

Optimising His-tags for purification and phasing

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

vorgelegt von
Christian Große
aus Pöbneck, Deutschland

Göttingen, 2010

D7

Referent: Prof. G. M. Sheldrick
Korreferent: Dr. B. Dittrich

Tag der mündlichen Prüfung: 5.10.2010

Acknowledgment

I am deeply indebted to Prof. George M. Sheldrick for my exciting project and interesting discussions. During my time in Göttingen I have greatly benefited from his advice and the familial environment in his work group.

A lot of work was done at the Max Planck Institute of Biophysical Chemistry in Göttingen in the group of Dr. S. Becker. So I am very thankful to Dr. Stefan Becker for all the work he did concerning the protein samples and his technical assistance.

I would like to thank the whole Sheldrick Group: Andrea Thorn, Kevin Pröpper, Navdeep Sidhu, Dr. Roland Pfoh, Dr. Tim Grüne, Tobias Beck, Dr. Regina Herbst-Irmer, Helmut Dehnhardt and Dalila Griffin. Of course I am very thankful to the Dittrich group as well: Dr. Birger Dittrich, Dr. Christian B. Hübschle and Julian Holstein. It was always a cooperative work and a friendly atmosphere. We had a lot of interesting wide-ranging discussions and a lot of fun. Thank you all.

I am thankful to the IRTG for funding and all IRTG students and associated supervisors for collaboration and discussions. Special thanks to two members of the group of Prof. U. Diederichsen: Dr. Philipp Schneggenburger and Karsten Meyenberg. You supported me with the solid phase peptide synthesiser and all the related chemistry.

Thanks to all supporting departments at the University of Göttingen: Dr. M. John and the NMR group for data collection and the analytic group for collecting mass spectra.

I am very thankful to my family: Heike and my lads Tim and Moritz, my parents and parents-in-law, my sister Ines and Tom as well. You gave me the power to complete my thesis, supported me all the time and believed in my work. Thank you.

Last but not least I would like to thank my correction readers for many constructive discussions about English grammar. Thank you so much for the hard work!

to Heike, Tim and Moritz.

Abbreviations

#	Number
Ac	Acetate
Boc	Tert. Butoxycarbonyl
CBD	Cellulose-binding domain
CD	Circular dichromism
CMA	Carboxymethylaspartate
C-term.	C-terminus
DCM	Dichloromethane
DEAE	Diethylaminoethyl
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol-bis(beta-amino-ethyl ether)
Eq.	Equation
eq.	Equivalents
ESI-MS	Electrospray ionisation mass spectrometry
Fmoc	9-Fluorenylmethoxycarbonyl
GE	GE Healthcare Life Sciences
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HOBt	Hydroxybenzotriazole
HT	High-Throughput
IDA	Iminodiacetic acid
IMAC	Immobilized metal ion affinity chromatography
L	Length
LBTs	Lanthanide-binding tag
mAb	Monoclonal antibodies
MAD	Multi wavelength anomalous dispersion
MBP	Maltose-binding protein
MES	2-(N-morpholino)ethanesulfonic acid
MME	Monomethyl ether
MR	Molecular replacement
NMP	N-Methylpyrrolidone
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NTA	Nitrilotriacetic acid
N-term.	N-terminus
OcDH	Octopine dehydrogenase

OD ₆₀₀	Optical density, collected at 600 nm
Pbf	2,2,4,6,7-Pentamethyl-dihydrobenzofuran-5-sulfonyl
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PMSF	Phenylmethylsulfonyl fluoride
RP-HPLC	Reversed phase high pressure liquid chromatography
Rs	Residue
RT	Room temperature
SAD	Single wavelength anomalous dispersion
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
Synth.	Synthetic
TBTU	O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate
tBu	Tert. butyl
TED	Tris carboxymethyl ethylene diamine
TFA	Trifluoro acetic acid
TFE	Trifluoro ethanol
TOCSY	Total correlation spectroscopy
Tris	Tris hydroxymethyl aminomethane
Trt	Trityl
UV-Vis	Ultraviolet-visible spectroscopy

Abstract

Optimising His-tags for purification and phasing

During the last decade, several affinity protein tags consisting of short peptides or whole domains have been developed, which are included as an additive protein sequence on the C- or/and N-terminus. Their highly specific strong affinity affords a one-step purification with minimal effect on biological activity. On the other hand, protein tags are often disordered in the crystal structure and complicate protein crystallisation.

Experimental phasing of macromolecules by anomalous dispersion requires well ordered atoms in the crystal lattice. A pre-organised protein tag with metal chelating properties can bind anomalous scatterers in a predefined way and enhance SAD and MAD data quality using synchrotron radiation. As a welcome side-effect such protein tags could adopt stable conformation in crystals and help the crystallisation process. Therefore, the tags have to be ordered in the crystal structure.

Here we report polypeptides synthesised by solid phase peptide synthesis which may prove useful for both metal affinity chromatography and macromolecular phasing as well. A distinct secondary structure in terms of β -hairpin could be confirmed by circular dichroism spectroscopy and NMR. Their chelating properties could be proved by Ni, Zn, Co, and Cu affinity columns.

Fusion proteins made of the Maltose-binding protein fused with our polypeptides were used to check IMAC behaviour under real conditions and to validate crystal growth promotion. First crystal structures from native protein could be solved but shows partly unfolded protein tags. To separate the structural impact of the fusion protein on the tag various linker sequences weretaken into account.

Contents

1	Introduction	1
1.1	Tags for fusion proteins	2
1.1.1	The His-tag	5
1.1.2	Maltose-binding protein	7
1.2	Protein tags in crystal structures	8
1.2.1	Impact on crystal structures	9
1.2.2	Protein tags for crystallisation and phasing	11
1.3	Aim of this thesis	12
2	Methods	15
2.1	Materials	15
2.2	Synthetical methods	15
2.2.1	Peptide synthesis	15
2.3	Biological methods	20
2.4	Chromatographic and spectroscopic methods	22
2.4.1	Purification by RP-HPLC	22
2.4.2	Purification by immobilised metal ion affinity chromatography	25
2.4.3	Determination of secondary structure by CD spectroscopy	27
2.4.4	Determination of secondary structure by 2D-NMR	29
2.5	Crystallographic methods	31
2.5.1	Manual crystallisation	31
2.5.2	Crystallisation supported by robots	33
2.5.3	Structure solution by anomalous dispersion	36
2.5.4	Structure solution by MR	41
2.5.5	Crystallisation, data collection, structure solution and structure refinement of MBP fusion proteins	44
3	Short peptides mimicking protein tags	53
3.1	Peptide engineering	53
3.2	α -Helix forming peptides	57
3.2.1	Mutants of Vps27	58
3.3	β -Hairpin forming peptides	60
3.3.1	Protein G B1 based β -hairpins	61
3.3.2	Mutants of Protein G B1 based β -hairpins	62
3.3.3	Analytical data of β -hairpin forming peptides	63

4	Fusion proteins of MBP with β-hairpin forming peptides	75
4.1	MBP as fusion partner	75
4.2	Fusion of MBP to peptides mimicking protein tags	76
4.2.1	Metal affinity chromatography of fusion proteins	77
4.3	Crystal structures of MBP fusion proteins	79
4.3.1	MBPCG1	79
4.3.2	MBPCG2	84
4.3.3	MBPCG3	89
5	Conclusion	93
6	References	99
7	Appendix	111
7.1	Protecting groups	111
7.2	List of synthesised peptides	112
7.3	Primary and secondary structure of MBP	114
7.4	Crystallographic tables	115

1. Introduction

Crystallography has become a powerful tool of structural research during the last decades, not only in the field of small molecules but also for protein science. After discovery of X-radiation by W. C. Röntgen [66] in 1895 and the work of W. H. Bragg [6] and M. von Laue in 1912 concerning X-ray diffraction in crystals, the foundation in X-ray structural analysis was laid. In the following period many important chemical/biological structures were solved. Because this research was so outstanding many researchers were decorated with Nobel prizes. In 1958 J. Kendrew [47] published the first crystal structure of myoglobin and one year later in 1959 M. Perutz [60] solved the structure of haemoglobin. Therefore J. Kendrew and M. Perutz shared the Nobel prize in chemistry in 1962. In the same year J. Watson and F. Crick [93] were honoured for discovering the double helix structure of DNA. Their work was based on X-ray diffraction images made by R. Franklin. Later on D. C. Hodgkin was decorated with the Nobel prize for medicine in 1964 for her structural research of vitamin B12 and other compounds using X-ray crystallography [21, 9]. All these papers represent milestones in the history of crystallography and substantiate the advancement of this method as a powerful tool in structure analysis. Today crystallography has become much more capable, which approves structural research of huge protein/DNA complexes. In 2009 V. Ramakrishnan, T. A. Steitz and A. E. Yonath [94, 69, 1] shared the Nobel prize in chemistry for structural researches of the ribosome. Since the first diffraction pattern were collected by M. von Laue a huge number of structures have been solved. They delivered insight into both structural environment in general and functional properties of matter. These days structural data bases accrue as fast as never before and scientific publications concerning novel compounds usually desire a structural characterisation by X-ray structure analysis.

Small molecules usually crystallise spontaneously from concentrated solutions, even when they are impure. Their big crystals diffract to high resolution and data can be processed and analysed by almost automated software. Solving the

crystallographic phase problem can be carried out in seconds by direct methods using enormous CPU power. In contrast, in the field of macromolecules like proteins, DNA/RNA or their complexes, highly pure and homogeneous samples, generated by various purification steps, are essential. However, getting well-diffracting crystals is strongly project dependent and solving the crystallographic phase problem of a native protein is impossible in the bulk of cases. Actually all these tasks are hard to overcome. So, several methods have been developed to facilitate or to simplify protein production including genetic modification, protein purification, crystallisation and finally generation and interpretation of X-ray diffraction data. In this thesis a novel class of protein tags is described that might be helpful to purify recombinantly expressed proteins, growing suitable diffracting protein crystals and solving the crystallographic phase problem.

1.1 Tags for fusion proteins

Since the early days of protein structural research purification became a major task in crystallography. To get well-diffracting crystals only samples of highest purity and homogeneity usually lend themselves to form crystals. Contrary to the early days, today a huge pool of sophisticated chromatographic purification methods are established. Moreover genetically modified mutants of native proteins and their overexpression in various cells are common practice. To increase expression rate, solubility of proteins and degree of purity, so-called protein tags were introduced. Several protein tags, consisting of short peptides, domains or whole proteins, can be added as an additive protein sequence at the C- or/and N-terminus for recombinantly produced proteins. In Fig. 1.1 a selection of most important tags is represented showing differences in usage. Facilitating purification is a major task of protein tags and therefore so-called affinity-tags were developed to share their special features with their fusion partner. These tags allow almost one-step purification using a simple and accurate assay of the recombinant protein during purification. Highly specific interactions of antigen/antibody, enzyme/substrate or receptor/ligand were used to bind protein of interest reversible to a chromatographic stationary support. After washing, the protein can be selectively cleaved to get a pure protein sample. Usually protein tags contain in their amino acid sequence a site for proteolytic cleavage. Because of this site tags are easy and specific to remove to retrieve native protein. Commonly protein tags have minimal effects on tertiary structure and biological activity of target protein (Bucher et al. [12]). Protein tags are applicable to a number of proteins expressed in different cells, but in general depending on target protein itself, expression system and application, it is difficult to decide about the best fusion system.

Tag	Length	Position	Matrix/Elution	Typical Use
Albumin-binding protein (ABP)	137	N-term. or C-term.	Albumin Low pH or denaturation (e.g. heat, urea)	Purification, expression, Increased solubility Increased solubility
Calmodulin binding peptide (CBP)	26	N-term. or C-term.	Calmodulin EGTA or EGTA and high salt	Purification
Cellulose binding domain (CBD)	27-189	N-term., C-term., or internal	Cellulose Denaturation (Family I CBDs) or ethylene glycol (Family II or III CBDs)	Purification, Immobilisation
FLAG	8 DYKDDDDK	N-term. or C-term.	mAb Low pH, EDTA, or FLAG peptide	Detection, Purification
Glutathione transferase (GST)	S-211	N-term. or C-term.	Glutathione, Reduced glutathione	Detection, Purification, Increased expression, Increased solubility, Immobilisation
Green fluorescent protein (GFP)	220	N-term. or C-term.	NA	Detection
Histidine-affinity tag (HAT)	19 KDHLIHNVHK- EFHAHAHNK	N-term. or C-term.	Divalent metal (Ni ²⁺ , Co ²⁺ , Cu ²⁺ , or Zn ²⁺) Imidazole or low pH	Purification
Maltose-binding protein (MBP)	396	N-term. or C-term.	Cross-linked amylose Maltose	Detection, Purification, Increased expression, Increased solubility
Myc	11 CEQKLISEEDL	N-term., C-term. or internal	mAb/ Low pH	Detection, Purification

Tag	Length	Position	Matrix/Elution	Typical Use
Polyarginine tag)	(Arg- 5-6 (usually 5) RRRRR	C-term.	Cation exchange resin High pH salt gradient	Purification
Polyhistidine tag)	(His- 2-12 (usually 6) HHHHHH	N-term. or C-term.	Divalent metal (Ni^{2+} , Co^{2+} , Cu^{2+} , or Zn^{2+}) Imidazole or low pH	Detection, Purification, Immobilisation
S-tag	15 KETAAAKFER-QHMDS	N-term., C-term. or internal	S-fragment of RNaseA/Low pH	Detection, Purification
Strep-tag	8 or 9 WSHPQFEK or AWAHPQPGG	N-term. or C-term.	Strep-Tactin (modified streptavidin) Biotin or desthiobiotin	Detection, Purification, Immobilisation
Streptavidin	159	N-term. or C-term.	Biotin, Biotin or denaturation (heat, urea)	Detection, Purification, Increased expression, Immobilisation

Table 1.1: List of most prominent protein tags used for recombinantly expressed proteins. Here size and amino acid composition are shown as well as possible localisations in the fusion protein and typical applications [85].

Often large protein tags increase solubility but have to be removed because of their size. Interaction of fusion partners cannot be excluded both in solution and crystal lattice. In general keeping a large protein tag can interfere with biological functionality. In contrast, very small protein tags mostly do not, depending on the localisation and amino acid composition, interfere with tertiary structure and therefore may not need to be cleaved. They are not as immunogenic as large tags and can be used as antigen for antibody interactions. To conclude, small protein tags seem to be more eligible, in modified form, to create a tag, that supports crystallising.

1.1.1 The His-tag

The His-tag is an amino acid motif in proteins that consists of several sequential histidine residues. Usually the number of residues is in the range of four to twelve, but six is most common. It is known as hexahistidine-tag, 6xHis-tag and by the trademarked name His-tag registered by EMD Biosciences. The tag was invented by Roche and its vectors are distributed by Qiagen. However, the use of this tag is unrestricted for academic users. To avoid royalties against Roche, sequences free for commercial users are available from various distributors. The choice of the end where His-tag is fused to is dependent mainly on both the characteristics of the protein and the methods of tag elimination. Some protein ends participate in protein structure or functionality and some are buried inside the hydrophobic protein core. Consequently fusion is limited to the opposite terminus. For removal of the affinity-tag after purification the tag is customarily equipped with a linker sequence suitable for proteolytic cleavage from the C-terminus. When the tag is fused N-terminal, this site is not needed because of cleavage by exopeptidases. To conclude, this protein tag is used in almost 60% of all recombinantly expressed proteins [24] and thus represents the most frequently used affinity-tag.

The His-tag is closely related to immobilised metal-affinity chromatography (IMAC), invented by Porath et al. [61] in 1975. But not before Hochuli invented immobilised nitrilotriacetic acid [37] as stationary support for purification IMAC become popular. This purification method based on complex forming of metal ions with an immobilised chelating matrix and amino acid sidechains with electron donor function. Sulphur, nitrogen and oxygen are suitable donors but histidine residues are the major targets. Histidine shows strong interactions and binds with micromolar affinity by forming coordination bonds and at once is easy to remove by imidazole displacement or low pH. A short affinity tag built of five to six residues binds with remarkable selectivity to such transition metal ion chelated chromatography matrices. In the early days only native protein containing consecutive histidine residues are purified by IMAC. Later on the synthetic polyhistidine sequence was introduced and retained more efficiently.

Affinity purification using His-tags usually results in relatively pure protein when the protein of interest was expressed in prokaryotic organisms. Purification of

proteins expressed in higher organisms like yeast, mammalian cells and baculovirus-infected insect cells is possible but requires additional purification steps due to a higher amount of unspecific bonded proteins. Several of these proteins are well known and one of them is *FKBP-type peptidyl prolyl isomerase*, which appears in the SDS-PAGE around 25 kDa (SlyD). There are two general ways to overcome this impurities, expression in *SlyD*-deficient *E. coli* strains or use of cobalt-loaded IMAC resins. In contrast to nickel-loaded resins, these resins do not bind *SlyD*.

In the case of purification of recombinant proteins under denaturing conditions his-tagging is the option of choice. Retaining of histidine residues is dependent only on the primary structure of the tag and the IMAC support tolerates high a concentration of urea and guanidinium salts. For elution, low pH instead of imidazole gradient is used. At $\text{pH} < 6$ for Co^{2+} and $\text{pH} < 4$ for Ni^{2+} the histidine residues become protonated and are released from the metal ion.

Several affinity supports providing different coordination spheres have been developed (see Fig. 1.1). Iminodiacetic acid (IDA) is a tridentate chelator which occupies three of the six binding sites in an octahedral coordination sphere of divalent metal ions. Three coordination sites are usually occupied by water molecules and are exchanged with histidine residues during purification. Nitrilotriacetic acid (NTA) and carboxymethylaspartate (CMA, TALONTM) are quadridentate chelators with two valencies remaining for histidine residues. Lowest binding capacity towards histidine residues is implemented by tris-carboxymethyl ethylenediamine (TED), a pentadentate chelator which offers one valency. All these chelators are connected to an agarose bead for immobilisation. These affinity resins share compatibility with many laboratory reagents. In general IMAC supports do not tolerate reducing agents like 1,4-dithiothreitol (DTT) or mercaptoethanol which reduce nickel or cobalt to oxidation state three. Strong chelators usually strip metal ions from the column, thus ethylene diamine tetraacetic acid (EDTA) or ethylene glycol-bis(beta-aminoethyl ether) (EGTA) remove metal ions.

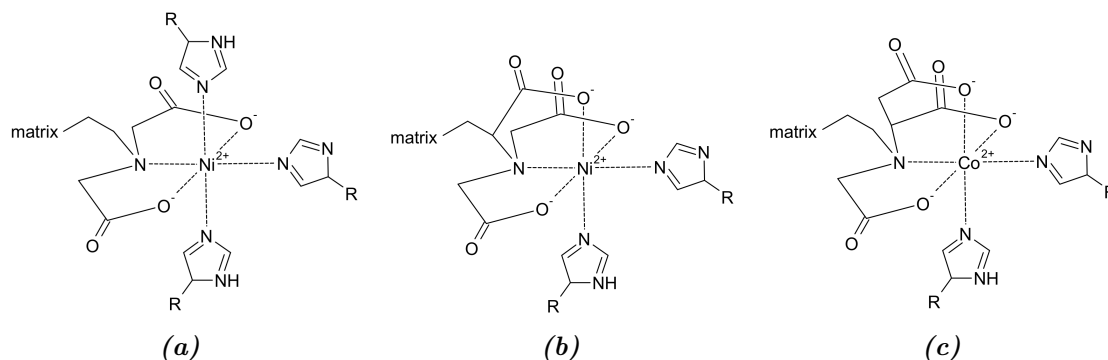


Figure 1.1: Examples of chelators, covalently bound to a solid matrix. A complex with metal ion and histidine residues is shown for: *a*) iminodiacetic acid (IDA), *b*) nitrilotriacetic acid (NTA), *c*) carboxymethylaspartate (TalonTM)

1.1.2 Maltose-binding protein

The maltose-binding protein (MBP) is part of the maltodextrin transport system in *Echirichia coli* which is responsible for the uptake and efficient catabolism of maltodextrins. MBP is a complex regulation and transport system involved in many proteins and protein complexes. It consists of 371 amino acids and has an approximate molecular mass of 42.5 kDa with an ellipsoid shape. MBP is divided in two domains (I and II) with similar shape and secondary structure. The two domains are connected by three bridges forming a deep groove where maltose is bound. A network of hydrogen bonds formed by amino acid side chains and water molecules are involved in the interaction of sugar and protein. Typical for sugar binding proteins are interactions between aromatic side chains in the binding groove and sugar molecules.

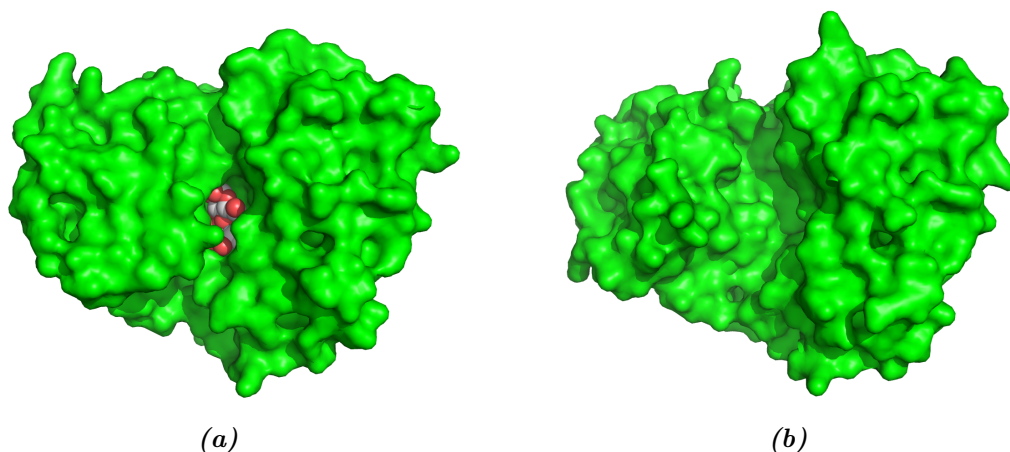


Figure 1.2: Surface of MBP in the (a) open and (b) closed state. The ligand (maltose) is highlighted as a sphere model. Domains I and II move towards each other and close the binding cleft due to ligand binding.

Due to sugar binding domains I and II are movable with respect to each other as highlighted in Fig. 1.2. In the sugar free state, MBP exists in equilibrium between the *closed* and *open* conformations. In the *open form* MBP is elongated and forms a broad binding groove. At the opposite side of the sugar-binding cleft an interface, the balancing interface, is formed. Assimilable to an articulation, in the sugar bound state (*closed form*) the two domain move towards one another. Consequently the binding groove is closed with a deeply fixed sugar. The closing is possible due to three linking bridges connecting domains I and II and broken contacts at the balancing interface. The binding constant of maltose is about 1.6×10^{-7} M. Depending on sugar concentration the open or closed form is present (see Fig. 1.3). Both states have been crystallographically described by Spurlino et al. [79] and Sharff et al. [71].

MBP is in recombinant proteins often expressed in *E. coli* as fusion protein. These complexes usually show an enhancement in solubility (Braun et al. [8], Hammarström et al. [34]). Furthermore it was reported that MBP has chaperone-like

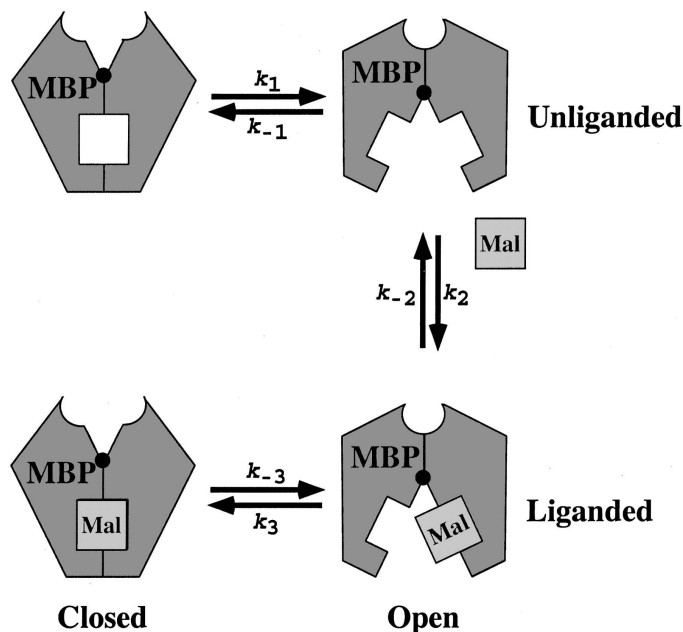


Figure 1.3: Illustration (taken from Telmer and Shilton [84]) of the transformation from *open* to *closed* state of MBP. The unliganded state exists in equilibrium between the closed and open conformations. Once the binding cleft is opened in the unligated protein state the ligand is able to enter. When the maltose (*Mal*) is bound to the sugar-binding site conformational changes in the MBP molecule occurs that close the cleft. All these activities are equilibria and correlate to ligand concentration.

qualities and is able to promote correct folding of its fusion partner (Kapust and Waugh [42], Sachdev and Chirgwin [67]). Because of maltose binding, MBP interacts strongly with an immobilised amylose matrix and can itself act as an affinity-tag for purification. While all other proteins in the crude cell lysate are unaffected by this matrix, MBP binds selectively and can be eluted with maltose to give pure protein. Thereby amylose, fixed in the maltose binding niche, is replaced with free maltose present in the elution buffer. Once the protein of interest is obtained in purified form MBP is usually cleaved off and separated in a second amylose affinity step. Alternatively proteolytic cleavage can be performed with bonded MBP during purification. Therefore, final elution of MBP fusion protein is skipped after washing untagged proteins.

1.2 Protein tags in crystal structures

The C- and N-terminus of a protein are usually the most flexible parts in a protein structure. There are cases where the termini are involved in maintenance of structure or functional activity and therefore restricted in motion, but these termini are usually not tagged. So in the crystal structure electron density of the protein ends is often missing. When a short tag, having no intrinsic endeavour to form a secondary structure, is fused to a protein termini the tag becomes invisible in the density as well. Usually small tags do not contribute to termini flexibility reduction and thus a higher tendency to form well-diffracting crystals. That is why protein

tags are most often cleaved from their fusion partner. But this procedure brings new challenges. Further purification steps can reduce yield of protein and cause protein precipitation due to decreased solubility. Proteases are often expensive and the obtained protein may not be active or structurally intact. Large protein tags behave differently because of their size compared to the fusion partner. However, there are reasonable advantages to keep the fusion protein intact.

1.2.1 Impact on crystal structures

It is known that only molecules having a low degree of structural freedom tend to crystallise due to forming of intermolecular contacts in the crystal. So commonly small protein tags are cleaved from the protein of interest to prevent a higher degree of terminal flexibility. This is the reason for the smaller number of cleaved than kept His-tags in crystal structures stored in the Protein Data Base (PDB). There are various examples of His-tagged proteins, showing a well-defined electron density of His-tag histidine residues in the crystal structure.

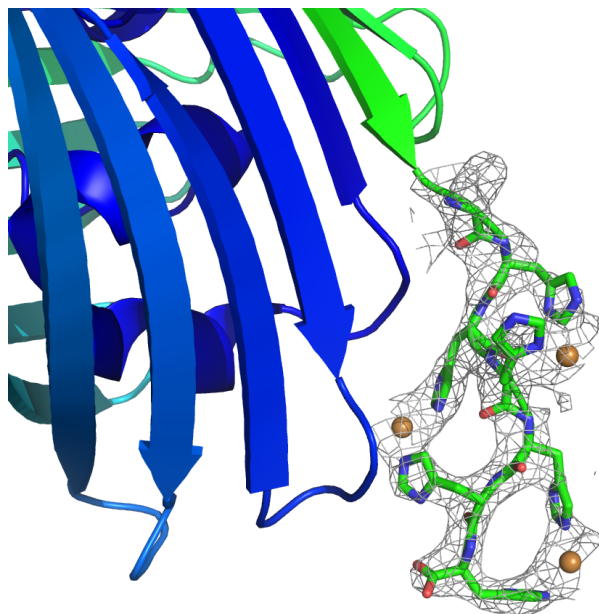


Figure 1.4: N-terminal His-tag region of the bacterial fatty acid transporter FadL from *Escherichia coli* (PDB-ID:1T16). Six histidine residues are chelated by three copper ions. The electron density is drawn as σ_A weighted map contoured at 1.5σ .

One impressive example of a His-tag that is well-defined in the crystal structure is the *Escherichia coli* bacterial fatty acid transporter FadL reported by van den Berg et al. [90] in 2004 (see Fig. 1.4). Here the protein was C-terminal tagged with a His₆-tag and even crystallised in that form. The crystal was grown from a solution containing CuCl₂ at pH 5.5 and thus metal ions that can be chelated by the histidine residues were still present during the crystal growth. So all six histidine residues are pairwise bound to Cu²⁺ ions. One copper ion is involved in a crystal contact too. Due to metal chelation the protein tag flexibility is effectively decreased. Consequently the electron density is well-defined.

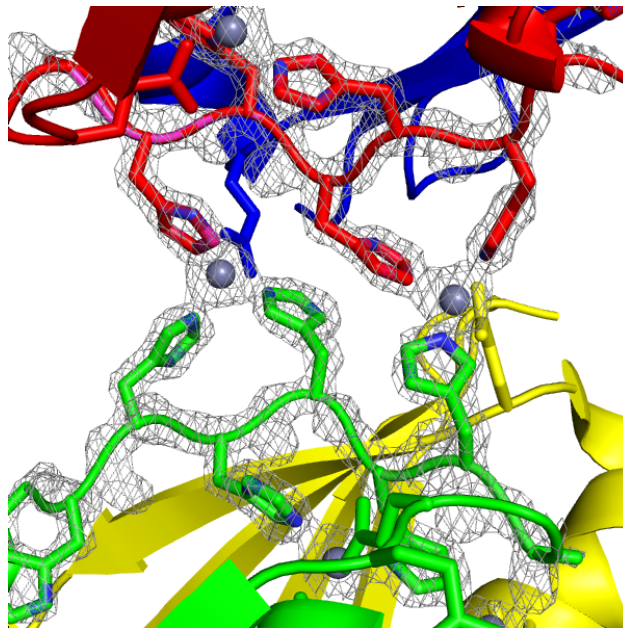


Figure 1.5: A His-tag that is strongly involved in crystal packing. B1 domain of protein L from *Peptostreptococcus magnus* (PDB-ID:1HZ5). Altogether four protein molecules are involved in chelation of two zinc ions by the His-tag. The electron density is drawn as σ_A weighted map contoured at 1.5σ .

Neill Kim et al. [49] presented a fusion protein of B1 Domain of Protein L including a well-behaved His-tag already in 2001 (see Fig. 1.5). In this particular case four protein molecules are linked to a tetrameric arrangement by linking of N-terminal His-tags by zinc ions. Two molecules form a dimer by interaction of three histidine residues each from their His-tags with two zinc ions in a tetrahedral arrangement. The fourth valence of the zinc ion is bonded to E46 of protein molecule three and four by forming the tetramer. Finally H(-7) of molecule three and four interact with H(-7) of molecule one and two again by zinc complexation. So we can see that crystal forming was strongly influenced by histidine-zinc interaction. Furthermore Smits et al. [78] demonstrated that in some cases not only the presence but also the length of the His-tag is important for crystal growing. The crystallisation of octopine dehydrogenase OcDH was induced by a His₅-tag. By using the wild type OcDH as well as a His₆-tagged OcDH they were not able to get crystals at all. A His₅-tagged crystal structure could be solved and shows that the tag sticks into the cleft between domains I and II of OcDH. Thus, a stable conformation due to lower flexibility is caused. Therefore crystal contacts are formed by various interactions of amino acid side chains between two monomers. He speculated that in the case of His₆-tagged OcDH the resulting distance between two protein monomers is too large to form crystal contacts. Similar observations were reported by Tajika et al. [83] who were able to get crystals of protein PH0828 from *Pyrococcus horikoshii* only in the presence of a His-tag. The influence of various protein tags on crystal structure was analysed by Carson et al. [14] who made five constructs of *Pyrococcus furiosus*. They used C-terminal Arg₅-, His₆-, FLAG-, Strep-II- and BAP-tags

as fusion partners and could demonstrate that the tag sequence and its length is crucial for crystal growing. Nevertheless they were able to solve crystal structures of all of these constructs with the exception of the Strep-II-tag. Furthermore Carson et al. [14] could show that His-tags have a negligible effect on the structure of its fusion partner because of the small size. The B factors are slightly higher in the case of His-tagging and disordered tags only effect the structure minimally. So, in most cases His-tags do not interfere with crystal formation [15]. To conclude, small protein tags do not have detrimental effects on crystallisation by nature. If there is sufficient space in the crystal lattice between the protein molecules interactions between tag residues and amino acids in the protein surface are not repulsive, small protein tags may not disturb crystal growing. Because of size of small tags and their flexibility voids in the crystal lattice are occupied by the tags.

Large tags like MBP, GST or GFP behave differently than small protein tags. Various papers report dramatically increased solubility. Furthermore, these proteins are well-structured and keep their appearance in the fused form. Many crystal structures of proteins fused with a large protein tags are reported (Kobe et al. [50], Liu et al. [54], Ke and Wolberger [44]). In these reports the structures contain *E.coli* MBP on their C-terminus as the fusion partner of small proteins. So the ratio between the tag and the protein of interest is large and therefore the tag dominates the crystal structure. Most crystal contacts are made between two protein tags. Amino acid residues from the protein of interest interact little with the protein tag. However, it was demonstrated [50, 54, 44] that although the protein tag dominates the crystal lattice, biologically relevant and active multimers could be formed. The reason for that was a long flexible linker sequence that separates the tag and the protein of interest but does not hinder or participate in the formation of multimers. These linkers are of special interest. If they are too long and floppy the motion is too high and crystal growing is hindered. On the other hand, a too rigid connection between the protein and the tag excludes the formation of biological active multimers. In the case of C-terminal fused MBP, a poly-Ala linker or linkers with a high amount of Ala have been often used to connect MBP's helix H20 and the protein of interest. A different group of linkers are sequences assessable for proteolytic cleavage. These sequences usually contain four to six amino acid residues and are therefore relatively flexible. There are also cases where the protein of interest and the protein tag are of equal size (Ullah et al. [87]). In that case it is possible that only contacts between the protein and the protein tag are generated. Therefore the protein is fully isolated and forms a monomer.

1.2.2 Protein tags for crystallisation and phasing

Growing well-diffracting crystals from large biomolecules is a major task in crystallography, a bottleneck that hinders progress in projects to get structural and functional information. As shown before, small protein tags have small effects on crystal growing, but a small protein tag which supports crystallisation and phasing is not available so far. Indeed some crystal structures show an impressive interaction

with their fusion partner, but these interactions strongly depend on the project and are not predictable. In general, a new tag that promotes crystallisation should be possible to design [14]. If it succeeds to make the small protein tags more rigid and let them interact with each other to generate crystal contacts, small protein tags would promote crystal growing more often.

As mentioned before large protein tags tend to interact with each other more strongly than small tags. They can dominate the crystal structure and thus help the protein crystal to grow [50, 54, 44, 87]. Once the crystal is formed one can collect SAD data using the anomalous signal of native sulfur of methionine and cysteine residues, but commonly they do not include strong anomalous scatterers by nature. On the other hand, large protein tags can be used as search model for molecular replacement. Strategies to introduce external anomalously scattering atoms are co-crystallisation and soaking. However, experimental phasing of macromolecules by anomalous dispersion of heavy metals requires that these atoms are well-ordered in the crystal lattice. Large protein tags usually have no facility to bind metal ions. In contrast, the His-tag strongly binds transition metal ions but has no intrinsic secondary structure. A pre-organised protein tag equipped with metal chelating properties can bind heavy metals in a predefined way. Metal ions from the solvent are incorporated in similar positions in the crystal lattice and give rise to high occupancy. A better crystal packing using co-crystallisation compared to soaking derivatives should enhance SAD and MAD data quality. Thus the anomalous scatterer should be contained in the solution for crystal growing.

In the field of NMR tags containing lanthanides are well-established. Such a lanthanide-binding tag (LBTs) is a short polypeptide that is able to bind metal ions and was reported by Franz et al. [28]. These peptides are derivatives from calcium-binding motifs. Martin et al. [55] were able to solve the structure of ubiquitin with a SAD experiment using a double-lanthanide-binding tag in 2000. Two Tb^{3+} ions were bound to this double tag and a data set at 1.1 Å wavelength was sufficient to solve the structure. Of course, one could measure a MAD data set as well because Tb has absorption edges at 1.4238 Å, 1.5025 Å and 1.6500 Å, respectively. This example shows that incorporation of anomalous scatterers into a protein by covalently bound tags is possible.

1.3 Aim of this thesis

The aim of this project is to develop a new protein tag which combines features for biochemistry and crystallography (see Fig.1.6). The novel protein tag should allow one-step protein purification using the established affinity system IMAC, known from His-tag purification and influence solubility positively. Because of its secondary structure and therefore stable conformation in crystals the tag should help the crystallisation process. Using chelated heavy metal ions in the crystallisation conditions metal induced protein contacts should appear that again help crystal forming but also can be used to determinate experimental phases with single- and

multi-wavelength anomalous dispersion. Thereby the major focus of this project is on crystallography and experimental phasing.

A new protein tag shall be designed which allows one-step protein purification using common IMAC technique, has positive influence on solubility of fusion protein, promotes or at least does not hinder the protein during the crystallisation process, adopts a stable conformation in the crystal by forming a secondary structure and finally enables experimental phasing by heavy atoms. By modification of the established His-tag by incorporation of other amino acids, short polypeptides mimicking tags shall be designed and synthesised by solid phase peptide synthesis. Folding in aqueous solution shall be analysed by CD spectroscopy and NMR. Their purification properties using IMAC techniques have to be determined as well as crystal structures of metal complexes. Peptides which show stable secondary structure in aqueous solutions and good metal ion chelating properties as well shall be used for fusion protein.

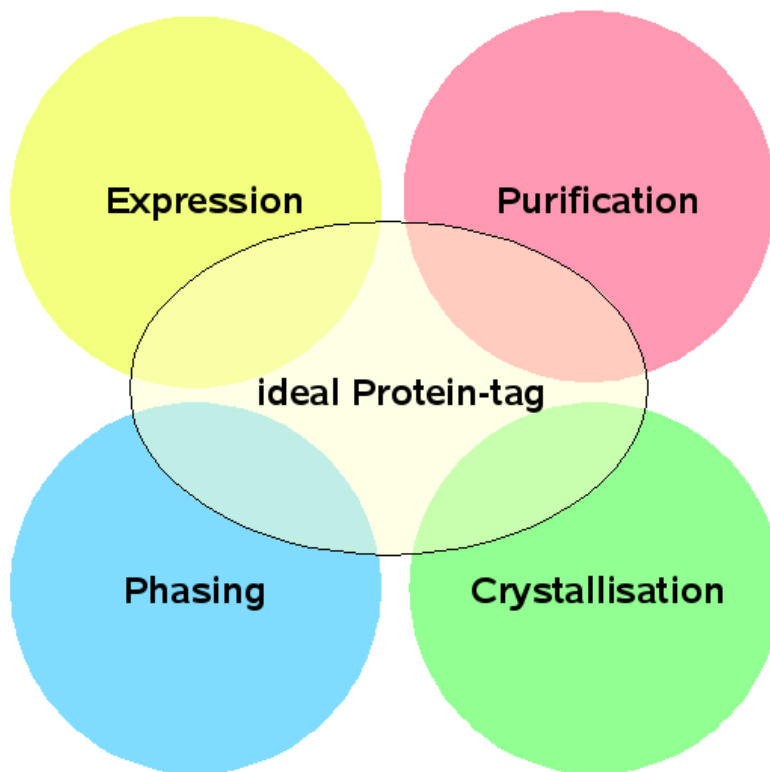


Figure 1.6: Illustration of topics that should be promoted by an ideal protein tag. In an ideal case protein overexpression in various hosts should not be hindered by the protein tag. Different purification methods should benefit from the tag resulting in a pure sample. The tag should support the crystallisation process and introduce anomalous scatterers in the crystal structure to get experimental phase information. Some of these topics are still bottlenecks in the project progress. An ideal protein tag should help to overcome these problems.

2. Methods

The present thesis comprises both synthetic and analytical work. For sample preparation microwave supported solid phase peptide synthesis (SPPS) and protein overexpression in *E.coli* were performed. Purification was realised combining RP-HPLC and IMAC. CD-spectroscopy, NMR methods, and crystallography were carried out for analysis. Main focus was structural research by X-ray protein crystallography.

2.1 Materials

All used chemicals were obtained from Baker B.V. (Phillipsburg, USA), Fluka (Buchs, Swiss), Grüssing (Filsum, Germany), Hampton Research (Aliso Viejo, USA), IRIS Biotech GmbH (Marktredwitz, Germany), Jena Bioscience (Jena, Germany), Merck (Darmstadt, Germany), Nova Biochem (Darmstadt, Germany), Roth (Karlsruhe, Germany), Sigma-Aldrich Chemie GmbH (Steinheim, Germany), and VWR (Darmstadt, Germany) in purity grade *pro analysis*. The used water was deionised and bi-distilled.

2.2 Synthetic methods

2.2.1 Peptide synthesis

Solid state peptide synthesis (SPPS) was introduced by Merrifield [57] in 1963 and is still the method of choice to produce polypeptides by chemical synthesis. In contrast to classical synthesis reaction products are fixed to an insoluble polymeric support (polystyrene resin). Between resin bead and polypeptide chain various linking sequences are available due to diverse chemical environments during synthesis. This leads to the fact that reactions do not take place in solution anymore and the reaction product can be separated easily from crude reaction mixture just by filtration. After several washing and filtration steps the next reaction step can

be carried out. Since reaction protocols are still optimised to a high yield ($\gg 95\%$) numerous coupling steps, manually or automated, can be carried out to finally get chain lengths of up to 100 amino acids.

In peptide synthesis, microwave irradiation has become a tool to optimise reaction conditions by completing long peptide sequences with high yields and low degrees of racemisation. During amino acid coupling microwave radiation catalyses reaction through the increase in temperature and alternating electromagnetic radiation to which the polar backbone is adapted incessantly. These interactions prevent polypeptide aggregation due to heavy thermal motion and increases the yield while coupling two amino acids. For a standard coupling procedure the reaction vessel was irradiated with microwaves for 300 s at a power of 20 W and a maximum temperature of 75 °C. More sensitive amino acids (Arg, Cys, and His) were irradiated with a decreased power. In a first step the vessel was just heated to 50 °C for about 120 s. Afterwards the vessel was irradiated for 240 s at 25 W power and 50 °C. Microwave radiation is able to support the N_α -Fmoc deprotection reaction as well. Therefore a procedure of 180 s irradiation at 40 W power and 75 °C was performed.

The amino acids are N-terminal as well as side chain protected to prevent side reactions during polypeptide synthesis. Protecting groups for amino acid side chains have to be orthogonal. During the step of N_α -Fmoc deprotection in an alkaline media these protecting groups have to stay intact. By changing the pH of the media to acidic the protecting groups can be cleaved to give an unprotected polypeptide. Thereby several protecting groups can be cleaved in slightly acidic media already whereas other groups are still stable. So the deprotection can be moderated by the pH of the media. Typical protecting group for amino acid side chains are listed in Tab. 7.1. For carboxyl and hydroxyl functions in the amino acids Asp, Glu, Ser, Thr, and Tyr the tBu group was used for protecting. Side chains containing amine functions as in the amino acids Lys and Trp are protected with the Boc group. In the case of the amino acid Arg the Pbf group was used. The extensively used amino acid His was protected with the Trt group. The side chains of Asn and Gln as well as all hydrophobic amino acid side chains were left unprotected. All these protecting groups are stable in basic media but can be cleaved in strongly acidic media using TFA to give the free amino acid side chains.

Reactions

Coupling of amino acid and solid support is the first step in solid phase synthesis. Therefore a ternary base (DIPEA) is added to an N-Fmoc amino acid to give a tritylester but keep the N-terminal protection intact. Having a pre-loaded resin solid phase peptide chemistry in general is divided into three major steps: deprotection of the peptide's N-terminus, coupling reaction and final washing. This cycle is repeated until the last amino acid of the polypeptide chain is added. Chemistry taking place during single coupling steps is shown in Fig. 2.1 to Fig. 2.3. In the early days trityl resin was used, introduced by Merrifield himself, and are shown exemplarily.

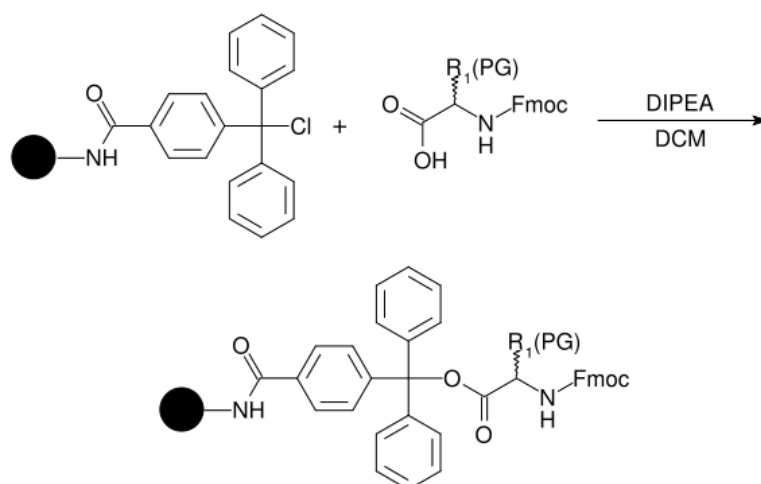


Figure 2.1: Coupling of protected amino acid to trityl linked resin, emblematic by black circles, to get a “loaded resin”. By using the ternary amine DIPEA in dry DCM a tritylester is formed.

For amino acid activation the base piperidine is used in a two-step reaction to cleave the Fmoc group. Piperidine is able to abstract the acid Fmoc β -hydrogen to form a carbanion intermediate. This compound which eliminates CO_2 in an E1 reaction gives the polypeptide and dibenzofulvene. Free dibenzofulvene is quenched by excessive piperidine. These reactions are supported and accelerated using polar solvents like NMP or DMF, thereby NMP is most favoured.

For chain elongation the TBTU/HOBt method is the method of choice to couple N-Fmoc protected amino acids. A base (DIPEA) dissolved in NMP is added to increase nucleophilicity of HOBt. An active ester of amino acid_{n+1} and HOBt (see Fig. 2.3) is generated intermediately which is nucleophilically attacked by the free amino group of amino acid_n to form an amid bond and elongate the peptide chain by one residue.

Taking a closer look to amino acid activation (see Fig. 2.4), the intermediately formed acyluronium derivative is in balance with its educts. Due to internal cyclisation and final base catalysed epimerisation, the equilibrium has to be pushed to the educt side by adding excessive HOBt.

Formation of pre-loaded resin

Wang and 2-Chlorotryl resins (*100-200 mesh*) were obtained already pre-loaded with the first Fmoc-coupled amino acid. They release the peptide with a C-terminal carboxyl group. Unloaded Rink-Amid resin (*100-200 mesh*), which release the peptide with an amid function, is not available already pre-loaded. In the author’s experience Tyr residues show a tendency to support crystal packing because of hydrophobic and aromatic interactions. Furthermore Tyr is easily detectable in the UV-Vis spectrum. So Tyr was used as pre-loaded amino acid for the Rink-Amid resin. The resin was pre-swelled in DCM for about 30 min. Fmoc protection was eliminated with 20% (v/v) piperidine in NMP for about 20 min and finally

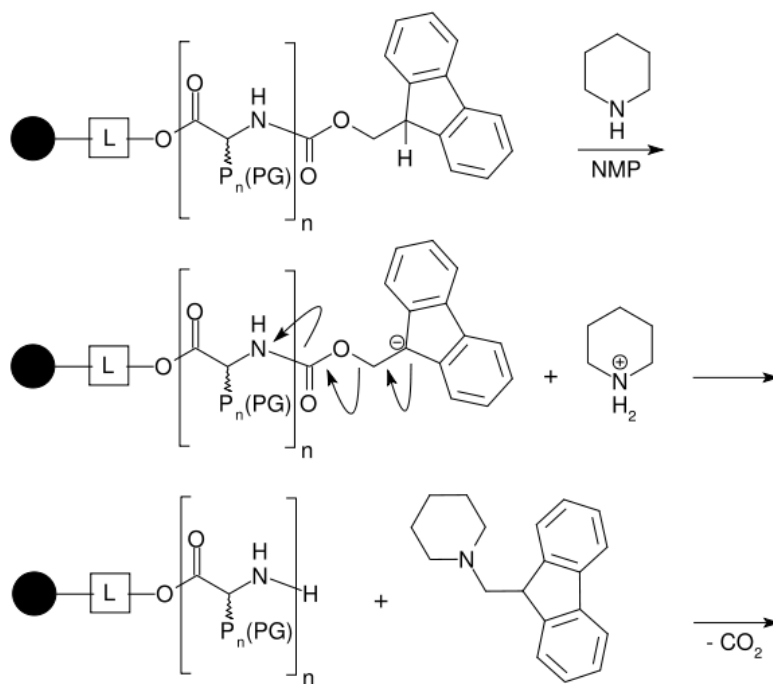


Figure 2.2: The Fmoc protecting group has to be removed for the next coupling step. Therefore the N-Fmoc group can be cleaved by adding 20% (v/v) piperidine dissolved in NMP. An α -hydrogen is abstracted to give a carbanion that stabilises by forming dibenzofulvene and free CO_2 . The free amino group is now prepared for coupling. Solid support and linker are symbolised by a black circle and a boxed L .

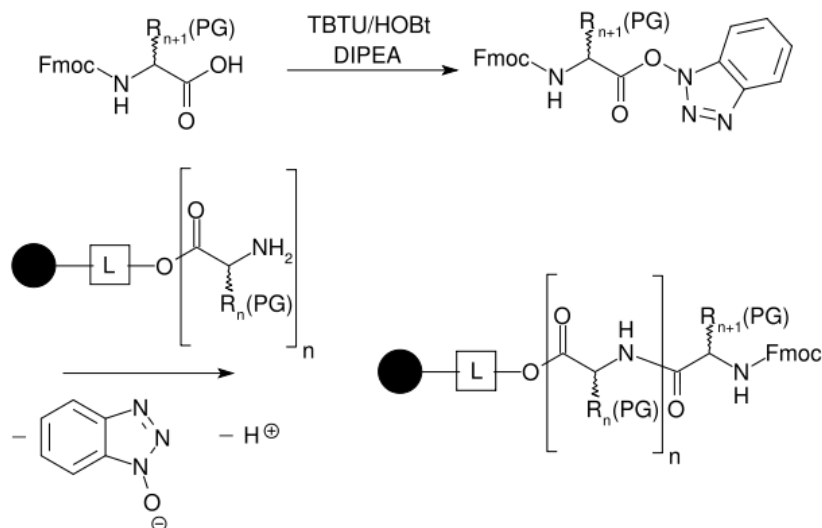


Figure 2.3: Coupling of two amino acids to elongate the polypeptide chain. A N-Fmoc and side chain protected amino acid was activated by the reaction with HBTU/HOBt to give an intermediate ester. By eliminating the O-benzotriazolium group the peptide bond is formed.

thoroughly washed with DMF. A pre-mixed solution of 5 eq. amino acid, 5 eq. HBTU and 5 eq. HOBt, all solved in NMP, was added to 10 eq. DIPEA and left to stand for 10 min. This mixture was added to the resin and reacted for about 1 h. The

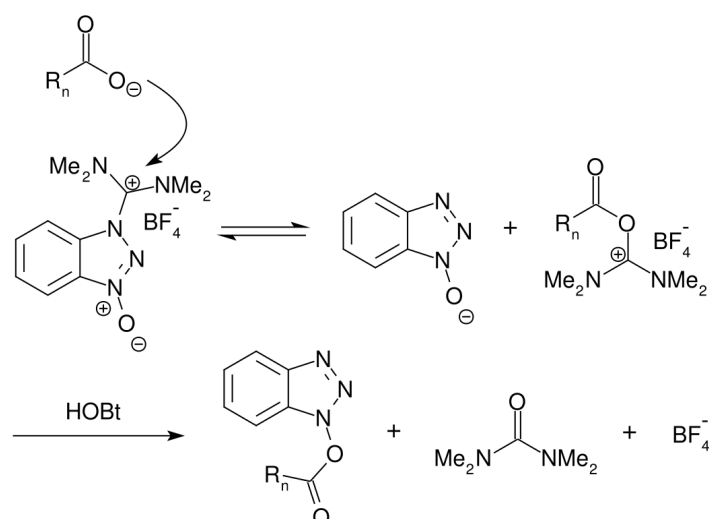


Figure 2.4: Equilibrium reaction of the acyluronium derivative. The activated amino acid can be coupled with the polypeptide chain.

coupling procedure was repeated three times to complete the rate of yield. Finally five washing steps each were carried out using NMP and DCM. All reactions were performed in acid-resistant syringes equipped with filters to prevent resin spilling.

Synthesis procedure

Polypeptides used in this thesis were synthesised using an automated Microwave Peptide Synthesiser (CEM Liberty) with standard fluorenylmethoxycarbonyl chemistry. Pre-loaded resin (0.1 μM , 100-200 mesh) was pre-swelled in DCM for about 30 min. After solvent removal the resin was transferred into the reaction vessel and washed with NMP three times. 7 ml Fmoc-deprotection solution consisting of 20% (v/v) piperidine in NMP was added. The reaction vessel was irradiated with microwaves for 3 min (40 W, 75 °C) followed by washing with NMP. The deprotection-reaction was repeated to complete reaction. For chain elongation 2.5 ml of an 0.2 M amino acid resolved in NMP was added. The vessel was irradiated for about 5 min (20 W, 75 °C). This procedure was repeated for every amino acid starting at the C-terminus. After synthesising the full chain length the reaction vessel was thoroughly washed with DMF and DCM. The peptide cleaving procedure is strongly depending on resin and linker characteristics. For Wang and Rink Amid resins usually a solution of 95% (v/v) trifluoroacetic acid (TFA), 2.5% (v/v) water and 2.5% (v/v) triisopropylsilane is sufficient to cleave the polypeptide from the resin. Some resins (2-Chlorotrityl resin) are not as acid resistant and the cleavage can be performed with solutions containing 1% (v/v) TFA. However, the crude peptide solution was evaporated and resolved in toluene to remove traces of TFA. Finally, the peptide was precipitated in cool ether, decanted and lyophilised.

2.3 Biological methods

All fusion proteins of MBP with various peptides were overexpressed and purified in the group Dr. Stefan Becker at the Max-Planck-Institute for Biophysical Chemistry in Göttingen.

Cloning strategy

The cDNA coding for MBP was amplified by PCR with the N-terminal primer MAL3 and the C-terminal primers listed in Tab. 2.1, resulting in sequences encoding MBP containing the C-terminal "tag genes" CG1, CG2, and CG3, respectively. These three PCR fragments were cloned into the commercial T7 promotor expression vector pET32a (see Fig. 2.5). The fusion protein MBPCG4 was generated by introducing the mutation G371P into MBPCG3 resulting a proline linked MBP-tag-fusion protein.

Table 2.1: Oligonucleotides used as PCR primers for PCR to generate the MBP mutants CG1, CG2, and CG3.

Fragment	Sequence
CG1	5' CAG GGA TCC TTA CAC TGT GTG CCA AGT GAC GTG ATG TGG GTG GTA CAC AGT GAC CTT GGT GAT ACG AGT CTG CGC 3'
CG2	5' CAG GGA TCC TTA CAC TGT GTG CCA AGT GAC GTG ATG TGG GTG GTA CAC AGT GAC ACT TGA ACC GCT AGA TCC CTT GGT GAT ACG AGT CTG CGC 3'
CG3	5' CAG GGA TCC TTA GCC GTG GAC AGT CCA GTG CTT GTG AGC TGG GTG ATA TGT CCA ATG TCC CTT GGT GAT ACG AGT CTG CGC 3'
MAL3	5' GTC GAA TTC CAT ATG AAA ATC GAA GAA GGT AAA CTG GTA ATC TGG ATT AAC 3'

Overexpression and Purification

Sequence verified clones were transformed into the *E.coli* expression strain BL21(DE3) and grown up in LB/Amp (Luria-Bertani) medium up to $OD_{600}=0.7$. Protein expression was then induced with 0.5 mM IPTG and cells were harvested 7 hours after induction at an OD_{600} between 4.5 and 5. After centrifugation the cells were re-suspended in lysis buffer containing 20 mM Tris/HCl pH 7.4, 200 mM NaCl, 1 mM EDTA, and 0.5 mM PMSF.

After seven cycles of sonication the crude extract was centrifuged for 45 min and 4°C at 48000 g. The supernatant was loaded on a 50 ml amylose resin column at 4°C at a flow rate of 1 ml/min and washed with lysis buffer until the baseline was reached. The protein was eluted with lysis buffer supplemented with 10 mM maltose. Ten 8 ml fractions were collected. Elution fractions were checked by SDS

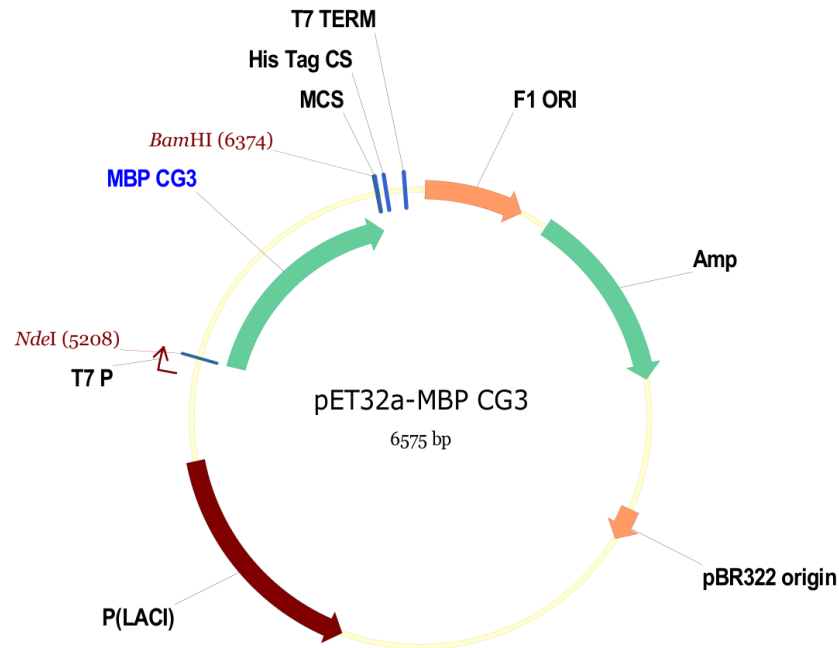


Figure 2.5: Plasmid pET32a constructs used for expression of C-terminal MBP-tag mutants. The cDNA coding the fusion protein was introduced between *NdeI* and *BamHI*.

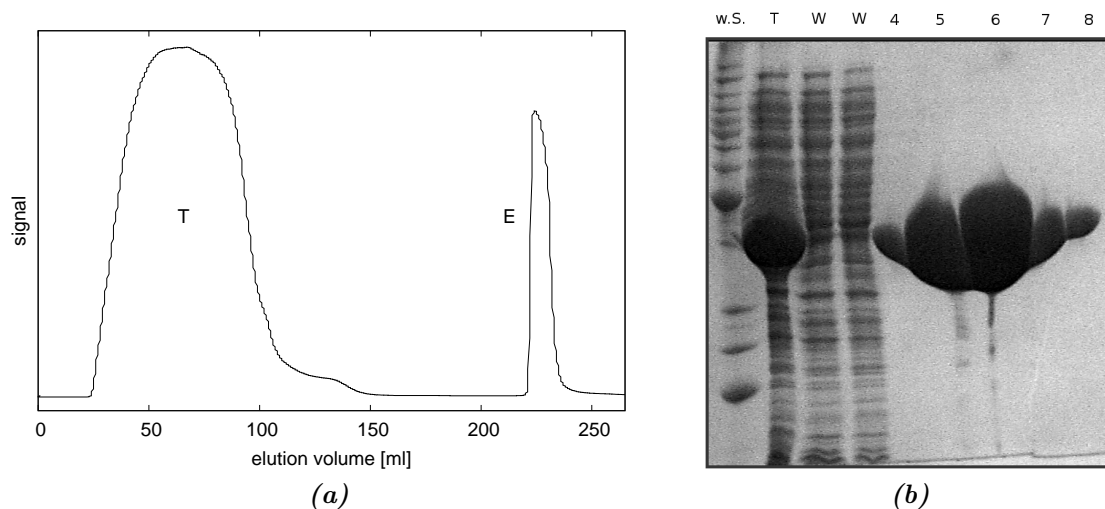


Figure 2.6: Amylose resin purification step: *a*) Chromatogram containing the flow through fraction **T** and the elution fraction **E**; *b*) SDS gel of fractions. Molecular weight standards **w.S.**, flow through fraction **T**, wash fractions **W**, elution fractions **4** to **8**.

polyacrylamide gel electrophoresis on a 12.5% SDS-gel. Fractions 4 to 8, containing the bulk of the protein (see Fig. 2.6), were pooled and dialysed against buffer A (20 mM Tris/HCl buffer pH 8.0, 25 mM NaCl) for ion exchange chromatography. After dialysis the protein was loaded onto a 5 ml DEAE-FF Sepharose columns with a flow rate of 2 ml/min and washed with 100 ml buffer A. The protein was eluted with a gradient from 25 to 500 mM NaCl in buffer A. The eluate was collected in

1ml fractions and analysed by a 12.5% SDS gel (see Fig. 2.7). Fractions 11 to 23 were pooled and dialysed again against buffer A.

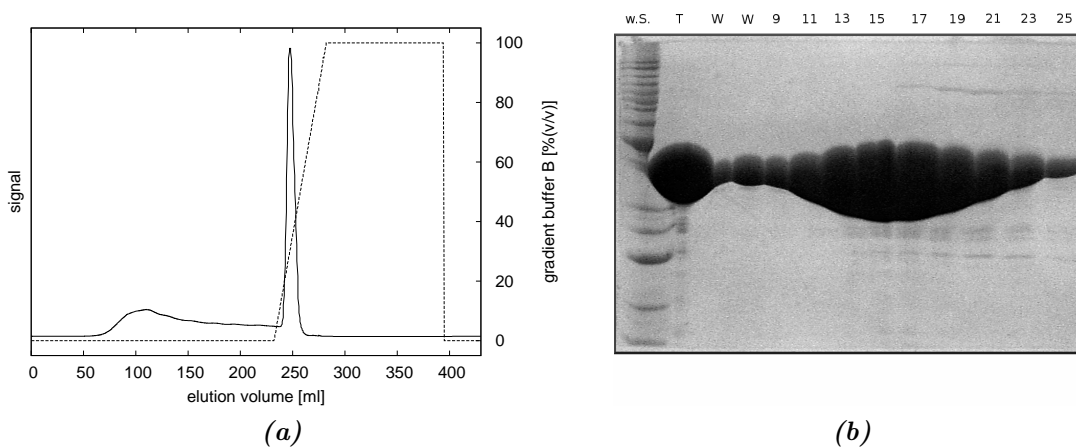


Figure 2.7: DEAE-FF Sepharose purification step: *a)* Chromatogram of DEAE-FF Sepharose column with gradient indicated as dashed line. *b)* SDS gel of fractions. Molecular weight standards **w.S.**, flow through fraction **T**, wash fractions **W**, elution fractions **9** to **25**.

To improve resolution of this anion chromatography step the purification was repeated on a MonoQ 10/10 column with a linear gradient from 25 to 500 mM NaCl in buffer A. In Fig. 2.8 the chromatogram shows several peaks following the major elution peak. SDS-PAGE analysis (see Fig. 2.8) showed that apparently pure protein was eluted in all these peaks, indicating multiple conformations of the eluted MBP. To proceed with uniformly folded MBP only the fractions of about the first half of the major peak (fractions 31 to 34) were combined. Further crystallisation trials showed that only this fraction could be crystallised (see Sec. 2.5.5) indicating protein of highest purity and conformational homogeneity. A gel filtration as final polishing step was performed to improve homogeneity. The pure protein solution was dialysed against 10 mM MES buffer pH 6.2 and concentrated to 20 mg/ml.

2.4 Chromatographic and spectroscopic methods

In this section methods for purification and spectroscopic analysis of the peptides are presented. Synthetic peptides were purified by High Performance Liquid Chromatography (HPLC) and Immobilised Metal Affinity Chromatography (IMAC). To get structural information in aqueous buffered solution both Circular Dichroism (CD) spectroscopy and two-dimensional Nuclear Magnetic Resonance (NMR) were carried out.

2.4.1 Purification by RP-HPLC

High Performance Liquid Chromatography (HPLC) is today a common and powerful technique in instrumental analysis. In a common reverse phase HPLC

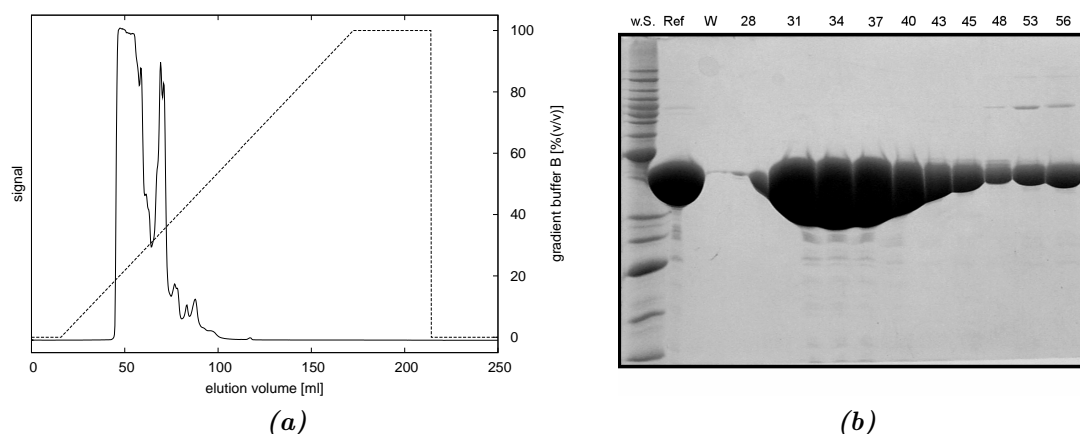


Figure 2.8: Anion chromatography purification step on a MonoQ 10/10 column: *a*) Chromatogram. The gradient is indicated by a dashed line. *b*) SDS gel. Molecular weight standards **w.S.**, sample before column **Ref**, wash fraction **W**, elution fractions **28** to **56**.

application compounds are separated between a hydrophobic stationary phase and an aqueous mobile phase. This interaction is modified by various concentrations of organic solvents added to the liquid phase. Suitable ways in sample purification are the so-called isocratic or gradient technique, where the proportion of buffer A and B is steady or concentration of organic buffer B is continuously increased during purification. Common columns are tubular with an internal diameter of 4.6 mm for analytical or even larger for semi-preparative applications and packed with silica particles. Those beads are characterised by particle and pore size. For peptides particles of around 100 μm particle size and 5 μm pore size are sufficient. In contrast, proteins are much more bulkier and would occasion clogs in the system. So particle size and pore size raised to 300 μm and 10 μm , which induce less system pressure but lower separation efficiency. Modern stationary material is short-term resistant to extreme pH values and generally made up of hydrophobic alkyl chains, sometimes further modified with aromatic, acidic or alkaline functionalities. Those aliphatic chains differ in length (usually C4, C8, and C18) with respect to hydrophobic properties of the analyte. Routinely C8 and C18 material is used for peptide and protein purification. Buffer A by convention is the aqueous solvent. Suitable solvents for buffer B are acetonitrile, methanol, and propyl alcohol. For charged biomolecules always a buffer that determines the pH and thus the total charge is needed. Trifluoroacetic acid (TFA), formic acid, acetic acid, and PO_4^{3-} buffer in a concentration of 0.1% (v/v) are used to improve the chromatographic peak shape. More exotic buffer compounds are heptafluorobutyric acid (HFBA) and ammonium hydroxide.

Peptide purity was determined using analytical PR-HPLC and later on improved by using this method semi-preparatively. A modular Shimadzu HPLC LC-20A prominence system consisting of a DGU-20A₃ degasser and two LC-20AD pumps, realising a high pressure gradient system, supported by a CBM-20A control unit were used. Sample injection was realised by a SIC-20AC autosampler equipped with a 2 ml injection loop for semi-preparative usage. A multi-wavelength

diode-array detector SPD-M20A collecting full UV-Vis spectra in the range of 190 to 800 nm simultaneously was used for detection. Automated sample collection was not implemented in the system and therefore purified samples had to be collected manually. Both analytical and semi-preparative RP-HPLC columns are applicable. For analytical purpose columns with 250×4.6 mm dimensions and 300 ODS $5 \mu\text{m}$ pore size (C18 material) or 120 C4 $5 \mu\text{m}$ pore size for low hydrophobic samples were used. To handle semi-preparative samples up to 40 mg a 250×10 mm 300 ODS $5 \mu\text{m}$ column with pre-column was used. All columns were obtained from MZ-Analysentechnik, Mainz.

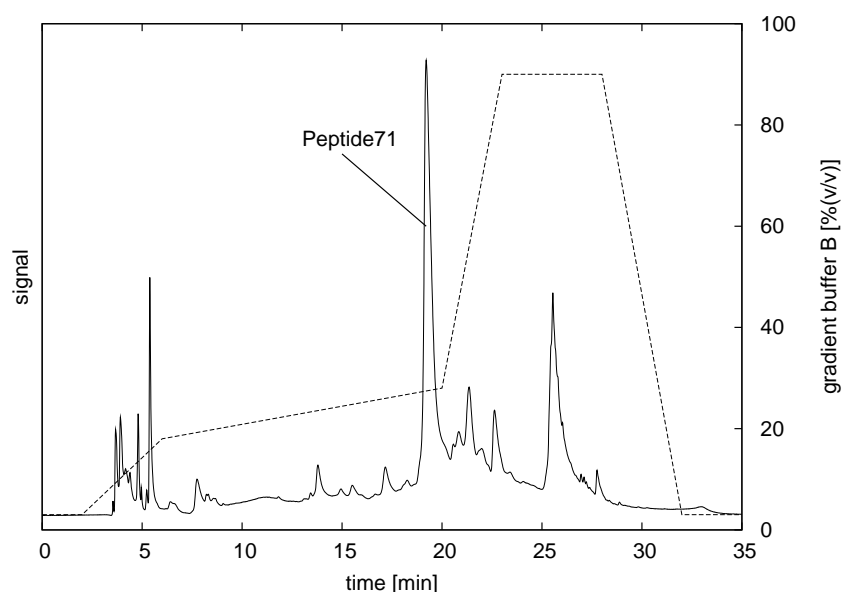


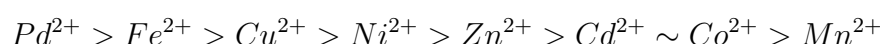
Figure 2.9: Semi-preparative RP-HPLC chromatogram of Peptide71 at 280 nm. Signals at low concentration of buffer B represent cleaved protecting groups whereas the major peak is Peptide71. The gradient (dashed line) was first increased very fast to 18% buffer B and further slowly increased to 28% buffer B to improve separation.

Buffer A consists of 99.9% water and 0.1% (v/v) TFA and organic buffer B was made of 89.9% (v/v) acetonitrile and 0.1% (v/v) TFA mixed with 10% (v/v) water, to prevent volume contraction and thermal effects. For purification of polypeptides TFA was established as additive to form ion pairs. Already small amounts of this strong organic acid are sufficient to keep the whole sample in the same protonation state at approximately pH 1. This acid usually forms highly soluble salts with amino acid side chains and therefore increases the solubility of the polypeptide. Surplus acid can be removed easily by evaporation. As an overview analysis a quick increasing gradient method starting at 3% to 90% buffer B in almost 30 min was carried out and shows major sample features. The majority of peptides were eluted when the gradient reaches 30% organic buffer. In the analytical mode small volumes (20 μl of a low concentrated sample) were injected to sharpen the chromatograms, although in the semi-preparative mode (see Fig. 2.9) up to 40 mg crude peptide solved in 2 ml buffer A were purified. If the the sample is poorly soluble in buffer A organic

solvents like methanol or acetonitrile should be added. Alternatively buffer A can be replaced by organic solvents.

2.4.2 Purification by immobilised metal ion affinity chromatography

For affinity purification beaded agarose supports can be derivatised with chelating groups like iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA) to immobilise transition metal ions. This technique was introduced by Porath et al. [61] as immobilised metal affinity chromatography (IMAC) in 1975. Once resins are prepared it has to be “loaded” with the desired ion. Suitable metal ions, sorted by their binding strength are:



The resulting resins are easy to distinguish because of typical colours. In general both batch methods, gravity-flow columns, and standard chromatography applications are available. Packed into columns or chromatography cartridges, 1 ml of such resins provides for purification of up to 20 mg of crude protein solution. For upscaling pre-packed 5 ml cartridges as well as loose support material to prepare large-volume cartridges are available for up to 100 mg protein solution. Finally, purity and amount of protein can be assessed by SDS-PAGE.

IMAC is compatible against high concentrations of salt and certain denaturants, so samples in various starting buffers can be used. Nevertheless chelators and reducing agents are not compatible due to decreased chelation of metal ions. Elution and recovery of captured His-tag fusion proteins from IMAC support is accomplished with an increasing concentration of imidazole up to 300 mM or a decrease to pH 2.5. However, most proteins are sensitive to high pH variations, so imidazole is the most commonly used elution agent. Unspecific binding of proteins is decreased by a low imidazole concentration (20 mM) during the washing step.

Metal ions used for IMAC show different binding properties during purification. Nickel is the most widely available metal ion and provides high chelating efficiency, but shows non-specific binding of native histidine residues. Cobalt is more specific and often used when high purity is a primary concern. Nevertheless cobalt and nickel are problematic metal ions because of their cancerogenicity and toxicity. However, the release of metal ions in the wash and elution buffer is weak (Ni-NTA \sim 1 ppm, Schäfer et al. [68]). Affinity of histidine to copper is stronger than to cobalt and nickel. Consequently its high binding affinity is coupled with poor selectivity. A rather weakly affine metal ion towards histidine residues is immobilised zinc. However, zinc is non toxic and flexible in coordination. It is found in tetrahedral as well as octahedral coordination, depending on the ligand.

A modular Shimadzu HPLC LC-20A prominence HPLC system was used for preparing IMAC. The component arrangement was equal to HPLC purification (see Sec. 2.4.1), but unloaded 1ml affinity columns obtained from Qiagen and GE

Table 2.2: Compositions of different IMAC cartridges.

	GE Healthcare		Qiagen	
	HiTrap-FF	HiTrap-HP	NTA	Super-flow
Support	Highly cross-linked spherical agarose, 6%			
Bead diameter	90 μm	34 μm	60-169 μm	
Chelating matrices	IDA	IDA	NTA	
Recommended flow rate	1 ml/min			

Healthcare were used (see Tab. 2.2). Free resin was loaded with 1 ml of 1 M sulphates of Ni^{2+} , Zn^{2+} , Cu^{2+} , and Co^{2+} . The buffer composition for washing and elution is shown in Tab. 2.3.

Table 2.3: Buffer compositions for IMAC purification.

	Wash buffer	Elution buffer
Buffer	50 mM NaH_2PO_4	50 mM NaH_2PO_4
Salt	300 mM NaCl	300 mM NaCl
Elution compound	20 mM imidazole	500 mM imidazole
	Adjust pH to 8.0 using NaOH	

Cells are harvested *via* centrifuging and the resulting cell pellet is resolved again and lysed by either physical stress or detergents and enzymes. The received lysate contains a crude mixture of tagged and host specific proteins. His-tags bind best to IMAC resins in near-neutral buffer conditions. It is common to use loading buffer containing PO_4^{3-} buffer at pH 8.0 and 300 mM sodium chloride to load the sample on the support. This loading buffer was not used in this work. The peptide and protein samples were loaded using the wash buffer which typically contains 10 to 20 mM imidazole. This low concentration of imidazole is necessary to remove proteins with native histidine residues that bind not specifically. This procedure may cause a leakage of recombinant protein and lower the total yield but increases the purity of the tagged protein. For elution of tagged protein an elution buffer containing PO_4^{3-} buffer pH 8.0, 300 mM sodium chloride but 500 mM imidazole was used. After loading the sample with washing buffer, a gradient consisting of washing and elution buffer (see Tab. 2.3) is able to elute tagged proteins. This procedure can be carried out by a continuous or stepwise gradient. Fig. 2.10 shows the performed standard stop-flow elution method. When the sample is loaded the solvent flow is stopped for 3 min to increase the binding rate of the peptide. After 12 min washing all impurities and unspecifically bound molecules are removed. Concentration of buffer B is increased to rise the imidazole concentration from 20 mM (0% buffer B) to 332 mM (65% buffer B). During this period elution of peptides or protein could be monitored using a UV-Vis diode-array-detector. Each time a full spectrum from 190 nm to 800 nm was measured. The signals of choice for detection are absorption

bands by aromatic residues around 270 to 280 nm, as shown in Tab. 2.4. Amino acids Tyr and Trp show high absorption coefficients and are often present in native proteins. Unfortunately impure imidazole itself shows a strong absorption in the range of 190 to 300 nm and thus generates a strong background signal. Consequently the sensitivity is decreased through a low signal to noise ratio.

Several problems may occur during purification. The affinity of the His-tag and thus final purity of the protein is related to the tag length. Logically short tags elute more easily than longer tags. So, one has to play with various lengths of His-tags and both C- and N-terminal constructs to find an optimum. In the case of incomplete elution of the protein a denaturing step or stripping the metal ions has to be achieved. An additional solution of 8 M urea or 6 M guanidine hydrochloride will denature the protein but result in pure protein samples. Stripping off chromatographic support can be achieved by buffer A containing additional 50 mM EDTA. This procedure can be easily monitored by changing cartridge colour. The resulting elute contains a high amount of imidazole which does not interfere with most downstream applications and therefore does not need to be removed. If removal of imidazole is necessary for further applications like sensitive enzyme assays, simple dialysis, precipitation, or ultra-centrifugation are sufficient.

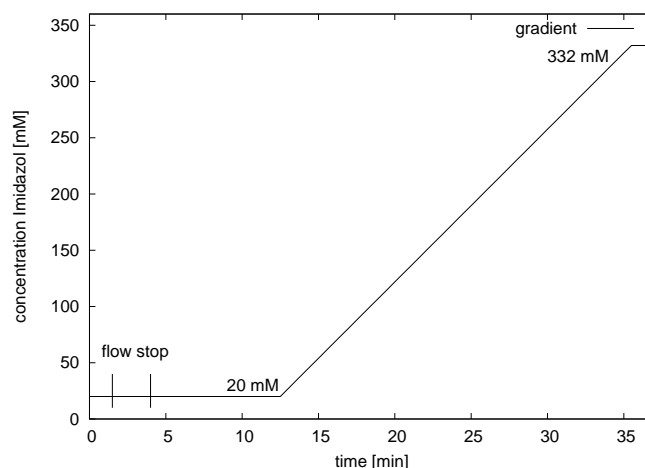


Figure 2.10: A stop-flow method is used for IMAC to increase the peptide yield. After loading the sample the flow is stopped for 3min followed by washing. For purification the concentration of imidazole is slowly rose from 20 mM to 332 mM.

2.4.3 Determination of secondary structure by CD spectroscopy

In structural biology circular dichroism spectroscopy can give detailed insights into tertiary structure and into the function of biological systems. Thereby proteins can be studied under native conditions whose size is not accessible for NMR and those who do not crystallise to well diffracting crystals. So CD has become increasingly recognised as a valuable technique for addressing structural and functional questions.

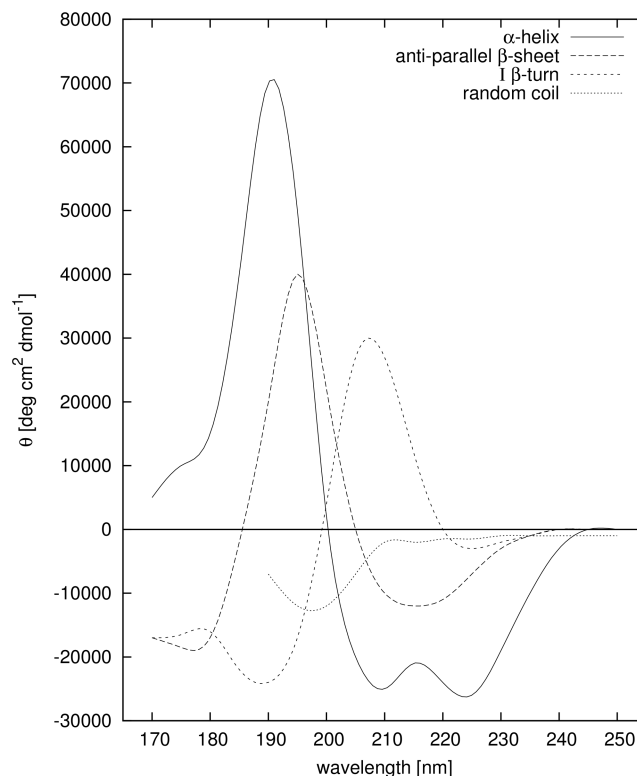


Figure 2.11: Typical CD spectra of the major secondary structures showing characteristic shapes and magnitudes (Davidson and Fasman [23]). The overall content of secondary structures of a protein can be calculated as sum of individual signals.

CD spectroscopy is a low resolution, non-destructive structural technique and can describe overall structural features, but it cannot determine which specific residues are involved. In contrast to NMR and X-ray crystallography it is a much less demanding technique both in terms of sample and time requirements. With CD instruments, also known as spectropolarimeter, the so-called ellipticity is measured as a function of wavelength. Thus, differences in the absorption of left-handed polarised light versus right-handed polarised light which arise due to structural asymmetry cause this effect. Circular polarised light is emitted and interacts with the dissolved sample in a way that a CD signal is observed when the chromophore is chiral. The chromophores of interest in proteins and polypeptides include the peptide bonds (absorption below 240 nm), aromatic amino acid side chains (absorption in the range of 260 to 320 nm, see Tab. 2.4), and disulfide bonds (weak broad absorption bands centred around 260 nm). In general CD signals strongly depend on the structural environment of these chromophores and therefore show characteristics in magnitude and wavelength as reported by Kelly et al. [46]. So different secondary structures give rise to specific CD spectra in the far UV, as illustrated in Fig. 2.11. A fully unstructured protein environment (random coil) generates a weak positive signal at 212 nm ($\pi \rightarrow \pi^*$ transition) and a low negative signal at 195 nm ($n \rightarrow \pi^*$ transition). In contrast β -strands in proteins are characterised by a negative signal at 218 nm ($\pi \rightarrow \pi^*$ transition) and a positive signal at 196 nm ($n \rightarrow \pi^*$ transition).

A helical environment is characterised by a positive signal ($\pi \rightarrow \pi^*$ transition) at 192 nm and two negative signals, at 209 nm ($\pi \rightarrow \pi^*$ transitions) and at 222 nm (red shifted $n \rightarrow \pi^*$ transition). Thus, by analysing the CD spectra of protein in the far UV range the overall signal is interpreted as the sum of fractional multiples of each structural type and the approximate fractions can be calculated. Conformational changes whether a ligand is bound and losing all or rather a portion of its secondary structure due to high temperature can be monitored as well.

Table 2.4: Absorption maxima and coefficient ϵ of aromatic amino acids are shown.

Amino acid	Absorption maximum [nm]	Absorption coefficient $\epsilon[\frac{1}{M \times cm}]$
Phe	257	200
Tyr	274	1400
Trp	280	5600

A Jasco J-810 CD-spectrometer and Jasco Spectra Manager™ software was used for data collection and analysis (see Tab. 2.5). Usually CD spectra of proteins are measured under native conditions in buffers with low absorption in the range below 200 nm. So measurements were made in 20 mM sodium phosphate buffer at pH 5.5 to compare CD and NMR data using 1 cm, 5 mm, and 1 mm cuvettes depending on peptide concentration. Therefore concentrations between 1 mM and 1 μ M of the peptides were used. All spectra were corrected for buffer background using buffer scans without protein. Water has a strong absorption effect and therefore the sample chamber was purged with dry nitrogen. The temperature was controlled using a thermostat.

Table 2.5: CD spectra method.

scan speed	100 nm/min
time constant	0.5 s
bandwidth	1 nm
accumulations	3
temperature [°C]	2-80

2.4.4 Determination of secondary structure by 2D-NMR

Analysing polypeptides and proteins by nuclear magnetic resonance (NMR) techniques provide advantages that cannot be covered by other methods. Data can be collected under native conditions to study dynamic processes due to ligand binding or structural changes due to higher temperature. Nevertheless deliberate changes in sample environment including buffer, additives, or salts are possible. Various methods were developed to initiate spin-spin coupling between different atom cores, for long distance coupling and even unbonded neighbours. For determining

polypeptide and small protein structures a combination of two-dimensional experiments, NOESY and TOCSY, has proven beneficial. In the TOCSY experiment, nuclei that are chemically bonded (scalar or J-coupled) give rise to signal. So cross peaks of coupled protons form the same spin system are shown. The signals are divided into groups of coupling networks which help interpreting the data especially in the case of overlapping multiplets. By analysing polypeptides coupling between H_α and H_N can be obtained, and because of its specific shifts, correlate more easily to a certain amino acid. By using the Nuclear Overhauser Effect (NOE) of nuclear spins in a NOESY experiment, signals can be determined which arise from protons that are close to each other in space but not bonded. In this spectrum space correlation *via* spin-lattice relaxation as well as chemical and conformational exchange can be detected. Therefore coupling of H_α and of amino acid n and H_N of amino acid $n+1$ is recognisable (Fig. 2.12). Distances up to 5 Å can be detected. The combination of both methods was introduced by Kessler et al. [48] in 1988 as *chain tracing*. He could show that by superimposing NOESY and TOCSY data in the range of H_N and H_α shifts an assignment of amino acid residues can be realised more easily. So following the signal trace, a full reconstruction of the polypeptide backbone is possible. Besides chain tracing, analysing chemical shifts δ can give a detailed insight in the three dimensional arrangement of atoms. Bundi and Wüthrich [13] and Wishart et al. [95] could show that $C_\alpha H$ conformational shifts strongly depend on structural environment. Differences in shift profiles ($\delta_{C\alpha H} = \delta_{\text{observed}} - \delta_{\text{random coil}}$) are significant for a specific secondary structure. In this manner helical, strand, loop, and random coil regions can be easily identified.

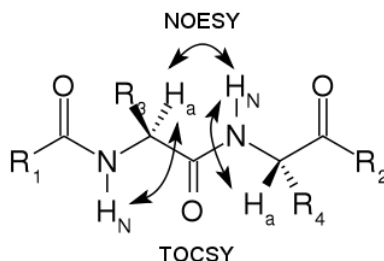


Figure 2.12: Here a segment of a polypeptide chain is shown. The spin-correlation between H_α and H_N of adjacent residues can be detected using NOESY technique. These atoms are not spin-coupled. Data for H_α and H_N coupling from the same residue can be determined in the TOCSY spectrum.

Structural information for some peptides were obtained by two-dimensional NMR using NOESY and TOCSY techniques in 50 mM PO_4^{3-} buffer at pH 5.5 and a peptide concentration of 2 mM (Peptide32 and Peptide56). In the case of Peptide71 the puffer concentration was decreased to 20 mM and the compound concentration was increased to 8.2 mM. Because of the high water content 10% D_2O were added. Total correlation spectroscopy (TOCSY) and NOE spectroscopy (NOESY) spectra were recorded by standard techniques using pre-saturation of the water signal. Solvent suppression allows to record a weak signal of the solute in the background of dominating solvent resonances and was performed with the WATERGATE pro-

cedure. For processing and data analysing the software Topspin v1.3 and Sparky v3.113 were used. The data for Peptide32 and Peptide56 were collected using a Bruker Advance DRX 500 instrument at 500.130 MHz and a temperature of 277 K. The NOESY spectra were recorded using the *noesygp19* pulse sequence with a mixing times of 500 ms (Peptide32) and 200 ms (Peptide56). The TOCSY spectra were performed with the *dipsi2gp19* pulse sequence and mixing times of 80 ms (Peptide32) and 60 ms (Peptide56). NMR data for Peptide71 were collected at a Bruker Advance DRX 400 instrument at 400.130 MHz and a temperature of 277 K. For the NOESY data the pulse sequence *neoesygp19* with a mixing time of 200 ms was used. The TOCSY spectra was performed with the *mlevgp19* pulse sequence and a mixing time of 80 ms.

2.5 Crystallographic methods

2.5.1 Manual crystallisation

Sample preparation is an essential step in every analytical method. In the case of X-ray crystallography the analyte has to be in solid, single crystalline form of sufficient size and quality. Thus growing of single crystals in contrast to solid amorphous and multi crystalline material is crucial for this technique. To overcome this bottleneck sundry methods have been developed and implemented during the last decades.

The basic concept is driving molecule concentration from under- to supersaturation. In this state first order aggregates are formed due to a higher rate of molecule collision. Not before a minimum size is reached these nuclei are stable. Their number affects the number of crystals and should be as few as possible. That is why the process of supersaturation has to be as slow as possible. To slow down the rate of supersaturation temperature during crystallisation is often decreased to 4°C.

Small molecules and proteins differ in their intrinsic tendency to form crystals. Usually small molecules are characterised by well defined Coulomb or dipole-dipole interactions as well as hydrogen bonding. Because of their small surface these interactions lead to only a few but attractive molecule contacts. Therefore small molecule crystals usually formed more easily. In the easiest way crystals grow by evaporation of sample solvent like water, alcohols, ethers, aromatic solvents, halide organic solvents etc. or mixtures of these. Crystal growth caused by decreasing solubility can occur by slow decrease of temperature, the freezing out technique. Nevertheless additives and buffers are usually not necessary. In contrast to small molecules, proteins feature various equal possible interactions on a big surface. Hydrophobic domains as well as polar and charged side chains able to form hydrogen bonds are dominant in proteins. Protein crystals usually have a high solvent content up to 70-80% which is essentially involved in crystal packing. For this and keeping their biological relevancy and activity, crystal growth is limited in aqueous native

solutions. A huge number of parameters are known that exert influence in the growing process that let this task rise to a multi dimensional problem. Depending on sample purity and homogeneity, initial protein concentration, ionic strength, pH, buffer type, viscosity, polarity, hydrophobicity of solution, additives and ligands can be of crucial importance, as reported by Benvenuti and Mangani [2]. The salting out method by using precipitant salt or PEG is a mild way of increasing the concentration of protein. Finally other physical factors as vibration and temperature effect this process.

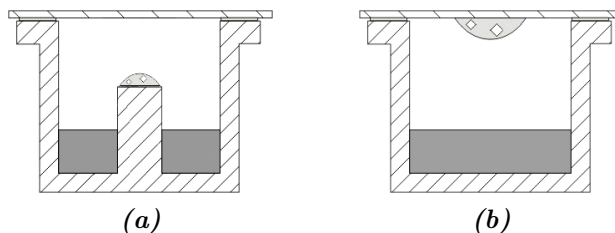


Figure 2.13: The two major methods for vapour diffusion are shown. Both methods share the fact that the drop gets supersaturated due to evaporation of solvent from the drop. *a)* In the *sitting drop* method the drop containing a mixture of protein and reservoir solution is placed on a microbridge. The crystals usually grow on the surface between drop and microbridge. *b)* In the case of *hanging drop* the droplet is placed headlong on the siliconised cover slide that hermetically seals the well.

Because of the high complexity semi-empirical investigations using trial-and-error methods were developed. Nowadays commercially available sparse matrix screens containing systematic gradients of pH or ingredient concentrations as well as wide-ranging conditions of high diversity are used. Crystallisation conditions are premixed and ready-to-use which were already successfully used in other projects. For manual preparation vapour diffusion methods are most widespread and can be carried out as sitting and hanging drop method, as shown in Fig.2.13. A hermetically sealed crystallisation well is filled with aqueous crystallisation solution containing buffer, salt and other additives. A small droplet of a few μl made of crystallisation solution and protein solution is placed either on a siliconised cover slides (*hanging drop*) or on a microbridge (*sitting drop*) sitting in the well. In the case of hanging drop crystals usually grow at the bottom side of the droplet and therefore are not in contact to the cover surface. Contrarily, crystals grow on the surface of the microbridge using sitting drops. Sometimes crystal forming is sensitive to such interactions. However, the droplet mixture is equilibrated against the reservoir solution and thus solvent molecules diffuse through evaporation from the droplet into the reservoir solution. As a consequence the concentration in the reservoir stays constant because of the huge volume compared to the droplet, and the concentration in the droplet increases slowly. By reaching supersaturation in the droplet crystallisation seeds are formed that lead to protein crystals.

Once initial crystals are formed seeding methods can be used to get crystals of sufficient quality. By using streak-seeding nuclei taken from the surface of an initial

crystal with a cat or horse whisker are transferred into a new, pre-equilibrated crystallisation condition. By touching the crystals surface micro crystals are dislodged by friction. Because crystal nuclei are now already present in a highly concentrated protein solution near to supersaturation crystals usually grow to larger size and higher quality. Usually the nuclei are transferred to various new wells one after another. So the concentration of nuclei is reduce to get a few big crystals. In the case of micro-seeding initial crystals are crushed with beads made of a polymeric material. The resulting solution contains microcrystals that can be transferred into new crystal trial containers as well. Again the concentration of nuclei should be reduced. So the nuclei solution should be diluted. By using macro-seeding an initial crystal itself which is free of defects can be transferred into a novel crystallisation well containing fresh solution. Here the initial crystal simply grows to bigger size. If an initial crystal is not accessible one can take a crystal of a similar compound. This cross-seeding named method works well if the compounds are as similar as possible. One can take crystals of a protein mutant or a protein with a different protein tag.

2.5.2 Crystallisation supported by robots

Nowadays the demand in getting crystals is increasing this much because of the high number of projects. Often proteins that crystallise more easily are already structurally resolved and proteins that could not be crystallised are still left. Moreover it is common praxis today to test several constructs of the same protein at the same time, made with different tags an different linker sequences fused to different termini of the proteins or proteins with mutations at the surface or in the active site. Often large complexes of proteins, protein ligand complexes or complexes of protein and DNA/RNA are in the focus of research activities. However, the number of proteins to be crystallised is increased be means of high throughput methods that were introduced in the field of crystallography as well. As mentioned before the step of getting crystals is the most important step and very often a bottleneck in crystallography. By knowing this and the fact that usually trial-and-error methods are used to find suitable crystallisation conditions the requirement to do fast as many trials as possible come up. So crystallisation robots were constructed that prepare these crystallisation trials (see Fig. 2.14). Plates used for preparation by hand usually consists of 24-wells and the well volume is 500 to 1000 μL . To increase speed of preparing and coincidentally decrease financial costs crystallisation plates were introduced consisting of 96-wells. These wells are much smaller and contain almost 100 μL of solution and several places for droplets were implemented into one well. Because of low mechanical requirements the most widely used method in robot supported crystallisation is the sitting drop technique. Here a bundle of syringes controlled by robot can be used to transfer aqueous solutions from a master reservoir into the plates. These reservoirs can contain different stock solutions that are mixed by the robot in the plate well as well as ready-to-use pre-mixed screens. Many different sparse matrix screens are available covering a wide range of very different conditions. Various pH ranges as well as precipitants, salts, organic solvents etc. are components in these solutions. However, it is easier to prepare all these

reservoirs and drops by a robot-controlled syringe direct on a plate and cover the well with a clear cover strip than prepare droplets on a sheet and turn it around to cover the well. So the hanging drop technique appears to be unpractical because preparing the droplets on the well cover is mechanically difficult to realise.

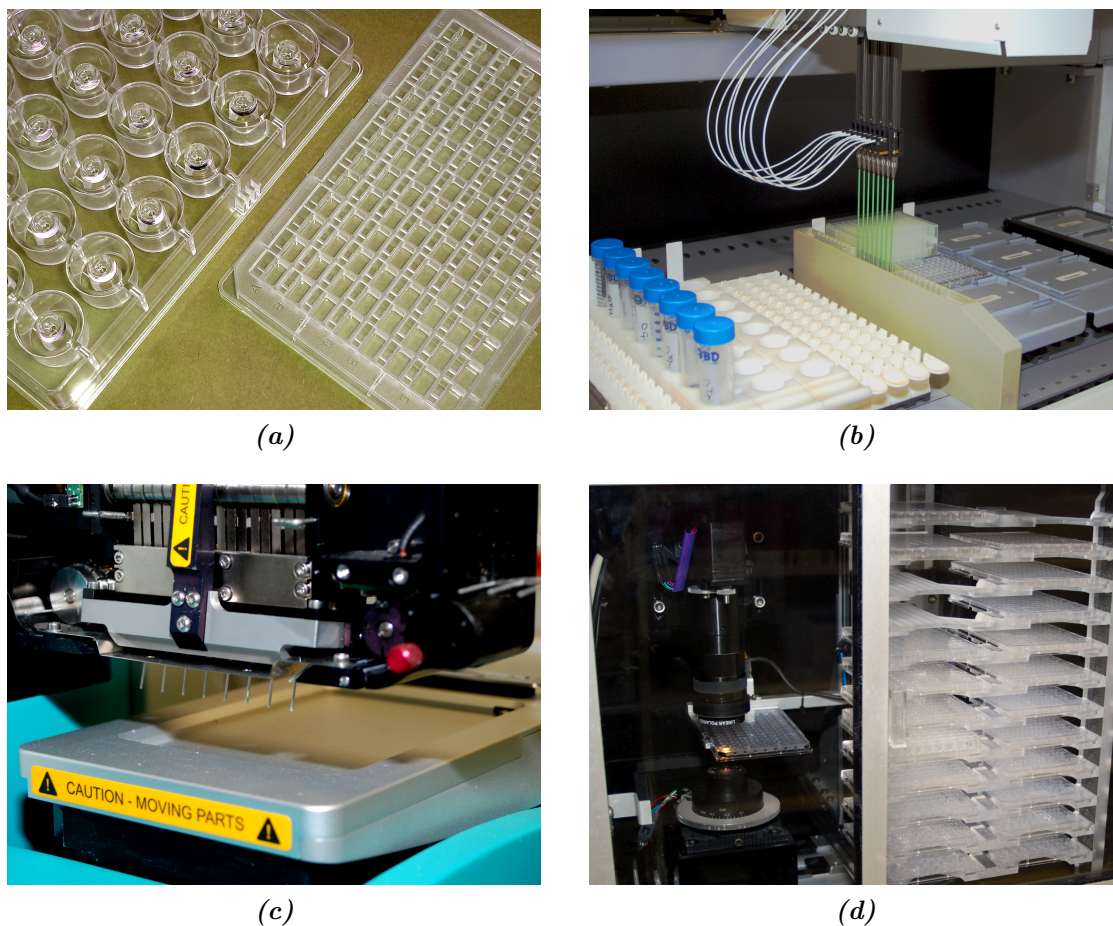


Figure 2.14: Crystallisation robots and equipment are shown. *a.)* In contrast to 24-well sitting drop plates for manually usage, modern high throughput techniques require 96-well plates. These wells (Greiner low profile) contain 100 μL solution, instead of 500 to 1000 μL , and a small area for one or more droplets. The Greiner low profile plate contains space for one droplet. *b.)* The TECAN Genesis RSP 150 is used to transfer crystallisation solutions into the wells by eight syringes. Those solutions can be presented in a pre-mixed form and transferred from a “master block” or mixed on-the-fly. *c.)* Once the wells are pre-filled protein solution and crystallisation condition have to be mixed in a droplet on the well surface. The TTP Laptech Mosquito transfers the protein solution with single-use syringes into the wells and mixes that droplet with an equal volume of reservoir solution. *d.)* The sealed plates are stored in a Formulatrix RockImager. In this tempered box a robot can drive the plates to a camera system for automated imaging. The plates are denoted by barcode and made images are stored in a database.

Crystallisation trials were made using a TECAN Genesis RSP 150 and Limbro 96-well low profile plates. Various sparse matrix screens were used as shown in Tab. 2.6. This robot can pipette 450 μL solution in each of his eight syringes at the

same time and was mostly used to transfer 100 μL of screen solution from the screen reservoir (master block) into the crystallisation wells. For preparing a 96-well plate the TECAN robot needs almost five minutes.

Table 2.6: Sparse matrix screens used for peptide and protein crystallisation. All HT screens consist of 96 conditions whereas the grid screen contains 24 conditions.

Manufacturer	Screen
Hampton	Index HT
	Crystal Screen HT
	PEGRx HT
	PeGIon HT
	SaltRX HT
	Grid Screen Ammonium Sulfate
Emerald BioSystems	JCSG+ suite
	Wizard HT

Once the crystallisation plate is filled with screen solution the protein solution has to be mixed with the reservoir solution and this drop has to be placed on the crystallisation plate as well. This work was done by a TTP Laptech Mosquito robot. Here a small volume (2 μL) of protein solution is presented and transferred column by column into the individual wells. So in each individual well 100 nL of protein solution is pipetted. Before transferring the next column the protein droplet is mixed with an equal volume of reservoir solution from the same well. By doing so the whole plate is filled with protein-reservoir droplets and finally hermetically sealed with a clear cover. To transfer the protein and the reservoir solution and mix it up to get the droplet the Mosquito robot almost needs two minutes. Anyway, the major benefit of using high throughput plates is the small amount of protein solution that is used to find an initial crystallisation condition. For 96 conditions only about 10 μL of protein solution is needed and therefore thousands of different conditions can be tested with a small amount of protein. Finally the sealed plate is stored in an automated imaging system, the Formulatrix RockImager. In a tempered box the crystallisation plates are placed in a plate storage that is able to identify individual plates by barcode and drive them to a camera system. With the help of that camera every single crystallisation trial in the plate can be monitored after a certain time period to check for crystals. The camera uses polarised light and takes photos at various levels in the droplet. Because of the polarised light crystals can be identified easy by extinction. Out of the pictures the best focused picture is elected and stored in a database. By controlling the droplet over a certain time period with pictures growing as well as dissolving of crystals can be monitored. Once a suitable crystallisation condition is found and the crystal was proven to diffract at all manual techniques can be used to increase size and quality of the crystal.

Beside the described robots there are a lot of similar robots available on the market that handle crystallisation trials. Thereby the technique of high throughput

is widely used and screening for conditions to grow crystals speed up crystallisation. Nevertheless high throughput is not a guarantee to get crystals at all. Only the number of trial can be easily increased by several orders of magnitude through usage of robots. At the end in the case one could not find any conditions to grow a crystal of a certain protein, various constructs and an improved purification has to be considered.

2.5.3 Structure solution by anomalous dispersion

In a usual X-ray diffraction experiment the primary photon beam is diffracted following Bragg's Law (see Eq. 2.1) in various directions and give single spots that can be collected on a detector. Thus the diffraction angle θ is a function of the used wavelength λ and distance d between the diffraction planes. So reflexes can be measured when d is a multiple n of the wavelength λ .

$$2d \sin \theta = n\lambda \quad (2.1)$$

Unfortunately only spot intensities I can be measured. As shown in Eq. 2.2 for an electron density reconstruction $\rho(xyz)$ for every point in the volume V of the unit cell the structure factor F_{hkl} as well as phase information $\Phi(hkl)$ are needed.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F_{hkl}| \exp(2\pi i(hx + ky + lz) + i\Phi(hkl)) \quad (2.2)$$

The modulus of the structure factor $|F_{hkl}|$ can be derived from the collected intensities I_{hkl} as shown in Eq. 2.3.

$$|F_{hkl}|^2 \sim I_{hkl} \quad (2.3)$$

The most informative part of the signal, the phase $\Phi(hkl)$, cannot be detected. Kevin Cowtan [20] could demonstrate that phase information from a picture combined with intensities from a second picture result in a reconstruction that looks more like the first picture, which the phases were taken from. As a consequent the image reconstructions is strongly dominated by the phase information. Unfortunately there is no mathematical relation between phases and intensities barring the electron density. So to overcome this problem of missing phases, the so called crystallographic phase problem, several methods were implemented to elicit this information from experimental data.

Nowadays methods using anomalous dispersion through native intrinsic or introduction of anomalous scatterers in the crystal lattice are accessible and become more and more popular. Up to now approximately 20% of all structures stored in the protein crystal database (PDB) are solved by this technique. In principle anomalous

scattering is an effect that occurs at element specific absorption edges and change the normal scattering signal marginally in magnitude. In a centrosymmetric structure Friedel's Law (see Eq. 2.4) is obeyed exactly because the structure factors $F_{(hkl)}$ and $F_{(-h-k-l)}$ as well as the phases are identical but with opposite signs.

$$F_{(hkl)} = F_{(-h-k-l)} \text{ and } \phi_{(hkl)} = -\phi_{(-h-k-l)} \quad (2.4)$$

However, in the presence of anomalous scattering Friedel's Law is broken and therefore the Friedel pairs are not equal anymore and small differences in $F_{(hkl)}$ and $F_{(-h-k-l)}$ occur (see Eq. 2.5). These differences are called Bijvoet differences.

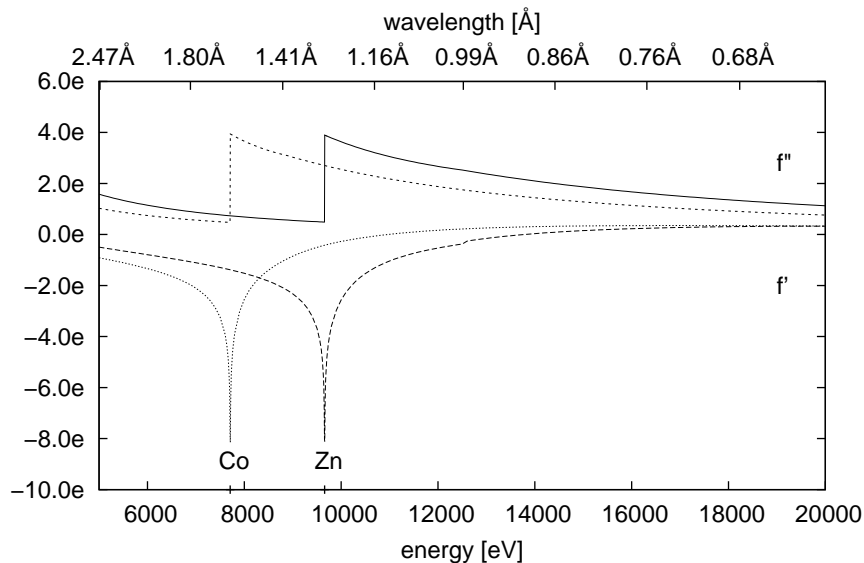
$$\Delta F = |F_{(hkl)}| - |F_{(-h-k-l)}| \quad (2.5)$$

At the element specific absorption edge radiation energy is sufficient to promote inner shell electrons. Radiation is absorbed and emitted back immediately. Thus in the scattering factor f_{anom} is changed (see Eq. 2.6). The effect of anomalous scattering provokes a magnitude shift in the real term f' and a 90° phase shift in the imaginary absorption term f'' .

$$f_{anom} = f_0 + f' + f'' \quad (2.6)$$

The normal scattering f_0 is only dependent on the Bragg angle and decreases at higher resolution (see Fig. 2.15). In contrast, the two terms f' and f'' are not resolution but wavelength dependent and are equal in magnitude in the whole range of resolution. Thus the ratio of anomalous and normal signal increases at high resolution. In theory the Bijvoet differences can be measured more easily at high resolution.

These differences are in the range of 5% or less (1-2% for S-SAD). Taking common collection errors into account Bijvoet differences often disappear in noise. So data have to be collected as precisely as possible but very carefully to prevent radiation damage. For this one needs a stable primary beam, good optics to get an accurate wavelength setting, a fast detector readout and complete, highly redundant data for a high statistical significance. The anomalous signal can be increased by choosing a wavelength in the vicinity of an absorption edge. This can be done at tunable beamlines at synchrotrons. On the other hand one can choose a proper anomalous scatterer that fits to home source radiation. In Tab. 2.7 common anomalous scatterers and their specific f' and f'' values are shown and demonstrate that an ingenious choice of elements give rise to a high anomalous signal using home source Cu- K_α radiation. By changing natural sulfur in the amino acid methionine to selenium as well as introducing of Co ions in the crystal the anomalous signal can be increased. In the case of small molecules the Bijvoet differences, caused by small anomalous scattering of light elements, can be used to analyse chiral compounds and determine their absolute structure.

Figure 2.15: Theoretical absorption spectra of Zn and Co showing f' and f'' .

Element	Mo-K $_{\alpha}$ $\lambda=0.7107$ Å		Cu-K $_{\alpha}$ $\lambda=1.5418$ Å	
	f' [e]	f'' [e]	f' [e]	f'' [e]
C	0.0033	0.0016	0.0181	0.0091
S	0.1246	0.1234	0.3331	0.5567
Se	-0.0929	2.2259	-0.7943	1.1372
Fe	0.3463	0.8444	-1.1336	3.1974
Co	0.3494	0.9721	-2.3653	3.6143
Ni	0.3393	1.1124	-3.0029	0.5091
Cu	0.3201	1.2651	-1.9646	0.5888
Zn	0.2839	1.4301	-1.5491	0.6778
I	-0.4742	1.8119	-0.3257	6.8362

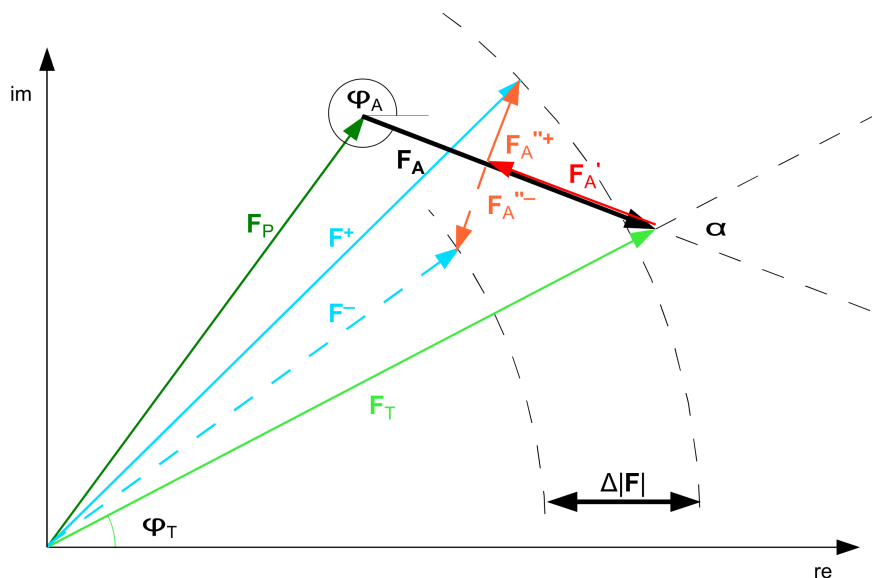
Table 2.7: Values for f' and f'' for common elements [62].

However, an anomalous scatterer has to be present in the crystal lattice whether through natural abundance, chemical modification or additional introduction during crystallisation by co-crystallisation or soaking techniques. Depending on project and type of experiment various anomalous scatterer can be chosen. Anomalous scattering experiments can be carried out using only one particular wavelength, single-wavelength anomalous dispersion (SAD), or sundry designated wavelengths which is called a multi-wavelength anomalous dispersion (MAD) experiment.

In the case of a rapid crystal decay due to radiation damage, an absorption edge not covered by the X-ray beam or an untunable X-ray source the SAD experiment is the method of choice to get experimental phase information. SAD data are collected at a single appropriate wavelength. Often in-house X-ray Cu-

K_α sources are used for analysing proteins containing sulfur or selenium, transition metals in metallo proteins and iodine. However, various heavy atoms are suitable weak anomalous scatterers at Cu- K_α wavelength. One advantage of this technique is the minimisation of time spent in the beam by the crystal, thus reducing potential radiation damage to the molecule while collecting data. In-house data can be collected using an extremely stable beam for a long time to high redundancy without radiation damage because of the reduced power. On the other hand, the phasing power is much weaker than using MAD, due to lower contrast between normal and anomalous signal because of single wavelength usage.

Figure 2.16: Argand diagram representing the Friedel pair F_+ and F_- caused by an anomalous scatterer. The structure factor F_T (total non-anomalous scattering) can be expressed as sum of F_P (structure factor of protein) and F_A (non-anomalous contribution of the anomalous scatterer)



By using SAD and MAD techniques and the approximation that there is only one anomalous scatterer, the phase of the native structure ϕ_T can be described as a sum of the phase from the anomalous scatterer substructure ϕ_A and an angle α that can be estimated from the experimental data (see Eq.2.7). Now the substructure of the anomalous scatterer has to be estimated from its structure factors F_A .

$$\phi_T = \phi_A + \alpha \quad (2.7)$$

Taking Eq.2.7 into account the structure factors F_\pm can be expressed as a sum of structure factors of the total non-anomalous scattering F_T and the non-anomalous contribution of the anomalous scatterer substructure F_A (see Fig. 2.16). Now $|F_+|$ and $|F_-|$ can be described in the following way (see Eq. 2.8) published by Hendrickson et al. [36].

$$|F_{\pm}|^2 = |F_T|^2 + a |F_A|^2 + b |F_T| |F_A| \cos \alpha \pm c |F_T| |F_A| \sin \alpha \quad (2.8)$$

with $a = ((f'')^2 + (f')^2)/f_0^2$, $b = 2f'/f_0$, $c = 2f''/f_0$ and $\alpha = \phi_T - \phi_A$

In a SAD experiment this equation is underdetermined because there are only two data points F_+ and F_- per reflection at one wavelength and one crystal. For the strongest reflections in a SAD dataset Eq. 2.8 can be simplified to Eq. 2.9.

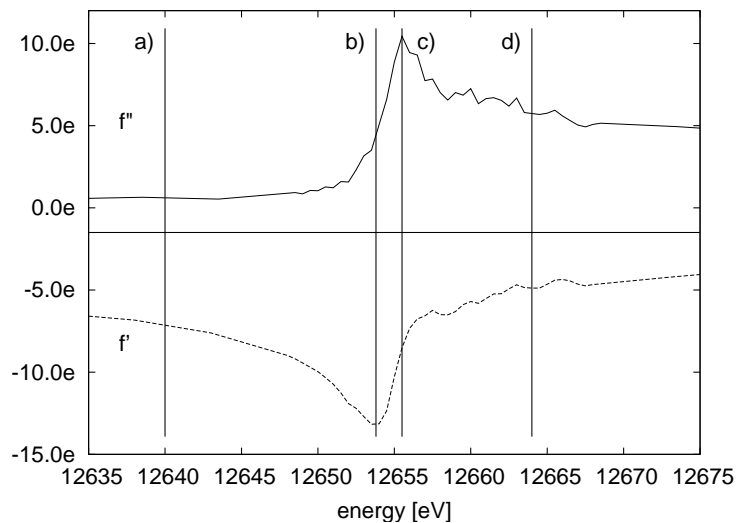
$$|F_+| - |F_-| = c |F_A| \sin \alpha \quad (2.9)$$

By using only the reflections with the largest normalised anomalous differences $|E_A|$ several phases can be estimated. As shown in Fig. 2.16, the angle α depends on $|F_+|$ and $|F_-|$. If $|F_+| \gg |F_-|$ the angle α must be close to $+90^\circ$. In the inverted case when $|F_+| \ll |F_-|$ was observed the angle α must be close to -90° . In the case of $|F_+| \approx |F_-|$ the angle α has to be close to 0° or 180° . Those reflections with $\alpha \approx 90^\circ$ and $\alpha \approx 270^\circ$ can be used as start phases to calculate the anomalous scatterer substructure in the program SHELXD using Patterson or Direct methods. Later on these phases can be improved by density modification and autotracing in SHELXE. Thereby the resolution is not critical as long as the distance between the anomalous scatterers is larger than the resolution. In the case of sulfur in a disulfide bridge of a protein both sulfur atoms can be detected as a single super-sulfur and a high resolution limit of almost 3 to 3.5 Å is sufficient. In general solving a structure by SAD results in two solutions that are inverted with respect to each other. Only one solution gives the correct electron density map and the wrong solution gives nonsense. Therefore always both solutions have to be checked. If the space group belongs to an enantiomorphic pair the space group has to be inverted as well (e.g. $P4_12_12 \leftrightarrow P4_32_12$).

The absolute value of an absorption edge on energy scale depends on the atomic number of the particular element. To a small amount oxidation state and chemical environment influence energy of absorption. However, an absorption spectrum cannot be calculated precisely enough and has to be measured for every crystal. Out of this fluorescence scan accurate wavelengths for data collection can be determined. In Fig. 2.17 a fluorescence spectrum for a protein that contains Se-Met residue instead of Met is shown. So the expected absorption edge was in the range of 0.98 Å (12651 eV). For a MAD experiment two to four datasets at slightly different wavelengths around an absorption edge are measured, caused by large differences in f' and f'' (see Fig. 2.17). Where the absorption f'' gains the maximum the "peak" dataset is collected. At the point of the inflection on absorption curve f'' the value of f' becomes minimal and the second dataset, the "inflection point", is measured. The final dataset is collected at a remote wavelength with a minimal dispersive difference compared to the inflection point dataset. This can be a "high energy remote" and/or "low energy remote" dataset. The major benefit of a MAD experiment is

more phase information. On the one hand the anomalous and the dispersive signal is usable and on the other hand all reflections can be taken into account not only just those with $\alpha \approx 90^\circ$ and $\alpha \approx 270^\circ$. Contrarily due to the usage of three datasets of the same crystal to prevent non-isomorphism the crystal is covering a high radiation dose and often decays rapidly. Therefore MAD datasets are usually not collected to maximum resolution using as little radiation as possible. MAD is widely used for Se-Met data at synchrotrons at approximately 0.98 \AA wavelength. Most protein crystallography synchrotrons beamlines are optimised for this wavelength range. In proteins the amino acid Met can be replaced by Se-Met without losing structural composition and biological functionality. Out of this anomalous diffraction can be measured from an intrinsic scatterer. Of course every other anomalous scatterer can be used that can be fixed in the crystal lattice and its absorption edge can be covered by the radiation.

Figure 2.17: Fluorescence scan of a protein containing Se-Met. Mostly datasets at three different wavelengths were collected for a MAD experiment, but it is possible to collect a fourth dataset too: *a)* low energy remote, *b)* inflection point, *c)* peak, *d)* high energy remote. Some scientists collect all four datasets, mostly low energy or high energy remote dataset are collected. The peak and inflection peak datasets are essential for MAD.



Contrarily to SAD, in a MAD experiment Eq. 2.8 is overdetermined because of six data points from three wavelengths and three unknowns F_A , F_T and α . Therefore a perfect MAD dataset should solve the phase problem and give a perfect electron density map. Again after localisation of the anomalous scatterer substructure the angle ϕ_A and final start phases for the protein can be calculated. Density modification should further improve all phases and lead to an electron density map of the whole protein.

2.5.4 Structure solution by MR

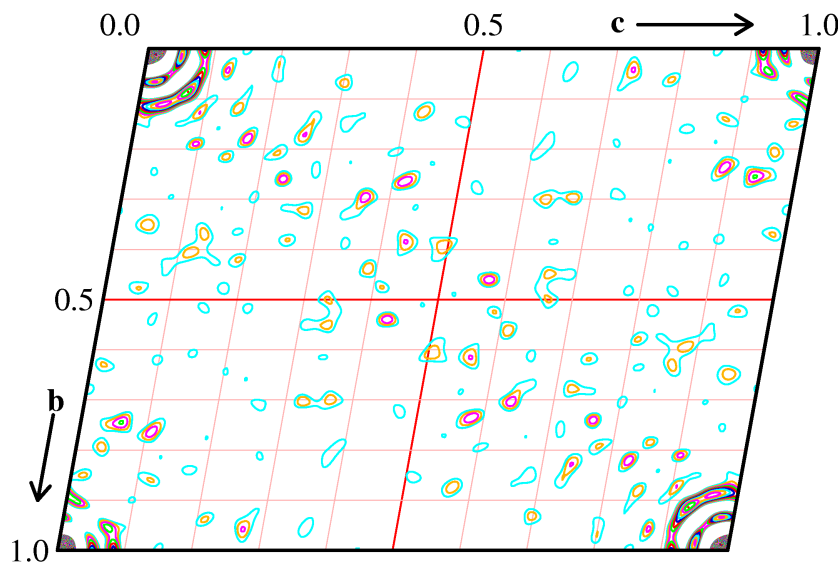
In the absence of anomalous scattering, phasing using experimental data is not possible. If experimental data are collected far away from an absorption edge, so

the phasing power is too small or the included scatterer has an edge not covered by the X-ray source the method of choice is molecular replacement (MR). This method is very popular due to less experimental complexity and therefore used in almost 80% of all crystal structure depositions in the protein crystal database. Only one complete dataset with a resolution of $2\text{-}3\text{ \AA}$, collected at a unique wavelength is sufficient. Due to missing phase information Patterson methods (see Fig. 2.18) are used to calculate a starting map. Here the structure factors F_{hkl} are replaced by the square root of the collected intensities I (see Eq. 2.3). By doing so the equation for the Patterson space $P(uvw)$ is losing its phase term (see Eq. 2.10).

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl}^2 \cos[(2\pi(hu + kv + lw))] \quad (2.10)$$

As a consequence, the function does not result in a classical electron density map but shows all inter-atomic vectors starting from the origin and weighted by the atomic numbers of the atoms, which the vector connects.

Figure 2.18: Detail of a Patterson map. Harker section ($a=1/2$) of a dataset of the maltose binding protein fused to Peptide56 in the space group P1. The Dataset was collected at an in-house Cu-K $_{\alpha}$ rotating anode and a SMART6000 CCD detector.



For a small molecule structure or a protein substructure of an anomalous scatterer the extraction of individual atom positions is possible. In the case of a native protein this vector diagram is too crowded to be interpreted directly but still has characteristics for the protein. This fact can be exploited to superimpose it with a Patterson map calculated of an homologue model. This model should be as homologue in sequence as possible. Some rules of thumb have shown up to be beneficial. The model has to have a sequence identity of at least 30% and an r.m.s. coordination deviation of up to 1.5 \AA . Furthermore the model structure has to be trimmed to its most conserved residues. Solvent, flexible side chains and loops

have to be removed. Also parts of a domain, single α helices, small motifs or even poly-Ala chains are suitable candidates.

However, it is not necessary that the protein of investigation and the model should not have the same space group or a similar orientation in space because all orientation calculations were performed in reciprocal space. So a molecular replacement calculation is divided into two major parts, the rotation and translation search. In a first step Patterson maps from the model data and from the experimental data are calculated and rotated in small increments with respect to each other to align relative orientation in space. Correlation coefficients and nowadays maximum likelihood-based algorithms are used to score this function. An overlap function R in the integration volume U of the Patterson function $P(u)$ from the collected data and the Patterson function $P_r(u_r)$ of the rotated search model is calculated taking the rotation matrix $[C]$ into account (see Eq. 2.11). In an equal orientation the self vectors in the Patterson function become small. A correlation coefficient can be calculated to validate the results.

$$R([C]) = \int_U P(u) \times P_r(u_r) du \quad (2.11)$$

Once a suitable orientation is found the Patterson maps have to be translated within the asymmetric unit to a concerted origin. Therefore by moving the model in real space, the Patterson map of the model is recalculated and compared to the Patterson map of the unknown protein. Since the calculation is done in the Patterson space only the phases of the reflections change when moving the molecule relative to the origin. Again a calculated correlation coefficient can show how well model structure and collected structure are superimposed. Because this trial-and-error method is time consuming and computationally expensive the more systematic translation function $T(t)$ was implemented (see Eq. 2.12). Cross Patterson vectors of the model and the structure of investigation are calculated. Thereby $P_{1,2}(u, t)$ is the set of cross Patterson vectors of two symmetry related molecules and $P(u)$ is a Patterson function calculated from the collected data. This function $T(t)$ becomes maximal if t is equal to an intermolecular vector.

$$T(t) = \int_V P_{1,2}(u, t) \times P(u) du \quad (2.12)$$

Molecular replacement is extensively used in the case of mutants of known structures, fusion proteins involving big protein tags, complexes with various ligands and structures of high structural similarity. Nevertheless the major drawback is the low data quality because a MR solution is strongly dominated by the given model structure. Most popular programs for molecular replacement calculations are Phaser [56], AMoRe [86] and MOLREP [88].

2.5.5 Crystallisation, data collection, structure solution and structure refinement of MBP fusion proteins

MBPCG1

Conditions to grow crystals of MBP fusion proteins were reported by Kobe et al. [50], Liu et al. [54] and Ke and Wolberger [44]. Those are similar to several MBP mutants and fusion constructs, stored in the PDB. Widespread are conditions containing PEG at a pH between 6 and 7 and a low salt concentration. Even though MBP is soluble to high concentration most often a protein concentration between 10 and 20 mg/ml is used.

MBPCG1, the construct which contains Peptide56 without a linker sequence, was concentrated to 40 mg/ml in 20 mM MES buffer pH 6.2 and 50 mM NaCl to get a highly concentrated protein sample. This sample was used for initial crystallisation trials using robots (see Sec. 2.5.2). Initial crystals could not be obtained with INDEX HT, CrystalScreen HT, and Wizard HT screens. Protein MBPCG1 purified by a single anion exchange chromatography step on a DEAD-FF Sepharose column was not sufficiently pure and structurally homogeneous enough. A second anion exchange chromatography step with a MonoQ 10/10 column which has a smaller support bead size and a much higher purification resolution was performed (see Fig. 2.8). However, the protein was again concentrated to 40 mg/ml in 20 mM MES buffer pH 6.2 and 50 mM NaCl. To keep the MBP molecules in a static state by closing the sugar-binding cleft 1 mM maltose was added. Crystallisation trials were carried out using crystallisation robots as described before. In the crystallisation well 100 μ l screen solution was transferred. The droplets were set up by mixing 100 nl protein solution and 100 nl well solution at 20 °C. By using the sparse-matrix screens INDEX HT, CrystalScreen HT, and Wizard HT initial crystals could be grown. In the case of Hampton Index HT bunches of needle crystals of MBPCG1 could be obtained already after 15 min in many conditions. It could be confirmed that conditions containing high molecular weight PEGs were suitable for crystal growth of the fusion protein. Nevertheless most of these crystals turned out to be grown into each other. Because crystals grew too fast the protein concentration was decreased to 20 mg/ml respectively. Crystallisation conditions #45 (0.1 M Tris/HCl pH 8.5, 25% (w/v) PEG3350) and #80 (0.2M (NH₄)₂SO₄, 0.1 M HEPES pH 7.5, 25% (w/v) PEG3350) from INDEX HT screen as well as #61 (0.2 M (NH₄)₂SO₄, 0.1 M NaAc pH 4.6, 30% (w/v) PEG2000-MME) from Crystal screen HT were selected for further manual refinement of crystal growth conditions. In 24-well sitting drop plates these initial conditions were systemically optimised by varying the precipitant concentration, the protein concentration, the drop size and the buffer. MBPCG1 could be crystallised in three different space groups, namely C2, P2₁, and P2₁2₁2₁.

The crystals were taken out of the drop and transferred into a droplet of a cryo-protectant solution containing glycerol to get a final concentration of PEG and glycerol of about 30% (v/v). After a short soaking time (1 min) the crystals were flash-cooled in liquid nitrogen. Diffraction data of all MBPCG1 crystals were

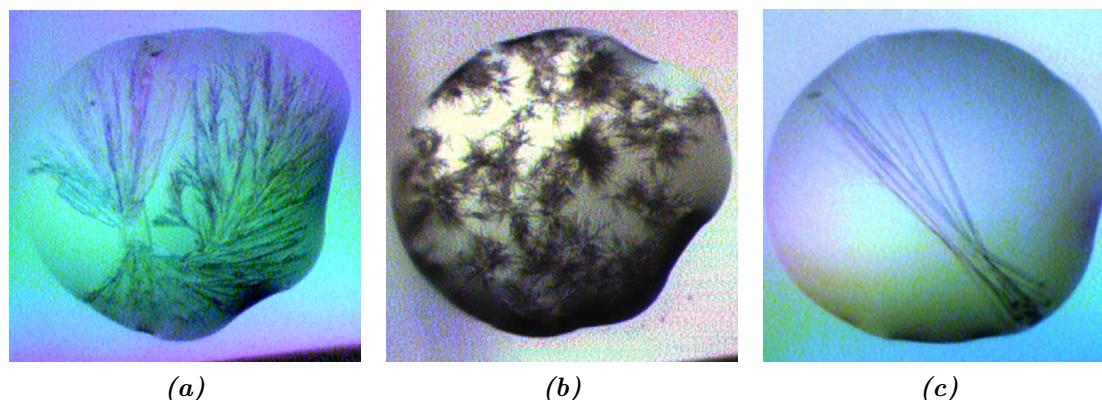


Figure 2.19: Initial crystals of MBPCG1: a) INDEX HT screen #45, b) INDEX HT screen #80, c) Crystal screen HT #61

collected in-house at 100 K at a Bruker Cu-K α rotating anode generator equipped with an INCOATEC multilayer optic, a three circle goniometer and a SMART6000 CCD detector. Data processing was performed with APEX2 [10] and reflections were integrated with SAINT [11]. SADABS [72] was used for scaling and absorption correction and the space group was determined with XPREP [75].

Thin plate-like crystals of **crystal form I** were grown from 0.2 M (NH $_4$) $_2$ SO $_4$, 30% PEG2000-MME and sodium acetate at pH 4.6 within five days. Because of the high PEG concentration additional cryo-protection of the crystal was not performed. The crystal was flash-cooled in liquid nitrogen and mounted in the nitrogen steam. Three ω -scans and one ϕ -scan were collected at a detector distance of 65 mm, a scan width of 0.2 $^\circ$ per image and 40s exposure time. The crystal diffracted to a resolution of 1.80 \AA and belongs to the monoclinic space group C2 with cell dimension of a=118.92 \AA , b=40.28 \AA , c=88.50 \AA , and β =115.88 $^\circ$. The data statistics are summarised in the appendix (see Sec. 7.3). Because of the absence of anomalous scatterers in the protein structure the method of choice to solve the crystallographic phase problem was molecular replacement by Phaser 2.1.4 (McCoy et al. [56]). As starting model a MBP structure published by Srinivasan et al. [80] (PDB-ID:1JVX) was used. This structure shows MBP in the *closed form* and contains maltose in the sugar-binding cleft. Solvent molecules, amino acid side chains, and the maltose molecule were removed from the starting model. Once the structure was successfully solved by Phaser a structure refinement was performed with SHELXL (Sheldrick [73]). The graphics program COOT (Emsley et al. [26]) was used for protein sequence docking and further manual model building. Maltose could be found in the sugar-binding cleft and was placed with COOT. By alternating use of SHELXL and COOT the protein structure was refined to $R_{\text{work}}=21.45\%$ and $R_{\text{free}}=26.85\%$.

Crystal form II was taken from a solution containing 25% PEG3350, and 0.1 M HEPES buffer at pH 8.5. These needle-shaped crystals grew within one week. The crystals were transferred into a cryo-protectant solution containing crystal growth solution and additional 5% glycerol. After 1 min soaking and flash cooling

the diffraction data were collected in-house at 100 K. The monoclinic space group $P2_1$ and a cell of $a=41.71 \text{ \AA}$, $b=63.42 \text{ \AA}$, $c=76.15 \text{ \AA}$, and $\beta=99.73^\circ$ were determined. One ω -scan and one ϕ -scan were collected at a detector distance of 65 mm, with a scan width of 0.25° per image and 60 s exposure time. The crystal diffracted to 2.56 \AA resolution. Again PDB-ID:1JVX [80] was used as starting model in Phaser [56] for molecular replacement as described before. Structure refinement was alternately performed by Refmac [89] and COOT [26]. MBP was again present in the *closed form* and maltose was manually placed with COOT. A TLS refinement was performed using 5 TLS domains which give final values of $R_{\text{work}}=25.97\%$ and $R_{\text{free}}=31.68\%$. These TLS domains were calculated by the TLSMD server (Painter and Merritt [59])

Crystal form III was grown in 0.2 M ammonium sulphate, 25% PEG3350, and 0.1 M HEPES buffer at pH 8.5 in the orthorhombic space group $P2_12_12_1$. Needle-shaped crystals grew within one week and were soaked in cryo-protectant solution containing 5% glycerol. Diffraction data were collected in-house using Cu- K_α radiation and sample cooling to 100 K. Because of technical problems only one ϕ -scan was performed at a detector distance of 45 mm, a scan width of 0.2° per image, and 120 s exposure time. The crystal diffracted to 2.76 \AA resolution and the cell dimensions were determined to $a=38.76 \text{ \AA}$, $b=74.61 \text{ \AA}$, and $c=111.84 \text{ \AA}$. This crystal was non-merohedrally twinned with a twin ratio of 0.55/0.45. The data set was de-twinned with the program TWINABS [74]. Because of the twinning a poor $R_{(\text{INT})}=0.294$ was calculated. The structure was solved by molecular replacement using Phaser [56] and again PDB-ID:1JVX [80] as starting model. The solved structure was alternately refined by Refmac [89] and manually edited with COOT [26]. MBP was again present in the *closed form* but unfortunately maltose could not be localised in the electron density. Because of the low data quality and resolution many amino acid side chains were omitted. The structure was refined to final values of $R_{\text{work}}=34.69\%$ and $R_{\text{free}}=45.09\%$.

MBPCG2

MBPCG2, the construct which contains Peptide56 and the linker sequence *GSSGSSG*, was overexpressed and purified as described before and concentrated to 22 mg/ml in 20 mM MES buffer pH 6.2 and 50 mM NaCl. Initial crystallisation trials were performed with crystallisation robots and sparse-matrix screens. 100 μl screen solution were transferred into the crystallisation well whereas the protein droplet contains 100 nl protein solution and 100 nl well solution. In contrast to MBPCG1, initial crystals of MBPCG2 could not be obtained using sparse-matrix screens. So crystal nuclei from a MBPCG1 crystal growth condition containing 0.1 M Tris/HCl pH 8.5 and 25% (w/v) PEG3350 were transferred with a cat whisker into a pre-equilibrated 24-well plate containing the same well solution (0.1 M Tris/HCl pH 8.5 and 25% (w/v) PEG3350) and MBPCG2. Under this condition initial crystals of MBPCG2 grew within three days. To make pure MBPCG2 initial crystals, nuclei from this cross seeding trial were used for further streak seeding experiments.

Nuclei were transferred again with a cat whisker in several new pre-equilibrated 24-well plates containing crystal growth conditions of MBPCG1. These conditions were taken from Hampton Index Screen HT, conditions #45 (0.1 M Tris/HCl pH 8.5, 25% (w/v) PEG3350) and #80 (0.2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES pH 7.5, 25% (w/v) PEG3350), as well as condition #61 (0.2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M NaAc pH 4.6, 30% (w/v) PEG2000-MME) from Crystal screen HT. Initial crystals could be obtained for all trials and were further manually optimised by varying the precipitant concentration and the pH. So four different crystal forms of MBPCG2 could be crystallised, namely P1, C2, P2₁, and P2₁2₁2₁.

Crystal form I was grown as *open form* of MBP in needle-shaped crystals from 25% PEG4000 and 0.1 M MES buffer pH 6.5 within one week. The crystals were transferred into a cryo-protectant solution containing crystal growth solution and additional 5% glycerol. The crystals were soaked for 1 min and flash-cooled in liquid nitrogen. Diffraction data were collected at the BESSY beamline X14.1 at a wavelength of 0.918 Å and 100 K. To ensure high data redundancy and precision three data sets were collected. For high resolution two 180° ϕ -scans with an angle offset of 0.5° at 145 mm detector distance, 7 s exposure time, and 1° oscillation range were collected. One 360/180° ϕ -scan at 290 mm detector distance, 3 s exposure time and 2° oscillation range was collected to get a low resolution data set. The crystal diffracted to 1.55 Å resolution. The space group P1 and a cell with $a=37.78$ Å, $b=43.86$ Å, $c=57.67$ Å, $\alpha=100.32^\circ$, $\beta=100.92^\circ$, and $\gamma=103.82^\circ$ were determined. All data sets were integrated with XDS (Kabsch [41]) and scaled with SADABS [72]. The MBP structure PDB-ID:1JVX [80] was used as starting model for Phaser [56]. Because MBP was present in the *open form* this trial failed. By analysing the electron density map only one domain was found by Phaser. The starting model 1JVX was separated into the two MBP domains I and II and again used for Phaser. Now the full structure could be phased. Further refinement was performed by SHELXL [73] and manual editing with COOT [26]. The structure was refined to $R_{\text{work}}=20.57\%$ and $R_{\text{free}}=24.65\%$.

Crystal form II was grown in 24% PEG3350 using 0.1 M Tris/HCl buffer at pH 8.5 as needle-shaped crystals within one week. The crystals were transferred into cryo-protection solution containing crystal growth solution and 6% glycerol. The crystals were flash-cooled in liquid nitrogen. One data set was collected at the BESSY beamline X14.1 at a wavelength of 0.918 Å and 100 K. A ϕ -scan containing 380 images was performed with a detector distance of 175 mm, 8 s exposure time, 1° oscillation range, and 0.5° overlap of images. The data were integrated with XDS [41] and the space group P2₁ and a cell of $a=49.43$ Å, $b=42.12$ Å, $c=88.21$ Å, and $\beta=92.61^\circ$ were determined. The crystal diffracted to a maximum resolution of 1.88 Å. For Phaser [56] the *closed form* of MBP (PDB-ID:1JVX, [80]) was used as starting model. Since Phaser was not able to solve the full structure the starting model 1JVX was again divided into two independent starting models, domain I and II. Now the full structure could be solved and the obtained electron density map shows the full structure in the *open form*. Structure refinement was alternatingly

performed by Refmac [89] and structure editing in COOT [26]. TLS refinement was performed using 7 TLS domains calculated by the TLSMD server (Painter and Merritt [59]). The fusion protein structure could be refined to a final $R_{\text{work}}=26.44\%$ and $R_{\text{free}}=32.46\%$.

Crystal form III was grown in 30% PEG2000-MME and 0.1 M Tris/HCl buffer pH 7.5 as plate-shaped crystals within five days. Because of the high PEG concentration already present in the crystal growth condition cryo-protection was not performed. Instead of cryo-protection the crystal was transferred into a soaking solution containing 100 mM CoCl_2 . After 30 s soaking time the crystal was flash-cooled in liquid nitrogen. At the BESSY beamline X14.1 one data set with a maximum resolution of 2.80 Å was collected at 100 K. Because of the low diffraction power of the crystal an anomalous diffraction experiment at the absorption edge of cobalt was not performed. Furthermore for structure solution using molecular replacement a highly homologous model was available. Therefore a wavelength of 0.918 Å was selected for data collection. In total 180 images, collected with a detector distance of 250 mm, 10 s exposure time and an oscillation range of 1° per image, were integrated with XDS [41]. The monoclinic space group C2 and cell dimensions of $a=110.21$ Å, $b=38.04$ Å, $c=90.36$ Å, and $\beta=111.90^\circ$ were determined. After scaling with SADABS [72] the MBP structure PDB-ID:1JVX [80] was used as starting model for a structure solution in Phaser [56]. Because this crystal form contained MBP in the *open form* the Phaser starting model had to be modified. By performing a structure solution with the *closed form* starting model only one domain was found in the initial electron density map. When the MBP domains I and II were separated into two search fragments the structure could be fully solved. Sequence docking and first refinement was performed with ARP/wARP (Langer et al. [52]). Structure refinement was alternately carried out by Refmac [89] and structure editing in COOT [26]. The TLSMD server [59] assigned 7 TLS domains which were used for refinement. The structure was refined to a final $R_{\text{work}}=24.80\%$ and $R_{\text{free}}=33.14\%$.

Crystal form IV was taken from a solution containing 24% PEG3350, and 0.1 M Tris/HCl buffer pH 8.0. Needle-shaped crystals grew within one week in the orthorhombic space group $P2_12_12_1$ with cell dimensions of $a=42.31$ Å, $b=95.33$ Å, and $c=96.68$ Å. The crystals were soaked in a solution containing the crystal growth condition, 6% glycerol for cryo-protection and 100 mM NiCl_2 for approximately 30 s. The crystal slowly dissolved during the soaking process. So the soaking procedure was performed as fast as possible followed by flash cooling in liquid nitrogen. Three $360^\circ\phi$ -scans were performed at the Bruker Cu- K_α home source at 100 K. Images were collected at a detector distance of 65 mm and an oscillation range of 0.25° per image. The crystal diffracted to a resolution of 2.24 Å. The data were integrated with SAINT [11] and scaled with SADABS [72]. The space group was determined with XPREP [75]. Because of the low anomalous signal of Ni^{2+} at Cu- K_α radiation the structure was solved by molecular replacement. Again the MBP structure PDB-ID:1JVX [80] was used as starting model for Phaser [56]. The structure was successfully solved and sequence docking and first refinement was performed with ARP/wARP ([52]). The

structure refinement was alternately carried out by Refmac [89] and COOT [26]. With the TLSMD server [59] in total five TLS domains were determined (Painter and Merritt [59]) for structure improvement. The structure was refined to $R_{\text{work}}=19.17\%$ and $R_{\text{free}}=24.85\%$.

MBPCG3

MBPCG3, the construct containing Peptide71 as fusion partner, was concentrated to 20 mg/ml in MES buffer pH 6.2 and 50 mM sodium chloride. By using crystallisation robots and sparse-matrix screens only a few crystallisation conditions could be found that gave initial crystals. In total 100 μl screen solution were transferred into 96-well crystallisation plates. The protein droplets were prepared by mixing 100 nl protein solution and 100 nl well solution. Initial crystals grew in condition #20 (0.1 M HEPES pH 7.5, 1.4 M Sodium citrate tribasic dihydrate) of the Hampton Research Index screen HT. Surprisingly a second initial crystal form grew in conditions containing Zn^{2+} ions. These crystals were found in #93 (0.05 M Zinc acetate 20% (w/v) PEG 3,350) of the Hampton Research Index screen HT and in condition #26 (0.2 M Zinc acetate 20% (w/v) PEG 3,350) and #92 (0.02 M Zinc chloride 20% (w/v) PEG 3,350). The crystals grown in sodium citrate were used for seeding experiments. Nuclei were transferred with a cat whisker into a pre-equilibrated 24-well plate containing sodium citrate at various concentrations up to 2 M and different buffers with a pH range from 6.5 to 8.5. Crystals from the Zinc containing initial condition could be reproduced in 24-well plates without seeding. It could be shown that those crystals would only grow if Zn^{2+} ions were present in the crystal growth solution in a concentration between 20 mM and 50 mM. When the concentration was higher as well as if Zn^{2+} ions are not present MBPCG3 did not crystallise. Therefore Zn^{2+} ions have to play a crucial role in the crystal-forming process. Various crystals were tested for diffraction but did not diffract at all. Further trials were performed by optimising the initial conditions. These conditions were modified by changing the pH, the PEG concentration, the transition metal salts and its concentrations (NiCl_2 , ZnAc_2 , CoAc_2 , and CuCl_2) as well as the drop size. Finally a condition to co-crystallise MBPCG3 with transition metals was found. A solution containing 15% PEG3350, 0.1 M MES buffer at pH 5.5 and 20 mM salt (ZnAc_2 , CdAc_2 , CoCl_2) gave reproducible needle shaped crystals as shown in Fig. 2.20. Unfortunately none of these crystals diffracted.

Further seeding experiments were extensively performed with conditions already known from MBPCG1 and MBPCG2. So conditions containing different PEGs in a concentration of 10% to 30% and buffers (MES, HEPES, Tris) at a pH range from 5.5 to 8.5 were inoculated with nuclei from MBPCG1. Because maltose was not added to the protein solution MBP crystals in the *open form* were expected. These seeding experiments were not successful and a new crystal form could not be found. Finally MBPCG3 could be crystallised in two different space groups, the triclinic space group P1 and the monoclinic space group C2. Diffraction data were collected in-house at the Bruker rotating anode generator using a SMART6000 CCD detector.

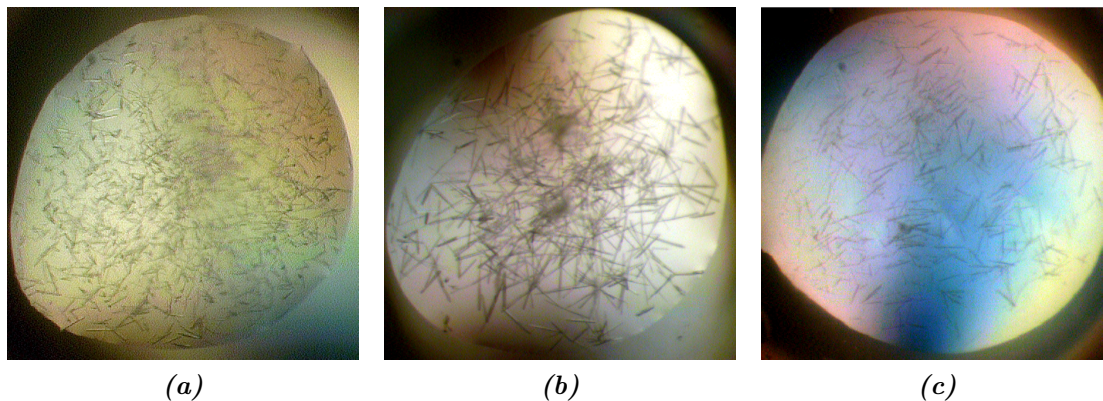


Figure 2.20: Crystals of MBPCG3 grown in crystallisation wells containing metal ions: a) 20 mM ZnAc_2 , b) 20 mM CdAc_2 , c) 20 mM CoCl_2 . However, none of these crystals diffracted.

Crystal form I appears in a condition containing 2 M sodium citrate and 0.1 M Tris at pH 7.5. Only one block-shaped crystal grew within two weeks. Because of the high salt concentration the crystal was flash-cooled without any cryo-protection and transferred to the goniometer. Data were collected in-house using Cu-K_α radiation and cooling to 100 K. The crystal diffracted to a maximum resolution of 2.00 Å. Images were indexed and processed with the program APEX2 [10]. It was found that the crystal was a non-merohedral twin. Two domains could be separated and were integrated with SAINT [11]. The processed data were scaled and de-twinned with TWINABS [74]. The twin ratio was calculated to 0.54/0.46. A cell of $a=55.66$ Å, $b=65.24$ Å, and $c=79.40$ Å with angles of $\alpha=73.14^\circ$, $\beta=64.33^\circ$, and $\gamma=64.69^\circ$ in the space group P1 was found. The structure solution was performed by molecular replacement. In a first trial again MBP structure PDB-ID:1JVX [80] was used as starting model in Phaser [56]. After dividing the MBP model in domains I and II structure solution with Phaser was performed. By using two search models (domains I and II) just a part of the structure could be solved. If the unit cell were to have one protein molecule, about 82% of the unit cell volume would be filled with water molecules which is very uncommon. It is more likely that the number of protein molecules is two (64% water) or three (45% water). The structure was solved by searching for two independent protein molecules with the MBP domains I and II as search fragments. The solution from Phaser was first refinement with ARP/wARP (Langer et al. [52]) and later by alternating usage of Refmac [89] and COOT [26]. TLS domains were calculated by the TLSMD server [59]. For the two MBP molecules 7 TLS domains in total were refined and gave $R_{\text{work}}=27.68\%$ and $R_{\text{free}}=30.57\%$.

Crystal form II grew in a solution containing 30% PEG2000-MME and 0.1 M Tris/HCl pH 7.5 as plate-shaped crystals within one week after seeding. Seeding nuclei were transferred with a cat whisker from an MBPCG1 crystal (#45 from Hampton Research Index HT screen) into pre-equilibrated crystallisation trials performed in a 24-well plate. Because already 30% of PEG was present in the crystal

growth condition the crystal was flash-cooled without cryo-protection. Diffraction images were collected in-house using Cu-K α radiation at 100 K. Four ω -scans at a detector distance of 65 mm and 120 s exposure time were performed. The crystal diffracted very weakly to a maximum resolution of 3.70 Å. Several crystals were tested but unfortunately this was the best diffracting crystal. A crystal cell was determined to have $a=106.86$ Å, $b=37.2$ Å, $c=91.72$ Å, and $\beta=110.86^\circ$. The data were processed with the program APEX2 [10] and integrated with SAINT [11]. The space group C2 was determined with XPREP [75]. The structure was solved by molecular replacement. MBP was present in the *open form* and again domains I and II of the search model PDB-ID:1JVX [80] were separated for Phaser [56]. The structure could be solved and was first refined with ARP/wARP (Langer et al. [52]) and later by alternating usage of Refmac [89] and COOT [26]. Because of the low resolution the structure could not be fully refined. Further refinement was stopped at $R_{\text{work}}=32.10\%$ and $R_{\text{free}}=53.37\%$.

MBPCG4

MBPCG4, the construct containing Peptide71 with the G372P mutation as fusion partner, was overexpressed and purified as described in Sec. 2.3. The protein solution was concentrated to 20 mg/ml in 20 mM MES buffer pH 6.2 and 50 mM NaCl. By using various sparse-matrix screens (see Tab. 2.6) crystallisation trials were performed supported by crystallisation robots. Several crystallisation trials were modified with 20 mM salt (NiCl $_2$, CoCl $_2$, ZnAc $_2$, and CdAc $_2$) in the crystallisation droplet. Surprisingly initial crystals of MBPCG4 could not be grown. Therefore, seeding experiments with crystals from MBPCG1 were performed. MBPCG1 crystal nuclei from condition #45 from Hampton Research Index Screen HT were transferred with a cat whisker into pre-equilibrated crystallisation 24-well plates. Conditions successfully applied for crystal growth of MBPCG1 were modified by changing the pH and the PEG concentration. Crystallisation trials were performed containing transition metal salts in various concentrations (NiCl $_2$, CoAc $_2$, ZnAc $_2$, CuCl $_2$, and CdAc $_2$). Finally same experiments were carried out at 4°C. However, only one crystallisation condition could be found containing crystals of MBPCG4. In a solution containing 15% PEG3350, 0.1 M MES buffer at pH 5.5 and 20 mM salt CdAc $_2$ needle shaped crystals could be grown (see Fig. 2.21). This condition was previously used to grow crystals of MBPCG3 (see Fig. 2.20). Unfortunately, as in the case of MBPCG3 these crystals did not diffract.

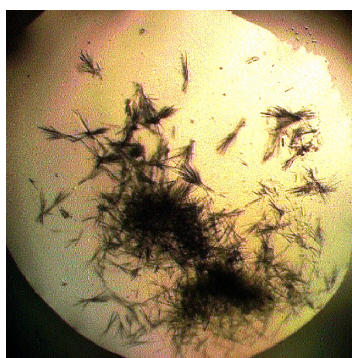


Figure 2.21: Crystals of MBPCG4 grown in a solution containing 20 mM CdAc₂.

3. Short peptides mimicking protein tags

Short polypeptide were designed by bioinformational methods to find suitable tag sequences for purification and crystallisation enhancement. These peptides were synthesised to prove the predicted structure in aqueous solution, to test affinity using the IMAC technique and finally to be crystallised in the presence of metal ions.

3.1 Peptide engineering

During the last decade various protein tags were established in daily protein engineering (see Sec. 1.1) but the His-tag is used by far most frequently. Its short tag sequence binds strongly to divalent metal ions for purification applications. Various expression systems are available commercially for different hosts and C-terminal as well as N-terminal fusion proteins. Usually linker sequences are included to facilitate the cleavage. The His-tag represents an easy-to-use and widely used protein tag system and is therefore the system of choice for modification and improvement.

In its present form the His-tag neither hinders nor promotes crystallisation of the fusion construct as shown in Sec. 1.2.1. Only approximately 6% of all structures with uncleaved His-tags have main chain atom coordinates deposited for His-tag histidine residues (four or more) in the PDB [14]. The reason for that is a high terminal motion. Furthermore the His-tag is by nature structurally flexible, a random coil motif that adjusts its structural characteristics in the crystal lattice to the fusion partner. Because of its histidine residue content bonding of transition metals in solution as well as in the crystal is possible and incorporation of anomalous scatterers in the crystal lattice is sometimes observed. Several crystal structures (Kim et al. [49], van den Berg et al. [90]) show metal ions bonded to a His-tag. These metal ions can be used in general for obtaining experimental phase information by

SAD or MAD methods. Various transition metals which support IMAC purification can also be incorporated in the crystal lattice as anomalous scatterer. Of special interest is zinc because of its low toxicity, low redox susceptibility and its flexible coordination sphere as divalent element. On the other hand cobalt has a high purification affinity and exhibits a stronger anomalous signal using Cu-K α radiation compared to nickel, copper, and zinc. Cobalt gives also rise to a stronger anomalous signal than selenium or sulfur data at a X-ray Cu-K α home source.

In various crystal structures secondary structures are known that contain histidine residues and chelate metal ions. As seen in Fig. 3.1 two histidine residues introduced in helices are able to chelate metal ions. Often these residues contribute to crystal packing by crystal contacts. That could be demonstrated by Reiling et al. [64] as well as Hazes et al. [35]. In theory there are four possible ways to incorporate histidine residues in a helix for metal complexation: two histidine residues adjacent in the protein sequence (HH), two histidine residues with a distance of three or four residues in the protein sequence ($HxxH$ or $HxxxH$), three histidine residues in a $HHxxH$ or $HxxHH$ arrangement, and finally one histidine residue at one end of the helix and a second histidine residue in a loop region next to it. In the last case the arrangement of histidine residues is less stable and therefore not relevant for further investigation. An arrangement of three histidine residues has a high affinity to the metal ion but is very uncommon. Histidine residues do not support α -helix formation. So this motif chelating a metal ion has not been found in crystal structures so far.

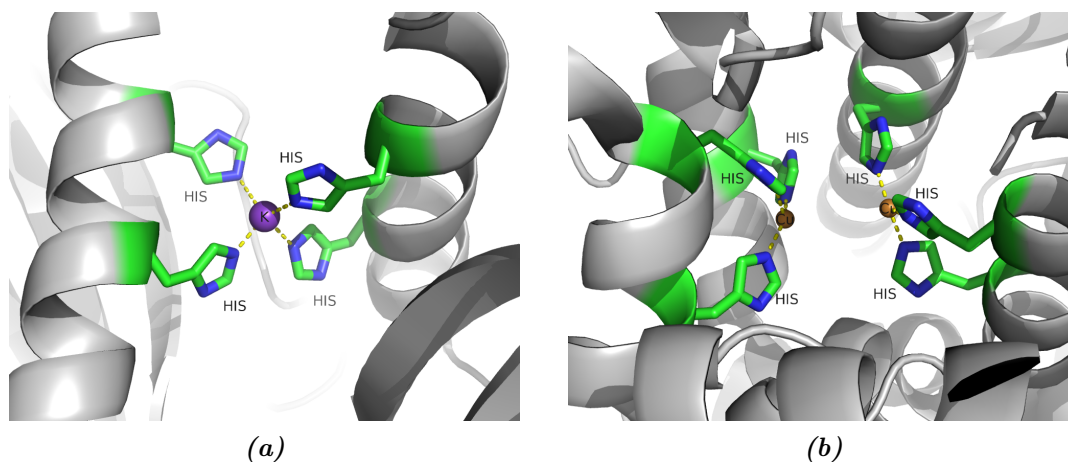


Figure 3.1: Histidine residues incorporated in α -helices that chelate metal ions in the crystal structure. *a)* Detail of the KSHV protease (PDB-ID:1FL1) published by Reiling et al. [64]. Two helices are linked by a tetrahedral chelation of a potassium ion. Two histidine residues of each helix are involved in interaction and form a crystal contact. *b)* Detail of the *Limulus Polyphemus* subunit II (PDB-ID:1LLA) published by Hazes et al. [35]. Six histidine residues from four different helices chelate two copper ions.

In contrast, in β -strands connected by a loop to form a β -hairpin the arrangement of histidine residues is different. Theoretically up to four histidine

residues can be present to chelate a metal ion in a planar geometry but this arrangement has never been observed so far. More common are two and three histidine residues that form tetrahedral and octahedral coordination spheres. Thus several valences of the metal ion are unoccupied and enable coordination of water molecules or side chains of adjacent residues. In the case of two histidine residues, they can be placed on the same strand or one in the strand region and one in the loop next to it or finally one residues each in one strand to be on opposite sites of the two strands. One has to keep in mind that in a strand adjacent residues are oriented to different sides of the strand. Therefore one amino acid residue has to be in between the histidine residues to keep the facility to chelate metal ions. If three histidine residues are present they can be placed on both strands with or without sharing the loop region. Placing the three histidine residues on a single strand is not sufficient for tridentate chelation of a metal ion. From the geometrical point of view the histidine residues have to be on both strands (see Fig. 3.2). This arrangement is common in various proteins as shown by Schreiter et al. [70] and Riboldi-Tunnicliffe et al. [65].

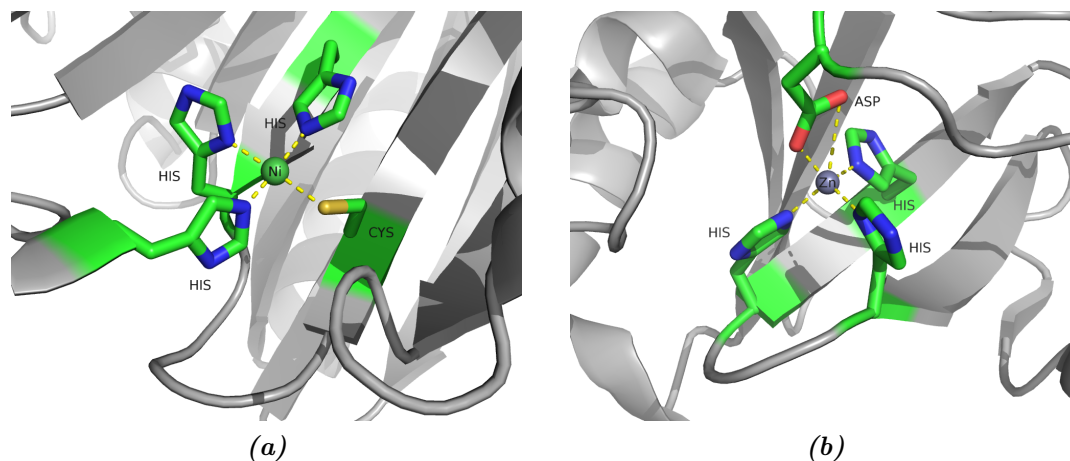


Figure 3.2: Histidine residues incorporated in β -hairpins that chelate a metal ion. *a)* Details of the C-terminal domain of NikR (PDB-ID:1Q5Y) published by Schreiter et al. [70]. A nickel ion is chelated by two histidine and one cysteine residues from the hairpin. The residues are close to the loop region. The fourth residue is a histidine from an adjacent α -helix. *b)* Detail of the *Streptococcus pneumoniae* PhtA histidine triad domain (PDB-ID:2CS7) published by Riboldi-Tunnicliffe et al. [65]. Near the loop region three histidine residues chelate a zinc ion.

The Protein Data Bank (PDB) was used as starting point for research activities. The basic task was to find a preferred amino acid sequence involving adjacent histidine residues in a folded environment, which bind metal ions. Small peptides with a well-characterised secondary structure, either α -helices or β -hairpins are possible candidates. Certainly α -helices are more strongly stabilised by hydrogen bonding than β -strands. As shown in Fig. 3.3 the protein tags should have a rigid backbone that is functionalised with histidine residues, which point toward on direction. Two of these tags could stick together *via* one or more metal centres. We have been looking for short peptides with a molar weight less than 3000 u to mimick stable protein

tags. Chou and Fasman [16, 17] showed that every amino acid has an individual tendency for integration in a secondary structure element. Thus, amino acids glutamate, methionine and alanine are high potential α -helix builder where as proline and glycine are so called helix “breaker” and are almost never observed in α -helices. It was shown that the amino acid histidine is indifferent against all analysed secondary structure elements (α -helix, β -sheets, and β -loops). That is why a large number of histidine residues in the polypeptide chain, needed for high IMAC affinity, usually involve a random coil structure like the classical His-tag. To overcome this dilemma residues have to be incorporated to enhance a stable secondary structure. Once a pre-organised backbone has formed additional stabilising through histidine residues that chelate metal ions can occur.

The number of histidine residues for sufficient purification can be estimated. Kozlov et al. [51] demonstrated that not only the amount of histidine but the sequence as well is important. When comparing the histidine content of an 11-mer peptide, one histidine requires 83 mM imidazole solution for the elution of the peptide. However, if the histidine content is increased to 10 the required imidazole concentration is dramatically increased to 460 mM. Furthermore if there are two histidine residues with only one amino acid in between this sequence is more affine than a sequence with two histidine residues and seperated by two or more amino acids. One could imagine that in the case of the sequence HxH the distance of the histidine side chains is better applicable to chelate metal ions on the IMAC support. To conclude, the amount of histidine residues can estimated to four to six for sufficient purification and the arrangement has to take coordination geometry of the metal ion into account.

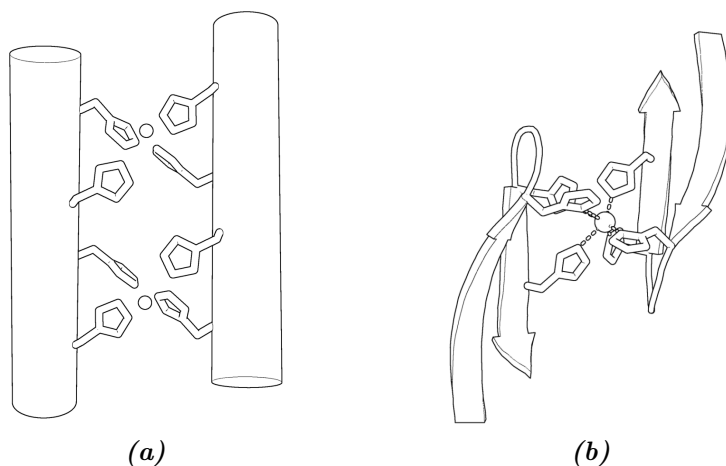


Figure 3.3: Illustration of dimer forming *via* metal ion chelation by the protein tag. *a)* The dimer build by two α -helices is fixed by two tetradentally bonded metal ions. *b)* The two β -hairpins are complexed hexadentately by a metal ion.

For subsequent application the protein tag has to be able to be overexpressed in various hosts and therefore a sequence composition of native *L*-amino acids is essential. Often linker sequences susceptible to proteolytic cleavage are incorporated in fusion proteins. These linkers promote a spartial separation of protein and tag.

Thus mutual influences are reduced and the degree of terminal motion is increased as well. This discrepancy is hard to overcome. A finetuning in the linker sequence length has to be performed.

In general, both α -helices and β -hairpins seem to be equally appropriate candidates and for this reason both secondary structures were considered for this project. Taking the Chou Fasman algorithm into account peptides were designed by incorporation of histidine residues in peptide sequences found in the PDB. Peptides, predicted to form either α -helices or β -hairpins, in total 45, have been synthesised with a solid phase peptide synthesisers using Fmoc-strategy with the chain lengths from 12 to 48 amino acids. The peptides were purified by RP-HPLC and IMAC technique and characterised by electrospray ionization mass spectrometry (ESI-MS), NMR, and CD-spectroscopy. Finally crystallisation trials were carried out to obtain well-diffracting crystals for X-ray structural analysis.

3.2 α -Helix forming peptides

One of the two major secondary structure elements present in proteins are α -helices. Due to chirality of natural L-amino acids α -helices are characterised by a right-handed helical structure, formed by a residue translation (pitch) along the helical axis of 1.5 Å and 3.6 residues per turn. As a consequence the vertical distance between two turns is 5.4 Å. Helices are characterised by a unique network of hydrogen bonds. The carbonyl group of amino acid n is connected to the amid hydrogen of amino acid $n+4$. Every peptide bond in a helical structure is involved in this stabilising network of 3 to 4 hydrogen bonds per turn. In an α -helix, all amino acid side chains are located on the helix outside. The tendency to form α -helices depend both on amino acid type and sequence. Sequences of several identical, charged amino acids usually destabilise a helical structure due to repulsing forces. However, if a positively charged amino acid is located 3 or 4 residues apart of a negatively charged amino acids, these residue become very close due to the helix turn. Such arrangement is able to form ion pairs that stabilise the helix. In the case of two aromatic amino acid residues hydrophobic interactions are formed. α -Helices have an overall dipole moment, caused by accumulation of equally oriented dipoles of single carbonyl groups, that become stronger with their length of the helix. The four residues in the helix terminus are not fully involved in the hydrogen bonding network and therefore the negative and positive partial charges are located in amino- and carbonyl groups. These destabilising forces can be compensated by amino acids with opposite charged side chains. So negatively charged Glu and Asp are more often located at the N-terminus where as positively charge side chains more frequently observed at the C-terminus of α -helices.

As mentioned before interaction of amino acid side chains contribute to helices backbone stabilisation. Several groups have shown that small polypeptides can form helices. Cnudde et al. [18] demonstrated by incorporation of the unnatural amino acid γ -carboxyglutame (Gla) and chelation by Ca^{2+} ions a metal induced

helix formation. This amino acid has a high intrinsic force to build α -helices. Furthermore Gelinsky and Vahrenkamp [30] and Ghadiri and Fernholz [32] could show that transition metal chelation can support helix stability. Finally covalently linked sulfur atoms in a disulfide bridge can force a peptide to keep a helical arrangement (Nicoll et al. [58], Fütterer et al. [29]). However, short polypeptide containing two histidine residues are rare in literature. To the author's knowledge, polypeptides that contain four histidine residues which would be necessary for sufficient IMAC affinity are not known so far. Therefore among others the Ubiquitin Interacting Motif Vps27p (Fisher et al. [27], PDB-ID:1O06), a 20 amino acid long helical peptide, was used for modification. The wild-type structure was determined in the hexagonal space group $P6_222$ and to a resolution of 1.45 Å. In the structure various Zn^{2+} ions bonded to acidic amino acid residues could be found. Thereby crystal contacts between the peptides were generated.

3.2.1 Mutants of Vps27

Systematical modifications to incorporate histidine residues were carried out. Vps27 (1O06_A) is already equipped with two metal centres realised by aspartate and glutamate residues as shown in Fig. 3.4. These residues were successively modified to histidine residues to increase affinity to metal ions, but without changing folding tendency. The aim of mutation was to introduce in a first step one bidentate and finally a second bidentate metal centre to get a metal affine plane as illustrated in Fig. 3.5 useful for purification as well as dimerisation.

	1	10	20
1006_A	EEDP	LKAAIQESLR	EAEEA
Peptide19	YEDP	LKHAIQESLR	EAEEA
Peptide27	YEDP	LKHAIQESLR	EAEEH

Figure 3.4: Sequence alignment of Vps27 family polypeptides based on PDB-ID:1O06. Highlighted are the positions in the wild-type sequence of Vps27 (*red*) which were mutated to histidine residues (*green*).

Two mutants (see Fig. 3.4) were synthesised by SPPS using standard Fmoc techniques and purified by RP-HPLC. To reduce electrostatic repulsion the N-terminus was blocked by acylation. Affinity to IMAC support was tested using Ni^{2+} loaded IDA cartridge from GE Healthcare at 1 ml/min and buffer A and B as described before. It could be shown that both mutants do not bind to Ni^{2+} loaded IDA support at all and flow through the HiTrap-FF cartridge without retardation. One has to speculate that the helical structure of Vps27 is lost due to the incorporation of histidine residues. In the unfolded state the histidine residues are too far apart to chelate the same metal ion on the IMAC support.

CD spectroscopy

Structural informations could be obtained by CD spectroscopy. Samples of Peptide19 and Peptide27 at 50 μ mol/l concentration and dissolved in 50 μ M

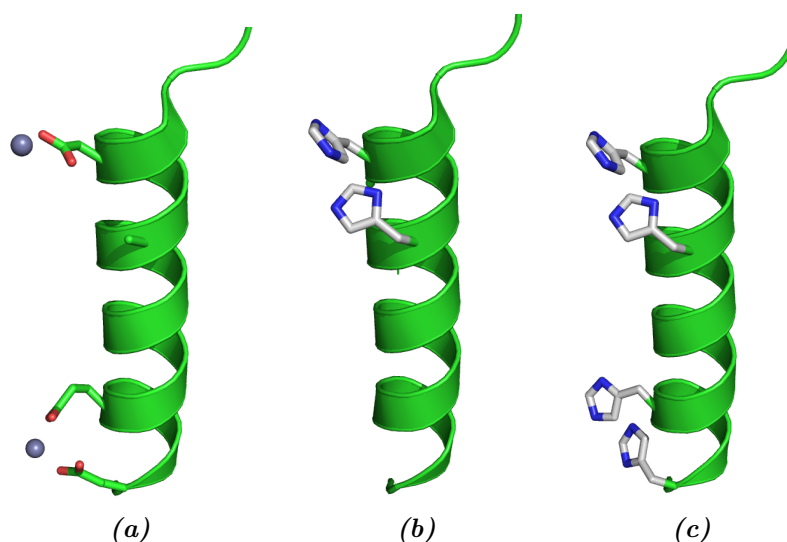


Figure 3.5: Determined structure of Vps27 and predicted structures of helical Peptide19 and Peptide27 are shown. *a)* Crystal structure of Vps27 presented by Fisher et al. [27] (PDB-ID:1O06) shows zinc ions bonded to aspartate and glutamate. Thus crystal contacts are formed. *b)* Predicted structure of Peptide19 consisting of two histidine residues are shown. *c)* The expected structure of Peptide27 should keep helical constitution despite incorporation of four histidine residues.

Na_2HPO_4 at pH 7.5 were analysed using a 1 ml cuvette. The metal ion (NiSO_4) concentration was increased from 0 to 5 eq. to check influence of chelation on folding. The resulting spectra are shown in Fig. 3.6 and characterise fully random coil peptides. One has to conclude that incorporation of histidine residues degrade polypeptide fold of Vps27. A folded protein tag using Vps27 mutants could not be obtained.

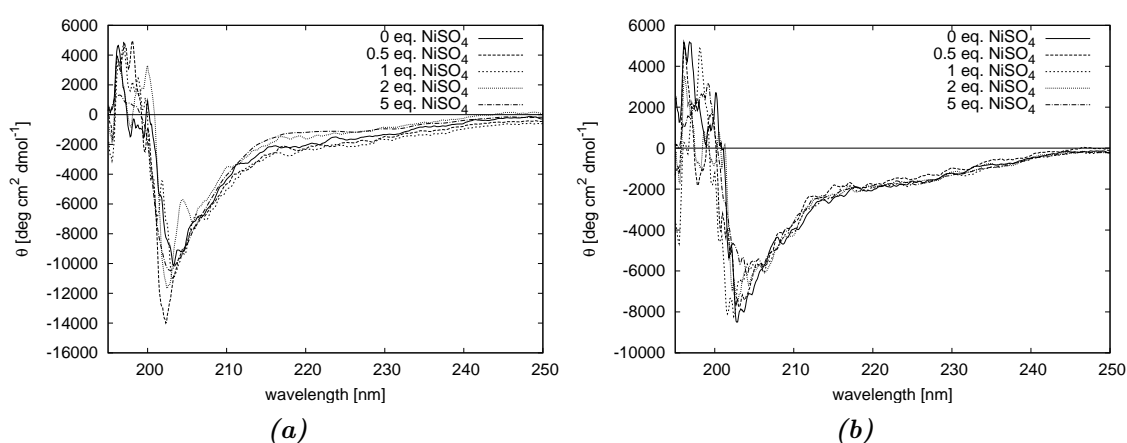


Figure 3.6: CD spectra of *a)* Peptide19 and *b)* Peptide27 at 50 $\mu\text{mol/l}$ peptide concentration and various NiSO_4 concentrations (0-5 eq.) at 20 $^\circ\text{C}$. The spectra show no bands significant for an α -helix but those typical for unfolded random coil peptides. One can exclude a partly helical folded peptide as well. Addition of metal ions for chelation shows no effect and indicates a missing metal chelation.

3.3 β -Hairpin forming peptides

In nature β -hairpins are widespread in globular proteins. Their task is to link the major elements helices and strands. A short loop region is embedded between two anti-parallel hydrogen bonded β -strands, where the carbonyl group of residue i forms a hydrogen bond to a NH group of residue $i+3$. Sibanda and Thornton [76, 77] have devised a system for classifying these hairpins which is based on two conventions for defining loop regions. They used two numbers [X:Y]: the first number denotes the number of residues in the loop, not linked by backbone hydrogen bonding. The second stands for the total number of residues between the first strand residue and the residue which is connected by two distal hydrogen bonds. There are several types of hairpins depending of the type of turn. As reported by Sibanda the short loop β -hairpins are most common. In proteins the [2:2] hairpins are the most frequently occurred, followed by [3:5] and [4:4] hairpins. There is another classification analysed by Venkatachalam [91] depending on conformation in the loop. He found three general types related to ϕ and ψ values. In contrast to type III which has ϕ and ψ values identical to 3_{10} -helices, type I and II turns differ by an 180° flip of the central peptide bond. Further two subtypes could be found the I' and II', respectively, which are characterised as the mirror images of type I and II. It was found, that these types are predominant in [2:2] loops.

In general, structures of β -hairpin forming peptides are rare in literature. It is known that the central loop region is more critical to folding and stability than in the strand region. A few sequences are reported (Gellman [31], Ramírez-Alvarado et al. [63], Blandl et al. [4]) that are more often found in naturally formed β -hairpins in proteins. These sequences tend to form loops (*NPDG*, *NSDG* and *NG*) and favour a strongly bended arrangement. Unfortunately none of these sequences contain histidine residues. Furthermore several groups were able to stabilise hairpin forming peptides by incorporation of conformationally restricted *D*-proline in the loop region (Karle et al. [43], Syud et al. [82]). This amino acid, in contrast to *L*-proline, is sterically more favourable to form loops due to its ϕ and ψ angles. Furthermore, stability is usually strongly dependent on chain length and decreases through more floppy termini (Stanger et al. [81]). To reduce this effect the strand can be covalently bonded *via* a disulfide bridge of two cysteine residues located at the strand ends (Brauer et al. [7], Bonvin and van Gunsteren [5]). These methods to stabilise β -hairpins are not sufficient for this project due to overexpression of the tag in living cells (*D*-amino acid) and the risk of disulfide bridge misfolding. A more elegant solution to stabilise a β -hairpin is through side chain-side chain interaction. It was demonstrated by Cochran et al. [19], Honda et al. [39] and Wu et al. [97] that hydrophobic interactions of aromatic amino acid side chains are able to stabilise β -hairpins. As known from the Trp-zipper motif hydrophobic as well as interactions of π -electrons are involved.

3.3.1 Protein G B1 based β -hairpins

One of the well-investigated peptides that show spontaneously folding in aqueous solution is the peptide G (Blanco et al. [3]). Source of this polypeptide is a dissected fragment of residues 41-56 of protein G domain B1, an immunoglobulin-binding protein expressed in group C and G *Streptococcal bacteria*. One member of the Protein G B1 based peptide family is Chignolin, presented by Honda et al. [39] in 2004. He was able to characterise β -hairpin formation by CD spectroscopy, NMR and X-ray crystallography (Honda et al. [40]). Unfortunately histidine residues are very rare in the natural peptide sequence with the one exception of the kelch domain of human Klhl (PDB-ID:2VPJ). In this structure [45] two histidine residues are incorporated in the loop region of the β -hairpin but do not chelate a metal ion. To the author's knowledge this is the only β -hairpin which contains two histidine residues. However, the type I [4:4] β -hairpin forming peptide presented by Honda et al. [39] could be shown to fold spontaneously in solution even at higher temperature. Due to its anti-parallel strands this peptide generates two opposite regions, one hydrophobic plane formed by aromatic amino acid side chains, responsible for strand stability, and a sticky side accessible for modification and installation of a metal ion affine surface. For this sticky side, residues located in the loop region have to be taken into account because of the short length of the peptide. Two metal centres, which means four to six histidine residues are needed, shall be incorporated to ensure high purification power and permit dimer forming. In a 10-mer peptide, as reported by Honda et al. [39], a ratio of almost 50% histidine in total was achieved and hindered hairpin forming. To decrease the content of histidine per residue the final tag should be slightly longer.

		1		10
		.Y.P....W.		
1k3i_A	252-262	LYDSSSDSWI		
1k3i_A	301-311	VYSPSSKTWT		
1zgk_A	344-354	AYNPSNGTWL		
1zgk_A	536-546	RYDVETETWT		
1zgk_A	583-593	CYDPTDTTWS		
1x2j_A	344-354	AYNPSNGSWL		
1x2j_A	395-405	CYNPMTNQWS		
2uvk_A	90-100	KYNPKTNSWV		
2vpj_A	301-311	KYDPKTQEWS		
2vpj_A	445-455	KYDPHTGHWI		
2vpj_A	539-549	CYDPIIDSWE		
1guq_A	13-23	RYNPLTGQWI		
1opm_A	252-262	GYRVRNGQWT		

Figure 3.7: Sequence alignment of β -hairpins in various proteins. The source of the sequence is illustrated by its PDB-ID, chain-ID and the residue numbers. Highlighted in *red* are highly conserved amino acids necessary for stability of the β -hairpin. To the author's knowledge only one case is known where histidine residues are included in the β -hairpin. These amino acids are highlighted in *green*.

In literature various proteins can be found that contain β -hairpin structures with a high sequence homology to the protein G domain B1, as shown in Fig. 3.7.

Highly conserved are the red highlighted tyrosine residue Y2 and W9 as well as the proline residue P4. Besides these three amino acids the sequence seems to be unspecific and not restricted. There are polar as well as hydrophobic and charged amino acids at almost every position. However, the conserved amino acids Tyr and Trp are necessary for side chain-side chain interaction and represent a minimal Trp-zipper motif.

3.3.2 Mutants of Protein G B1 based β -hairpins

Taking previous results into account several mutants were designed (see Fig. 3.8) which should fold spontaneously into β -hairpin structure, so keep their natural fold, and comprise replacements of various amino acids to histidine residues to incorporate metal ions by chelating. As known from studies by Kozlov et al. [51] two metal centres should be enough for purification. One centre should be located in the loop region whereas the second has to be divided in two parts located in the strand region. These second centre give rise to link the two strands and consequently, once formed, stabilise the β -hairpin.

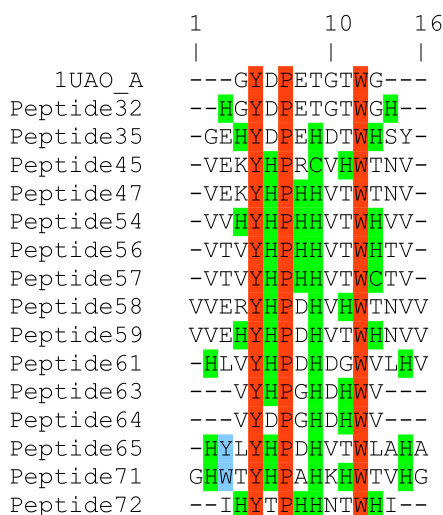


Figure 3.8: Sequence alignment of Chignolin mutants prepared for this study. Highlighted are mutated positions to introduce histidine residues (*green*), the conserved amino acids (*red*) as well as additional aromatic amino acids (*blue*) to make the β -hairpin more stable.

In total 15 mutants of peptide G were designed which differ in sequence composition of histidine residues (*green*) in contrast to conserved (*red*) amino acids, as illustrated in Fig. 3.8. In *blue* additional aromatic amino residues are highlighted. The major task was to determine amino acids in the central loop region that can be modified to histidine (or even cysteine) without losing loop structure. As mentioned before the amino acid histidine is indifferent to any secondary structure and cannot stabilise β -hairpins as well as β -strands in its uncomplexed form. On the contrary histidine is very flexible in its protonation state over a wide pH range. Due to this phenomenon and the large number of histidine residues repulsing forces should be expected that counteract secondary structure forming.

After this analysis 20 different mutants were synthesised by microwave supported SPPS and purified with RP-HPLC. They were characterised by ESI-MS. All peptides were tested for their IMAC affinity but only the most powerful peptides were analysed by NMR and CD spectroscopy. The first synthesised mutant, named Peptide32, was a peptide corresponding to peptide G published by Honda et al. [38] with one additional histidine residue on C- and N-terminus to elongate the peptide to 12 residues. To check stability and fold the sequence composition was not modified. Additional histidine should afford metal chelation. A second peptide, very similar to peptide G, was designed involving 14 amino acids with one mutation T8H in the loop region and two His in the strand at position 3 and 12. Various mutants were synthesised with a large number of histidine residues to increase IMAC affinity. To force dimer forming a self-complementary sequence involving a tridentate and a monodentate chelation centre was designed. Last but not least to stabilise the strand region a third aromatic amino acid side chain was introduced on the hydrophobic side of the strand. Inspired by Trp-zipper motifs and the natural protein G B1 the amino acid Trp was chosen because of its size and aromatic interaction potential. The most promising position for this residues is position $-B_4$ (see Fig. 3.9), two residues ahead of Tyr at position $-B_2$.

It was found that Peptide71 is the most promising candidate in terms of folding and purification properties. As illustrated in Fig. 3.9 and 3.10 the β -hairpin is stabilised by three effects: two highly potential loop former proline and lysine, hydrophobic and aromatic interactions at the back side of the anti-parallel β -strands and finally two metal chelation centres on the sticky front side realised by five histidine residues linking the two anti-parallel β -strands. In contrast to all other peptides, Peptide71 consists of three amino acids with aromatic side chains that are able to interact with each other. Taking the large number of histidine residues into account this is essential for keeping the β -hairpin folded.

3.3.3 Analytical data of β -hairpin forming peptides

CD spectroscopy

Several mutants of protein G B1 based peptides show well structured CD spectra. Data of Peptide32 and Peptide35 were collected in PO_4^{3-} buffer pH 5.5 at a peptide concentration of 100 μM . The spectra in Fig. 3.11 show a strong band at 229 nm as well as an intersection point at 222 nm (*isosbestic point*). These results are strongly related to a cross-strand interaction of Tyr and Trp in the hydrophobic strand side and confirm hairpin forming. As shown by Cochran et al. [19], Honda et al. [39] and Wu et al. [96] edge-to-face coupled aromatic side chains have such strong negative-positive $\pi - \pi^*$ exciton-coupled bands. There is some controversy in the literature which interactions, Trp-Trp or Trp-Tyr, is the cause for this band at 299 nm. It is possible that a edge-to-face Trp-Trp interaction but also a edge-to-face Trp-Tyr induces this effect. Interestingly spectra of peptides containing only Trp and Tyr are identical with spectra of peptides involving two residues each. However, this

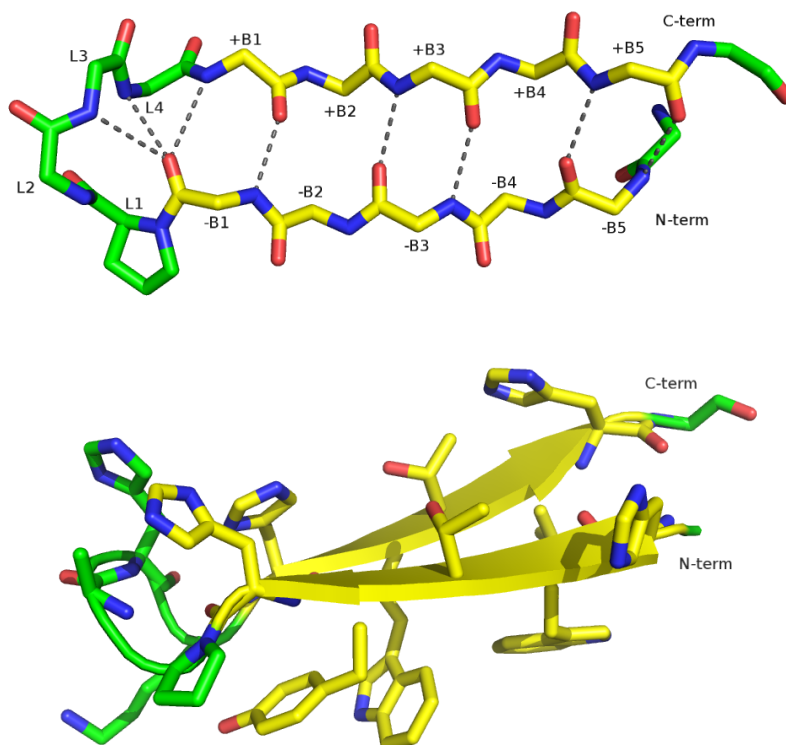


Figure 3.9: *Top*) Illustration of the Peptide71 backbone. The antiparallel β -stands (-B1 to -B5, +B1 to +B5) are coloured in *yellow*. The loop region (L1 to L4) is coloured in *green*. The main chain hydrogen bond are shown and indicate a [4:4] β -hairpin. *Bottom*) Illustration of the expected structure of Peptide71, reconstructed from residue 41-56 of protein G B1 domain (PDB-ID: 2ZW0, Watanabe et al. [92]). Visible are the aromatic amino acid side chains on the hydrophobic back side and the sticky metal ion affine front side built by five histidine residues.

phenomenon is characteristic for aromatic interactions and can occur for Peptide32 and Peptide35 only in the case of cross-strand coupling.

In contrast, Peptide56 shows a CD spectrum (see Fig. 3.12) similar to our synthesised helical peptides and indicates a fully random coil structure. The typical features of a folded peptide, a strong band at 229 nm in combination with the zero crossing at 222 nm and a strong negative signal in the region below 222 nm could not be observed. The signals are weak and close to the baseline and show a weak minima around 210 to 215 nm. One has to speculate that through the large number of histidine residues in the loop region and repulsing forces between their side chains, caused by the flexible protonation state of histidine, secondary structure is lost. So the sequence arrangement in the loop has to change dramatically. Additional stabilising interactions to overcompensate repulsing forces have to be introduced in the strand sequences. Both hydrophobic residues as well as polar interacting amino acid sidechains can contribute to secondary structure and have to be tested.

The most promising candidate is Peptide71, which possesses only one histidine residue in the loop but four histidine in the strands. Appearance of secondary

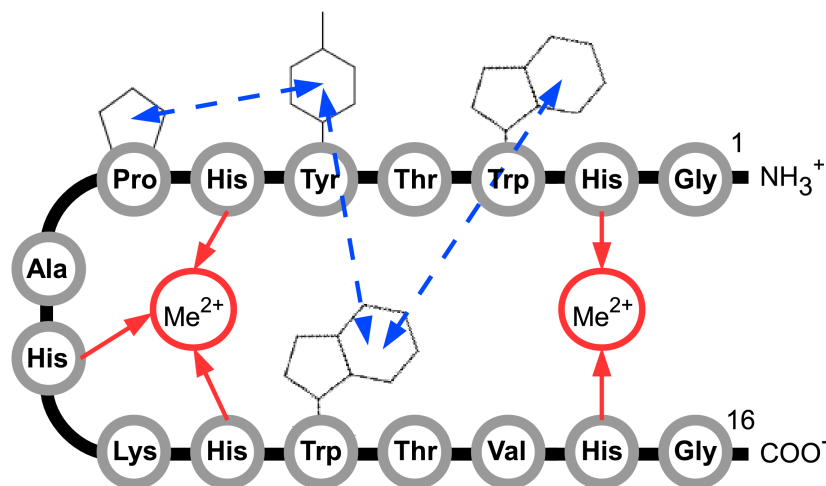


Figure 3.10: Illustration of the interactions of amino acid side chains that influences Peptide71 folding to a β -hairpin. The aromatic amino acids (and proline) are shown as sketch. The chelated divalent metal ions are shown as Me^{2+} . The interactions are highlighted by arrows.

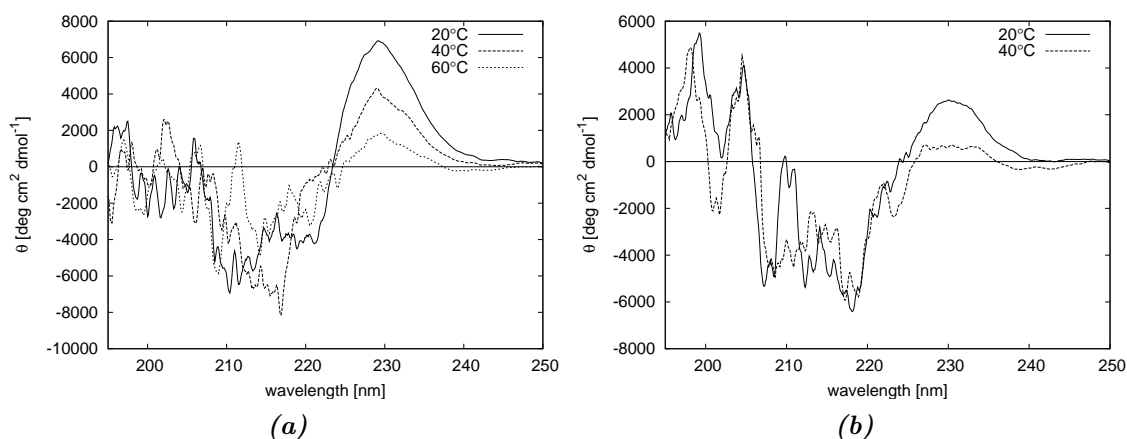


Figure 3.11: CD spectra of a) Peptide32 and b) Peptide35 were collected at pH 5.5 in 20 mM PO_4^{3-} buffer at various temperatures. The solution was free of metal ions for complexation. The decreasing signal at 229 nm indicates peptide melting.

structure was analysed by CD spectroscopy (Fig. 3.13) in the absence as well as the presence of Ni^{2+} ions. First, samples of native Peptide71 (100 μM) dissolved in PO_4^{3-} buffer at pH 5.5 were analysed and show the characteristic small band at 229 nm and the temperature-independent intersection point at 222 nm. The spectra attest again edge-to-face aromatic interaction of Trp and Tyr. Therefore a cross-strand coupling has to exist. This result indicates an equal appearance compared with Peptide32 and Peptide35 in aqueous solution. Surprisingly by adding only 1 eq. of NiSO_4 the signal at 229 nm is increased dramatically. Additionally a minimum at 215 nm could be found in the spectrum which indicates a anti-parallel β -sheet (see Fig. 2.11). The CD spectrum of the metal bound Peptide71 is almost identical to the spectra of the peptides Chignolin and CLN025 reported by Honda et al. [39, 40].

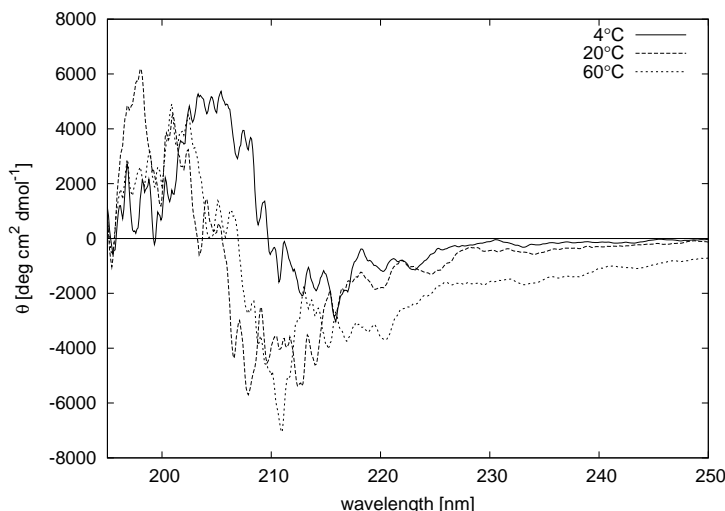


Figure 3.12: CD spectra of Peptide56 were collected at pH 5.5 in 20 mM PO_4^{3+} buffer at various temperatures. The data shown were collected in the absence of metal ions for complexation. The signal is typical for a random coil peptide. Spectra of samples containing up to 5 eq. NiSO_4 show similar shape (data not presented). One has to assess that the folding is lost.

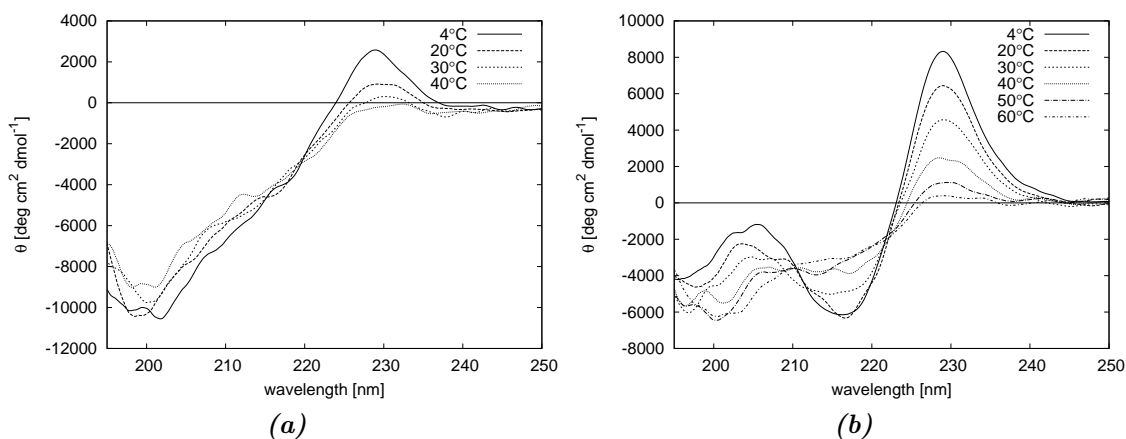


Figure 3.13: The Peptide71 (100 μM) was dissolved in PO_4^{3+} buffer at pH 5.5. These CD spectra of Peptide71 were collected *a*) without NiSO_2 and *b*) with 1 eq. NiSO_4 in the sample. The band at 229 nm is increased dramatically due to stronger interaction of the aromatic side chains Tyr and Trp. The folding of Peptide71 gets more pronounced and therefore the melting temperature of the peptide is increased.

They unequivocally demonstrated by X-ray crystallography and 2D-NMR that these peptides form a β -hairpin. However, the peptides Chignolin and CLN025 did not contain any histidine residues.

For Peptide71 it was not clear if the metal is bonded to histidine residues in just the loop or in the strand region as well, but chelation of Ni^{2+} that makes the interaction of the aromatic side chains more efficient have to occur. However, bonding of Ni^{2+} has to have influence the folding of the peptide backbone. Fig. 3.13b shows a metal depending folding of the peptide. One has to speculate that in aqueous

solution the peptide is pre-organised in a bended arrangement because Tyr and Trp show weak interacts in the CD spectrum. Once the metal ion is bonded Peptide71 is fully folded in a β -hairpin arrangement. So the folding becomes well-structured by chelation of metal ions. To conclude, Peptide71 is to the author's knowledge the only example of a histidine-metal chelation induced β -hairpin folded short polypeptide.

σ_A weighted map

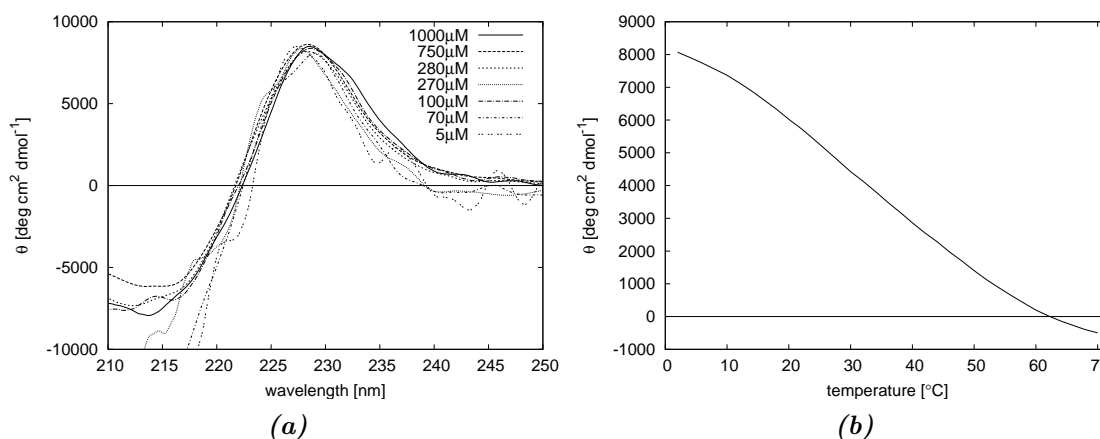


Figure 3.14: a) The CD spectrum was collected at 20 mM PO_4^{3+} buffer at pH 5.5 and 10 mM ZnCl_2 at various concentrations of Peptide71. The shape and magnitude of the spectra indicate a concentration-independent effect of cross-strand $\pi - \pi$ interaction. b) Melting curve of Peptide71 monitored at 229 nm respectively. Peptide71 ($\sim 100 \mu\text{M}$) was dissolved in 20 mM PO_4^{3+} buffer at pH 5.5 and 1 eq. NiSO_2 was added. The temperature was slowly increased from 4 to 70 °C in steps of 2 °C.

CD spectra at various concentrations were collected to exclude interactions between aromatic side chains of different peptides. As shown in Fig. 3.14a the signal at 229 nm is equal in wavelength and magnitude at concentrations from 5 μM up to 1 mM and indicates that an effect depending on concentration does not significantly perturb. Therefore the peptide is monomeric in aqueous solution and aromatic side chains interact intramolecularly with each other. Furthermore thermal stability of Peptide71 was analysed at the most significant band of 229 nm, indicating the β -hairpin formation, using a thermal gradient from 2 °C to 70 °C. The sample was heated up in 2 °C steps and allowed to equilibrate for 5 min. A decreasing signal could be detected indicating destruction of the aromatic interaction and therefore a melting of Peptide71. In contrast the unchelated peptide melts much faster than the unchelated peptide and indicated a strong structural stabilisation through histidine chelation (see Fig. 3.13a and 3.13b. By analysing the melting curve of metal chelated Peptide71 (see Fig. 3.14b) a two-state equilibrium have to be assumed because of a shallow transition. The melting temperature was calculated to $T_T=33.2^\circ\text{C}$ and is therefore 5.8 °C lower than the melting point of peptide Chignolin, reported by Honda et al. [39]. The fact of peptide melting is not fully reversible due to specific peptide composition. As shown in Fig. 3.15 the signal at 229 nm is significant lower after

the heating cycle. The minimum at about 215 nm which indicates an anti-parallel β -sheet (see Fig. 2.11) is lost. One has to assume that in the step of temperature decreasing and refolding the chelation of histidine residues by the Ni^{2+} is much faster than folding of the backbone in a correct way. Therefore annealing induces misfolding. The metal ion is still chelated by the histidine residues but the peptide backbone is defolded.

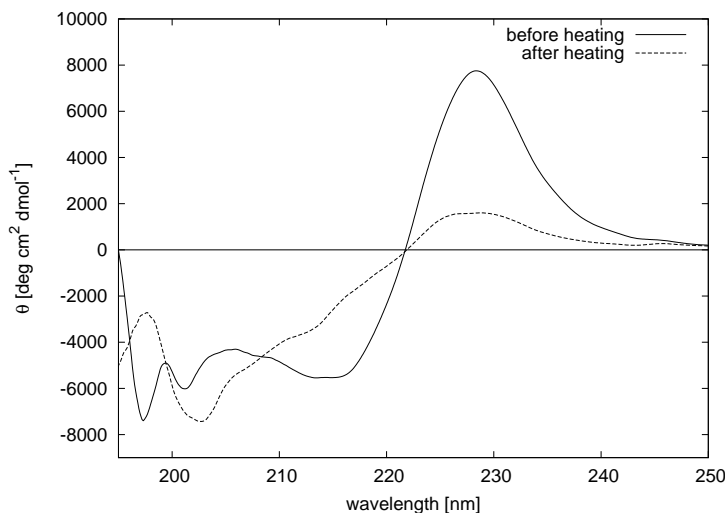


Figure 3.15: CD spectra of Peptide71 (100 μM) collected in 20 mM PO_4^{3-} buffer at pH 5.5 and 1 eq. NiSO_4 . The spectra were taken at 2 $^\circ\text{C}$ and show Peptide71 before and after heating to 70 $^\circ\text{C}$.

2D-NMR

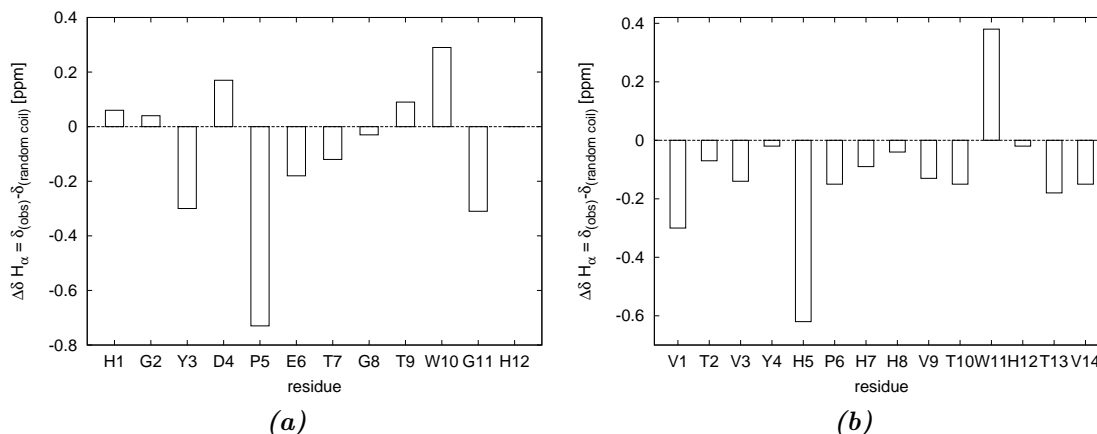


Figure 3.16: Differences in NMR C_αH shifts ($\delta_{\text{C}\alpha\text{H}} = \delta_{\text{observed}} - \delta_{\text{random coil}}$) for Peptide32 and Peptide56.

The differences in C_αH -shifts from random coil structure shifts, reported by Wishart et al. [95], were calculated for the TOCSY data of Peptide32 and Peptide56 and are shown in Fig. 3.16. In the case of Peptide32 negative values at residue 5-8 are indicative for a chain-bend region. Here the [4:4] loop region is located. Positive

values indicate strand regions and can be detected for residue 1-4 as well as 9-12. With single exception of residues at position 3 and 11 the profiles are characteristic for a [4:4] hairpin. The NMR data of Peptide32 are in good agreement with the data obtained by CD spectroscopy and fit very well to results reported by Honda et al. [39].

In the case of Peptide56 all shifts are negative with the single exception of Trp11. A little discrimination of a bent loop region could be found. Compared to the $C_{\alpha}H$ -shifts for Peptide32 the region of residue H5 to H8 is very similar. At this position the proposed β -hairpin loop region is located. However, $C_{\alpha}H$ -shifts which indicate a strand region or any other secondary structure could not be found. Therefore Peptide56 has to have a random coil structure with a bent region involving the proline residue. So, NMR data are consistent with the CD spectroscopy data. One has to assess that Peptide56 is unfolded in aqueous solution.

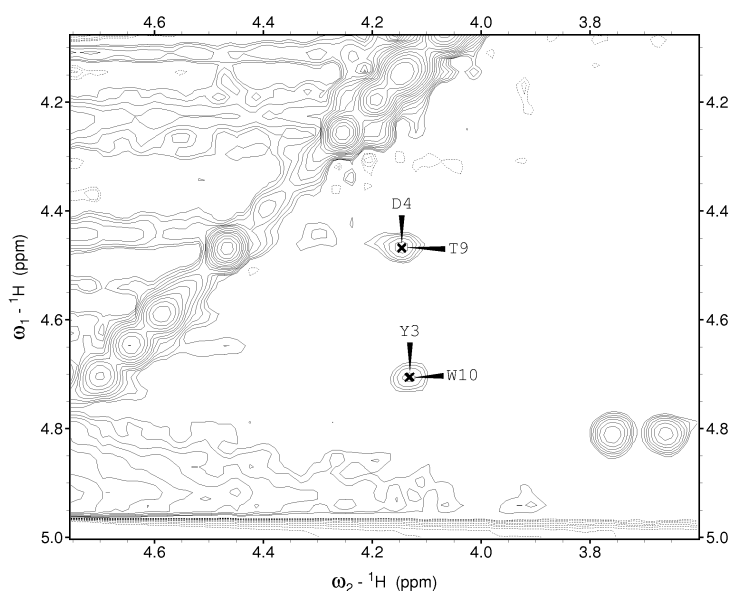


Figure 3.17: Region of the NOESY spectrum of Peptide32. Non-sequential $C_{\alpha}H$ - $C_{\alpha}H$ cross peaks are shown.

Beside the $C_{\alpha}H$ -shifts long-range NOEs of $C_{\alpha}H$ are indicative for hydrogen atoms that are close to each other. For Peptide32 non-sequential cross peaks could be found in the NOESY spectrum. These long-range $C_{\alpha}H$ - $C_{\alpha}H$ NOEs between residue D4 and T9 as well as Y3 and W10 attest a formed antiparallel β -sheet (see Fig. 3.17). In the case of our short peptides this antiparallel sheet involving both termini of the peptide can only be formed if the peptide is strongly bent, so it has to be a β -hairpin. Further NOEs could be found in that indicate a close distance between the aromatic side chains of residue Y3 and W10.

TOCSY and NOESY data of native Peptide71 were collected as described above (see Sec. 2.4.4). As already known from the CD spectroscopy Peptide71 is partly unfolded in the metal ion free state. Unfortunately the NMR data show that metal ion free Peptide71 is highly dynamic and the peptide appears to be in

at least two different conformations. Additionally the amount of histidine residues which show signals at similar shifts is very high. In the peptide structure two Thr, Trp and Gly residues are present as well. However, the TOCSY and NOESY spectra of Peptide71 are crowded. NMR data collected at a higher temperature ($\sim 40^\circ\text{C}$) are usually more sharp, because of the higher thermal motion signals that are close together and represent independent conformational states are merged in only one signal. Thus the spectrum becomes more simple and more clear. So NMR data of peptides and proteins are often collected at higher temperature. On the other hand the conformational shape of Peptide71 was proven to be dependent on the temperature. The higher the temperature the lower the percentage of peptide that shows β -hairpin folding. From that point of view the temperature for a data collection in aqueous solution should be as low as possible. However, the resolution was too low and therefore the spectra could not be fully interpreted.

TOCSY and NOESY spectra of Peptide71 were also collected in the presence of transition metal salts. A low concentration of 10 mM CoCl_2 and in an other experiment 10 mM CdAc_2 were added. Because these metal salts were not isotopically pure paramagnetic nuclei were introduced in the NMR experiment. It is a fact that paramagnetic nuclei are responsible for a strong broadening of the NMR signals. Also in this case the spectra are very crowded and could not be fully analysed so far.

IMAC

Binding affinity towards IMAC support was analysed as described above (see Sec. 2.4.2). One has to keep in mind that all these cartridges have maximal loading capacities. Therefore for analytical data total amount of peptide was always much lower than 1 mg. Both Peptide32 and Peptide35 show low IMAC affinity using various transition metals immobilised on the chromatography support and could be eluted with 20 mM imidazole, already present in the washing buffer (data not presented). These result indicate that two respectively three histidine residues are not sufficient for purification applications. In the case of Peptide32 the histidine residues are located at the termini of the peptide and are very flexible. One has to speculate that these residues are therefore not able to chelate a metal ion. They are too far apart. In Peptide35 a high content of Asp and Glu increase the concentration of imidazole necessary to elution. The phenomenon was already reported by Kozlov et al. [51]. However, the content of histidine residues is not sufficient for purification.

IMAC properties of unfolded Peptide56 were obtained using Ni^{2+} , Zn^{2+} and Cu^{2+} loaded chromatography support. As a standard a synthesised His-tag peptide (Peptide49) with the sequence (*MKHHHHHHQVRY*) invented by Qiagen was used. The chromatograms (3.18) show a high affinity for the His-tag (200 mM) using Ni^{2+} as expected. The concentration of imidazole needed to elute Peptide56 depends on the metal ions and decreases as follows: $\text{Ni}^{2+} > \text{Cu}^{2+} > \text{Zn}^{2+}$. For HiTrap-FF cartridges loaded with Ni^{2+} two peaks for Peptide56 could be found indicating two different chelation states of the peptide. Peptide56 should be able to complex a

metal ion with two different centres, one containing three histidine residues in the expected loop region and a second centre with one histidine residue on a strand. In fact, the absolute concentration of imidazole is high enough (150 mM) in the case of Ni^{2+} to ensure sufficient purification.

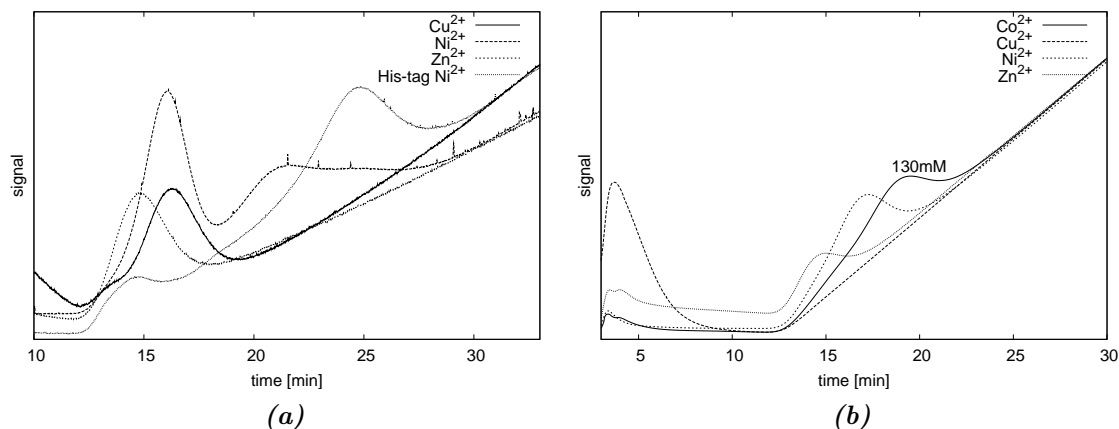


Figure 3.18: a) IMAC chromatogram of Peptide56 using HiTrap-FF cartridges loaded with various metal ions. Elution conditions are depending on metal ion and differ in concentration of imidazole. Ni^{2+} shows highest binding affinity and binds equal strong than a His-tag peptide. b) Peptide71 was analysed using Qiagen NTA-Sepharose. Here IMAC chromatograms of NTA-Sepharose loaded with Cu^{2+} , Zn^{2+} , Ni^{2+} , and Co^{2+} are shown. Using Cu^{2+} loaded cartridges Peptide71 was not retarded.

For the folded Peptide71 the IMAC data look similar to Peptide56. In the case of Qiagen NTA-Sepharose cartridges for Co^{2+} 130 mM imidazole is needed whereas for Ni^{2+} and for Zn^{2+} the peptide is much weaker bonded. For Cu^{2+} loaded resin the peptide was eluted during the washing step using 20 mM imidazole. Further experiments were carried out comparing influence of smaller bead diameter, resulting in closer support packing, using GE HiTrap-HP and HiTrap-FF material (see Fig. 3.19). In the case of GE HiTrap-FF cartridges for Ni^{2+} loaded support almost 130 mM imidazole is needed to elute. This is due to only four histidine residues compared to six for the His-tag. Cu^{2+} loaded support binds Peptide71 much weaker and is not sufficient for purification. Using GE HiTrap-HP cartridges loaded with Cu^{2+} , Zn^{2+} , Ni^{2+} , and Co^{2+} Peptide71 shows very high affinity. For all four metal more than 130 mM imidazole, in the case of Co^{2+} almost 200 mM imidazole, is needed to elute. A fine structure is observed in the chromatogram indicating again different chelation states. To conclude, the affinity properties in the case of HiTrap-FF are slightly different than using HiTrap-HP because elution concentration is decreased by almost 50 mM imidazole. In general, the concentration to elute Peptide71 is slightly lower in the case of Qiagen Sepharose cartridges than by using GE IDA chromatography support. This is caused by using a bidentate chelater instead of tridentate chelater like the GE IDA chromatography support. Due to a high support density of Qiagen Sepharose support, much higher than HiTrap-FF and HiTrap-HP, absolute concentration of imidazole are comparable. One has to conclude that Pep-

tide71 is sufficient for almost all commonly used metal ions. The concentration of imidazole to elute the peptide is high enough for high purification resolution and comparable with the classical His-tag.

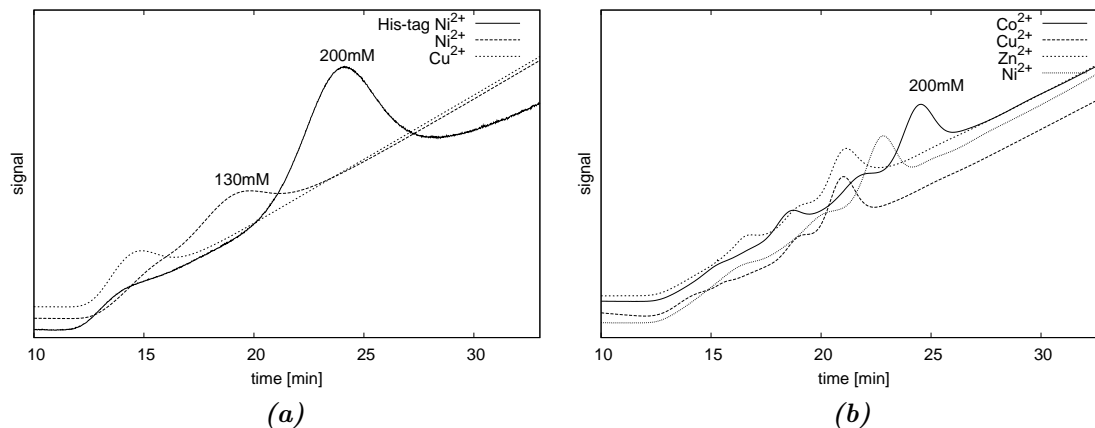


Figure 3.19: IMAC chromatograms of Peptide71 using various loaded metal ions are shown. The peptide sample was loaded using the wash buffer containing 50 mM NaH₂PO₄, 300 mM NaCl and 20 mM imidazole. For elution the concentration of imidazole was slowly increased up to 500 mM. For *a*) HiTrap-FF support the bonding of Peptide71 is weaker than using *b*) HiTrap-HP support due to lower support density.

Crystal structures

Crystal structures of polypeptides are very rare in structural databases. Most structural studies are carried out in solution using NMR techniques. Polypeptides are often more flexible. In solution an equilibrium between different conformation states can often be observed. A high degree of conformational freedom caused by many similar energetic minima are present in peptides. The molecule size is in a range between small molecules and proteins and interactions are often not specific. Because of short chain length polypeptide are not able to form hydrophobic cores (with stabilising salt bridges) like globular proteins.

For crystallisation trials small molecule as well as protein methods were used. Various organic solvents (methanol, ethanol, ether, THF, toluene, acetonitrile, methylene chloride and dioxane) and water as well as mixtures of those solvents were used to grow crystals by slowly evaporation. Additional trials were carried out by slowly cooling to -85 °C. Thus supersaturation of the solvent is reached by decreasing the solubility. However, native peptide crystals did not grow using organic solvents. Trial to get peptide crystals with additional metal ions (ZnCl₂, ZnAc₂, NiCl₂, CuCl₂) at various concentrations always gave inorganic salt precipitation or crystals.

Methods usually used for protein crystallisation were used as well. Peptide samples up to 50 mg/ml dissolved in buffered solutions were prepared. By using crystallisation robots and sparse-matrix screens as described above the number of crystallisation trials could be increased. These trials were carried out at

room temperature as well as 4°C. Furthermore systematic grid-screens containing pH or concentration gradients were designed and applied for the trials. Trials were also carried out by using additional transition metals that should be complexed by histidine residues. Several times crystal structures of ZnNH_4PO_4 were accidentally determined. However, several 10,000 trials were carried out without any success of obtaining peptide crystals. Therefore fusion proteins containing the most promising peptides were designed.

4. Fusion proteins of MBP with β -hairpin forming peptides

Crystals of the synthesised peptides were not obtained and therefore getting structural information using X-ray crystallography was not possible. To validate structural data collected by CD spectroscopy and NMR, fusion proteins made of a peptide and a carrier protein were designed. As fusion partner the well-described globular protein MBP was used. By using these protein-peptide complexes, growing crystals that contain the peptide in well-ordered form is more likely.

4.1 MBP as fusion partner

The maltose-binding protein (MBP) was chosen as a fusion partner due to various advantages. MBP is a highly soluble, easy-to-handle globular protein with exposed termini. Therefore it is suitable for tests for both C- and N-terminal fusion proteins with various protein tags. For cloning and expression at enough yields an established protocol could be used. Optimised purification procedures using MBP-selective maltose columns in combination with ion chromatography and polishing using gel filtration ensure high purity. And finally, crystallisation conditions of MBP and manyfold fusion proteins are known already and only differ marginally from each other. Usually conditions including PEG at a pH range from 6 to 7 were used for crystallisation. Since MBP is large compared to the peptide tag one could expect that at least some of these conditions produce crystals of the fusion constructs. Concluding all these advantages, MBP was the platform of choice to test my peptides.

4.2 Fusion of MBP to peptides mimicking protein tags

Based on analytical data two peptides, Peptide56 and Peptide71, were chosen to be fused with MBP. The designed fusion proteins were recombinantly expressed and only differ in their linkage towards MBP (see Fig. 4.2). Of special interest was a comparison with already crystallised MBP-His-tag fusion proteins (see Fig. 4.1).

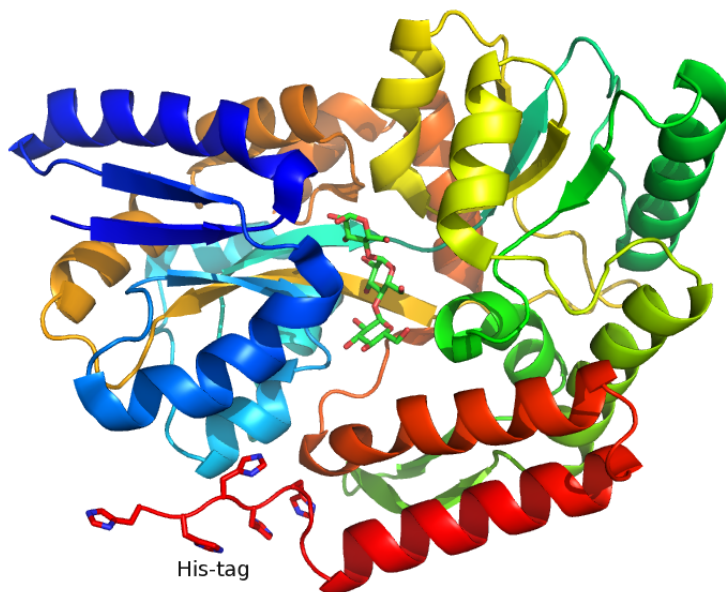


Figure 4.1: Crystal structure of a *closed form* MBP fusion protein published by Cuneo et al. [22] containing an uncleaved His-tag (PDB-ID:2FNC) and bound maltotriose. The structure was solved in the monoclinic space group $P2_1$ to a resolution of 1.7 Å. The affinity tag was fused on the C-terminus and shows an ordered random coil structure in the crystal. The tag is located in the voids in the crystal lattice.

Peptide56 is a potential candidate because of its high IMAC affinity, although expected not to have an intrinsic β -hairpin structure in solution as shown in CD spectroscopy and NMR. Because short polypeptides tend to be flexible and unfolded, in complex with a protein Peptide56 should probably form a secondary structure. Nevertheless as a first try an N-terminal fusion with MBP with Peptide56 (MBPCG1) was carried out to validate both structural arrangement and affinity properties. With respect to the analytic results of MBPCG1 a second fusion protein was designed. It was shown that MBP's helix H20 influences the secondary structure of Peptide56 in a way that the very first amino acids of Peptide56, that are linked to MBP, adopt a helical arrangement. Thus the formation of a β -strand or β -hairpin region in the peptide is strongly inhibited. There are two different ways to prevent this interaction, firstly by introduction of a proline or a glycine residue between helix H20 and peptide or secondly by addition of a long linker sequence. Proline and glycine are very uncommon in α -helices and are known as helix "breakers" (Chou and Fasman [17]). A long linker separates protein and peptide in space and dis-

ables structural interactions. Various linkers, different in length and composition, are known in the literature. Most common are sequences for proteolytic cleavage and poly-Ala sequences. Because a helix “breaker” should be introduced the linker sequence *GSSGSSG* was selected for testing flexibility requirement of Peptide56 producing fusion protein MBPCG2. The used linker sequence was reported by Gilbreth et al. [33] in a Monobody YSX1 and MBP fusion complex (PDB-ID:3CSB).

		360	370	380	390
1FQA_A		ALKDAQTRIT	K-----	-----	----
MBPCG1	Peptide56	ALKDAQTRIT	K-----VT	VYHPHHVTWH	TV--
MBPCG2	Peptide56	ALKDAQTRIT	KGSSGSSGVT	VYHPHHVTWH	TV--
MBPCG3	Peptide71	ALKDAQTRIT	K-----GHWT	-YHEAKKHW-	TVHG
MBPCG4	Peptide71	ALKDAQTRIT	K-----PHWT	-YHEAKKHW-	TVHG

Figure 4.2: Sequence alignment of MBP fusion constructs. The C-termini of the fusion proteins, starting at residue 360, and the fused peptides is shown. The wild-type MBP structure (PDB-ID:1FQA, Duan et al. [25]) contains 371 residues. The linker sequences are highlighted in *yellow*, the histidine residues in *green*, conserved residues from the model peptide Chignolin in *red* and the additional Trp residue in Peptide71 in *blue*.

In the next step Peptide71 (MBPCG3) was fused with MBP without a linker sequence. Peptide71 is known to form a β -hairpin in aqueous solution. This behaviour should be validated in the crystal. The protein was expressed and purified as described above (see Sec. 2.3). After analysing IMAC data and crystal structures of MBPCG3, it was concluded that, again, a linker is needed to separate the peptide from the protein. The linker sequence in MBPCG2 (*GSSGSSG*) appeared to be too long and conformationally flexible. For this reason no additional linker sequence was introduced in MBPCG4 but the first amino acid of the tag was mutated to proline (G372P). In the fusion protein MBPCG4 this proline residue separates the protein and the peptide and hinders the peptide to be influenced by MBP’s helix H20.

4.2.1 Metal affinity chromatography of fusion proteins

Metal ion affinity data were collected as described before (see Sec. 2.4.2). Unloaded 1 ml HiTrap-HP cartridges from GE Healthcare were loaded with Ni^{2+} , Zn^{2+} , Co^{2+} , or Cu^{2+} . Protein samples were eluted with the standard stop-flow method using buffer A and buffer B.

In contrast to Peptide56, its fusion proteins MBPCG1 and MBPCG2 show a dramatically lower IMAC affinity using Ni^{2+} loaded HiTrap-HP columns. As shown in Fig. 4.3 the protein samples are eluted with 20 mM imidazole already present in the wash buffer A. One has to conclude that MBPCG1 as well as MBPCG2 did not interact well with the Ni-IDA support. Since Peptide56 is a highly affine peptide in the unfused form (see Sec. 3.3.3) in its fused form (MBPCG1 and MBPCG2) the tag has to be inhibited to chelate Ni^{2+} ions. One could speculated that in the fusion protein without a linker sequence (MBPCG1) Peptide56 is badly orientated and sticks to the protein surface, so that the peptide is not accessible for the IMAC

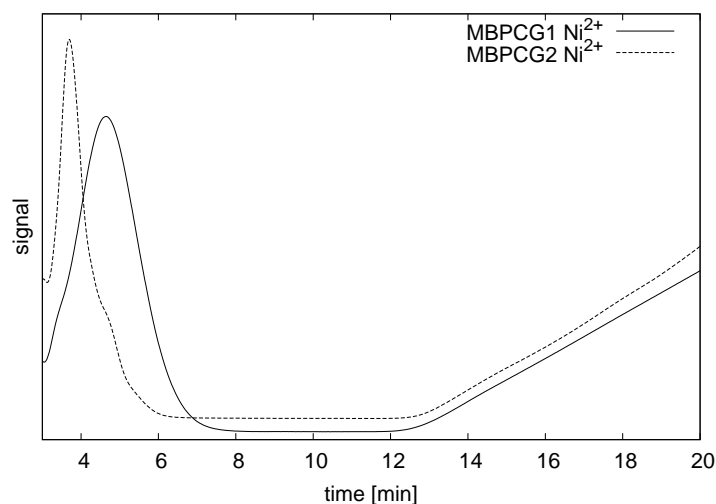


Figure 4.3: IMAC data of fusion proteins MBPCG1 and MBPCG2 using Ni^{2+} loaded HiTrap-HP cartridges. These proteins are fused with Peptide56 which was shown to be highly affine to IMAC stationary support. Surprisingly the fusion proteins show no affinity and run through the column with the wash buffer. In the fused form the affinity of Peptide56 is lost.

support. To overcome this problem MBPCG2 was designed with a linker sequence inbetween MBP and the tag. Since MBPCG2 shows a similar IMAC affinity as MBPCG1 and runs through the column without interaction the influence of the linker on the tag affinity seems to be low.

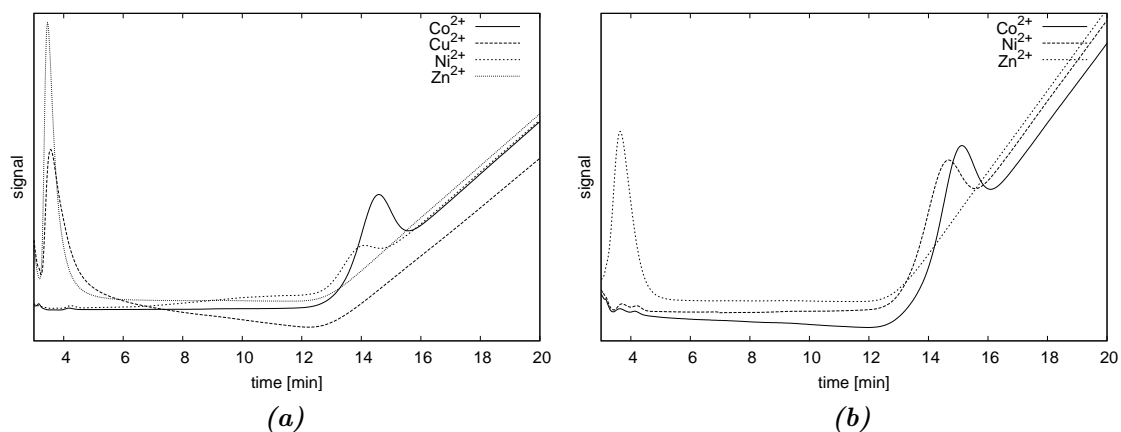


Figure 4.4: IMAC data of fusion proteins MBPCG3 and MBPCG4 on HiTrap-HP cartridges. *a)* MBPCG3 show affinity for Ni^{2+} and Co^{2+} . The concentration is much lower than for Peptide71. Zn^{2+} and Cu^{2+} loaded support did not interact well with the protein tag. *b)* For MBPCG4 the concentration of imidazole is increased by almost 10 mM for Ni^{2+} and Co^{2+} . MBPCG4 runs through the column using Zn^{2+} loaded HiTrap-HP cartridges.

Metal ion affinity data of MBPCG3 and MBPCG4 were collected using the standard IMAC system with HiTrap-HP cartridges and the stop-flow method at 1 ml/min (see Fig. 4.4). In the case of MBPCG3 the affinity of the fusion protein

is much lower than for Peptide71. In contrast to MBPCG1 and MBPCG2 the affinity is commensurably higher. Nevertheless the concentration of imidazole to elute MBPCG3 is increased by almost 110 mM compared to Peptide71. For the metal ions Ni^{2+} and Co^{2+} the elution concentration of imidazole is about 60 mM. In the case of Zn^{2+} and Cu^{2+} the protein MBPCG3 elutes at an imidazole concentration that is 20 mM or lower. For both metal ions imidazole in the wash buffer was sufficient to elute. In a β -strand adjacent residues face opposite sides. So one could speculate that the sticky side of the folded peptide is not facing towards the solvent but to the protein surface. In this case only the histidine residues near the loop region could be bound by the IMAC support which could be a good reason for the low affinity. But by analysing the crystal structures of MBPCG3 this hypothesis could not be confirmed (see Fig. 4.19). However, for five histidine residues a concentration of about 200 mM imidazole would be expected for elution (Kozlov et al. [51]). The second metal chelation centre which is located at the end of the β -strands seems to be inaccessible for the IMAC support. For MBPCG4 the first amino acid of the tag sequence was mutated to proline G372P to lower the flexibility of the tag. By comparing IMAC data of MBPCG3 and MBPCG4 the concentration of imidazole to elute MBPCG4 is increased by about 10 mM to almost 70 mM for Co^{2+} and Ni^{2+} . Again a Zn^{2+} loaded support was not able to retard the fusion protein. However, the mutation G372P improves the accessibility of the tag.

4.3 Crystal structures of MBP fusion proteins

4.3.1 MBPCG1

Crystal form I

The determined protein structure of **crystal form I** (C2) shows MBP in the *closed form* with one maltose molecule bound in the sugar-binding cleft (see Fig. 4.7). In this crystal form eleven amino acids of the tag (Peptide56) are well ordered in the crystal. Only the last three C-terminal amino acids are missing. The tag shows an elongated "L" shaped form with a bent region at the expected loop region near the proline residue P377. The angle of the bent region is close to 90° and a β -hairpin structure of the fused peptide is not present in the crystal structure. The protein tag is involved in crystal contacts. Any two MBP molecules are linked by a fused peptide *via* three hydrogen bonds and hydrophobic interactions (see Fig. 4.6). On the MBP surface a shallow hydrophobic pocket is formed by the α -helices H6, H18 and the β -sheet A (see Fig. 7.1). Here a number of hydrophobic and aromatic amino acids are located. Several residues of the C-terminal tag interact with this pocket. Amino acid T381 interacts with the side chain hydroxy group and the backbone carbonyl oxygen *via* the water molecule O1 with the amid hydrogen of F93. Furthermore the backbone carbonyl oxygen of H378 is bonded to I330 *via* a hydrogen bonding of the amid hydrogen. And the backbone carboxyl oxygen of amino acid H376 of the tag forms a hydrogen bond to the acid hydroxy hydrogen of MBP's Y172. Additionally aromatic interactions are generated by an edge-to-face orientation of tag amino acid

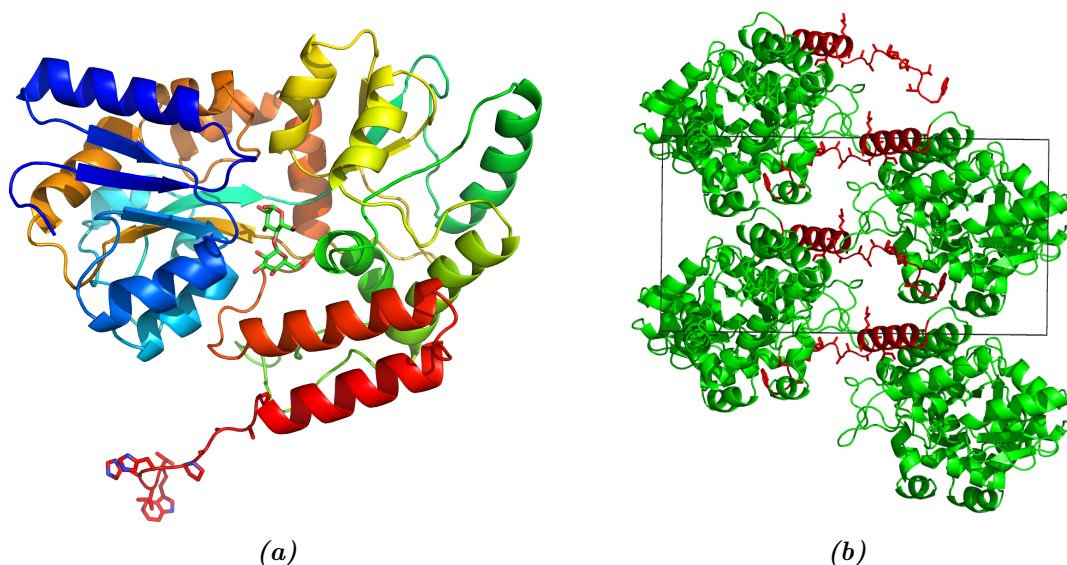


Figure 4.5: Crystal structure of MBPCG1 in the monoclinic space group C2. The C-terminus of MBP is highlighted in *red*. *a)* Closed form of MBP with maltose bound in the sugar-binding cleft. The tag is almost completely ordered and elongated in an “L” shaped form. *b)* Every protein tag interacts with one MBP molecule by hydrogen bonding and hydrophobic interactions. Thus MBP molecules are connected in columns along the b-axis.

W382 towards Y91 and F93 and hydrophobic interactions are formed by P377 and V380.

By linking MBP molecules by the fused peptides multimers are formed. These multimers are biologically not relevant. In these complexes MBP is column-like arranged along the b-axis. These columns are arranged side by side in the crystal to give the space group C2.

Crystal form II

The determined MBP structure ($P2_1$) contains MBP in the *closed form*. One maltose molecule is located in the sugar-binding cleft. The fused peptide is well-defined and eleven residues could be refined. So only three tag residues are missing. An elongated “L” shaped form of the fused peptide was observed. In the crystal the MBP molecules are linked by the C-terminal tags and form multimers. The arrangement is very similar to the multimers found in *crystal form I* (see Fig. 4.8). The tag is again bound to the shallow hydrophobic pocket formed by α -helix H6, α -helix H18 and β -sheet A (see Fig. 7.1). Besides hydrophobic interactions of W382, V380, and P377 several hydrogen bonds between amino acid residues of MBP and the tag are formed. Hydrogen bonds are formed between the amid hydrogen of residue I329 and the carbonyl oxygen of tag residue H378 and the hydroxy group of tyrosine Y176 and the carbonyl oxygen of P377. Additionally tag residue H376 generates two hydrogen bonds involving the backbone. The carbonyl oxygen of H376

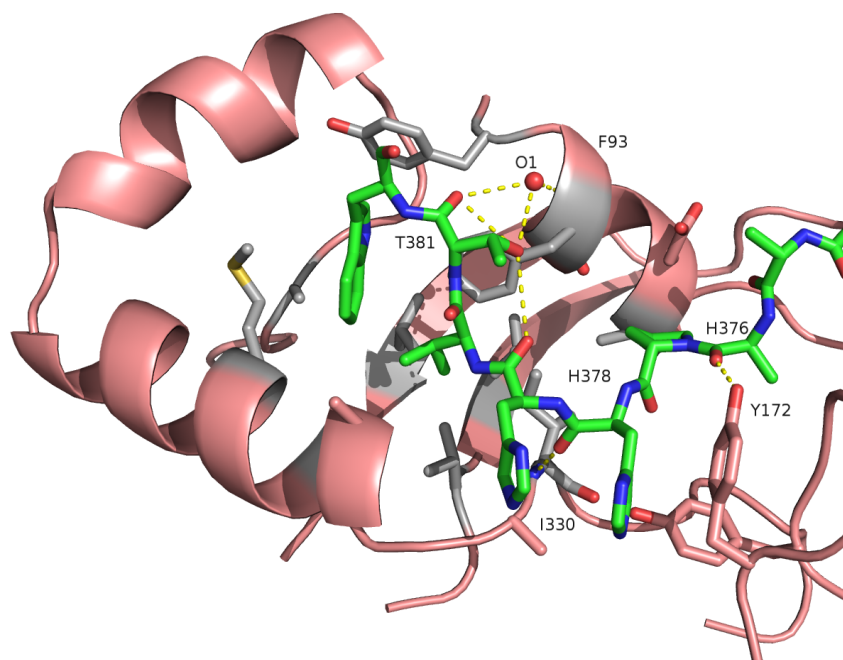


Figure 4.6: Detail of the MBPCG1 crystal structure in the monoclinic space group C2. Hydrophobic amino acid residues (V380, W382, P377) of the fused peptide (*green*) interacts with a hydrophobic pocket of a second MBP molecule. Hydrophobic MBP residues are highlighted in grey. Furthermore, hydrogen bonds between the tag and MBP residues are formed that fix the tag too.

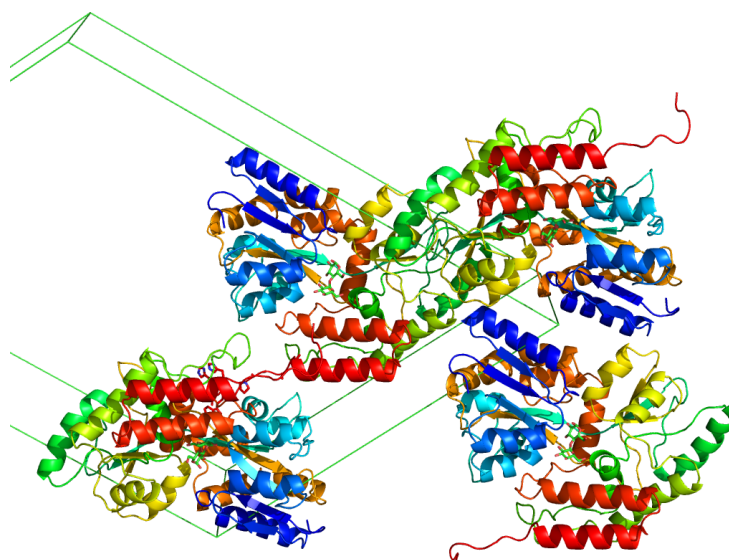


Figure 4.7: Crystal packing of MBPCG1 in the monoclinic space group C2. The C-terminus of the fusion protein is highlighted in *red*. The C-terminal fused peptides interact with the MBP molecules and form crystal contacts.

is bound to Y176 and the amid hydrogen of H376 generates a hydrogen bond to the carbonyl oxygen of G174. Thus the tag is fixed to MBP by four hydrogen bonds.

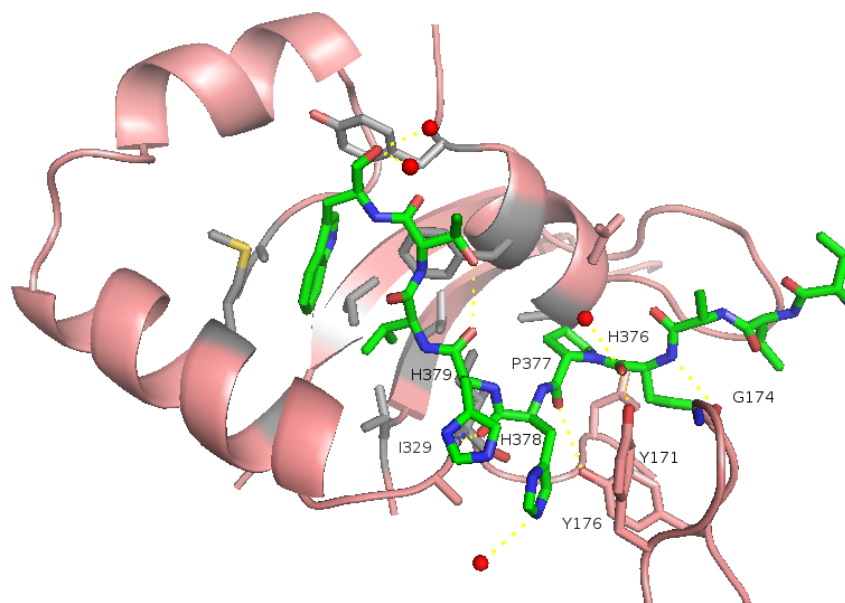


Figure 4.8: Detail of the MBPCG1 crystal structure in the monoclinic space group $P2_1$. The crystal structure in the space group $P2_1$ is very similar to the crystal structure in the space group $C2$. The fused tag (*green*), again, interacts with a hydrophobic pocket (*grey*) at the surface of MBP. The tag is fixed by hydrophobic interactions (V380, W382, and P377) and hydrogen bonding.

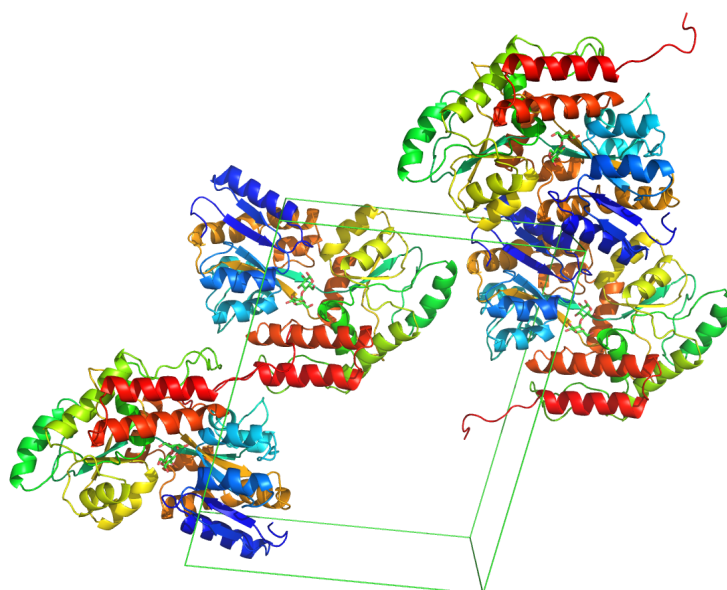


Figure 4.9: Crystal packing of MBPCG2 in the monoclinic space group $P2_1$. The C-terminus of the fusion protein is highlighted in *red*. The C-terminal fused peptide interacts with a MBP molecule and forms crystal contacts.

Not only the tag arrangement but also crystal packing of this crystal form is similar to the crystal packing found in crystal form I. MBP multimers are formed that are arranged in columns around a 2_1 axes in the crystal. Every MBP molecule interacts with its C-terminal tag with the hydrophobic pocket of its neighbour along

the b-axis. The difference between crystal form I and crystal form II is the arrangement of the columns to each other.

Crystal form III

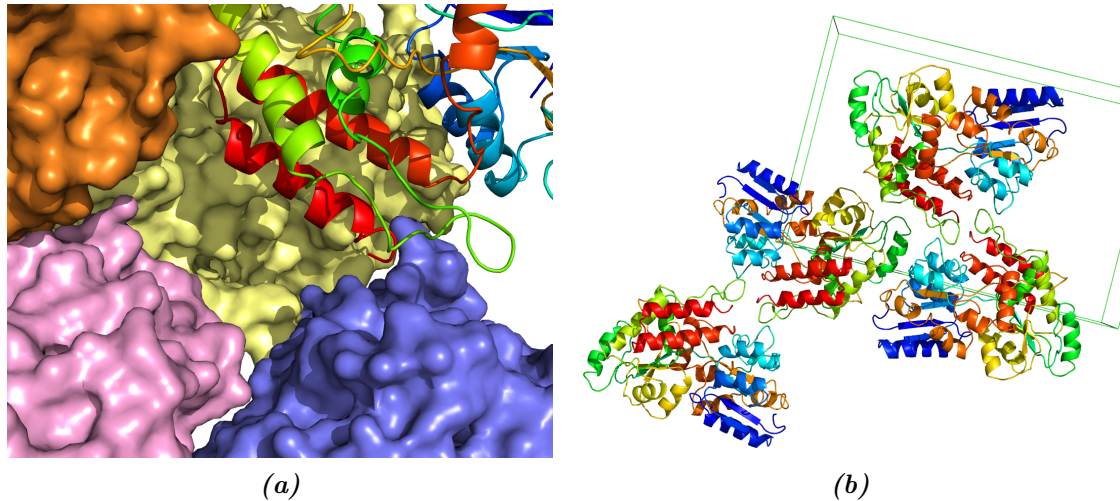


Figure 4.10: Crystal structure of MBPCG1 in the orthorhombic space group $P2_12_12_1$. The C-terminus of the fusion protein is highlighted in *red*. *a)* The tag region is exposed into voids in the crystal lattice. Thus the tag is very flexible and did not interact with the protein. *b)* Crystal packing: The tag region is missing and therefore linking of MBP molecules by the tag could not be observed in the crystal structure.

In this crystal structure ($P2_12_12_1$) MBP is present in *closed form* but without maltose in the sugar-binding cleft. At the position usually occupied by maltose Y155 sticks into the cleft. Thus the position of the ligand is blocked and maltose cannot enter the cleft. It could not be clarified whether the blocked ligand position is a consequence of the missing maltose or if tyrosine blocks the ligand to move into the sugar-binding cleft.

The fused peptide is not present in full length in the crystal structure. Only the first three amino acids could be localised. The electron density of the following tag residues is missing. The visible part of the tag is helically arranged (see Fig. 4.11). One can speculate that the structure of MBP's helix H20 influences the secondary structure of the fused peptide. Tag and protein are not separated by a linker sequence and so interaction of the secondary structures cannot be excluded. However, the major part of the tag is not visible in the electron density and therefore the tag has to be flexible. Because of the flexibility interactions with MBP that generate crystal contacts can be excluded. Compared to the crystal forms I and II multimers of MBP that are linked by the tag could not be found. The MBP molecules are arranged in a way that the fused peptide sticks into a void of the MBP lattice. As shown in Fig. 4.10a the voids in the crystal lattice are so big that the tag is not restricted in motion and is conformationally inhomogeneous. Thus a secondary structure of the fused peptide could not be determined.

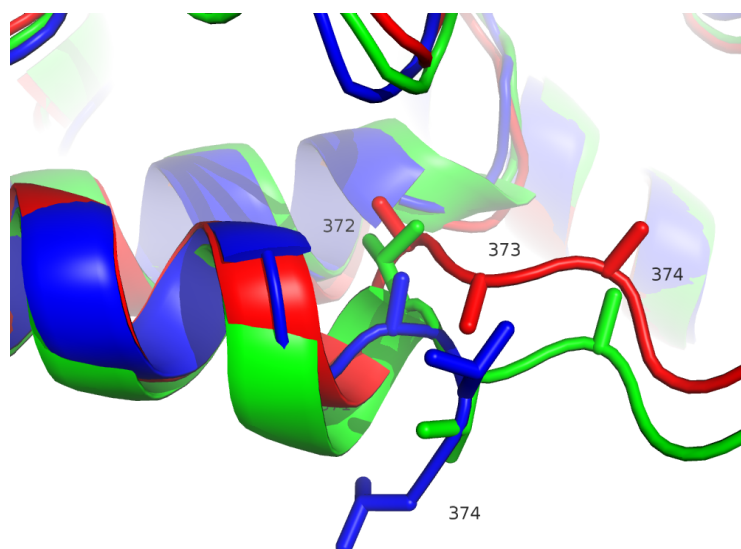


Figure 4.11: Overlay of the C-termini (fusion region) of three MBPCG1 crystal structures. The different crystal forms are coloured as follows: C2 in *red*, P2₁ in *green*, and P2₁2₁2₁ in *blue*. The residue numbers are shown.

Crystals of MBPCG1 grown in crystallisation conditions containing transition metal ions could not be obtained. Also co-crystallisation of a pre-mixed solution of 20 mg/ml MBPCG1 and 100 mM transition metal salts (NiCl₂, ZnAc₂, CoAc₂, CuCl₂) was not successful. Therefore soaking experiments by transferring MBPCG1 crystals in a solution containing transition metal ions were performed. Unfortunately crystals of MBPCG1 dissolved immediately. One could speculate that a tag which chelates a metal ion is not able to link two MBP molecules anymore. As a consequence the crystal contacts are lost and the crystal lattice starts to dissolve.

4.3.2 MBPCG2

Crystal form I

In the **crystal form I** (P1) MBP is present in the *open form*. The fused protein and the linker sequence are not visible in the electron density map. An interaction of MBP with the tag as well as an interaction of two tags could not be observed in the crystal structure. As shown in Fig. 4.12 between the MBP molecules big voids are located that provide a lot of space for the fused peptide. Because of that conformational freedom the tag is not well-ordered.

Crystal form II

In the **crystal form II** (P2₁) MBP is present in the *open form* and only the first residue of the linker sequence (G372) could be observed. Also in this case one has to assume that because of the long flexible linker a localisation of the tag in the crystal structure is not possible. As shown in Fig. 4.13 the C-termini of the

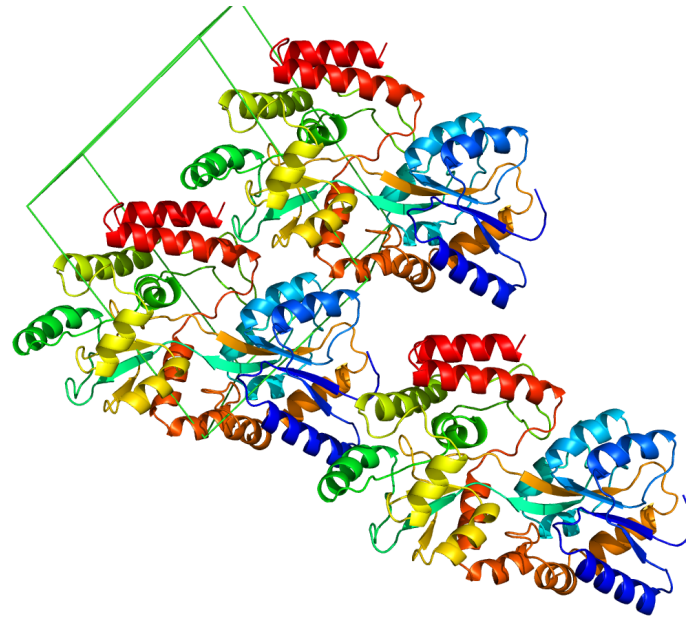


Figure 4.12: Crystal packing of MBPCG2 in the triclinic space group P1. The C-terminus of the fusion protein is highlighted in *red*. The tag and the linker sequence are not visible in the electron density map.

protein molecules are not oriented towards one another. So an interaction of the tags with each other can be excluded. Inbetween the MBP molecules in the crystal lattice big voids can be occupied by the disordered C-termini residues.

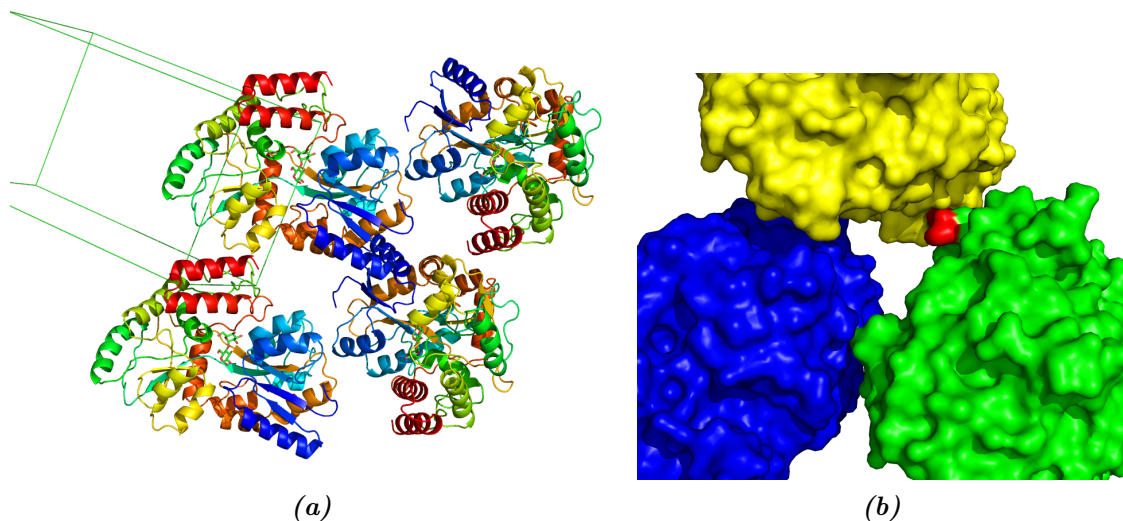


Figure 4.13: Crystal structure of MBPCG2 in the monoclinic space group P2₁. The C-terminus of the fusion protein is highlighted in *red*. *a)* The C-terminus is exposed into voids in the crystal lattice. Unfortunately none of the linker and tag residues could be found in the crystal structure. Thus the tag is very flexible because it did not interact with the protein or another fused peptide. *b)* Crystal packing: The tag region is missing and therefore linking of MBP molecules by the tag is missing as well.

Crystal form III

The refined **crystal form III** (C2) contains MBP in the ligand free *open form*. Surprisingly a cobalt ion was successfully introduced in the crystal lattice due to the soaking procedure. In preceding soaking experiments the protein crystals were rapidly destroyed. Despite the high metal chelation power of the fused Peptide56 the cobalt ion was bound to MBP's residue H39 at the protein surface (see Fig. 4.14). There is no Co^{2+} chelated by the fused peptide in the crystal structure. The bound metal ion is surrounded by three water molecules, a histidine side chain and the carbonyl oxygen atom in an octahedral arrangement. This coordination is typical for Co^{2+} complexes. One can speculate that the tag is still able to chelate a Co^{2+} ion but due to its unrestricted location in a crystal lattice void various conformations are present. Moreover the long linker sequence increases the flexibility of the tag. So one major location and conformation of the tag-metal complex cannot be refined.

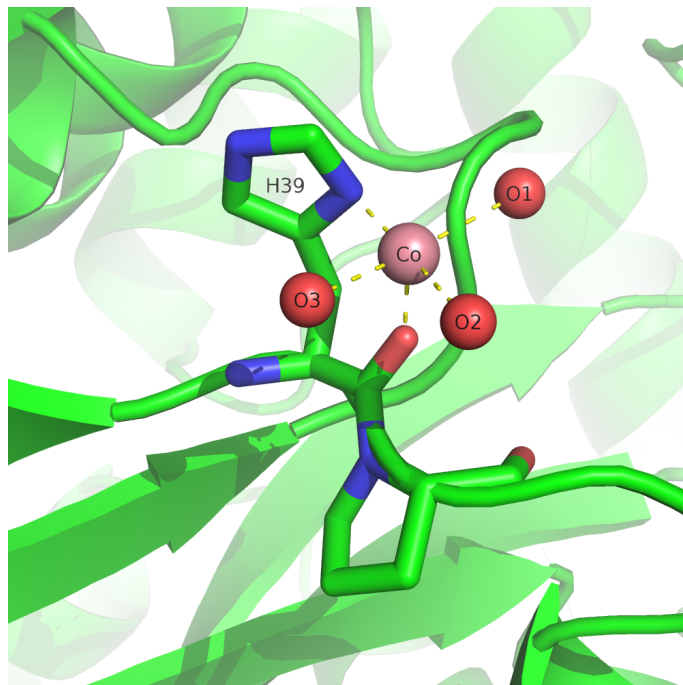


Figure 4.14: One Co^{2+} in an octahedral arrangement is present in the crystal structure. The cobalt ion is bound to residue H39 at the MBP surface and three water molecules.

Once the Co^{2+} was found in the crystal structure the anomalous signal was calculated in XPREP [75]. Because of the short wavelength the calculated anomalous signal was expected to be very low.

Inf	8.0	6.0	5.0	4.4	4.2	4.0	3.8	3.6	3.4	3.2	3.0	2.8	A
1.09	1.05	1.20	1.16	1.17	1.12	1.14	1.27	1.12	1.06	1.08	1.02		
1.45	1.29	1.27	1.16	1.16	1.05	1.25	1.13	1.18	1.10	1.07	1.00		

The anomalous signal divided by the noise ratio is shown for selected resolution shells. The anomalous signal in the second line is based on the input sigma

whereas the third line takes the variances of F^+ and F^- into account. Values higher than 1.3 indicate a reasonable anomalous signal. It is obvious that there is no significant anomalous signal, and no attempt to solve the crystallographic phase problem by an anomalous dispersion experiment was carried out.

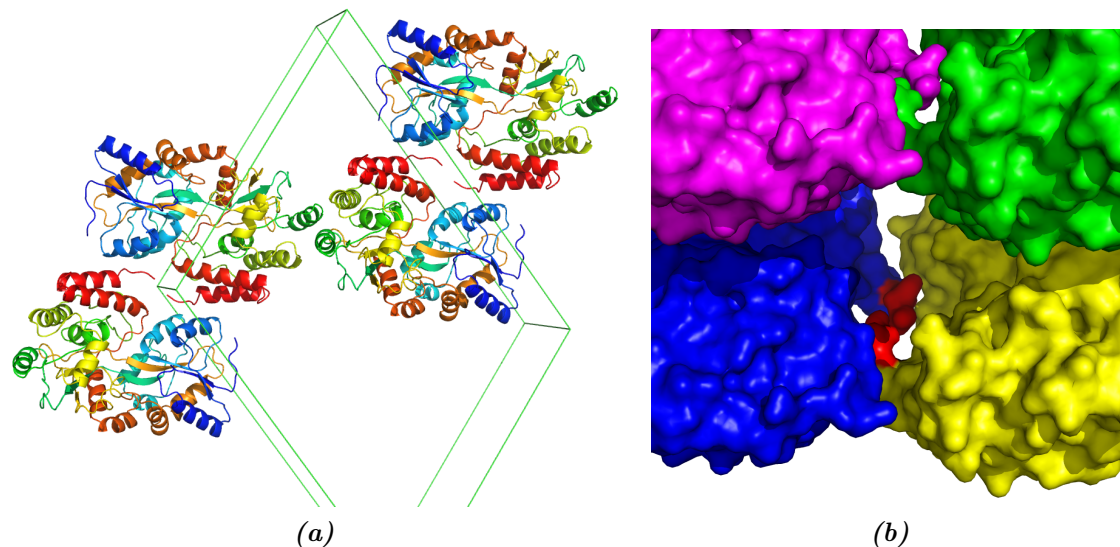


Figure 4.15: Crystal packing of MBPCG2 in the monoclinic space group $C2$. The C-terminus of the fusion protein is highlighted in red. *a)* The MBP molecules are arranged around a 2-fold axis. The C-terminal tag regions come close to each other. *b)* Even though the tags are only partially visible, it can be concluded from their position that two tags sticks to the large voids in the crystal lattice.

In the crystal structure two MBP molecules are arranged in a way that the C-terminal tag regions are close to each other as shown in Fig. 4.15. Thereby the MBP molecules are located around the 2-fold axis. Only the first three residues of the linker sequence are well-ordered in the structure. One residue in the linker sequence is involved in a crystal contact. The carbonyl oxygen of linker residue G372 forms a hydrogen bond to the hydroxy group of S74 of a second MBP molecule (see Fig. 4.16). The missing 18 residues at the C-terminus of MBP have to be very flexible. These residues are located in large voids inbetween the protein layers. Because of missing interaction of the C-terminus (tag and linker) its conformation is unrestricted and therefore inhomogeneous. As shown before in NMR and CD data, Peptide56 is predicted to be rather conformationally unstructured. In the crystal structure the distance between the last visible residues of the linker sequence is about 3.5 Å. Probably the protein chains pass each other without any interaction.

Crystal form IV

In the **crystal form IV** ($P2_12_12_1$) it was found that MBP is present in the *closed form* with a bound maltose molecule. Additionally one Ni^{2+} ion is coordinated to the same histidine H39 which coordinates a Co^{2+} in crystal form III of MBPCG2 (see Fig. 4.17). This histidine residue is located at the surface of MBP and is easily accessible. The nickel ion is not involved in crystal contacts but bound to the

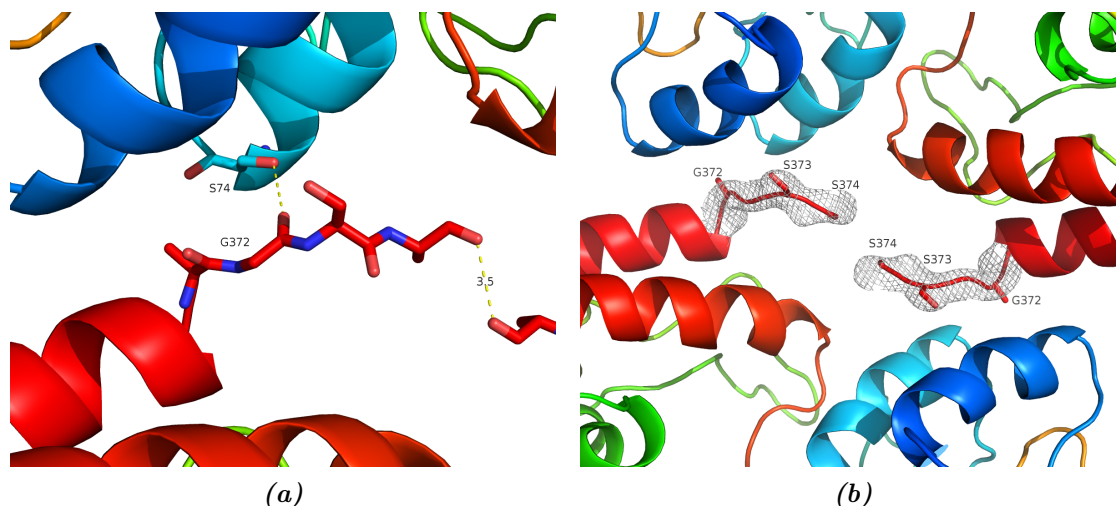


Figure 4.16: Detail of the MBPCG2 structure in the monoclinic space group C2. The C-terminus of the fusion protein is highlighted in red. a) Only three residues of the C-terminal tag could be detected. The residues belong to the linker sequence and form a crystal contact to a second MBP molecule. S74 and G372 form a hydrogen bond. b) The electron density is drawn as a σ_A weighted map contoured at 1.5σ .

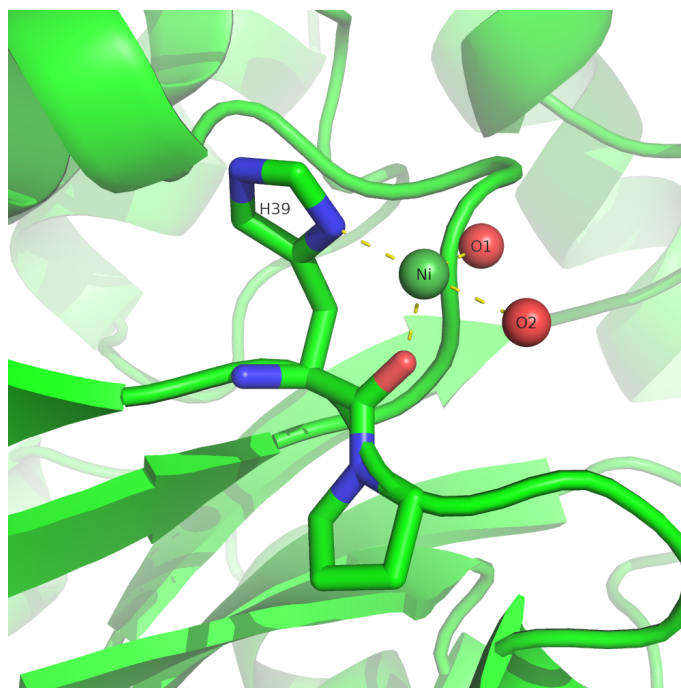


Figure 4.17: A Ni^{2+} ion in an octahedral arrangement was found in the crystal form IV of MBPCG2. The metal ion is bonded to residue H39 and two water molecules.

histidine side chain and the backbone carbonyl oxygen as well. Furthermore it is surrounded by two water molecules O1 and O2 in an octahedral arrangement. Since nickel shows a very low anomalous signal at the Cu- K_α radiation wavelength a single anomalous dispersion experiment (SAD) was not performed. As shown in the output of XPREP [75] an anomalous signal could not be determined. When an anomalous

Inf	- 8.0	- 6.0	- 5.0	- 4.0	- 3.6	- 3.4	- 3.2	- 3.0	- 2.8	- 2.6	- 2.4	- 2.2	Å
	0.98	0.76	0.76	0.90	1.19	1.25	1.10	0.93	0.96	1.18	1.25	1.02	
	1.23	1.08	1.07	1.06	0.95	0.94	0.87	0.87	0.96	0.98	1.01	0.96	

signal is present, the values in the given table should be distinctly above 1.3 and decrease steadily from low to high resolution data.

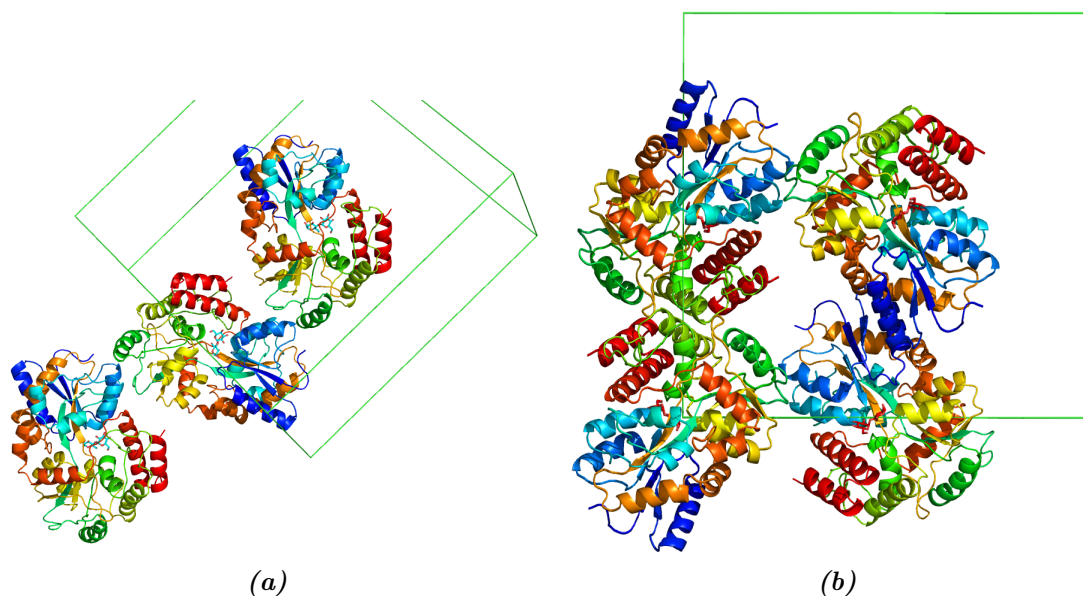


Figure 4.18: Crystal structure of MBPCG2 in the orthorhombic space group $P2_12_12_1$. The C-terminus of the fusion protein is highlighted in *red*. *a*) Overall structure of the *closed form* with maltose bound in the sugar-binding cleft. Non of the C-terminal tag and linker residues could be refined. *b*) Crystal packing: The protein tag (*red*) sticks into large voids in the crystal lattice formed by six MBP molecules. Interaction of the tag with MBP molecules could not be determined.

In this crystal structure the tag as well as the linker are not well-ordered. Non of the tag and linker residues could be localised in this crystal form. The residues could not be refined because of missing electron density. In this crystal structure the arrangement of the MBP molecules indicates that the C-terminal tags are not able to interact with each other. The C-terminus of MBP is exposed to big voids in the crystal lattice (see Fig. 4.18). Consequently the tag, caused by the long linker sequence, is not conformationally restricted. Therefore distinct atom positions of the tag could not be determined.

4.3.3 MBPCG3

Crystal form I

In the triclinic cell of the **crystal form I** ($P1$) two independent MBP molecules in the *open form* are present (see Fig. 4.19). Surprisingly additional electron density was found in the open sugar-binding cleft and was refined as maltose.

Two MBP molecules are arranged such that their C-termini are close to each other. Thereby the molecules form a platter with the C-terminal fused peptides at the bottom. Unfortunately only two residues of the tag could be refined in one of the MBP molecules. So the area above the bottom of the formed platter did not restrict the tags conformationally. The visible C-termini are roughly 20 Å apart from each other. But taking the missing 30 amino acids into account an interaction of the two tags cannot be excluded. On the other hand interactions between the tags should result in a well-defined arrangement which is visible in the electron density map. Furthermore metal ions that chelate the tags by forming a dimer were not present in the crystal growth condition. And as shown in the CD spectroscopy data (see Sec. 3.3.3) the metal ion is essential for the formation of the secondary structure of the tag.

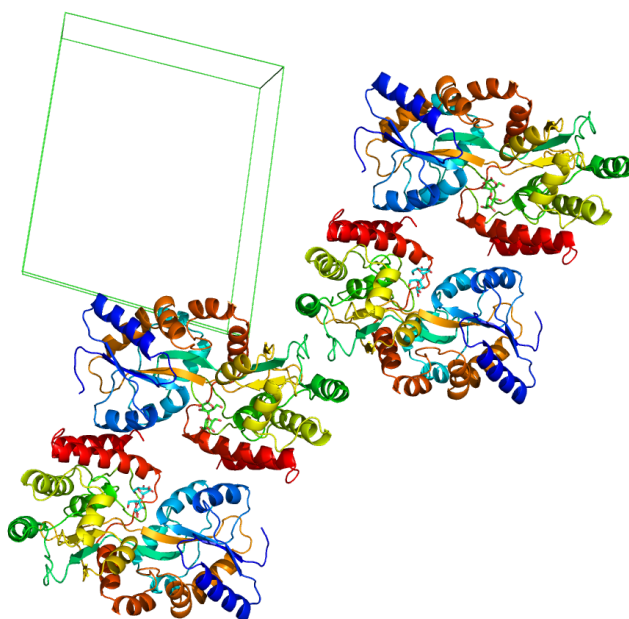


Figure 4.19: Crystal packing of MBPCG3 in the triclinic space group P1. The C-terminus of the fusion protein is highlighted in red. The C-termini are oriented towards each other but interaction of the fused peptides could not be determined because of missing tag residues.

It is uncommon that maltose is present in the *open form* of MBP. There are hints that the tag is folded into the open sugar-binding cleft and maltose is not bound. In that case the tag of one MBP molecule is bound in the open cleft of a second MBP molecule. This is in good agreement with the overall arrangement of the MBP molecules in the crystal lattice. Near the electron density for the maltose molecule various unmodelled electron density blobs in the difference map are present in the cleft region. These blobs can be interpreted as tag residues. One blob is as big as the tryptophan side chain. A trial to build a tag model in the cleft region was unsuccessful as the density is very fragmented. However, maltose was modelled in the open cleft with an occupancy of 50% which gave a better crystal structure refinement with a lower R_{work} ($\sim 1.5\%$) than refining the tag in the binding cleft.

Crystal form I

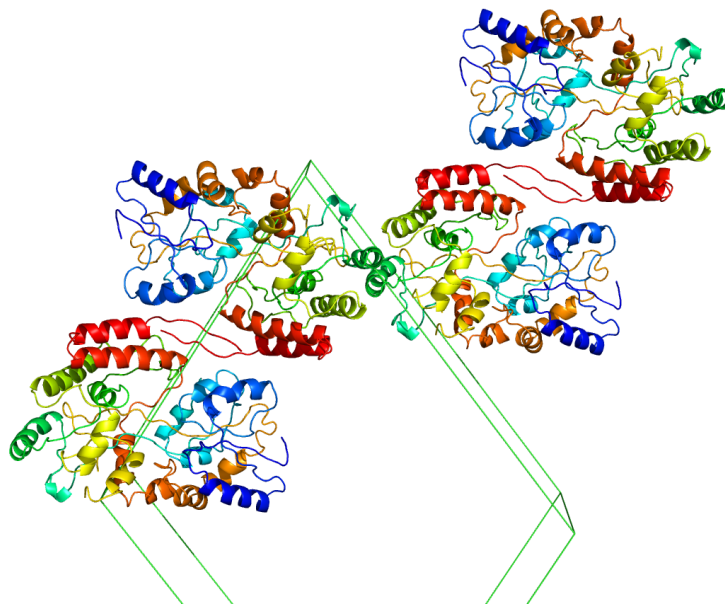


Figure 4.20: Crystal packing of MBPCG3 in the monoclinic space group C2. The C-terminus of the fusion protein is highlighted in *red*. The C-terminal fused peptides interact with each other by linking two MBP molecules together. Because of the low resolution, details are not shown.

In the **crystal form II** (C2) MBP is present in the *open form*. Contrary to crystal form I there is no maltose molecule in the binding cleft. The C-termini of two MBP molecules are very close and interact with each other (see Fig. 4.20). Because of the low resolution of the structure, amino acid side chains of the fused peptide as well as details of the interaction are not shown. However, the interaction of the C-terminal tag can be unequivocally determined (see Fig. 4.21). Only the first seven amino acid residues of the fused peptide could be built. As in crystal form I of MBPCG3, the platter-like arrangement of the MBP molecules in the crystal forms big voids between the MBP layers in which the missing residues are located. Here these residues are conformationally unrestricted and therefore not visible in the electron density map. In the crystal structure the interaction of the seven well-ordered tag residues cannot be provoked by the chelation of a metal ion. Interactions involving aromatic side chains are more plausible. As shown in crystal form I, in the absence of a chelated transition metal ion the secondary structure of the peptide is not fully formed. Therefore the structured and unstructured forms are in balance. In the case of this crystal form the unfolded form can be stabilised by forming a crystal contact with a second unfolded tag.

In conclusion, the crystal structure of MBPCG3 could only be solved in the absence of transition metal ions. In these crystal structures the C-terminal fused peptides are not folded in a β -hairpin arrangement. Conditions containing transition metal ions gave crystals that did not diffract. It could be shown that these crystals are protein crystals. Inorganic salt crystals usually diffract very well and have short

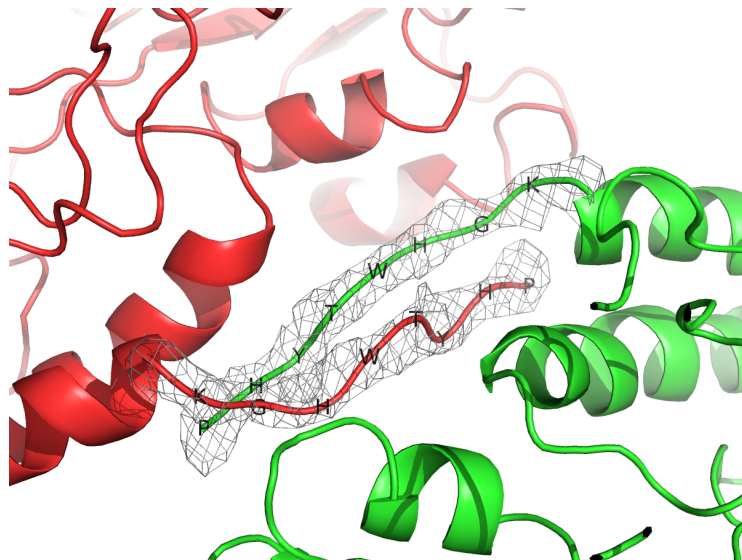


Figure 4.21: Detail in the crystal structure of MBPCG3, which is fused to Peptide71, in the monoclinic space group C2. The electron density of the interacting C-termini is drawn as σ_A weighted map contoured at 1.5σ . Because of the low resolution, the structure is poor in details.

unit cells. Unlike salt crystals, which are robust and rigid, crystals taken from the crystallisation trials are soft and easy to destroy. However, a crystal structure of MBPCG3 co-crystallised or soaked with a transition metal ion could not be solved.

5. Conclusion

Protein tags have been proven to be a very useful tool in the field of protein engineering and structural research. They offer the opportunity to equip the protein of interest with custom-made properties. The major task of protein tags is to launch the accessibility for highly selective purification. Thereby the most widely used protein tag is the His-tag. This short protein tag binds highly selectively to a stationary support equipped with immobilised transition metal ions because of its large number of histidine residues. In a crystal structure this tag and also other short protein tags are usually present either disordered or in a random coil conformation. In general short protein tags are structurally very flexible and align to their fusion partners. To the author's knowledge there is no short protein tag available that has an intrinsic tendency to build a secondary structure. Therefore they are pliable and commonly fit into crystal lattice voids. So in the crystal lattice empty cavities are occupied by the tag. Contingent on the size of these voids and interactions of the tag residues with the surface of the fusion protein, the protein tag is well-ordered or not. In general small tags are often not present in the electron density map. Only in a few cases, small protein tags interact with their fusion partner to form crystal contacts. So their conformational flexibility is responsible for the successful usage of small protein tags in structural biology.

There are cases where transition metal ions are bonded to the His-tag in crystal structures. If so, the crystallographic phase problem can be solved by collecting anomalous scattering data as shown before. Unfortunately the introduction of metal ions cannot be predicted. The major problem for anomalous scattering experiments using small tags is that these tags are flexible and often disordered because of their missing secondary structure. On the other hand, large protein tags are often well structured in solution as well as in the crystal lattice because of their size and rigid appearance. They possess a secondary and tertiary structure and often support the crystallisation process. Large protein tags tend to dominate the crystal lattice and often form crystal contacts. Thereby they are able to enforce crystal growth but

can hinder the formation of biologically relevant multimers of their fusion partner. This happens e.g. in the case of a too short linker sequence. Additionally large protein tags are commonly not equipped with an anomalous scatterer with the exception of sulfur in cysteine and methionine residues. So a combination of the rigid structure of large protein tags with the adaptability and metal chelating property of the small His-tag would be beneficial for a new class of protein tags.

Because several Protein G B1 based peptides are described as stable β -hairpin forming peptides their sequential and structural construction was the initial point for our peptide design. The peptide Chignolin which was shown to form a β -hairpin presented by Honda et al. [39] in 2004 was systematically modified.

Starting with Peptide32, which was Chignolin extended by two histidine residues at both termini and Peptide35, which was equipped with three histidine residues the amount of histidine residues in the peptide could be increased without losing its secondary structure. The overall folding of Peptide32 was determined by CD spectroscopy and 2D-NMR whereas for Peptide35 only CD spectroscopy data were collected. In both cases the data are very similar to data presented in the literature [19, 39, 96] and clearly show the cross-strand interactions of the aromatic side chains of Tyr and Trp which attests β -hairpin folding. Unfortunately the number of histidines was not sufficient for IMAC purification. The histidine residues are too far apart from each other to chelate a metal ion effectively. Peptide56 was created by an increment of histidine residues in the peptide to four histidines. This peptide shows high affinity in an IMAC purification. Because the predicted loop region was mainly modified the mutations hinder the β -hairpin folding of the peptide. CD spectroscopy and NMR data document a random coil arrangement of Peptide56 in aqueous solution. Even if metal ions are present in the media the peptide is still unfolded. In fusion protein crystal structures it was confirmed that Peptide56 did not form a β -hairpin. This fits in with the notion that the sequence in the loop region is essential for the bending of the peptide. So the Histidine residues which electrostatically repel each other and do not promote any secondary structure [17] hinder a proper folding of the tag.

Protein G B1 based peptides primarily form a β -hairpin because of the interaction of at least to aromatic amino acid side chains. To overcome the dilemma of misfolding of Peptide56 a new peptides was designed. In Peptide71 the aromatic interaction on the hydrophobic side of the tag was reinforced by introducing a second tryptophan residue in the strand region. Additionally Peptide71 was designed to contain two metal centres, a tridentate and a bidentate centre. The metal centres are far apart from each other to prevent mixed states of chelated metal ions. Near the expected loop region three amino acids were chosen which interact just with each other. All three residues were mutated to histidine residues. This part of the tag was designed to be bent because of the metal ion chelation. Additionally two histidine residues are placed face to face each in one strand to stick the two strand together by metal chelation. The expected folding of Peptide71 in the metal ion free and metal ion chelated state was verified by CD spectroscopy. In the metal ion

free state the bent form of Peptide71 was attested. Here the bent of the peptide is only supported by three aromatic side chains of the residues W3, Y5 and W12. In the Chignolin sequence only two aromatic residues are present. The side chains interact with each other by π - π interactions. The exact mode of interaction could not be clarified so far. However, CD spectroscopy could unambiguously attest interactions that are very similar to interactions found in the Trp-zipper motif [96, 97]. It seems that a well-formed β -hairpin structure is not present in that state. In the metal bound state the secondary structure of Peptide71 is much more pronounced. The CD spectrum of Peptide71 is very similar to the data obtained from Peptide32, Peptide35 and Chignolin. These data strikingly attest π - π interaction of the aromatic amino acids but additionally the presence of a β -strand region in Peptide71. The melting curve of Peptide71, calculated from CD spectra at various temperatures, is comparable to those collected for Chignolin. Despite the large number of histidine residues in the sequence of Peptide71 CD spectra indicate a similar thermodynamic stability of this peptide as for Chignolin because the melting temperature of Peptide71 is only about 6°C lower. Histidine residues occur in different protonation states depending on the pH of the media. In Peptide71 those repulsing forces of the histidine residues are reversed into stabilising forces by the chelation of a metal ion. So in the metal ion chelated state the folding of Peptide71 can be assumed to be a β -hairpin. Peptide71 was shown to be able to chelate various transition metal ions: Ni^{2+} , Co^{2+} , Zn^{2+} , Cu^{2+} , and Cd^{2+} . Therefore this peptide shows a high affinity in the IMAC purification. The retardation of Peptide71 is only slightly lower than the classical His-tag.

Because crystals of the synthesised peptides could not be obtained, the peptides were fused to the C-terminus of MBP from which crystal structures of the peptides could be determined. It was known already that MBP crystallises very well. Therefore this protein was fused with Peptide56 and Peptide71. Finally nine crystal structures could be solved.

In the case of MBPCG1, MBP fused with Peptide56, crystals grew already 15min after preparing the 96-well crystallisation plate. So the peptide did not hinder the protein in the crystallisation process. In two of the three structures the tag is well-ordered but unfolded. When a linker sequence is used to separate Peptide56 and MBP the protein's C-terminus becomes more flexible. Consequently the tag could not be seen in any of the crystal structures of MBPCG2. In the crystal structures of MBPCG2 the linker sequences interact with each other but not the tags.

It was much more difficult to get crystal structures of the Peptide71 which was the fusion partner in MBPCG3. Although there is no linker sequence the tag is not well-ordered in the crystal lattice. In the presence of metal ions well-diffracting crystals could not be obtained. One has to conclude that the crystal growth of MBP is obstructed by metal ions and/or the metal chelated form of Peptide71. In the case of MBPCG4 which was designed as fusion protein of MBP and Peptide71 with a proline residue linking the protein and the tag well-diffracting crystals could not be obtained either.

Peptide56 and Peptide71 are not of equal length. Fusion constructs of Peptide56 crystallise much easier than fusion constructs of Peptide71. While Peptide56 is conformationally unrestricted it promotes crystal contacts between the MBP molecules (see Sec. 2.5.5). The interacting residues fit to a hydrophobic pocket at the surface of MBP because this peptide is two amino acids shorter at the N-terminus compared to Peptide71. In contrast, Peptide71 in the metal free state is strong bent and β -hairpin shaped. This peptide is not as pliable as Peptide56. Additionally, Peptide71 does not fit to the hydrophobic pocket of MBP because the N-terminus of Peptide71 is two residues longer. So the distance of two fusion proteins is not compatible. Unfortunately well-diffracting crystal of MBPCG3 and MBPCG4 with chelated metal ions could not be obtained. Therefore a structure showing the secondary structure of Peptide71 is not available. However, crystals of Peptide71 fusion proteins from crystallisation trials containing Co^{2+} , Zn^{2+} , and Cd^{2+} could be grown whereas fusion proteins containing Peptide56 never crystallised in the presence of transition metal ions.

To the author's knowledge Peptide71 is the only short polypeptide with a large number of histidine residues and a metal ion chelation induced β -hairpin structure so far. The sequence of Peptide71 is free of unnatural as well as D-amino acids. In Peptide71 the hairpin is formed without any linkage of the β -strands by disulfide bridges. So Peptide71 can be easily expressed in various hosts without the risk of misfolding.

It was shown that Peptide71 is still useful for the purification of proteins. Through the introduction of a sixth histidine residue in the tag sequence a second trident metal centre will emerge and consequently the affinity of the tag in an IMAC purification step will be increased. This extended residue should be placed, depending on the protein terminus the tag is fused to, at the C- or N-terminus of the protein tag (see Fig. 5.1). So the sixth histidine residues should be at the residue to which the linker sequence is fused to. A problem could arise because of glycine residues G1 and G16. Glycine is not common in β -strands. Glycine is less restricted in the ϕ and ψ angles and could induce conformational freedom. In that case G1 and G16 should be replaced by two more hydrophobic amino acids like valine, leucine or isoleucine. It is possible as well to replace just G16 by an aromatic amino acid to fix both strands of the tag by the π - π interactions of a fourth aromatic amino acid side chain. Furthermore the loop region could be stabilised by the introduction of amino acids that are more common in loops. So A8 could be mutated to a glycine residue whereas K10 could be mutated to asparagine.

Linker sequences are of crucial relevance for the linkage between tag and protein. One of the major function of the linker is to separate the fusion partners and thus to interrupt the influence of their secondary structures. It was shown in the structure of MBPCG1 that without any linker sequence the conformational freedom of the tag is low. So the crystal structures of MBPCG1 showed a nearly complete tag. Moreover the fused protein introduces a helical conformation on a few residues of Peptide56. When Peptide56 was fused to MBP with a very long linker sequence in

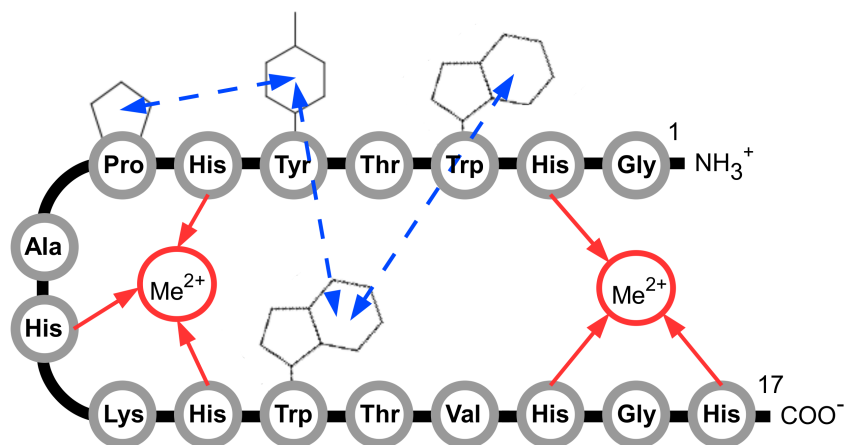


Figure 5.1: Illustration of a modified Peptide71. One histidine residue was added at the C-terminus of Peptide71. The amino acid side chains that influences the peptide folding to a β -hairpin are shown. The aromatic amino acids (and proline) are shown as sketch. The chelated divalent metal ions are shown as Me^{2+} . The interactions are highlighted by arrows.

between, the conformational flexibility was increased so that in all crystal structures only a few residues of the linker sequence could be found. The tag sequence was always missing. In the case of Peptide71 the introduction of a proline residue as linker could not be confirmed to be beneficial in the crystal. Well-diffracting crystals of MBPCG4 could not be obtained. However, this mutation is the reason for an increased affinity of the fusion protein in the IMAC purification. It was speculated that the tag is more exposed to the media because of the conformationally restricted proline residue. So the tag is more accessible for the IMAC support.

As reported by Chou and Fasman [17] the amino acids proline and glycine are very uncommon in α -helices and β -strands because they usually hinder the formation of these secondary structures. This fact should be taken into account since the secondary structure of the protein can influence several residues of the fused tag. In the case of Peptide71 this phenomenon would inhibit the formation of the β -hairpin and therefore the structure and the function of the tag. By introducing proline and glycine in the linker sequence a transfer of the secondary structure from the protein on the tag is blocked. The linker sequence should also allow slight motion of the tag to find its best position in the crystal lattice. In further experiments an appropriate length of the linker sequence should be systematically tested. The proposed best linker length is between two and four residues. This is a good compromise between flexibility and rigidity of the tag. Moreover an interaction of the tag with the fusion partner or another tag can prevent heavy motion and disorder of the tag. For further analytical applications of the fusion protein a proteolytic cleavage site might be necessary to cleave the tag. Appropriate sequences are the TEV or factor Xa site. The TEV site is a seven amino acid consensus sequence (Glu-X-X-Tyr-X-Gln/Ser) that is cleaved by the protease between residues Gln and Ser. This sequence might be too long and flexible and therefore the factor Xa cleavage site is

more interesting. This only four amino acid containing site (Ile-(Glu or Asp)-Gly-Arg/) is cleaved by the Factor Xa Protease after the Arg residue. Moreover there is a glycine residue in the factor Xa sequence that hinders the transfer of secondary structure between protein and tag.

Table 5.1: Sequences of tags based on Peptide71 which was modified by a sixth histidine residue and various linker sequences. The linker sequence and also the protein terminus to which the tag should be fused to is shown.

Sequence	Linker	Protein terminus
VHWTYHPAHKHWTVHTHPP	PP	N
PPHTHWTYHPAHKHWTVHV	PP	C
VHWTYHPAHKHWTVHTHAPA	APA	N
APAHTHWTYHPAHKHWTVHV	APA	C
VHWTYHPAHKHWTVHTHPPA	PPA	N
APPHTHWTYHPAHKHWTVHV	APP	C
VHWTYHPAHKHWTVHTHIDGR	Factor Xa	N
IDGRHVHWTYHPAHKHWTVHT	Factor Xa	C

Finally the new tag has to be able to be fused to the C-terminus as well as the N-terminus (see Tab. 5.1). When the tag is used C-terminally the linker and the sixth histidine residue should be placed to the N-terminus of the tag. In case of a N-terminal fusion of the tag to a protein the histidine residue and the linker are located at the C-terminus of the tag.

Anomalous scattering experiments with a fusion protein containing Peptide71 could not be performed so far. Because Peptide71 is able to chelate metal ions in aqueous solutions there is no doubt that Peptide71 can bind metal ions in a crystal structure as well. In the author's opinion this could not been proven so far because of an inappropriate linker sequences. To analyse the ability to fix anomalous scatterers further fusion proteins should be prepared. Besides MBP other easy to crystallise proteins like GST and GFP should be tested. These proteins give rise to get well-diffracting crystals.

6. References

- [1] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science (New York, N.Y.)*, 289(5481):905–920, 2000.
- [2] M. Benvenuti and S. Mangani. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols*, 2(7):1633–1651, 2007.
- [3] F. J. Blanco, G. Rivas, and L. Serrano. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Structural Biology*, 1(9):584–590, 1994.
- [4] T. Blandl, A. G. Cochran, and N. J. Skelton. Turn stability in beta-hairpin peptides: Investigation of peptides containing [3:5] type I g1 bulge turns. *Protein Science: A Publication of the Protein Society*, 12(2):237–247, 2003.
- [5] A. M. Bonvin and W. F. van Gunsteren. beta-hairpin stability and folding: molecular dynamics studies of the first beta-hairpin of tendamistat. *Journal of Molecular Biology*, 296(1):255–268, 2000.
- [6] W. H. Bragg. X-RAYS AND CRYSTALLINE STRUCTURE. *Science (New York, N.Y.)*, 40(1040):795–802, 1914.
- [7] A. B. Brauer, G. Kelly, J. D. McBride, R. M. Cooke, S. J. Matthews, and R. J. Leatherbarrow. The Bowman-Birk inhibitor reactive site loop sequence represents an independent structural beta-hairpin motif. *Journal of Molecular Biology*, 306(4):799–807, 2001.
- [8] P. Braun, Y. Hu, B. Shen, A. Halleck, M. Koundinya, E. Harlow, and J. LaBaer. Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 99(5):2654–2659, 2002.

- [9] C. Brink, D. C. Hodgkin, J. Lindsey, J. Pickworth, J. R. Robertson, and J. G. White. X-ray crystallographic evidence on the structure of vitamin b12. *Nature*, 174(4443):1169–1171, 1954.
- [10] Bruker AXS. APEX2: Programm zur Diffraktometersteuerung und Datensammlung mit Flächenzählerdaten. 2009.
- [11] Bruker AXS. SAINT: Programm zur Integration von Flächenzählerdaten. 2009.
- [12] M. H. Bucher, A. G. Evdokimov, and D. S. Waugh. Differential effects of short affinity tags on the crystallization of pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallographica. Section D, Biological Crystallography*, 58:392–397, 2002.
- [13] A. Bundi and K. Wüthrich. ¹H-nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, 18(2):285–297, 1979.
- [14] M. Carson, D. H. Johnson, H. McDonald, C. Brouillette, and L. J. Delucas. His-tag impact on structure. *Acta Crystallographica. Section D, Biological Crystallography*, 63:295–301, 2007.
- [15] A. Chant, C. M. Kraemer-Pecore, R. Watkin, and G. G. Kneale. Attachment of a histidine tag to the minimal zinc finger protein of the aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expression and Purification*, 39(2):152–159, 2005.
- [16] P. Y. Chou and G. D. Fasman. Beta-turns in proteins. *Journal of Molecular Biology*, 115(2):135–175, 1977.
- [17] P. Y. Chou and G. D. Fasman. Prediction of beta-turns. *Biophysical Journal*, 26(3):367–383, 1979.
- [18] S. E. Cnudde, M. Prorok, Q. Dai, F. J. Castellino, and J. H. Geiger. The crystal structures of the calcium-bound con-G and con-T[K7gamma] dimeric peptides demonstrate a metal-dependent helix-forming motif. *Journal of the American Chemical Society*, 129(6):1586–1593, 2007.
- [19] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik. Tryptophan zippers: stable, monomeric beta -hairpins. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10):5578–5583, 2001.
- [20] K. Cowtan. <http://www.ytbl.york.ac.uk/~cowtan/fourier/fourier.html>, . URL <http://www.ytbl.york.ac.uk/~cowtan/fourier/fourier.html>.
- [21] D. Crowfoot, C. Brunn, B. Rogers-Low, and A. Turner-Jones. X-ray crystallographic investigation of the structure of penicillin. In *Chemistry of penicillin.*, pages 310–367. New Jersey: Princeton University Press., 1945.

-
- [22] M. Cuneo, A. Changela, B. Hocker, L. Beese, and H. Hellinga. The crystal structure of redundant maltotriose binding proteins from the thermophile *thermotoga maritima*. *To be Published*, 2006.
- [23] B. Davidson and G. D. Fasman. The conformational transitions of uncharged poly-L-lysine. alpha helix-random coil-beta structure. *Biochemistry*, 6(6):1616–1629, 1967.
- [24] Z. S. Derewenda. The use of recombinant methods and molecular engineering in protein crystallization. *Methods (San Diego, Calif.)*, 34(3):354–363, 2004.
- [25] X. Duan, J. A. Hall, H. Nikaido, and F. A. Quiocho. Crystal structures of the maltodextrin/maltose-binding protein complexed with reduced oligosaccharides: flexibility of tertiary structure and ligand binding. *Journal of Molecular Biology*, 306(5):1115–1126, Mar. 2001.
- [26] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of coot. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4):486–501, 2010.
- [27] R. D. Fisher, B. Wang, S. L. Alam, D. S. Higginson, H. Robinson, W. I. Sundquist, and C. P. Hill. Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.*, 278:28976–28984, 2003.
- [28] K. J. Franz, M. Nitz, and B. Imperiali. Lanthanide-binding tags as versatile protein coexpression probes. *Chembiochem: A European Journal of Chemical Biology*, 4(4):265–271, 2003.
- [29] K. Fütterer, R. B. G. Ravelli, S. A. White, A. J. Nicoll, and R. K. Allemann. Differential specific radiation damage in the cu II-bound and pd II-bound forms of an alpha-helical foldamer: a case study of crystallographic phasing by RIP and SAD. *Acta Crystallographica. Section D, Biological Crystallography*, 64:264–272, 2008.
- [30] M. Gelinsky and H. Vahrenkamp. Zinc complexes of a helical 22-mer peptide with two histidine donors. *European Journal of Inorganic Chemistry*, 2002(9):2458–2462, 2002.
- [31] S. H. Gellman. Minimal model systems for beta sheet secondary structure in proteins. *Current Opinion in Chemical Biology*, 2(6):717–725, 1998.
- [32] M. R. Ghadiri and A. K. Fernholz. Peptide architecture. design of stable .alpha.-helical metallopeptides via a novel exchange-inert ruthenium(III) complex. *Journal of the American Chemical Society*, 112(26):9633–9635, 1990.
- [33] R. N. Gilbreth, K. Esaki, A. Koide, S. S. Sidhu, and S. Koide. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *Journal of Molecular Biology*, 381(2):407–418, 2008.

- [34] M. Hammarström, N. Hellgren, S. van Den Berg, H. Berglund, and T. Härd. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Protein Science: A Publication of the Protein Society*, 11(2):313–321, 2002.
- [35] B. Hazes, K. A. Magnus, C. Bonaventura, J. Bonaventura, Z. Dauter, K. H. Kalk, and W. G. Hol. Crystal structure of deoxygenated limulus polyphemus subunit II hemocyanin at 2.18 Å resolution: clues for a mechanism for allosteric regulation. *Protein Science: A Publication of the Protein Society*, 2(4):597–619, 1993.
- [36] W. A. Hendrickson, J. L. Smith, and S. Sheriff. Direct phase determination based on anomalous scattering. *Methods in Enzymology*, 115:41–55, 1985.
- [37] E. Hochuli, H. Döbeli, and A. Schacher. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography*, 411:177–184, 1987.
- [38] S. Honda, N. Kobayashi, and E. Munekata. Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. *Journal of Molecular Biology*, 295(2):269–278, 2000.
- [39] S. Honda, K. Yamasaki, Y. Sawada, and H. Morii. 10 residue folded peptide designed by segment statistics. *Structure (London, England: 1993)*, 12(8):1507–1518, 2004.
- [40] S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Ooishi, H. Watanabe, T. Odahara, and K. Harata. Crystal structure of a ten-amino acid protein. *Journal of the American Chemical Society*, 130(46):15327–15331, 2008.
- [41] W. Kabsch. XDS. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2):125–132, 2010.
- [42] R. B. Kapust and D. S. Waugh. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science: A Publication of the Protein Society*, 8(8):1668–1674, 1999.
- [43] I. L. Karle, S. K. Awasthi, and P. Balaram. A designed beta-hairpin peptide in crystals. *Proceedings of the National Academy of Sciences of the United States of America*, 93(16):8189–8193, 1996.
- [44] A. Ke and C. Wolberger. Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein Science: A Publication of the Protein Society*, 12(2):306–312, 2003.

-
- [45] T. Keates, A. C. W. Pike, A. Bullock, E. Salah, P. Filippakopoulos, A. K. Roos, F. V. Delft, P. Savitsky, J. Weigelt, A. Edwards, C. H. Arrowsmith, C. Bountra, and S. Knapp. To be published.
- [46] S. M. Kelly, T. J. Jess, and N. C. Price. How to study proteins by circular dichroism. *Biochim. Biophys. Acta*, 1751:119–139, 2005.
- [47] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, 1958.
- [48] H. Kessler, G. Gemmecker, and S. Steuernagel. NOESY-TOCSY, eine vorteilhafte 2D-NMR-Technik zur analyse von peptidsequenzen. *Angewandte Chemie*, 100(4):600–603, 1988.
- [49] K. M. Kim, E. C. Yi, D. Baker, and K. Y. Zhang. Post-translational modification of the n-terminal his tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. *Acta Crystallogr. D Biol. Crystallogr.*, 57:759–762, 2001.
- [50] B. Kobe, R. J. Center, B. E. Kemp, and P. Pombourios. Crystal structure of human t cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8):4319–4324, 1999.
- [51] I. A. Kozlov, B. G. Kermani, P. C. Melnyk, D. L. Barker, C. Zhao, J. P. Hachmann, and M. Lebl. Retention of histidine-containing peptides on a nickel affinity column. *Journal of Chromatographic Science*, 45(4):207–211, 2007.
- [52] G. Langer, S. X. Cohen, V. S. Lamzin, and A. Perrakis. Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7. *Nature Protocols*, 3(7):1171–1179, 2008.
- [53] R. A. Laskowski. PDBsum new things. *Nucleic Acids Research*, 37(Database issue):D355–359, 2009.
- [54] Y. Liu, A. Manna, R. Li, W. E. Martin, R. C. Murphy, A. L. Cheung, and G. Zhang. Crystal structure of the SarR protein from staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America*, 98(12):6877–6882, 2001.
- [55] L. J. Martin, M. J. Hähnke, M. Nitz, J. Wöhnert, N. R. Silvaggi, K. N. Allen, H. Schwalbe, and B. Imperiali. Double-lanthanide-binding tags: design, photophysical properties, and NMR applications. *J Am Chem Soc*, 129(22):7106–7113, 2007.

- [56] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, 40:658–674, 2007.
- [57] R. B. Merrifield. Solid phase peptide synthesis. i. the synthesis of a tetrapeptide. *Journal of the American Chemical Society*, 85(14):2149–2154, 1963.
- [58] A. J. Nicoll, D. J. Miller, K. Fütterer, R. Ravelli, and R. K. Allemann. Designed high affinity Cu^{2+} -binding α -helical foldamer. *J. Am. Chem. Soc.*, 128:9187–9193, 2006.
- [59] J. Painter and E. A. Merritt. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica. Section D, Biological Crystallography*, 62(Pt 4):439–450, 2006.
- [60] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, and A. C. North. Structure of haemoglobin: a three-dimensional fourier synthesis at 5.5-Å resolution, obtained by x-ray analysis. *Nature*, 185(4711):416–422, 1960.
- [61] J. Porath, J. Carlsson, I. Olsson, and G. Belfrage. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, 258(5536):598–599, 1975.
- [62] E. Prince, H. Fuess, T. Hahn, H. Wondratschek, U. Müller, U. Shmueli, E. Prince, A. Authier, V. Kopský, D. B. Litvin, M. G. Rossmann, E. Arnold, S. Hall, and B. McMahon, editors. *International Tables for Crystallography*, volume C. International Union of Crystallography, Chester, England, 1 edition, 2006. ISBN 978-1-4020-1900-5.
- [63] M. Ramírez-Alvarado, T. Kortemme, F. J. Blanco, and L. Serrano. Beta-hairpin and beta-sheet formation in designed linear peptides. *Bioorganic & Medicinal Chemistry*, 7(1):93–103, 1999.
- [64] K. K. Reiling, T. R. Pray, C. S. Craik, and R. M. Stroud. Functional consequences of the kaposi’s sarcoma-associated herpesvirus protease structure: regulation of activity and dimerization by conserved structural elements. *Biochemistry*, 39(42):12796–12803, 2000.
- [65] A. Riboldi-Tunnicliffe, N. W. Isaacs, and T. J. Mitchell. 1.2 angstroms crystal structure of the s. pneumoniae PhtA histidine triad domain a novel zinc binding fold. *FEBS Lett.*, 579:5353–5360, 2005.
- [66] W. C. Röntgen. ON a NEW KIND OF RAYS. *Science (New York, N.Y.)*, 3(59):227–231, 1896.
- [67] D. Sachdev and J. M. Chirgwin. Fusions to maltose-binding protein: control of folding and solubility in protein purification. *Methods in Enzymology*, 326:312–321, 2000.

-
- [68] F. Schäfer, J. Blümer, U. Römer, and K. Steinert. Ni-nta for large-scale imac processes. *QIAGEN News*, 4(4):11–15, 2000.
- [69] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102(5):615–623, 2000.
- [70] E. R. Schreiter, M. D. Sintchak, Y. Guo, P. T. Chivers, R. T. Sauer, and C. L. Drennan. Crystal structure of the nickel-responsive transcription factor NikR. *Nature Structural Biology*, 10(10):794–799, 2003.
- [71] A. J. Sharff, L. E. Rodseth, J. C. Spurlino, and F. A. Quioco. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry*, 31(44):10657–10663, 1992.
- [72] G. M. Sheldrick. SADABS: Program for absorption and other correction. *University Göttingen*, 2010.
- [73] G. M. Sheldrick. A short history of SHELX. *Acta Crystallographica. Section A, Foundations of Crystallography*, 64:112–122, 2008.
- [74] G. M. Sheldrick. TWINABS: Program for absorption and other correction of twinned structures. *University Göttingen*, 2010.
- [75] G. M. Sheldrick. XPREP: Programm zur Datenanalyse und Erkundung des reziproken Raumes. *Bruker AXS*, 2010.
- [76] B. L. Sibanda and J. M. Thornton. Beta-hairpin families in globular proteins. *Nature*, 316(6024):170–174, 1985.
- [77] B. L. Sibanda and J. M. Thornton. Conformation of beta hairpins in protein structures: classification and diversity in homologous structures. *Methods in Enzymology*, 202:59–82, 1991.
- [78] S. H. J. Smits, A. Mueller, M. K. Grieshaber, and L. Schmitt. Coenzyme- and his-tag-induced crystallization of octopine dehydrogenase. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, 64:836–839, 2008.
- [79] J. C. Spurlino, G. Y. Lu, and F. A. Quioco. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *The Journal of Biological Chemistry*, 266(8):5202–5219, 1991.
- [80] U. Srinivasan, G. H. Iyer, T. A. Przybycien, W. A. Samsonoff, and J. A. Bell. Crystine: fibrous biomolecular material from protein crystals cross-linked in a specific geometry. *Protein Engineering*, 15(11):895–902, 2002.

- [81] H. E. Stanger, F. A. Syud, J. F. Espinosa, I. Girit, T. Muir, and S. H. Gellman. Length-dependent stability and strand length limits in antiparallel beta-sheet secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12015–12020, 2001.
- [82] F. A. Syud, H. E. Stanger, H. S. Mortell, J. F. Espinosa, J. D. Fisk, C. G. Fry, and S. H. Gellman. Influence of strand number on antiparallel beta-sheet stability in designed three- and four-stranded beta-sheets. *Journal of Molecular Biology*, 326(2):553–568, 2003.
- [83] Y. Tajika, N. Sakai, T. Tamura, M. Yao, N. Watanabe, and I. Tanaka. Crystal structure of hypothetical protein PH0828 from *pyrococcus horikoshii*. *Proteins*, 57(4):862–865, 2004.
- [84] P. G. Telmer and B. H. Shilton. Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *The Journal of Biological Chemistry*, 278(36):34555–34567, Sept. 2003.
- [85] K. Terpe. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72(2):211–222, 2006.
- [86] S. Trapani and J. Navaza. AMoRe: classical and modern. *Acta Crystallographica. Section D, Biological Crystallography*, 64:11–16, 2008.
- [87] H. Ullah, E. L. Scappini, A. F. Moon, L. V. Williams, D. L. Armstrong, and L. C. Pedersen. Structure of a signal transduction regulator, RACK1, from *arabidopsis thaliana*. *Protein Science: A Publication of the Protein Society*, 17(10):1771–1780, 2008.
- [88] A. Vagin and A. Teplyakov. Molecular replacement with MOLREP. *Acta Crystallographica. Section D, Biological Crystallography*, 66:22–25, 2010.
- [89] A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, and G. N. Murshudov. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica. Section D, Biological Crystallography*, 60:2184–2195, 2004.
- [90] B. van den Berg, P. N. Black, W. M. Clemons, and T. A. Rapoport. Crystal structure of the long-chain fatty acid transporter FadL. *Science (New York, N.Y.)*, 304(5676):1506–1509, 2004.
- [91] C. M. Venkatachalam. Stereochemical criteria for polypeptides and proteins. v. conformation of a system of three linked peptide units. *Biopolymers*, 6(10):1425–1436, 1968.
- [92] H. Watanabe, H. Matsumaru, A. Ooishi, Y. Feng, T. Odahara, K. Suto, and S. Honda. Optimizing pH response of affinity between protein g and IgG fc:

-
- how electrostatic modulations affect protein-protein interactions. *The Journal of Biological Chemistry*, 284(18):12373–12383, May 2009.
- [93] J. D. Watson and F. H. Crick. The structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology*, 18:123–131, 1953.
- [94] B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, T. Hartsch, and V. Ramakrishnan. Structure of the 30S ribosomal subunit. *Nature*, 407(6802):327–339, 2000.
- [95] D. S. Wishart, B. D. Sykes, and F. M. Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of Molecular Biology*, 222(2):311–333, 1991.
- [96] L. Wu, D. McElheny, R. Huang, and T. A. Keiderling. Role of tryptophan-tryptophan interactions in trpzip beta-hairpin formation, structure, and stability. *Biochemistry*, 48(43):10362–10371, 2009.
- [97] L. Wu, D. McElheny, T. Takekiyo, and T. A. Keiderling. Geometry and efficacy of cross-strand Trp/Trp, Trp/Tyr, and Tyr/Tyr aromatic interaction in a beta-hairpin peptide. *Biochemistry*, 49(22):4705–4714, 2010.

List of publication

- [98] M. Stollenz, C. Grosse, and F. Meyer. An unusually stable octanuclear sigma-mesityl-bridged μ_4 -oxo-copper(I) complex encapsulated by a pyrazolate-based compartmental ligand scaffold. *Chemical Communications (Cambridge, England)*, (15):1744–1746, Apr. 2008.
- [99] L. F. Tietze, D. A. Spiegl, F. Stecker, J. Major, C. Raith, and C. Grosse. Stereoselective synthesis of 4-dehydroxydiversonol employing enantioselective palladium-catalysed domino reactions. *Chemistry*, 14(29):8956–8963, 2008.
- [100] G. B. Nikiforov, H. W. Roesky, B. C. Heisen, C. Grosse, and R. B. Oswald. Formation of a titanium complex with a $\text{Ti}=\text{CHAl}_2$ structural unit from LTiMe_3 and trimethylaluminum. *Organometallics*, 27(11):2544–2548, June 2008.
- [101] A. Prokofieva, S. Dechert, C. Grosse, G. M. Sheldrick, and F. Meyer. Oxidative ortho C-C coupling of 4-ethylphenol by dual substrate activation at a bioinspired dicopper complex—trapping of an unusual oligophenolic Cu_6 species. *Chemistry*, 15(20):4994–4997, 2009.
- [102] M. Stollenz, M. John, H. Gehring, S. Dechert, C. Grosse, and F. Meyer. Oligonuclear homoleptic copper(I) pyrazolates with multinucleating ligand scaffolds: high structural diversity in solid-state and solution. *Inorganic Chemistry*, 48(21):10049–10059, Nov. 2009.
- [103] D. D. Rodríguez, C. Grosse, S. Himmel, C. González, I. M. de Ilarduya, S. Becker, G. M. Sheldrick, and I. Usón. Crystallographic ab initio protein structure solution below atomic resolution. *Nature Methods*, 6(9):651–653, Sept. 2009.

Conference contribution

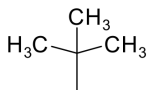
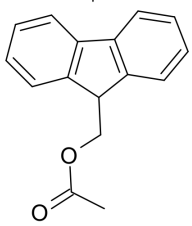
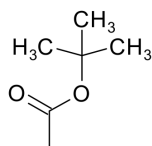
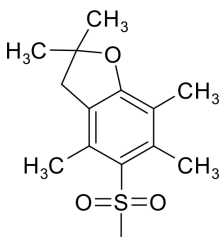
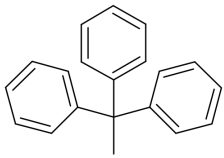
- [104] *Optimisation of the His-tag for Purification and Crystallisation.*, Göttingen, Germany, 2007. International Research Training Group 1422 Biometals Workshop, presentation.
- [105] *Optimisation of the His-tag for Purification and Crystallisation.*, Lund, Sweden, 2007. International Research Training Group 1422 Biometals Workshop, poster.
- [106] *Optimisation of the His-tag for Purification and Crystallisation.*, Como, Italy, 2008. 9th International School on the Crystallography of Biological Macromolecules, poster.
- [107] *How to assemble a protein tag for purification, crystallisation and phasing.*, Berlin, Germany, 2009. Winter School on Soft X-rays in Macromolecular Crystallography, presentation.

- [108] *How to assemble a protein tag for purification, crystallisation and phasing.*, Copenhagen, Denmark, 2009. International Research Training Group 1422 Biometals Workshop, presentation.
- [109] *How to assemble a protein tag for purification, crystallisation and phasing.*, Vienna, Austria, 2009. 11th International Congress on Amino Acids, Peptides and Proteins, poster.
- [110] *Novel short β -hairpin forming polypeptides as protein tags for purification, crystallisation and phasing.*, Goslar, Germany, 2010. International Research Training Group 1422 Biometals Workshop, poster.
- [111] *Sticky β -hairpin – a novel tool in protein crystallography.*, Göttingen, Germany, 2010. 12th JCF-Frühjahrssymposium, poster.

7. Appendix

7.1 Protecting groups

Table 7.1: Commonly used protecting groups in peptide synthesis.

	Tert. butyl (tBu)
	9-Fluorenylmethoxycarbonyl (Fmoc)
	Tert. Butoxycarbonyl (Boc)
	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl (Pbf)
	Trityl (Trt)

7.2 List of synthesised peptides

#	Sequence	Fold	Source	L	#	N- his term.	C- term.
1	PHTHAHLH	α -helix	Synth.	8	4	Fmoc	OH
2	PHTHYAHLH	α -helix	Synth.	9	4	Fmoc	OH
3	PHTHLYAHLH	α -helix	Synth.	10	4	Fmoc	OH
19	YEDPHLKHAIQESLREAEAA	α -helix	1O06	20	2	Ac	OH
21	GEHELQHNQHLIRHKSNI	α -helix	3HMX	18	4	Ac	OH
23	YHKIGHGRQTFTHKIGHA	α -helix	Rs 164-176 PrtC, 1GO7	18	4	Ac	OH
27	YEDPHLKHAIQESLRHAEHL	α -helix	1O06	20	4	Ac	OH
36	AHAYAHEAAAKEAAAKHAY AHA	α -helix	Synth.	22	4	H	OH
37	AHAYAHEAAAAKHAYAHA	α -helix	Synth.	18	4	H	OH
41	AHAYCHEAAAAKHAYHCA	α -helix	Synth.	18	4	H	OH
43	AHCYAHEAAAKEAAAKHCY AHA	α -helix	Synth.	22	4	H	OH
44	AHAYCHEAAAKEAAAKHAY CHA	α -helix	Synth.	22	4	H	OH
49	MKHHHHHHQVRY	His-tag	Qiagen	12	6	Ac	NH ₂
4	AYIVPHGDHYHYIP	β -hairpin	Rs 23-36, 2CS7	14	3	Fmoc	OH
5	YAYFCSFPGHWAMHKGTL	β -hairpin	Rs 108-126 Azurine, 1A4A	18	2	Fmoc	OH
9	RSMNPRTHATTRSNMHHAAG ASRS	β -hairpin	Synth.	24	3	Ac	OH
11	TVHGHSCHKHKSSEHGHT	β -hairpin	Synth.	18	6	Ac	NH ₂
12	IVHGHSCHKHYL	β -hairpin	Synth.	12	4	Ac	OH
15	WHEHGHSCHKHNHGHEHK	β -hairpin	Synth.	19	9	Ac	NH ₂
16	THEHEPGKHWHPGHRHTG	β -hairpin	Synth.	18	6	Ac	NH ₂
17	THLHPDGHVHLPDGVHLHT	β -hairpin	Synth.	19	6	Ac	NH ₂
18	TWHVHNGHYHNGKHTHG	β -hairpin	Synth.	18	6	Ac	NH ₂
30	GDAYIIPGDHYHYIPD	β -hairpin	Rs 23-36, 2CS7	17	3	Ac	NH ₂
31	HGHGYDDETKTVGHGH	β -hairpin	Protein G B1 domain	16	4	Ac	NH ₂
32	HGYDPETGTWGH	β -hairpin	Protein G B1 domain	12	2	Ac	NH ₂
35	GEHYDPEHDTWHSY	β -hairpin	Protein G B1 domain	14	3	Ac	NH ₂
45	VEKYHPRCVHWTVN	β -hairpin	Protein G B1 domain	14	2	Ac	OH

#	Sequence	Fold	Source	L	#	N- his term.	C- term.
46	VEKYHPHCVTWTV	β -hairpin	Protein B1 domain	G	14	2	Ac OH
47	VEKYHPHHVTWTV	β -hairpin	Protein B1 domain	G	14	3	Ac OH
48	TGVHYHPCHKTWHVCG	β -hairpin	Protein B1 domain	G	16	4	Ac OH
54	VVHYHPHHVTWHVV	β -hairpin	Protein B1 domain	G	14	5	H OH
55	VTHYHPHHVTWVTVEHYHP DHVTWHNVV	β -hairpin	Protein B1 domain	G	28	8	H OH
56	VTVYHPHHVTWHTV	β -hairpin	Protein B1 domain	G	14	4	H OH
57	VTVYHPHHVTWCTV	β -hairpin	Protein B1 domain	G	14	3	H OH
58	VVERYHPDHVHWTNVV	β -hairpin	Protein B1 domain	G	16	3	H OH
59	VVEHYHPDHVTWHNVV	β -hairpin	Protein B1 domain	G	16	4	H OH
60	VTHYHPHHVTWVTV	β -hairpin	Protein B1 domain	G	14	4	H OH
61	HLVYHPDHDGWVLHG	β -hairpin	Protein B1 domain	G	15	4	Ac OH
63	YVYHPGHDHWVA	β -hairpin	Protein B1 domain	G	12	3	Ac OH
64	YVYDPHDHGWVA	β -hairpin	Protein B1 domain	G	12	2	Ac OH
65	HYLYHPDHVTWLAHG	β -hairpin	Protein B1 domain	G	15	4	Ac OH
68	HYLYHPHHVTWLAHGPGTL QIRGRERFEMFRELNEALE LKDAQAG	β -hairpin	Fusion peptide with 1AIE		45	5	Ac OH
67	HYLYHPDHVTWLAHGPGKF NARRKLGAILTTMLATR N	β -hairpin	Fusion peptide with 2WEL		40	4	Ac OH
71	GHWTYHPAHKHWTVHG	β -hairpin	Protein B1 domain	G	16	5	H OH
72	IHYTPHHNTWHI	β -hairpin	Protein B1 domain	G	12	4	H OH

7.3 Primary and secondary structure of MBP

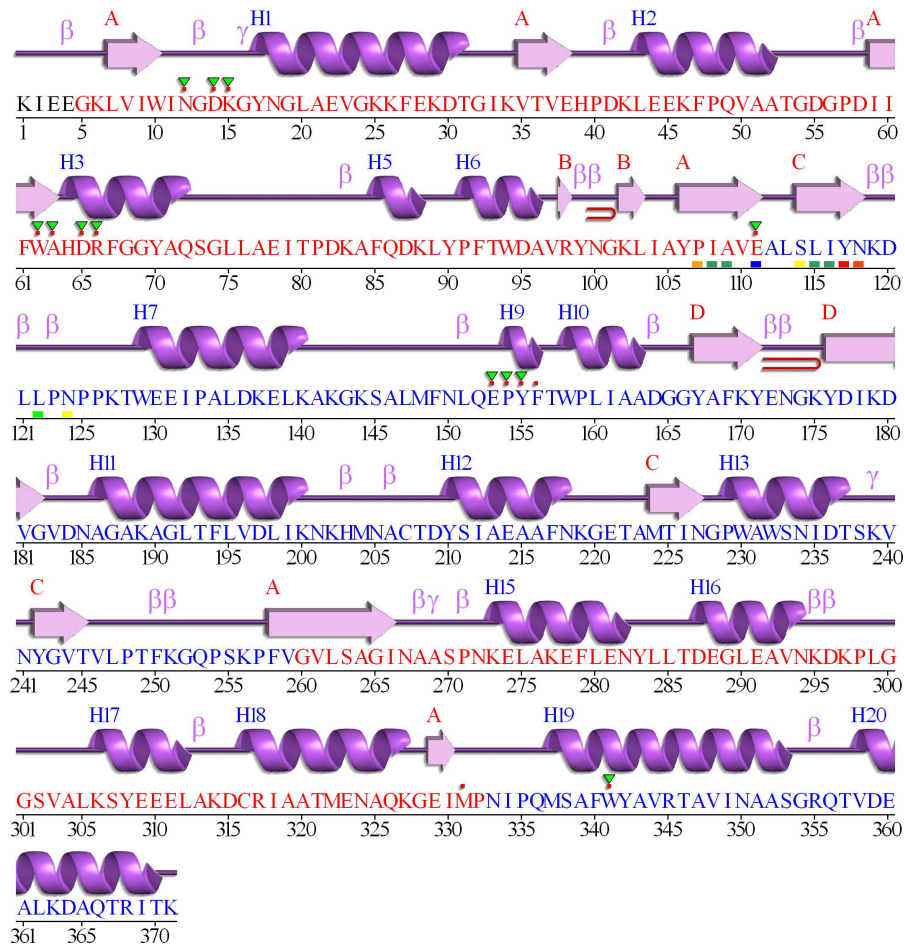


Figure 7.1: Sequence and secondary structure of the Maltose-binding protein (PDB-ID:1JVX) published by Srinivasan et al. [80]. The figure was taken from the PDBSum homepage (Laskowski [53]).

7.4 Crystallographic tables

Table 7.3: Several spacegroups of MBP mutant MBPCG1 could be found and crystallographic detail are shown.

	Native closed form	Native closed form	Native closed form
Cell dimensions			
(Å)	a=118.92	a=41.70	a=38.76
	b=40.29	b=63.42	b=74.61
	c=88.50	c=76.15	c=111.84
(°)	β =115.88	β =99.73	
Beamline	Smart6000	Smart6000	Smart6000
Wavelength (Å)	1.54178	1.54178	1.54178
Spacegroup	C2	P2 ₁	P2 ₁ 2 ₁ 2 ₁
Resolution (Å)	1.80	2.56	2.76
Data collected	241312	106997	123938
Unique reflexes	35357	12670	9110
Completeness (%)			
total	99.9	97.8	99.7
last shell	100(1.89- 1.80)	85.3(2.66- 2.56)	100(2.86- 2.76)
Redundancy	6.82	8.26	13.60
R _{INT}			
total	0.0529	0.102	0.2944
last shell	0.4046	0.3743	-
I/sigma (I)			
total	19.13	13.26	12.84
last shell	2.49	2.78	6.71
R _{work} /R _{free}	21.45/26.85	25.97/31.68	34.45/46.41
Ramachandran plot			
Preferred region	362	358	328
Allowed region	15	19	37
Disallowed region	1	2	6
Comment			non- merohedrally twinned

Table 7.4: Both *open* and *closed form* of MBP could be crystallised in the fusion protein MBPCG2. Crystallographic data are shown.

	Native open form	Open form, Co²⁺ soak	Native closed form	Closed form, Ni²⁺ soak
Cell dimensions				
(Å)	a=37.78 b=43.86 c=57.67	a=110.21 b=38.04 c=90.36	a=49.43 b=42.12 c=88.21	a=42.31 b=95.33 c=96.68
(°)	α =100.32 β =100.92 γ =103.82	β =111.90	β =92.61	
Beamline	Bessy 14.1	Bessy 14.1	Bessy 14.1	Smart6000
Wavelength (Å)	0.91801	0.91801	0.91801	1.54178
Spacegroup	P1	C2	P2 ₁	P2 ₁ 2 ₁ 2 ₁
Resolution (Å)	1.55	2.80	1.88	2.24
Data collected	213482	36189	115881	220378
Unique reflexes	46174	8849	27820	17489
Completeness (%)				
total	96.4	99.9	99.8	94.8
last shell	95(1.67- 1.57)	99.9(2.89- 2.80)	99.1(1.98- 1.88)	75.0(2.34- 2.24)
Redundancy	4.31	3.69	3.75	11.30
R _{INT}				
total	0.0697	0.0449	0.1488	0.1122
last shell	0.3652	0.3697	0.5076	0.2921
I/sigma (I)				
total	12.58	17.26	8.33	17.56
last shell	3.18	2.23	2.89	3.59
R _{work} /R _{free}	20.57/24.65	24.693/32.73	26.44/32.46	19.17/24.849
Ramachandran plot				
Preferred region	350	344	352	355
Allowed region	17	25	13	12
Disallowed region	0	3	0	0
Comment				

Table 7.5: Crystallographic table for MBPCG3

	Native open form	Native open form
Cell dimensions		
(Å)	a=55.66 b=65.24 c=79.40	a=106.86 b=37.32 c=91.72
(°)	α =73.14 β =64.33 γ =64.69	β =110.86
Beamline	Smart6000	Smart6000
Wavelength (Å)	1.54178	1.54178
Spacegroup	P1	C2
Resolution (Å)	1.98	3.68
Data collected	593424	44583
Unique reflexes	62853	3345
Completeness (%)		
total	99.3	96.1
last shell	98.1(2.08- 1.98)	69.4(3.78- 3.68)
Redundancy	9.44	11.07
R _{INT}		
total	0.2072	0.2225
last shell	-	0.4239
I/sigma (I)		
total	15.29	8.27
last shell	2.22	4.07
R _{work} /R _{free}	27.67/30.56	39.21/49.99
Ramachandran plot		
Preferred region	702	345
Allowed region	26	18
Disallowed region	6	6
Comment	non- merohedrally twinned	

Lebenslauf

Name:	Christian Große
Geburtstag:	29.09.1977
Geburtsort:	Pößneck
Staatsangehörigkeit:	deutsch
Familienstand:	ledig in fester Partnerschaft, 2 Kinder
9.1984-8.1991	29.polytechnische Oberschule in Berlin
9.1991-5.1997	Walter Felsenstein Oberschule in Berlin
5.1997	Abitur
7.1997-4.1998	Grundwehrdienst in Nienburg und Lingen
9.1998-7.2001	Berufsausbildung zum Chemielaboranten bei der Schering AG
7.2001	Berufsabschluß Chemielaborant, vorzeitig mit Auszeichnung
10.2001-7.2006	Studium Diplom Chemie an der HU Berlin
9.2003	Vordiplom Chemie
8.2006	Diplom Chemie
10.2006-9.2010	Doktorand an der Georg August Universität Göttingen
10.2010	Dissertation