12 Ile H ^{γ2*} | 56 Thr H ^α |
| 12 Ile H ^{γ2*} | 56 Thr H ^{γ2*} |
| 12 Ile H ^{γ2*} | 59 Glu (H ^{γb} /H ^{βb}) |
| 12 Ile H ^{γ2*} | 59 Glu H |
| 12 Ile H ^{γ2*} | 59 Glu H ^{βa} |
| 12 Ile H ^{γ2*} | 60 Leu H |
| 29 Leu H ^{δb*} | 64 Tyr H ^{ε*} |
| 48 Pro H ^{βa} | 49 Gln H ^α |
| 48 Pro H ^{βa} | 49 Gln H ^{ε2b} |
| 48 Pro H ^{βa} | 49 Gln H ^{γa} |
| 48 Pro H ^{βa} | 49 Gln H ^{γb} |
| 48 Pro H ^{βb} | 49 Gln H ^α |
| 48 Pro H ^{βb} | 49 Gln H ^{γa} |
| 49 Gln H ^{βa} | 52 Ala H ^{β*} |
| 49 Gln H ^{βa} | (49 Gln/52 Ala) H ^α |
| 49 Gln H ^{βb} | 49 Gln H |
| 49 Gln H ^{βb} | 52 Ala H ^{β*} |
| 49 Gln H ^{γa} | 48 Pro H ^{βa} |
| 49 Gln H ^{γa} | 52 Ala H |
| 49 Gln H ^{γa} | 52 Ala H ^{β*} |
| 49 Gln H ^{γa} | (49 Gln/48 Pro) H ^α |
| 49 Gln H ^{γb} | 48 Pro H ^{βa} |
| 49 Gln H ^{γb} | 52 Ala H |
| 49 Gln H ^{γb} | 52 Ala H ^{β*} |
| 49 Gln H ^{γb} | (48 Pro/49 Gln) H ^α |
| 52 Ala H ^{β*} | 49 Gln H |
| 52 Ala H ^{β*} | 49 Gln H ^α |
| 52 Ala H ^{β*} | 49 Gln H ^{βa} |
| 52 Ala H ^{β*} | 49 Gln H ^{ε2a} |
| 52 Ala H ^{β*} | 49 Gln H ^{ε2b} |
| 52 Ala H ^{β*} | 49 Gln H ^{γa} |
| 52 Ala H ^{β*} | 49 Gln H ^{γb} |
| 52 Ala H ^{β*} | 52 Ala H |
| 52 Ala H ^{β*} | 53 Lys (H ^{βa} /H ^{δ*}) |

8. Appendix

| | |
|-------------------------|--|
| 52 Ala H ^{β*} | 53 Lys H |
| 52 Ala H ^{β*} | 53 Lys H ^{ε2} |
| 52 Ala H ^{β*} | 53 Lys H ^{γa} |
| 52 Ala H ^{β*} | 53 Lys H ^{γb} |
| 53 Lys H ^{βa} | 56 Thr H ^{γ2*} |
| 53 Lys H ^{βb} | (53 Lys/54 Leu/57 Ala) H |
| 53 Lys H ^{δ*} | 52 Ala H ^α |
| 53 Lys H ^{δ*} | 52 Ala H ^{β*} |
| 53 Lys H ^{δ*} | 56 Thr H ^{γ2*} |
| 53 Lys H ^{δ*} | (Glu 37/ 51 Arg/53 Lys/59 Glu) (H ^{βa} /H ^{βa} /H ^{βb} /H ^{γb}) |
| 53 Lys H ^{ε*} | 55 Leu H ^{βb} |
| 53 Lys H ^{ε*} | 55 Leu H ^{δb*} |
| 53 Lys H ^{ε*} | 56 Thr H |
| 53 Lys H ^{ε*} | 56 Thr H ^β |
| 53 Lys H ^{ε*} | (53 Lys/57 Ala/60 Leu/54 Leu) H |
| 53 Lys H ^{ε2} | 56 Thr H ^a |
| 53 Lys H ^{ε2} | 56 Thr H ^{ε2*} |
| 53 Lys H ^{ε3} | 52 Ala H ^a |
| 53 Lys H ^{ε3} | 52 Ala H ^{b*} |
| 53 Lys H ^{γa} | 52 Ala H ^a |
| 53 Lys H ^{γa} | 52 Ala H ^{b*} |
| 53 Lys H ^{γa} | 53 Lys H |
| 53 Lys H ^{γb} | 52 Ala H ^{β*} |
| 55 Leu H ^{δa*} | 53 Lys (H ^{δ*} /H ^{βa}) |
| 55 Leu H ^{δa*} | (60 Leu/14 Leu) (H ^{d*} /H ^{db*}) |
| 55 Leu H ^{δb*} | 49 Gln (H ^{βa} /H ^β /H ^{γa}) |
| 55 Leu H ^{δb*} | 53 Lys H ^{δ*} |
| 55 Leu H ^{δb*} | 53 Lys H ^{ε2} |
| 56 Thr H ^{γ2*} | 8 Leu H ^{δa*} |
| 56 Thr H ^{γ2*} | 2 Ile H ^β |
| 56 Thr H ^{γ2*} | 12 Ile H ^{δ1*} |
| 56 Thr H ^{γ2*} | 12 Ile H ^{γ2*} |
| 56 Thr H ^{γ2*} | 53 Lys (H ^{δ2} /H ^{βa}) |
| 56 Thr H ^{γ2*} | 53 Lys H ^{βb} |
| 56 Thr H ^{γ2*} | 53 Lys H ^{ε1} |
| 56 Thr H ^{γ2*} | 56 Thr H |
| 56 Thr H ^{γ2*} | 56 Thr H ^β |
| 56 Thr H ^{γ2*} | 57 Ala H |
| 56 Thr H ^{γ2*} | 57 Ala H ^{b*} |
| 56 Thr H ^{γ2*} | (57 Ala/53 Lys) H ^α |
| 57 Ala H ^{β*} | 56 Thr H ^β |

| | |
|-------------------------|---------------------------------|
| 57 Ala H ^{β*} | 56 Thr H ^{γ2*} |
| 57 Ala H ^{β*} | 60 Leu H ^{δb*} |
| 57 Ala H ^{β*} | 60 Leu H ^γ |
| 57 Ala H ^{β*} | (57 Ala/56 Thr) H ^α |
| 57 Ala H ^{β*} | (57 Ala/60 Leu) H |
| 59 Glu H ^{βa} | 12 Ile H ^α |
| 59 Glu H ^{βa} | 12 Ile H ^{γ2*} |
| 59 Glu H ^{βb} | 12 Ile H ^{γ2*} |
| 59 Glu H ^{γa} | 12 Ile H ^{γ2*} |
| 59 Glu H ^{γb} | 12 Ile H ^{γ2*} |
| 60 Leu H ^{βa} | 57 Ala H ^α |
| 60 Leu H ^{βa} | (60 Leu/61 Leu)H ^{δb*} |
| 60 Leu H ^{βb} | 57 Ala H ^α |
| 60 Leu H ^{βb} | 60 Leu H ^{δb*} |
| 60 Leu H ^{βb} | (61 Leu/60 Leu) H |
| 60 Leu H ^{δa*} | 2 Pro H ^{δ*} |
| 60 Leu H ^{δa*} | 2 Pro H ^{γb} |
| 60 Leu H ^{δa*} | 8 Leu H ^{δa*} |
| 60 Leu H ^{δa*} | 8 Leu H ^γ |
| 60 Leu H ^{δa*} | 11 Arg H |
| 60 Leu H ^{δa*} | 11 Arg H ^{β*} |
| 60 Leu H ^{δa*} | 11 Arg H ^α |
| 60 Leu H ^{δa*} | 11 Arg H ^{γb} |
| 60 Leu H ^{δa*} | 57 Ala H |
| 60 Leu H ^{δa*} | 57 Ala H ^{β*} |
| 60 Leu H ^{δa*} | 61 Leu H |
| 60 Leu H ^{δa*} | (11 Arg/61 Leu) H ^α |
| 60 Leu H ^{δa*} | (57 Ala/8 Leu) H ^α |
| 60 Leu H ^{δb*} | 8 Leu H ^{δa*} |
| 60 Leu H ^{δb*} | 8 Leu H ^γ |
| 60 Leu H ^{δb*} | 56 Thr H ^{γ2*} |
| 60 Leu H ^{δb*} | 57 Ala H ^α |
| 60 Leu H ^{δb*} | 57 Ala H ^{β*} |
| 60 Leu H ^{δb*} | 61 Leu H |
| 60 Leu H ^{δb*} | 61 Leu H ^{βa} |
| 60 Leu H ^{δb*} | (60 Leu/61 Leu) H ^α |
| 60 Leu H ^{δb*} | (58 Ala/57 Ala/60 Leu) H |
| 60 Leu H ^γ | 57 Ala H ^{β*} |
| 60 Leu H ^γ | 61 Leu H ^{δb*} |
| 60 Leu H ^γ | (57 Ala/8 Leu) H ^α |
| 60 Leu H ^γ | (58 Ala/60 Leu/ 61 Leu) H |

8. Appendix

| | |
|-------------------------------|---|
| 61 Leu H ^{βa} | 60 Leu H ^{δb*} |
| 61 Leu H ^{βb} | 60 Leu H ^{δb*} |
| 61 Leu H ^{δa*} | 64 Tyr H ^{δ*} |
| 61 Leu H ^{δa*} | 64 Tyr H ^{ε*} |
| 61 Leu H ^{δb*} | 60 Leu H ^{βa} |
| 61 Leu H ^{δb*} | 64 Tyr H ^{δ*} |
| 61 Leu H ^{δb*} | 64 Tyr H ^{ε*} |
| 62 Leu H ^{δb*} | 11 Arg H ^a |
| 62 Leu H ^{δb*} | 11 Arg H ^{γ*} |
| 62 Leu H ^{δb*} | (2Pro/11Arg) H ^{γ*} |
| 64 Tyr H ^{βb} | 64 Tyr H ^{δ*} |
| 66 Thr H ^{γ2*} | 64 Tyr H ^{δ*} |
| 66 Thr H ^{γ2*} | 64 Tyr H ^{ε*} |
| 66 Thr H ^{γ2*} | 66 Thr H |
| 66 Thr H ^{γ2*} | 66 Thr H ^a |
| 66 Thr H ^{γ2*} | (65 Asp/63 Asp) H ^{β*} |
| 68 Cys H ^{βa} | 68 Cys H ^{βb} |
| 68 Cys H ^{βb} | 68 Cys H ^a |
| 68 Cys H ^{βb} | 68 Cys H ^{βa} |
| 69 Glu H ^{γ*} | 66 Thr H |
| (60Leu/61Leu)H ^{δb*} | 60 Leu (H ^γ /H ^{βb}) |

Publications and presentations

Publications

Laura Kukuk, Andrew J. Dingley, Joachim Granzin, Luitgard Nagel-Steger, Pallavi Thiagarajan-Rosenkranz, Daniel Ciupka, Karen Hänel, Renu Batra-Safferling, Victor Pacheco, Matthias Stoldt, Klaus Pfeffer, Sandra Beer-Hammer, Dieter Willbold & Bernd W. Koenig (2018). Structure of the SLy1 SAM homodimer reveals a new interface for SAM domain self-association, *Scientific reports* (manuscript submitted)

Oral Presentations

Laura Kukuk (2016). Structural characterization of the SLY1 SAM dimer. *Annual meeting of the Bio-NMR-Network-NRW (bio-N3MR), Düsseldorf, Germany*

Danksagung

Allen voran danke ich Prof. Dr. Dieter Willbold, dass er es mir ermöglicht hat, in diesem Institut meine Doktorarbeit anzufertigen, und für die Überlassung des sehr interessanten Themas. Die angenehme Arbeitsatmosphäre und die hervorragende Ausstattung des Instituts haben die Arbeit hier zu einer besonderen Erfahrung gemacht. Des Weiteren möchte ich mich für seine Unterstützung und seinen Optimismus bedanken.

In besonderem Maße danke ich Dr. Bernd König für die Übernahme des Koreferats, sowie für seine engagierte Betreuung und seine Diskussionsbereitschaft. Seine konstruktiven Ratschläge und Denkanstöße haben maßgeblich zum Zustandekommen dieser Arbeit beigetragen.

Mein herzlichster Dank gilt Dr. Joachim Granzin und Dr. Andrew Dingley, die mir mit fachlicher Kompetenz als auch mit großem Enthusiasmus zur Seite gestanden haben. Ich danke für die Unterstützung bei der Durchführung und Auswertung der anspruchsvollen strukturbioologischen Experimente, die vielen interessanten Gespräche und nicht zuletzt für die Durchsicht dieses Manuskripts.

Dr. Luitgard Nagel-Steger danke ich für die angenehme Zusammenarbeit und die Unterstützung bei der Durchführung und Auswertung der analytischen Ultrazentrifugations-Experimente.

Des Weiteren danke ich Dr. Silke Hoffmann, Dr. Renu Batra-Safferling, Dr. Karen Hänel, Dr. Oliver Weiergräber und Dr. Gunnar Schröder für die vielen wertvollen Ratschläge und die ständige Hilfsbereitschaft.

Meinen ehemaligen Bürokolleginnen Dr. Marina Pavlidou und Dr. Alexandra Boeske, meiner ‚Austauschkollegin‘ Nicole Herr und Dr. Stephan Thies danke ich in allergrößtem Maße für den endlosen Enthusiasmus, die Motivation und die zahlreichen Gespräche wissenschaftlicher und nicht wissenschaftlicher Natur, ohne die ich nicht so lange durchgehalten hätte. In diesem Zusammenhang möchte ich auch Nicole Bleffert und Nicole Kremer von ganzem Herzen danken, ohne die der Laboralltag deutlich

weniger effizient, deutlich weniger organisiert und deutlich weniger herzlich gewesen wäre.

Darüber hinaus danke ich all meinen Kollegen für die angenehme Zusammenarbeit und die ständige Hilfsbereitschaft.

Nicht zuletzt danke ich all meinen Freunden und meiner wunderbaren Familie, die in den letzten Jahren viel zu häufig haben auf mich verzichten müssen, und trotzdem nie aufgehört haben mich zu unterstützen und an mich zu glauben. Ohne sie wäre diese Arbeit nicht möglich gewesen.

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass die Dissertation von mir selbstständig und ohne unzulässige Hilfsmittel unter Beachtung der „Grundlagen zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Datum, Ort

Unterschrift