Crystal structure of a human U5 snRNP specific binary complex and crystal structure of a histone deacetylase-like bacterial amidohydrolase

Dissertation

Zur Erlangung des Doktorgrades

der Matematisch-Naturwissenshaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Tine Kragh Nielsen

aus Pusan

Göttingen 2005

D7

Referent: Prof. Dr. Ralf Ficner

Korreferent: Prof. Dr. Oliver Einsle

Tag der mündlichen Prüfung: 29.06.2005

1

1. Summary

The crystal structure of the complex comprising the GYF-domain of U5-52K and the U5-15K protein has been determined at 2.35 Å resolution. The protein U5-52K (Lin1p) is a component of yeast and human U5 snRNPs, which are involved in pre-mRNA splicing. In addition U5-52K plays a role in immune response as CD2 receptor binding protein 2 (CD2BP2), as it binds to the CD2 receptor via its GYF-domain specifically recognising a proline-rich motif on the cytoplasmic surface of the receptor. This GYF-domain also mediates the interaction of the proteins U5-52K and U5-15K within the spliceosomal U5 snRNP. The structure unravels novel interaction sites on both proteins, as neither the poly-proline binding site of the GYF-domain nor the common ligand binding cleft of thioredoxin-like proteins, to which U5-15K belongs, participate in this interaction. The mode of interaction is of particular interest with respect of the GYF-domain, since the U5-15K protein lacks the proline-rich motif. The structure furthermore reveals the structural basis for the bifunctionality of the GYF-domain.

The structure of the first HDAC class 2 homologue has been determined by X-ray crystallography. The apo-enzyme was refined at a resolution of 1.6 Å and complexes with two inhibitors: SAHA and CypX at a resolution of 1.57 Å and 1.75 Å, respectively. The HDAH (histone deacetylase-like amidohydrolase) from *Bordetella/Alcaligenes* Strain FB188 shows functional and sequential homology to human HDACs. HDACs are responsible for the removal of ε-amino groups of lysine residues on the amino terminus of nucleosomal histones. HDAC inhibitors have been shown to inhibit tumour activity in both pre-clinical models and in clinical trials and are among the most promising candidates for future anti-cancer drugs. The structure reveals the canonical class 1 fold with a zinc ion and two potassium ions bound. The highest diversity compared to known structures of the class 1 enzymes is found in the loop regions especially in the area around the entrance of the active site, indicating significant differences of the interacting partners to the class 1 and 2 enzymes. The structures of HDACs and HDAC homologues are important for understanding inhibitor binding and for the development of new inhibitors.

2. Zusammenfassung

Die Kristallstruktur der GYF-Domäne von U5-52K im Komplex mit U5-15 wurde bei einer Auflösung von 2,35 Å ermittelt. Das Protein U5-52K (Lin1p) ist sowohl in der Hefe als auch beim Menschen Bestandteil der U5 snRNPs, welche am Spleißen von mRNAs beteiligt sind. Weiterhin spielt U5-52K als CD2 Rezeptor-Bindeprotein 2 (CD2BP2) eine Rolle bei der Immunantwort, da es mit seiner GYF-Domäne spezifisch an ein prolinreiches Sequenzmotiv der zytoplasmatischen Oberfläche des CD2 Rezeptor bindet. Die GYF-Domäne vermittelt ebenfalls die Wechselwirkung der Proteine U5-52K und U5-15K innerhalb der spleißosomalen U5 snRNPs. Anhand der Struktur konnten neue Wechselwirkungsbereiche identifiziert werden, da weder Polyprolin-Bindungsstelle der GYF-Domäne noch die allgemeine Liganden-Bindungsstelle der Thioredoxin ähnlichen Proteine, zu denen U5-15K gehört, an dieser Wechselwirkung beteiligt sind. Die Art der Wechselwirkung ist insbesondere seitens der GYF-Domäne interessant, da das U5-15K Protein kein prolinreiches Sequenzmotiv besitzt. Des weiteren gibt die Struktur Aufschluss über die strukturelle Basis der Bifunktionalität der GYF-Domäne.

In dieser Arbeit konnte die Struktur einer bakteriellen Histondeacetylase-ähnlichen Amidohydrolase (HDAH) mittels Röntgenkristallographie ermittelt werden. Dieses Enzym aus *Bordetella/Alcaligenes* FB188 ist ein Histondeacetylase (HDAC)-Klasse 2 Homolog und stellt die erste Struktur eines HDAC-Klasse 2 Enzyms dar. Im allgemeinen sind HDAC's für die Prozessierung von Acetylresten an ε-Aminogruppen von Lysinen an den Aminotermini von Histonschwänzen verantwortlich. Das Holo-Enzym wurde bei 1,6 Å, das Enzym im Komplex mit den Histondeacetylase-Inhibitoren SAHA und CypX bei 1,57 Å bzw. 1,75 Å aufgelöst. Die Struktur zeigt eine kanonischen Klasse 1 Faltung mit einen Zink- und zwei Kalium-Ionen gebunden. Der größte Unterschied zwischen der HDAH und bereits bekannten Strukturen der Klasse 1 Enzyme liegt in den loop-Regionen, speziell in der näheren Umgebung der Substratbindungstasche, was auf deutlich unterschiedliche Interaktionspartner der Klasse 1-und 2-Enzyme Schließen lässt. Aus vor-klinischen sowie klinischen Studien haben sich Inhibitoren dieser Enzymfamilie als zukünftige Anti-Krebs-Therapeutika hervorgehoben. Die Strukturaufklärung dieser HDAH bzw. von HDAC's im allgemeinen dient in diesem Zusammenhang zu einem besseren Verständnis von Inhibitorbindung und spielt eine tragende Rolle bei der Entwicklung von neuen Inhibitoren.

3. Abbreviations

APAH: AcetylPolyamine-AmidoHydrolase

ATP: Adenosine TriPhosphate

bp: base pair

°C: Degree Celsius

CHAPS: 3,3-Cholamidopropyl-dimethylAmmonio-1-PropaneSulfonate

CypX: Cyclopentyle-Propionyle-Hydroxamic acid

Da: Dalton

DESY: Deutsche-Elektronen-SYnchrotron

DTT: DiThioThretiol

E. coli: *Escherichia coli*

EDTA: EthyleneDiamine-N,N,N',N'-TetraAcetate

EMBL: European Molecular Biology Laboratory

h: hour

HDAC: Histone DeACetylase

HDAH: Histone Deacetylase-like AmidoHydrolase

IC_{50}:   Inhibitor concentration at 50% inhibition of the enzyme

IPTG: IsoPropyl-β-D-ThioGalactopyranoside

K: Kelvin

LB: Luria Bertani medium

M: Molarity

MES: 2-MorpholinoEthaneSulfonic acid

mRNA: messenger RNA

3. Abbreviations

OD$_{600}$:   Optical Density at 600 nm

PDB: The RCSB Protein Data Bank

PEG: PolyEthylene Glycol

PEG mme: PolyEthylene Glycol mono methyl ether

PMSF: Poly Methane Sulfonyl Fluoride

Prp: Pre-RNA processing

rmsd: root mean square deviations

RNA: RiboNucleic Acid

S: Svedberg

S. cerevisiae: Saccharomyces cerevisiae

S. pombe: Schizosaccharomyces pombe

SAHA: SuberoylAnilide Hydroxamic Acid

SDS-PAGE: Sodium Dodecyl Sulphate Poly Acrylamide Gel Electrophoresis

snRNA: small nuclear RNA

snRNP: small nuclear Ribo Nucleo Protein Particle

Tris: Tris(hydroxymethyl)aminomethane

TSA: TrichoStatin A

(v/v%):   volume/volume percent

(w/v%):   weight/volume percent

xg: Times gravity

Å: Angstrom ($1\text{Å} = 10^{-10}$ m)

Crystal structure of a U5 snRNP binary complex

4. Introduction

4.1. Splicing

4.1.1. Components of the spliceosome

The spliceosome is a huge and complicated macromolecular complex responsible for the comprehensive splicing of pre-mRNA resulting in the removal of the non-coding introns from the coding exons. The major spliceosome is generally formed by the five snRNPs, U1, U2, U4, U5, U6 and several non-snRNP proteins (Staley and Guthrie 1998, Will and Lührmann 2001, Jurica and Moore 2003). Each UsnRNP is composed of a small nuclear RNA (snRNA), seven common proteins (Sm proteins: B/B', D1, D2, D3, E, F and G, except in U6, where the Sm proteins are replaced by Sm-like proteins) and a variable amount of UsnRNP-specific proteins.

4.1.2. Spliceosomal assembly

Introns contain highly conserved bases at the 3' and 5' splice site and at the branch point in order to provide specific recognition sites for the spliceosome. The splicing reaction initiates after the recognition of the 5' splice site by the U1 snRNA (Figure 4-1). The U2 snRNP subsequently interacts with the intron at the branch point forming the pre-spliceosomal complex A. The U5 snRNP assembles with the U4/U6 snRNP and the resulting U4/U6.U5 tri-snRNP joins U1 and U2 snRNPs and the other proteins of the pre-spliceosome forming the fully active spliceosome (the B complex). During the assembly of the tri-snRNP and the spliceosome, respectively, several proteins having fulfilled their function are leaving the complex (Makarov *et al.* 2002, Mougin *et al.* 2002). In the spliceosome, major rearrangements take place within the snRNPs, which result in U1 and U4 snRNPs leaving the spliceosome (the B* complex) and the two transesterification reactions can take place (the stage subsequent to the first transesterification process is known as the C complex). The rearrangements include several RNA rearrangements within the snRNPs (Nilsen 1998), an

example is the replacement of U6 snRNA with U1 snRNA at the 5' splice site. This rearrangement causes the disruption of the U4 and U6 snRNA interactions and the disturbance of the interactions between U1 snRNA and the splice site. These reactions are catalysed by RNA helicases of the DEXD/H box family, which require ATP for helicase activity (Staley and Guthrie 1998).
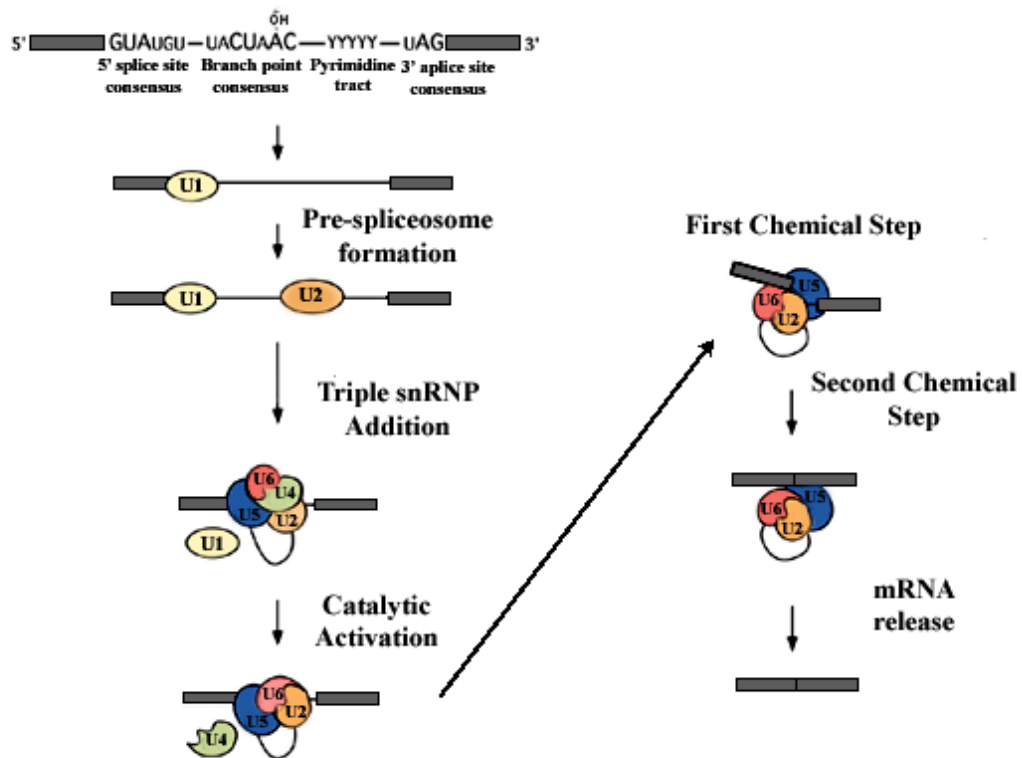
Figure 4-1 The spliceosomal assembly cycle. Different steps in the splicing of pre-mRNA starting with the initial recognition of the 5' splice site by the U1 snRNA and pre-spliceosome formation. The tri-snRNP joins the pre-spliceosome and after the disassembly of U1 and U4 snRNPs the actual splicing reactions can take place. Finally after the two reactions, U2, U5, U6 snRNPs and the intron as lariat structure leave the exons and the snRNPs are recycled for a new round of splicing. This figure is adapted from http://www.UCSF.edu/guthrie/projects.htm.

Besides the RNA helicases there is a single GTPase identified, the U5 snRNP specific 116K protein. Interestingly it shows homology to the translational elongation factor, EF-2 (Fabrizio

11

et al. 1997). U5-116K probably has a role in unwinding U4/U6 RNA or to stabilise Brr2, a characterised helicase (Bartels *et al.* 2003). The splicing reaction itself does not require energy. After the two transesterification reactions the intron leaves as a lariat structure together with U2, U5 and U6 snRNPs. The snRNPs disassemble and are ready to participate in another round of splicing.

4.1.3.   The catalytic reaction

The actual splicing takes place through two transesterification processes (Figure 4-2).
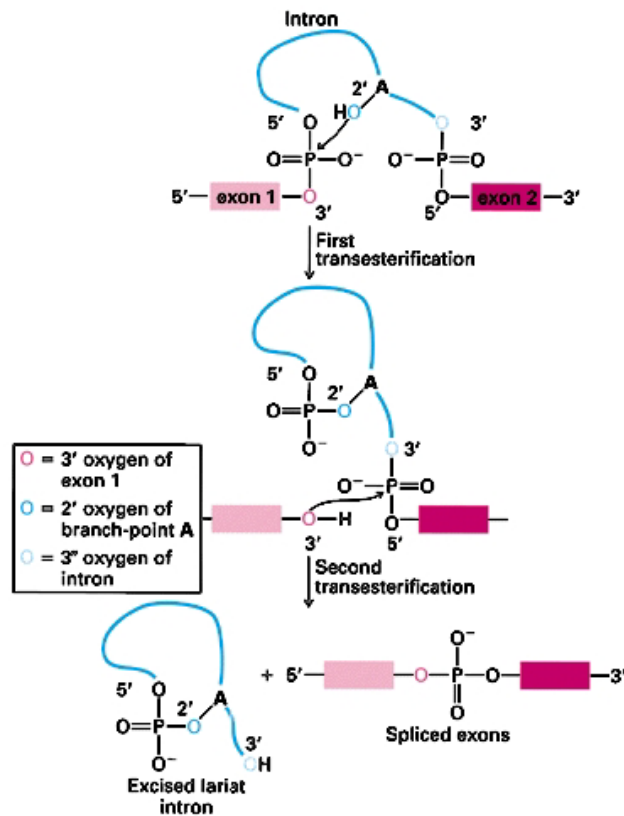
Figure 4-2 The two transesterification reactions splicing the exons. The figure is adapted from Lodish *et al.* (1995).

The intron contains conserved residues that direct and identify the splice sites and the branch site adenosine. The first transesterification reaction is initiated by an attack of the 2' hydroxyl group from the branch site adenosine on the 5' phosphate and a phosphoester bond is formed. The 3' oxygen of exon 1 attacks the phosphate on the 5' exon and a second phosphoester bond is formed. The intron leaves the spliced exon as a branched lariat structure. An interesting and yet unanswered point in the splicing process is whether the catalytic steps are mediated exclusively by spliceosomal RNA (as in the case of the ribosome where the RNA is responsible for peptide bond formation (Nissen *et al.* 2000)) or if proteins also take part in the catalysis (see Collins and Guthrie 2000 for a review).

4.2. U5 snRNP

4.2.1. Composition of the U5 snRNP

The U5 snRNP is a major component of the spliceosome. It contains, besides the U5 snRNA, the common Sm proteins (B/B', D1, D2, D3, E, F and G, Figure 4-3 B), which are arranged in a seven-membered ring (Kambach *et al.* 1999). The human 20S U5 snRNP contains the nine particle-specific proteins 15kD, 40kD, 52kD, 100kD, 102kD, 110kD, 116kD, 200kD and 220kD (see Figure 4-3 for a representation of the proteins in the human tri-snRNP) (Bach *et al.* 1989, Behrens and Lührmann 1991, Gottschalk *et al.* 1999). The U5 snRNP specific proteins are highly conserved between human and yeast. In yeast (*S. cerevisiae*) seven particle specific proteins are found. Those are the Prp8, Brr2, Snu114, Prp6, Prp28, Snu40 and Dib1 (corresponding to the 220K, 200K, 116K, 100K, 102K, 52K and 15K in human) (Gottschalk *et al.* 1999, Stevens *et al.* 2001). The U5 snRNP specific proteins mediate important steps during assembly, maturation and dissociation of the spliceosome (Staley and Guthrie 1998). So far structural information about the U5 snRNP is limited to the following: the X-ray crystallography structure of two Sm protein complexes (Kambach *et al.* 1999) and the U5-15K (Reuter *et al.* 1999) have been determined and the NMR structure of the C-terminal 86 amino acids of U5-52K is known (Freund *et al.* 1999).
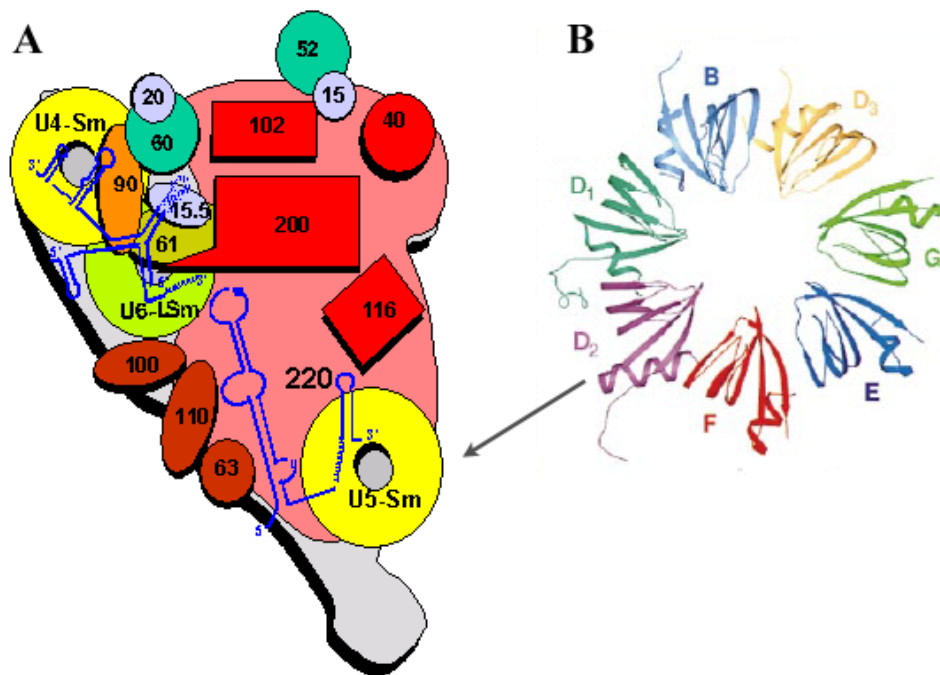
Figure 4-3 (A) A schematic drawing of the human tri-snRNP. All U5 snRNP specific proteins are shown. This representation is only a model since some of the proteins are no longer present in the tri-snRNP after assembly. Adapted from B. Kastner, MPI for Biophysical Chemistry, Göttingen, Germany. (B) Ribbon diagram of the heptamer model of the Sm proteins (Kambach *et al.* 1999).

4.2.2.   Function of U5 snRNP

The U5-snRNP interacts with the U4/U6 di-snRNP and afterwards the tri-snRNP joins the spliceosome, which now become the catalytically active spliceosome. The U5 snRNP, together with U2 and U6 snRNPs, is present during the two transesterification reactions. This indicates an important function of all three snRNPs or parts of the snRNPs during the catalytic reaction. E.g. the U5 snRNA contains a well conserved loop that could be involved in positioning the exon in the second transesterification step (Newman 1997). U5 snRNP proteins mediate many of the RNA rearrangement steps and one of the best suggestions of a spliceosomal protein co-factor is found within the U5 snRNP. This is the U5-220K (Prp8 in yeast), which is very well conserved through all organisms (62% identity from yeast to humans). U5-220K is a huge protein (2413 amino acids in yeast and 2335 amino acids in

human), it however lacks well-known sequence motifs and the domains are therefore not classified in known domains, which makes functional predictions of the U5-220K more difficult. The protein makes extensive contacts to U5 and U6 snRNAs, to the pre-mRNA at the branch site and splice sites, thus probably plays a role as a stabiliser of the U5 snRNA interactions with the exon (Teigelkamp *et al.* 1995). Furthermore it interacts with many proteins in the U5 and U4/U6 snRNPs.

4.2.3.   The U5-15K protein

The U5-15K is an evolutionarily highly conserved component of the U5 snRNP and U4/U6.U5 tri-snRNP containing 142 amino acids. The U5-15K homologue in yeast has been shown to be essential for viability (Berry and Gould 1997). The U5-15K homologue in *S. pombe* is denoted Dim1 and in *S. cerevisiae* Dib1.

The crystal structure of the U5-15K was the first structure of a U5 specific protein to be determined (Reuter *et al.* 1999). The structure revealed a thioredoxin-like fold that is characterised by a four-stranded β-sheet consisting of pairs of parallel and anti-parallel strands flanked by three α-helices. Compared to thioredoxin the U5-15K has additional 37 residues of which 22 residues extend the C-terminus. The canonical Cys-X-X-Cys motif found in thioredoxin and forming the functional disulfide bridge is not present in U5-15K. However, in U5-15K Cys38 makes an equivalent disulfide bridge to Cys79, this nevertheless has no influence on the function of the U5-15K protein. The U5-15K interacts with the U5-102K, another U5 snRNP-specific protein (Uetz *et al.* 2000). The U5-102K is known as a U4/U6.U5 tri-snRNP bridging protein making contacts to the U4/U6-specific 61K protein (Makarova *et al.* 2002, Schaffert *et al.* 2004).

4.2.4.   The U5-52K protein

The U5-52K protein is a U5 snRNP specific protein found in both human and yeast containing 341 amino acids (Bach *et al.* 1989, Stevens *et al.* 2001). The U5-52K protein belongs to the group of proteins leaving the U5 snRNP during the process, as it is thought to dissociate from the U5 snRNP during assembly of the U4/U6.U5 tri-snRNP (Mougin *et al.* 2002). The U5-52K homologues in yeast are called Snu40 in *S. cerevisiae* and Lin1p in *S.*

pombe. U5-52K associates with the U5 snRNP by interaction of its GYF-domain with the U5-15K protein (Laggerbauer *et al.* 2005). U5-52K was also identified to interact with another U5 specific protein, the U5-102K protein. The interaction takes place through the N-terminus of U5-52K. The U5-52K protein (Laggerbauer *et al.* 2005) is an interesting example of a bifunctional protein. This protein was first identified as a CD2 binding protein (CD2BP2; CD refers to a cluster of monoclonal antibodies) (Nishizawa *et al.* 1998). CD2 is present on the surface of T cells, thymocytes and natural killer cells, and contains proline-rich regions allowing interaction with several putative partners. Among them three interaction partners are well defined, namely the two SH3 domain-containing proteins, the CD2-associated protein (CD2AP) (Dustin *et al.* 1998) and the CD2 cytoplasmic tail-binding protein (CD2BP1) (Li *et al.* 1998). The third one is CD2BP2, which interacts with CD2 through two PPPPGHR motifs located in the cytoplasmic tail of CD2. Overexpression of a C-terminal CD2BP2 fragment, that binds to CD2, was reported to enhance the production of interleukin 2, which in turn is involved in immune response reactions (Nishizawa *et al.* 1998). The NMR structure of the C-terminal 86 amino acids revealed a novel poly-proline interacting domain, which was denoted the GYF-domain (Freund *et al.* 1999). Subsequently, the structure of the GYF-domain in complex with a peptide containing the cognate proline-rich PPPPGHR sequence present in CD2 has been determined by NMR (Freund *et al.* 2002) (Figure 4-4). This structure displays a different binding mode for a poly-proline motif with regard to the known mode for the SH3 domains. In contrast to the "classical" SH3 domain, which has two hydrophobic binding pockets the GYF-domain contains a single, central deep binding pocket. Recently, sequence analysis demonstrated that CD2BP2 is identical to the 52K protein associated with the spliceosomal 20S U5 snRNP (Laggerbauer *et al.* 2005).
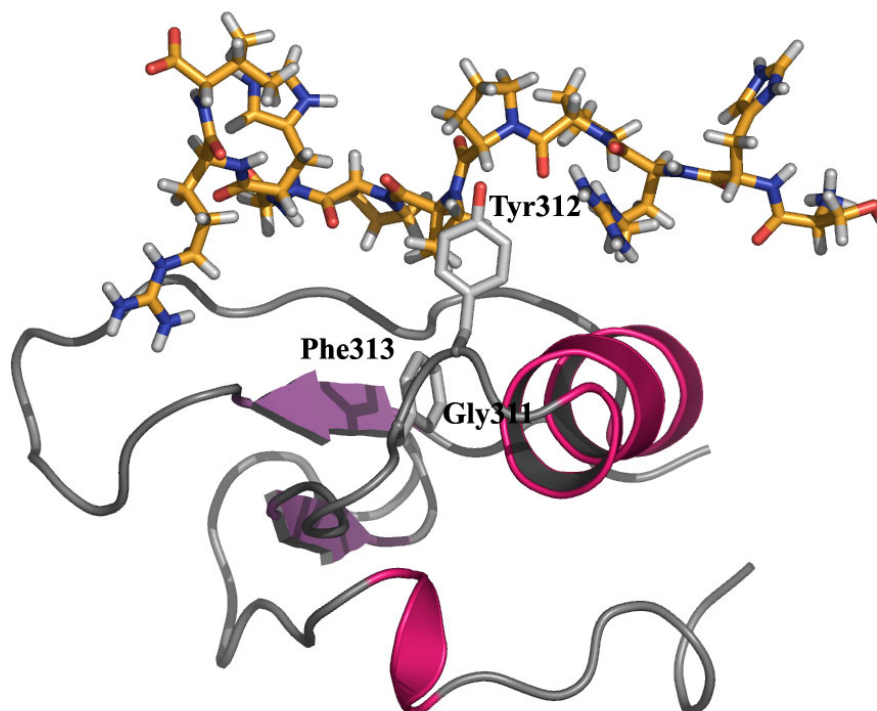
Figure 4-4 Structure of the GYF-domain in complex with the proline rich peptide from CD2. The GYF-domain is shown in cartoon representation with violet β-sheets and magenta α-helices. The CD2 peptide is shown in orange stick mode. The glycine, tyrosine and phenylalanine interacting with the peptide are marked (PDB code: 1L2Z).

4.3. Moonlighting proteins

4.3.1. Moonlighting

Due to an increasing number of bi- and multi-functional proteins the 'one protein - one structure - one function' paradigm has become outdated. The functionality of proteins may vary with respect to the intracellular localization, the occurrence in different cell types, the binding of different substrates, co-factors, or other proteins (for reviews see Copley 2003, Jeffery 2003, for reviews see Moore 2004). Bi- or multi-functional proteins, also called 'moonlighting' proteins, participate in many diverse cellular processes.

4.3.2. Examples of moonlighting proteins

Various ribosomal proteins also have a function in DNA replication, as well as in transcription, RNA processing and DNA repair (Wool 1996). Eukaryotic elongation factors have additional functions in signal transduction, translational control, apoptosis, F-actin building and severing (Ejiri 2002). CD26/DPPIV is a T cell activation antigen, which also functions as serine protease, receptor, costimulatory protein, adhesion protein and is furthermore involved in apoptosis (Boonacker and Van Noorden 2003). The crystallins are a group of proteins associated in a high concentration with the lens of the eye. Several of the proteins were shown to originally posses a function among the ubiquitous enzymes and have not lost that function (Piatigorsky 2003). The crystal structure of η-crystallin/retinal dehydrogenase has recently been determined (Bateman *et al.* 2003) and it reveals that in comparison to other dehydrogenases the structure is less flexible, has a change in co-factor binding site and in the tunnel to the active site. η-crystallin is thus an example of a moonlighting protein that has lost some features from the initial function in order to adapt to a second function, a so called 'adaptive conflict' (Jeffery 2004). An example of a bifunctional protein involved in pre-mRNA splicing is Clf1p from *S. cerevisiae*. It interacts with U5 and U6 snRNAs and plays a role in the assembly of the U4/U6.U5 tri-snRNP to the U1/U2 snRNP containing pre-spliceosome (Wang *et al.* 2003). Furthermore, Clf1p is involved in DNA replication, where it contacts the origin of replication complex (ORC) through the DNA replication initiation protein Orc2p (Zhu *et al.* 2002).

With the two characterised functions of U5-52K, it is a novel and exciting example of a moonlighting protein. It both participates in the large and composite spliceosomal machinery and furthermore has another function in the immune response (section 4.2.4).

5. Materials and Methods

5.1. Overexpression and purification

Recombinant human U5-15K protein and the GYF-domain of U5-52K were produced and purified separately as described earlier (Reuter and Ficner 1999, Laggerbauer *et al.* 2005) with minor changes.

5.1.1.   Production and purification of the U5-15K protein

The gene was inserted in the temperature sensitive induction vector, pXC35 (Cheng and Patterson 1992) and transferred into the *E. coli* strain, TAP106. 200 ml of LB medium with 100 mg/L ampicillin and 50 mg/L kanamycin was inoculated with a single colony and grown overnight. The overnight culture was 50-fold diluted in 5 L LB medium with antibiotics. The cells grew until an OD_{600} of 1.0 in a bioreactor (Applikon Biotechnology, Remsfeld, Germany) at 30 °C. The expression of the gene was induced by adding 5 L pre-warmed LB medium (60 °C) to the bioreactor, after which the cells were grown for another 4 h at 42 °C before harvesting.

Cells from 2 L culture were opened with a microfluidiser S100 (Microfluidics^{TM}, Newton, USA) in lysis buffer (100 mM Tris-HCl pH 7.8, 2 mM DTT, 1 mM EDTA, DNaseI, 10 mM $MgCl_2$ and 1 EDTA-free Protease Inhibitor Tablet (Roche Diagnostics, Mannheim, Germany)). The lysate was centrifuged at 30.000 xg for 1 h in an Avanti centrifuge (Beckman Coulter, Krefeld, Germany). The supernatant was then further centrifuged at 100.000 xg for 1 h. The supernatant was loaded on a Q-sepharose FF (Amersham Biosciences), which was equilibrated with buffer A (100 mM Tris-HCl pH 7.8 and 2 mM DTT). U5-15K was eluted with a linear gradient of buffer A and buffer B (100 mM Tris-HCl pH 7.8, 1 M NaCl and 2 mM DTT). As visualised on a 15% SDS-PAGE, the fractions containing U5-15K were pooled and $(NH_4)_2SO_4$ was added to a final concentration of 1 M. The protein was further purified on a HiLoad phenyl sepharose column (16/10 Amersham Biosciences), which was equilibrated with buffer A (100 mM Tris-HCl pH 7.8, 1 M $(NH_4)_2SO_4$, 2 mM DTT and 1 mM EDTA). The bound protein was eluted with a linear gradient of buffer A and buffer B (100 mM Tris-HCl pH 7.8, 2 mM DTT and 1 mM EDTA).

19

5.1.2. Production and purification of the GYF-domain

The 86 amino acid C-terminal of U5-52K (the GYF-domain) was cloned into the pET28a vector with an N-terminal His$_6$-tag. The gene was expressed in the *E. coli* strain BL21(DE3)RP (Novagen) after analysing the rare codon content of the gene. A single colony was added to 500 ml LB medium with 50 mg/L kanamycin and 20 mg/L chloramphenicol and grown overnight. The overnight culture was diluted 50-fold in 10 L of LB medium with antibiotics and the cells were grown in a bioreactor (Applikon Biotechnology) at 30 °C until the OD$_{600}$ reached 0.6. The production of the GYF-domain was induced with 1 mM IPTG and the cells were grown for 4 h before harvesting.

The cells were lysed in lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM β-mercaptoethanol and 1 EDTA-free Protease Inhibitor Tablet (Roche Diagnostics)) with a microfluidiser S100 (Microfluidics^{TM}) and the lysate was centrifuged for 1 h in an Avanti centrifuge at 100.000 xg (Beckman Coulter). The supernatant was incubated for 1 h with Talon beads (Clontech) at 4 °C. After three washes with buffer (lysis buffer with 10 mM imidazole), the bound protein was eluted with elution buffer (lysis buffer with 100 mM imidazole). The column was finally washed with lysis buffer containing 500 mM imidazole to make sure that everything was eluted from the column. Eluted fractions were analysed with SDS-PAGE.

5.1.3. Reconstitution of the complex

The *in vitro* reconstitution of the final complex was performed in two different ways. Initially the purified U5-15K protein was added in excess to the GYF-domain still bound to the Talon beads. The complex was formed overnight at 4 °C. After several rounds of wash, the complex was eluted from the beads as described above. Afterwards the complex was purified on a gel filtration column (Superdex 75 26/60 Amersham Biosciences) to get rid of uncomplexed protein and to transfer the complex in a suitable crystallisation buffer (20 mM Tris-HCl pH 7.8, 120 mM NaCl, 2 mM DTT and 1 mM EDTA). The pure complex was concentrated (Vivaspin) to 9.5 mg/ml, 0.3% (v/v) H$_2$O$_2$ was added and the sample was centrifuged at 23.000 xg for 5 minutes before crystallisation set up.

Another approach for the reconstitution of the complex was also tested, where the His_6-tag of the GYF-domain was cleaved by thrombin at 4°C overnight. Adding PMSF to a final concentration of 0.5 mM inhibited thrombin cleavage. The proteins were mixed in molar ratio and incubated overnight at 4°C. Afterwards the complex was purified on a gel filtration as described above to get rid of excess unbound proteins and the His_6-peptide.

5.2. Crystallisation and data collection

5.2.1. Crystallisation

Initial crystallisation screening with the complex was performed with Hampton Screen 1-2 (Hampton Research, Riverside, USA) and a screen of different precipitants (known from the literature to be successful in crystallisation like sodium chloride, ammonium sulphate and sodium malonate) versus pH. Initial condition for crystal growth was Hampton Screen 1, condition 46 with 18% (w/v) PEG 8000, 100 mM sodium cacodylate pH 6.5 and 200 mM calcium acetate. The crystals were improved by screening with respect to pH, buffering agent and PEG. The additive screens 1-3 (Hampton Research, Riverside, USA) were tested after finding a condition giving three-dimensional crystals.

The final condition for the crystallisation of the complex as plate bundles was 14% (w/v) PEG 2000 mme, 100 mM MES pH 6.5, 100 mM calcium acetate and 4% (v/v) 1,4-butanediol at 4°C.

5.2.2. Cryo-cooling

The crystals were flash frozen in liquid nitrogen and several different solutions were tested as cryoprotectant: in the beginning the best results were achieved with perfluoropolyether. After having established the final crystallisation condition, the cryoprotectant used was 17% (w/v) PEG 2000 mme, 100 mM MES pH 6.5, 100 mM calcium acetate and 15% (v/v) 1,4-butanediol.

5.2.3. Data collection

X-ray diffraction data were collected at 100K on a Bruker (Delft, Netherlands) copper rotating anode generator with Osmic mirrors, equipped with a Mar345 imaging plate detector

(MarResearch, Eppendorf, Germany) to a resolution of 2.85 Å. In order to get higher resolution data, another dataset was collected at the EMBL beam line X13, (DESY, Hamburg, Germany) to a resolution of 2.35 Å. The data were processed with DENZO and SCALEPACK (HKL Research, Charlottesville, USA). Both crystals belong to space group C2 (unit cell parameters: a = 136.7 Å, b = 76.9 Å, c = 77.0 Å, $\alpha = \gamma = 90°$ and $\beta = 94.3°$). There are three complexes in the asymmetric unit, resulting in a Matthews (Matthews 1968) coefficient of 2.6 $Å^3$/Da and a solvent content of 53%.

5.3. Structure determination, refinement and structure analysis

5.3.1. Protein crystallography

The following parts are based on Jan Drenth's book: *Principles of Protein X-ray Crystallography* (Drenth 1999).

The structure determination of a macromolecular structure with the method of X-ray crystallography requires suitable crystals for performing the experiment. During the experiment the crystal is exposed to X-rays. The X-ray waves will change direction after hitting the crystal and from the resulting diffraction pattern a model of the three-dimensional structure of the macromolecule can be built. The immediate result from the diffraction pattern is an electron density map, into which a model of the structure is fitted. The calculation of the electron density ρ at every position x, y, z in the unit cell is done with a Fourier transform and is defined as:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

where $\mathbf{F}$(hkl) is the structure factor and $|\mathbf{F}$(hkl)$|$ can be derived from the intensities of the spots on the diffraction pattern. The phase angle of the diffracting wave α(hkl) cannot be derived from the diffraction pattern. However, several methods have been developed in order to circumvent this problem known as the phase problem.

5.3.2. Solving the phase problem

The three most often used methods to obtain phase information in macromolecular crystallography are: MR (Molecular Replacement), MIR (Multiple Isomorphous Replacement) and MAD (Multiple Wavelength Anomalous Dispersion).

In molecular replacement[1] the knowledge of a known similar structure is used to search for the new and unknown structure, since related proteins are believed to have a similar three-dimensional fold. The phases of the known structure are then used as initial phases for the unknown structure. For placing the known structure in the new unit cell both the proper orientation and position are required, therefore both a rotation and translation step is performed. There are many different programs to look for molecular replacement solutions: MOLREP (Vagin and Teplyakov 1997), CNS (Brunger *et al.* 1998), AmoRe (Navaza 1994), EPMR (Kissinger *et al.* 1999), COMO (Jogl *et al.* 2001) and PHASER (Storoni *et al.* 2004). An important success criterion for the method is the identity between the two structures and a rule of thumb is that they should have at least 30% sequence identity.

MIR (Multiple Isomorphous Replacement) requires the presence of a heavy atom attached to the protein. In order to associate heavy metal atoms with protein molecules, the crystal is traditionally soaked in a heavy atom solution before exposing it to the X-rays. The intensity differences between the native crystal and the heavy metal atom containing crystal are used to determine the positions of the heavy atoms, which then serve as a starting point for determining the protein phases. An important requirement for this method is isomorphism between the native crystal and derivative crystal(s). Frequently more than one derivative is necessary because of the ambiguity of the phase angle.

MAD[2] (Multiple Wavelength Anomalous Dispersion) uses the break down of Friedels law for anomalous scatters at certain energies. The arising differences of the intensities between the Bijvoet mates can be used for the phase angle determination. Especially the substitution of methionine with selenomethionine during the expression of the protein has been very

---

[1] Pioneered by Rossman and Blow (Rossman and Blow 1962).

[2] Pioneered by Hendrickson (Hendrickson and Ogata 1997).

significant for the success rate of the method. Anomalous scattering contains a real (f') and an imaginary (f'') part. The strategy for a data collection is to collect data at different wavelengths: Peak wavelength (λ_1) where f'' has its maximum, inflection point (λ_2) where f' has it maximum and either high or low energy remote (λ_3) where both f' and f'' are small. However, the signal obtained from anomalous scattering is normally quite small (compared to conventional heavy atom phasing), therefore it is important to have a high occurrence of scatters (with selenomethionine a rule of thumb is one Se per 15-20 kDa) and a precise measurement of the signal, hence preferably a high multiplicity. A major advantage of the MAD method is the inherent isomorphism achieved, since only one crystal is used for the entire experiment. Among widely used programs for determining heavy atom sites and for calculating the phases in the MIR and MAD methods are: CNS (Brunger *et al.* 1998), SHELXD (Schneider and Sheldrick 2002), SnB (Miller *et al.* 1994), SOLVE (Terwilliger and Berendzen 1999) and SHARP (Global Phasing, Cambridge, UK).

Many structures have been determined by a combination of the three methods and especially MIRAS (Multiple Isomorphous Replacement using Anomalous Scattering) and SIRAS (Single Isomorphous Replacement using Anomalous Scattering) have found wide application for experimental phasing. Finally, SAD (Single wavelength Anomalous Dispersion) has been used, where data collected at the peak wavelength are sufficient for determining the sub-structure and hence the protein structure.

5.3.3.   Determining the structure

The structure of the U5-15K-GYF-domain complex was determined by molecular replacement using the program MOLREP from the CCP4 (Collaborative Computational Project 1994) program package. The known crystal and NMR structures of the U5-15K (PDB code: 1QGV) and the GYF-domain (PDB code: 1GYF) respectively, were used as search models.

5.3.4.   Refinement of the complex

The model was refined with CNS (Brunger *et al.* 1998) using simulated annealing, individual B-factor refinement and positional refinement. 5% of the reflections were omitted in the

refinement for the calculation of R_{free} (Brunger 1993). The water molecules were added automatically with the ARP/wARP (Perrakis *et al.* 1999) function in REFMAC (Murshudov *et al.* 1999) and afterwards examined manually with respect of electron density and reasonable hydrogen bonding possibilities. The model was visualised and manually corrected with the program O (Jones and Kjeldgaard 1997). The last refinement step was a TLS refinement in REFMAC.

5.4. Structure analysis

5.4.1. Validation and figures

The quality of the final model was verified with the validation program PROCHECK (Laskowski *et al.* 1993). For the comparison of the GYF-domain structure, the most representative NMR structure (out of the 16 available) was calculated on the OLDERADO server (Kelley and Sutcliffe 1997). The total buried surface area upon complex formation was calculated with the program AREAIMOL and the shape complementarity coefficient (sc) was calculated with SC (Lawrence and Colman 1993), both programs are included in the CCP4 program package (Collaborative Computational Project 1994). Figures were made in PYMOL (DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, USA). The electrostatic potential was calculated by the Poisson-Boltzmann equation as implemented in the program DelPhi (Honig and Nicholls 1995), (Nicholls and Honig 1991), and visualised with PYMOL. The amino acid sequence alignment was done using CLUSTALW (Thompson *et al.* 1994) and displayed with ESPript (Gouet *et al.* 1999).

6. Results

6.1. Purification of the complex

6.1.1.   Production and purification of the U5-15K protein

U5-15K was expressed from the pXC35 vector in TAP106 cells without any affinity tag. The final yield of purified U5-15K was 20-25 mg from 2 L of culture. From the anion exchange column (Figure 6-1) U5-15K elutes approximately at 15% buffer B (150 mM NaCl). Based on SDS-PAGE of eluted fractions, fractions containing U5-15K (39-50) were pooled (Figure 6-1 D) and loaded on the phenyl sepharose column. After elution, the protein was highly pure (Figure 6-1 E) and fractions containing the protein (52-59) were pooled. Attempts were made to freeze half of the protein and store it at -80 °C for later use, but the protein tended to precipitate. Therefore subsequently a smaller amount of protein (cells from 1 L culture) was purified so that fresh protein always was used for reconstitution of the complex.

6.1.2.   Production and purification of the GYF-domain

The GYF-domain was expressed from the pET28a vector in BL21(DE3)RP cells with a His$_6$-tag. The protein was batch purified on Talon beads (Figure 6-1 F) and approximately 5-10 mg of GYF-domain was obtained from 1 L of culture.

6.1.3.   Reconstitution of the complex

After the separate purification of the human U5-15K protein and the GYF-domain the binary protein complex was obtained by *in vitro* reconstitution using two different approaches. In the first approach, U5-15K was added in molar excess to the GYF-domain bound to the talon beads and the formed complex was eluted from the beads after three washing steps (Figure 6-1 F+G) and fractions containing the complex were pooled. In the second approach, U5-15K and the GYF-domain were allowed to form the complex after the removal of the affinity tag of the GYF-domain. Gel filtration chromatography as the last step in both approaches revealed an apparent molecular weight of 25 kDa corresponding to a heterodimeric complex composed of the two proteins in a 1:1 molar ratio (Figure 6-1 C).

26

Figure 6-1 Chromatograms and SDS-PAGEs from the purification of the U5-15K-GYF-domain complex. The lanes on the SDS-PAGE are labelled according to the fraction numbers on the profiles. (A) The first step of the purification of U5-15K was an anion exchange column. (B) A phenyl sepharose column was the second step after the anion exchange column. (C) The reconstituted complex was purified with gel filtration chromatography as a final step. (D) SDS-PAGE of fractions from the ion exchange column. The markers were previously purified U5-15K and GYF-domain proteins. (E) SDS-PAGE of fractions from the phenyl sepharose column. The U5-15K protein was highly pure after elution from the column. (F) Fractions from the initial steps of GYF-domain purification and formation of the complex on the Talon beads. (G) Eluted fractions of the complex from the Talon beads. (H) Fractions from the final purification step on the gel filtration column

On the chromatogram a shoulder was seen on the peak from the gel filtration. This was due to insufficient separation between the complex and unbound GYF-domain as a result of the small difference in molecular weight. This is also observed on the SDS-PAGE from the gel filtration (Figure 6-1 H). Fraction 31 appeared to be a mixture between the binary complex and uncomplexed GYF-domain and in fraction 32 there is a higher amount of uncomplexed GYF-domain than of the complex. Fractions 28-30 were pooled and used for crystallisation experiments. As seen from all SDS-PAGEs the GYF-domain migrated slower than U5-15K despite the fact that the molecular weight of the GYF-domain is smaller than for U5-15K. The reason for this is a high number of acidic amino acids in the GYF-domain and in U5-52K in general (calculated pI of 4.0). This is also the case for the full-length protein (341 amino acids with an experimental molecular weight of 52 kDa) (Laggerbauer *et al.* 2005). The second approach of the reconstitution of the complex was initiated in order to remove the often very flexible His$_6$-tag, which can disturb crystal formation and diffracting quality.

6.2. Crystallisation and structure determination

6.2.1.　Crystallisation

Initial thin needles were obtained in Hampton Screen 1 condition 46. A screen of pH and buffers resulted in slightly better crystals with MES instead of sodium cacodylate (Figure 6-2 A+B). At an early stage the importance of hydrogen peroxide in the drop was acknowledged, since it produced significantly better crystals (see Figure 6-2). Hydrogen peroxide has been

found to improve crystal quality in other studies but its function is not known (Ole Kristensen, personal communication). It is not clear whether the effect of hydrogen peroxide is mainly to aid in a more homogenous protein sample for crystallisation or if it helps by improving actual crystal contacts. No hydrogen peroxide molecules were found later in the electron density. When changing the PEG 8000 to PEG 2000 mme (other kind of PEGs were also tested but did not improve the crystals) the crystals clearly improved (see Figure 6-2 C+D) and became more three-dimensional. The obtained crystals in Figure 6-2 C only diffracted to approximately 5 Å in a capillary on a rotating anode and turned out to be very difficult to reproduce. The crystals in Figure 6-2 D were separated in the common point of nucleation with an acupuncture needle (see Figure 6-3 D for a look of the drop after separation of the crystals).
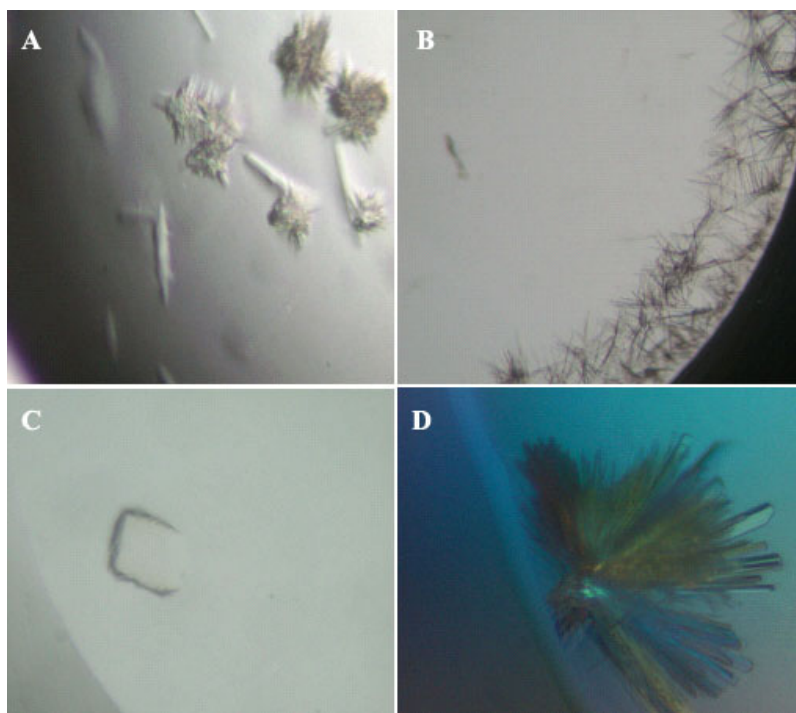
Figure 6-2 Crystals obtained in two different kind of PEG with and without hydrogen peroxide. (A) PEG 8000 and without hydrogen peroxide. (B) PEG 8000 with 0.3% hydrogen peroxide added to the crystallisation drop. (C) PEG 2000 mme without hydrogen peroxide. (D) PEG 2000 mme with 0.3% hydrogen peroxide

The crystals diffracted to a maximum of 3 Å on the synchrotron, which was achieved after annealing. Streak seeding improved neither the quality nor the singularity of the crystals. All the crystals in Figure 6-2 were grown at 20 °C.

The final improvement of the crystal quality was obtained with the Additive screens 1-3. Three-dimensional crystals were found in the drop containing 1,4-butanediol; the crystals however were not single but diffracted better than any crystals obtained so far (Figure 6-3 A+B). The crystals grew better at 4 °C (however the temperature change alone was not responsible for the change, since crystals without 1,4-butanediol did not improve at 4 °C), so the remaining drops were placed at 4 °C. One single crystal was obtained (see Figure 6-3 C), however the crystal did not diffract at all. It is not clear whether the inability of diffraction was due to poor crystal quality or problems with handling, cryoprotectant etc.
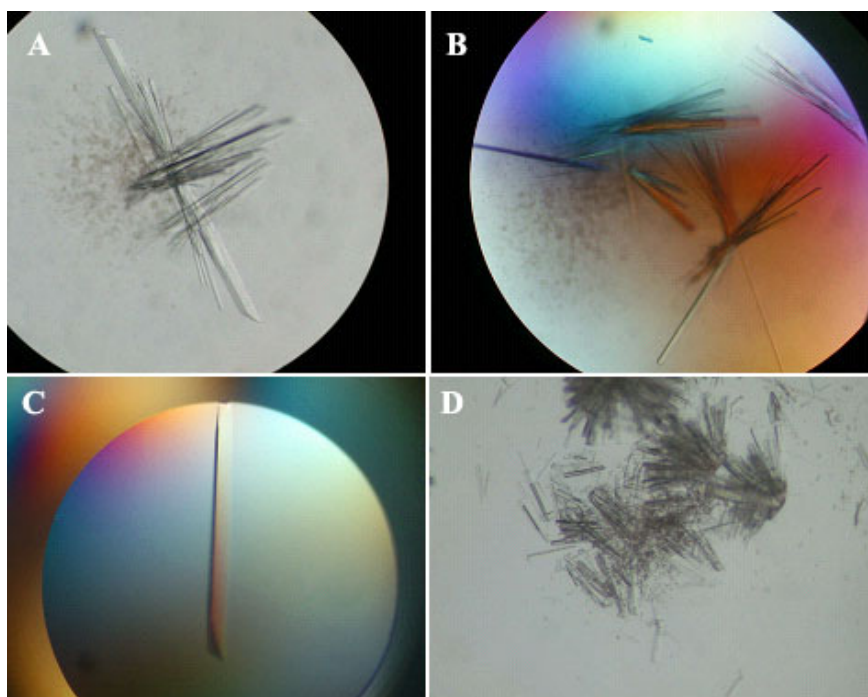
Figure 6-3 The final crystals used for data collection. (A)+(B) Crystals obtained after additive screening with 1,4-butanediol. (C) Only single crystal obtained. (D) The crystals after separation at the point of nucleation.

In general, a vast number of crystals were tested before the two well-diffracting crystals were found. These were used to collect the datasets on the rotating anode and the synchrotron. This randomisation and lack of diffracting ability of the crystals were most likely due to the crystals being very fragile and hence, probably did not survive the handling with the acupuncture needle.

6.2.2. Data collection

Data were tested and collected on a rotating anode and a higher resolution dataset was collected at the X13 beam line at DESY, Hamburg, Germany. Crystals of the complex belong to space group C2 and contain three hetero-dimeric complexes in the asymmetric unit. Data statistics are summarised in Table 6-1.

Table 6-1 Data collection statistics from the dataset collected at the rotating anode and at the synchrotron.

Data set	Rotating anode	X13 at DESY
Wavelength (Å)	1.5418	0.8
Resolution range (Å)[1]	50-2.85 (2.98-2.85)	50-2.35 (2.39-2.35)
Space group	C2	C2
Cell Dimensions (Å/°)	a = 136.5	a = 136.7
	b = 76.9	b = 76.9
	c = 77.5	c = 77.0
	β = 94.5	β = 94.3
Observed reflections	115983	72742
Unique reflections	23437	30397
Completeness (%)[1]	99.8 (99.7)	92.4 (95.2)
R_{sym}^2 (%)[1]	9 (39)	3.1 (18.4)
Average I/σ(I)[1]	10.7 (2.7)	21 (2.4)
Mosaicity (°)	0.703	0.893

[1]Numbers in parenthesis refer to the highest resolution shell. [2)] $R_{sym} = 100 \cdot \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h I(h)$ where $I_i(h)$ is the ith measurement of the h reflection and $\langle I(h) \rangle$ is the average value of the reflection intensity.

6.2.3. Structure determination

The crystal structure was determined by means of molecular replacement. An initial rotation and translation search was performed with the known structure of the U5-15K protein (PDB code: 1QGV) using the diffraction data obtained from the rotating anode X-ray source. As expected, a solution with three U5-15K molecules placed in the asymmetric unit was obtained, and after rigid body refinement, a $2|F_o\text{-}F_c|$ electron density map was calculated (Figure 6-4). This map clearly showed electron density for the missing GYF-domain molecules, although only 60% of the atoms present in the asymmetric unit were used for phase calculation. A second molecular replacement calculation was performed using the known NMR structure of the GYF-domain (PDB code: 1GYF) as search model and taking the known coordinates of the three U5-15K molecules into account. Using either all 16 deposited NMR models of the GYF-domain or alternatively a single model, representing the average of all 16 NMR structures, yielded the same solution for the three GYF domain molecules. However, only one of the three GYF-domain molecules fitted the electron density, although the differences in the correlation coefficients were not significant. Hence, the other two GYF-domain molecules were generated using the NCS operators defined by the positions of the three U5-15K molecules, resulting in a perfect fit of these GYF-domains with regard to the $2|F_o\text{-}F_c|$ electron density map that was calculated using only the U5-15K model.
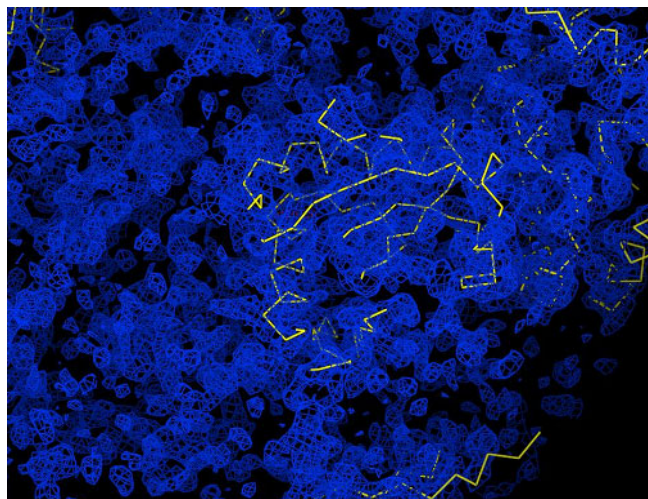
Figure 6-4 Initial electron density after the molecular replacement with the U5-15K as a model. The model solution from MOLREP is seen and from the electron density it is clear that extra density is present arising from the GYF-domain.

6.3. Refinement and structure analysis

6.3.1. Refinement of the structure

The crystal structure was refined by means of simulated annealing, positional and TLS refinement steps (see Table 6-2 for statistics). As a final step water molecules were added to the model. The structures of the U5-15K and GYF-domain are well defined in the electron density map (Figure 6-5), although both proteins are lacking a few residues at their C- and/or N-termini due to flexibility and disorder in the crystal. The main reason for the excellent map also before the refinement (Figure 6-5 A) is probably due to the fact that the phases were achieved from a rigid body refinement with finished refined identical structures. In principle model bias could be a problem, however no major rearrangements were expected upon complex formation. Furthermore, the calculation of an omit map did not hint model bias.
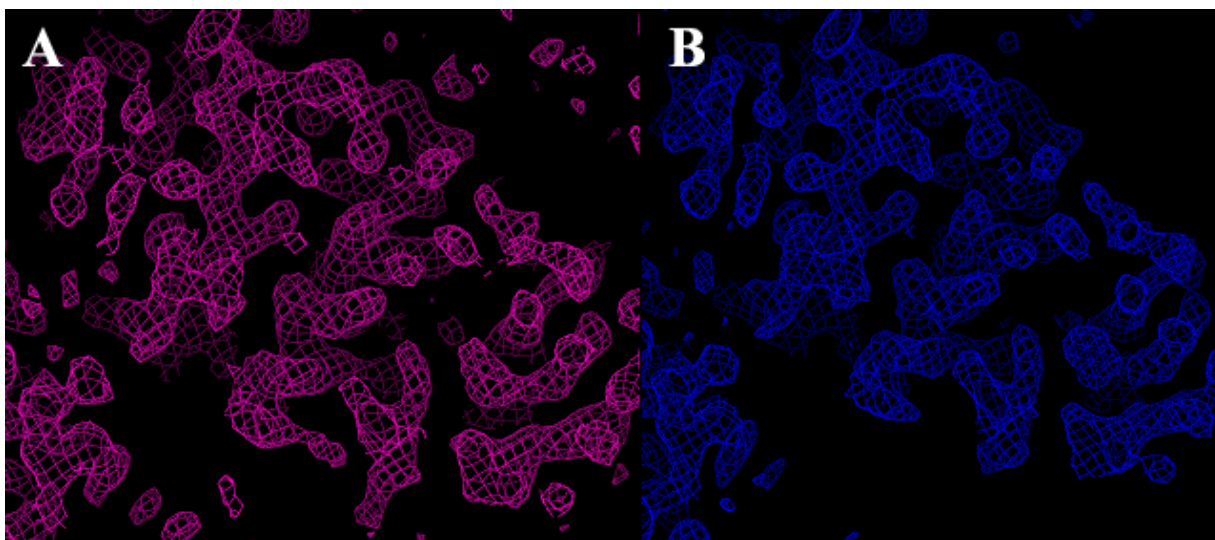
Figure 6-5 Example of the electron density before and after the refining contoured to 1σ. (A) Electron density map before the refining. (B) Electron density map after the refining.

The refined model of the U5-15K present in the complex consists of amino acids 3-137 out of 142 residues. The structure of the GYF-domain of U5-52K (residues 256-341) lacks the N-terminal 24 residues due to conformational flexibility. The three molecules of both proteins present in the asymmetric unit are almost identical. The rmsd values of all superimposed main

chain atoms of the three complexes are very small (< 0.6 Å and even smaller for the GYF-domain), and the few observed differences concern only side chain conformations mostly caused by different crystal contacts.

Table 6-2 Data statistics from the refinement of the synchrotron data.

Refinement statistics	X13 at DESY
Resolution range (Å)[1]	49.39-2.35 (2.39-2.35)
R_{cryst} (%) / R_{free} (%)[2]	21.6 (28)/26.2 (31)
Coordinate error (Å)[3]	0.185
# protein atoms	4893
# water molecules	192
Ramachandran plot[4]	
Most favourable regions (%)	92.8
Additionally allowed regions (%)	6.6
Generously allowed regions (%)	0.6
r.m.s. deviations from ideality	
Bonds (Å)/Angles (º)	0.008/1.07
Average B values (Å²)	40.592
Protein residues/ Waters	783/192

[1]**Numbers in parenthesis refer to the highest resolution shell.** [2] **$R_{cryst}=100·\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the structure factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} with 5 % of test set structure factors.** [3] **Based on Maximum Likelihood** [4] **Calculated using PROCHECK (Laskowski *et al.* 1993).**

6.3.2.   Quality of the model

The structure of the U5-15K-GYF-domain complex was refined at 2.35 Å with a final R-free of 26.2% with excellent stereochemistry as verified with PROCHECK (more than 90% of the

residues are found in the most favourable region). Figure 6-6 shows the Ramachandran plot of the final three molecules in the asymmetric unit.
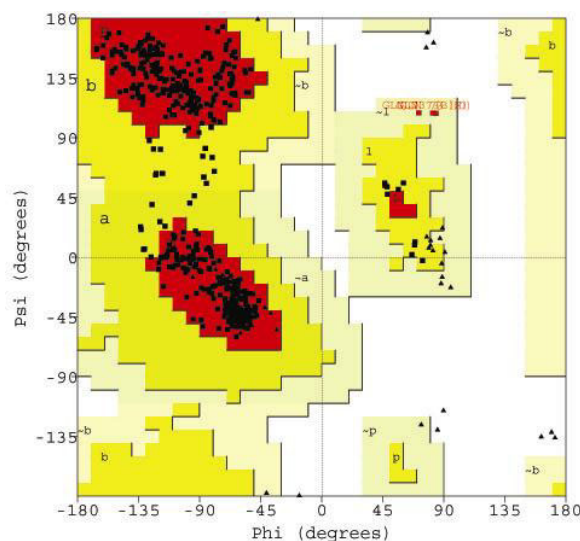
Figure 6-6 Ramachandran plot from the U5-15K-GYF-domain structure. The red area depicts the most favourable region, the yellow area is the additional allowed region and the beige region is the generously allowed region. The triangles symbolise glycines or prolines. It is desirable to have more than 90% of the residues in the most favourable region. The plot is for all three molecules in the asymmetric unit.

6.4. The complex structure

The structure of the binary complex of U5-15K and the GYF-domain of U5-52K has an elongated overall shape with the dimensions 50Å x 25Å x 25Å. It consists of a five stranded β-sheet surrounded by four α-helices (U5-15K) and an anti-parallel β-sheet flanked by three α-helices (GYF-domain) (Figure 6-7). The interaction surface of U5-15K comprises the 22 C-terminal residues Glu111-Ser132, which belong to helix α4 and the eight C-terminal residues extending the canonical thioredoxin-fold (Figure 6-8). The C-terminal residues Lys321-Thr341 of the GYF-domain are involved in the interaction with U5-15K.
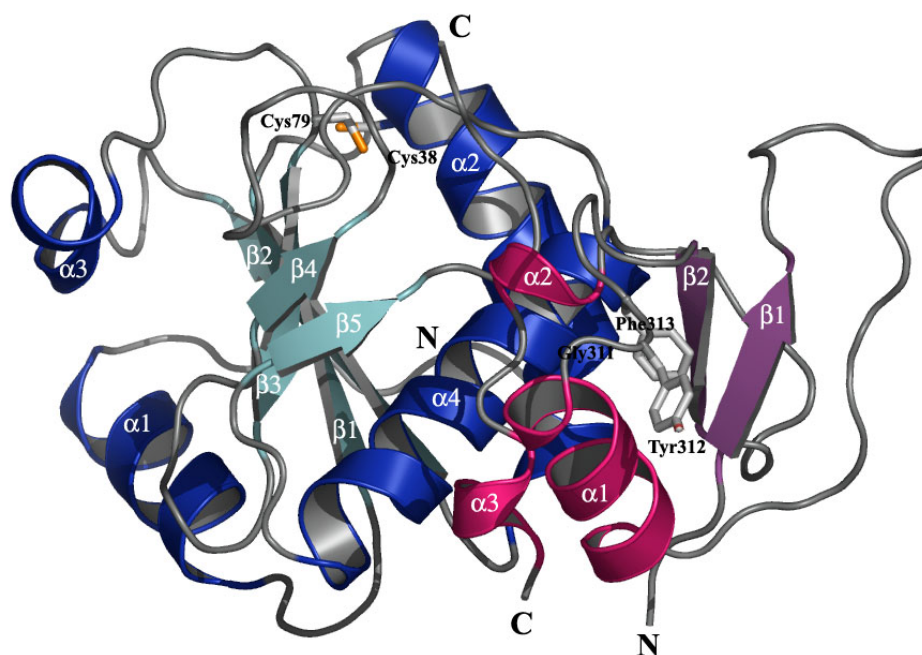
Figure 6-7 Cartoon representation of the U5-15K-GYF-domain complex. U5-15K is shown in cyan and dark blue. Cys38 and Cys79 are shown in stick mode. The GYF domain of U5-52K is shown in magenta and pink. Gly311, Tyr312 and Phe313 are displayed in stick mode.

The description of residues mediating the interaction of the two proteins is restricted to those, which are consistent for all three complex molecules present in the asymmetric unit. The GYF-domain and the U5-15K protein are bound mainly by polar interactions (Figure 6-8) including a salt bridge between Glu111, Oε1 (U5-15K) and Gln328, Nε2 (U5-52K), hydrogen bonds between side chains of Asp114, Oδ2, Glu117, Oε2 and Ser132, Oγ (U5-15K) and side chains of Tyr330, OH, Lys321, NZ and Arg334, NH2 (U5-52K) and between main chains Ser132, O (U5-15K) and side chains Arg334, Nε (U5-52K). Finally there is a hydrogen bond between the side chain of Lys125, NZ (U5-15K) and the C-terminal oxygen of Thr341 (U5-52K). Cδ1 and Cδ2 from residue Leu339 (U5-52K) are forming hydrophobic contacts to Cγ2 of Val130 and Cα of Gly122 (U5-15K). In order to confirm that the complex observed in the crystal is actually of biological relevance the complex was further analysed. The total surface areas buried upon complex formation for the three complex molecules in the asymmetric unit are 664 Å^2, 645 Å^2 and 655 Å^2 (per molecule for complexes AB, CD and EF), which gives an average of 654 Å^2. These values are in the lower end of the values suggested in Lawrence and

Colman (1993) however still supporting that the complex shown in Figure 6-7 is indeed a functional complex.
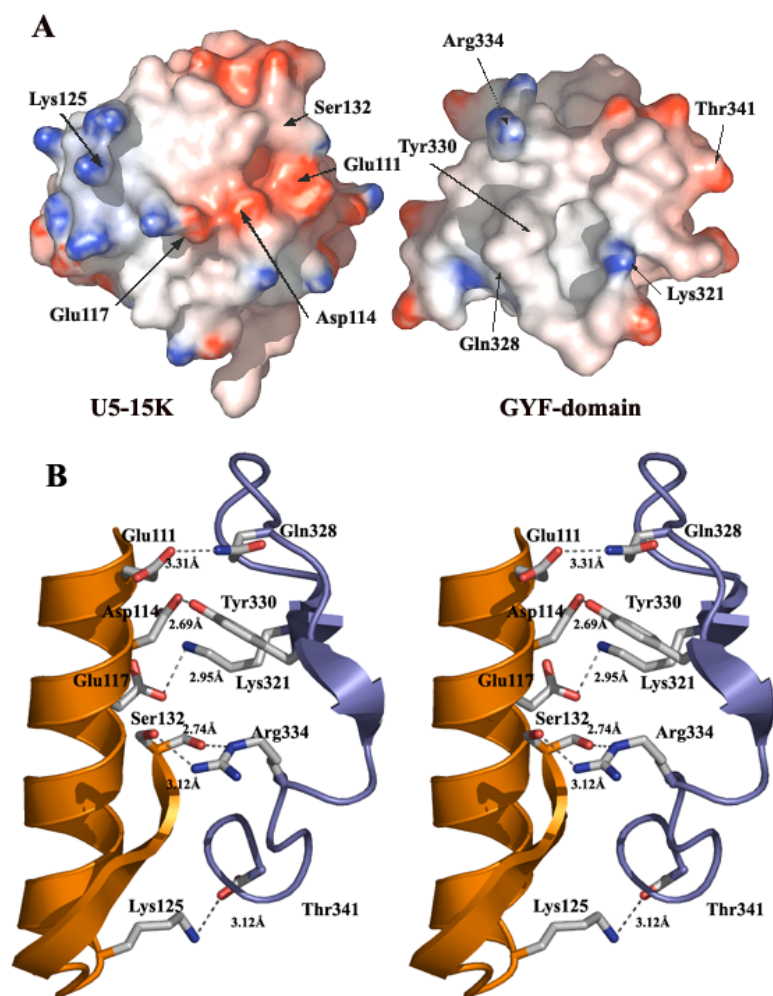
Figure 6-8 Interactions between the U5-15K protein and the GYF-domain. (A) Surface representation coloured according to the electrostatic potential. The blue patches represent the basic parts on the surface, whereas the red patches represent the acidic parts on the protein surface. The surface complementarity between the two molecules is clearly seen in the complex, which is opened like a book. The residues involved in the direct interaction are labelled. (B) A stereo picture of the polar interactions and the distances between the interacting atoms. U5-15K is shown in orange and U5-52K in blue (see text for description). The contacts between the two molecules were determined with the program CONTACT from the CCP4. The interactions are found in all three molecules. The distances are from molecules A+B, but the distances in all three complexes are similar.

37

Inspection of the molecule packing in the crystal (see Figure 6-9) reveals that both proteins of the complex mainly interact with their respective symmetry mates, as U5-15K proteins pack against other U5-15K proteins and likewise the GYF-domains are in contact with other GYF-domains. There is only one crystal contact between a U5-15K protein and a GYF-domain molecule belonging to a different complex in the unit cell. The corresponding total buried surface area for this contact between U5-15K and GYF-domain is only 45 Å^2.
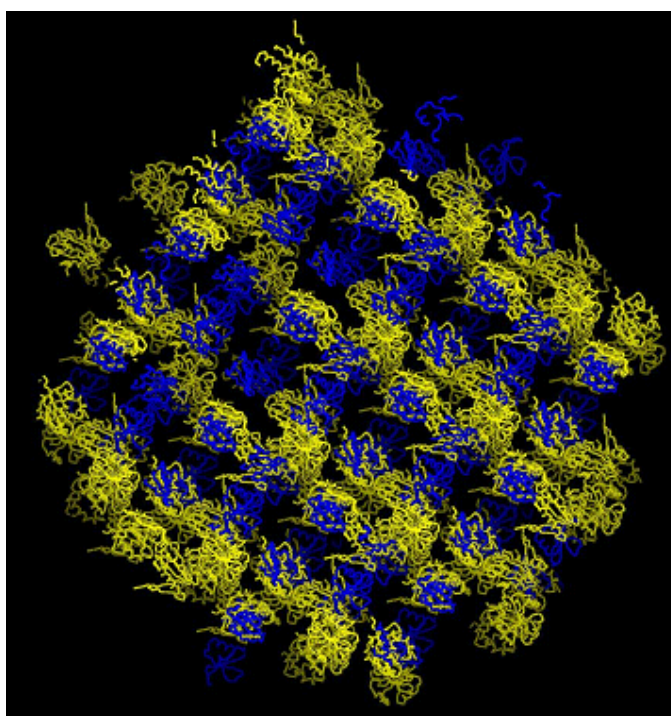
Figure 6-9 The packing of the molecules in the crystals belonging to space group C2. The U5-15K molecules are shown in yellow and the GYF-domain in blue.

Furthermore, the shape complementarity coefficient of the functional complex is 0.700 (average value for the three complex molecules in the asymmetric unit), which also strongly supports that this complex is truly the functional complex.

6.5. The U5-15K and GYF-domain structures

Superposing the previously determined structures of the sole proteins onto the complex structure reveals that the overall conformation of each protein does not change upon complex formation (Figure 6-10).

There are only minor changes compared to the crystal structure of the uncomplexed U5-15K (Reuter *et al.* 1999) (Figure 6-10 A). Residues 96-101 are missing in the crystal structure of the uncomplexed U5-15K, but they are well defined in the electron density of the complex due to stabilisation through crystal contacts. Residues 72-77 of U5-15K have a different conformation, where Tyr73-Leu75 is shifted one position in the direction of the C-terminus. Finally, Cys38 and Cys79 are in the reduced state.
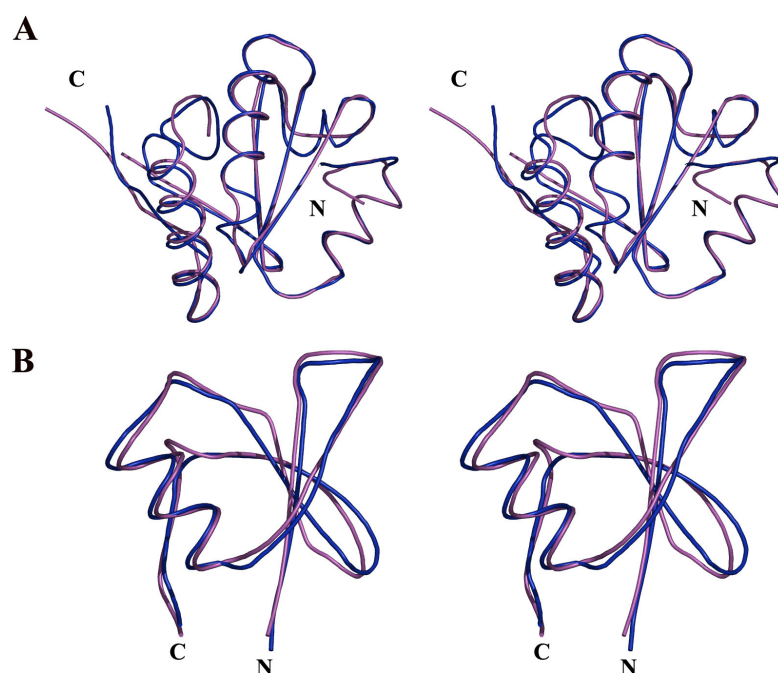
Figure 6-10 Stereo representations of the superposed components of the complex structure with the corresponding uncomplexed structures. **The structures of the two proteins in the complex are shown in blue, whereas the uncomplexed structures are depicted in magenta. The superpositions were done using the lsq-commands in O. (A) The superposition of complexed 15K and uncomplexed 15K (PDB code: 1QGV). (B) The superposition of the GYF-domain from the complex and from the NMR structure (PDB code: 1GYF)**

The structure of the GYF-domain of U5-52K in the complex is also almost identical to the structure of the uncomplexed GYF-domain (Figure 6-10 B). In the superposition shown in Figure 6-10 the crystal structure is compared with the most representative NMR structure (1GYF) (Freund *et al.* 1999). It consists of an anti-parallel β-sheet followed by an α-helix. As observed in the NMR structure of the uncomplexed GYF-domain (Freund *et al.* 1999), the crystal structure of the complex lacks the N-terminal 24 residues due to conformational flexibility and disorder in the crystal.

7. Discussion

7.1. *A novel protein-protein interaction motif for a thioredoxin-like protein*

The interaction between the U5-15K protein and the GYF-domain of U5-52K involves regions on both protein surfaces different than predicted by the previous structural studies on both proteins. Based on the crystal structure of the sole U5-15K (Reuter *et al.* 1999), two areas were previously proposed as interaction surfaces with other proteins. Compared to thioredoxin, U5-15K has a twisted fourth strand in its four-stranded-β-sheet, which forms a sheet with parts of the extended C-terminus. This feature opens up the four-stranded-β-sheet and creates a hydrophobic cavity suitable for protein-protein interactions. Another potential interaction area was identified by the mutation Gly126Asp, which was found to be lethal in yeast (Berry and Gould 1997). This amino acid is positioned in an area with several basic amino acids, which could be part of a binding surface to another protein. Hence, the introduction of a large negatively charged amino acid was suggested to impede the interaction between two proteins essential for the splicing process. In the U5-15K-GYF-domain complex, the GYF-domain interacts with the side chain of Lys125 (U5-15K).
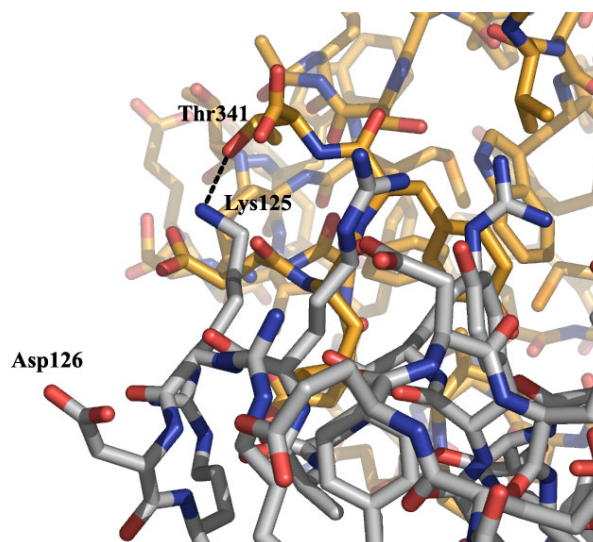
Figure 7-1 The interaction between Lys125 (U5-15K) and Thr341 (GYF-domain). U5-15K is shown in grey stick mode, whereas the GYF-domain is shown in orange stick mode.

This interaction could indeed be interrupted or affected by the mutation of the neighbouring Gly126 to an aspartate. Modelling this mutation *in silico* reveals that the side chain of the aspartate is solvent exposed (Figure 7-1). The reason for the lethal effect of the Gly126Asp mutation could be steric demands for a flexible amino acid in that position. However, Gly126 is found in a generously allowed region of the Ramachandran plot, which sterically accepts all other amino acids, hence the mutation of Gly126 should not change the conformation of the loop 125-128 and no obvious distortion of the interaction surface is expected. The fact that this mutation Gly126Asp in U5-15K is lethal in yeast cannot be related to the interaction with the U5-15K, since the U5-52K protein has been proven to be nonessential for viability in yeast (Stevens *et al.* 2001). Therefore it is more likely that the phenotype of the Gly126Asp mutation is related to an unidentified interaction with another spliceosomal protein.

Noteworthy, the structure of the U5-15K-GYF-domain complex reveals a new binding motif for thioredoxin-like proteins. Not many complex structures of thioredoxin or thioredoxin-like proteins have been determined so far. In the complex between thioredoxin and the DNA polymerase of bacteriophage T7 (Figure 7-2 A) (Kumar *et al.* 2001), the binding site of thioredoxin is near the catalytically active disulfide-bridge (with Cys32 making a hydrogen bond to Thr327 in the T7 polymerase) and the polymerase wraps around these active cysteines and a stretch of 4 amino acids (residues Arg73-Pro76). A similar mode of binding is seen in two other complexes, thioredoxin and a Ref-1 peptide (Figure 7-2 B) (Qin *et al.* 1996) as well as thioredoxin and a peptide containing its target from the transcription factor NFκB (Qin *et al.* 1995), where the peptides interact with residues in the area around the catalytic active Cys32 (both complexes have been determined with NMR). Also the crystal structure of a complex (Figure 7-2 C) comprising thioredoxin, thioredoxin reductase, FAD and AADP$^+$ reveals an interaction between the two proteins in the area around the catalytic active Cys32 (Lennon *et al.* 2000). The NADH domain of thioredoxin reductase is complementary to a loop (containing residues Tyr70 to Ile75) in thioredoxin and two phenylalanines (Phe141 and Phe142) from thioredoxin reductase interact with a hydrophobic pocket consisting of Trp31, Ile60, Gly74 and Ile75 in thioredoxin. The residues Asp104-Ala109 of the thioredoxin fit into a groove at the SPE7 (an antibody) binding-site surface in the thioredoxin-SPE7 complex (James *et al.* 2003). Further contact is established through a conserved region of thioredoxin

(residues Cys35-Ile45) to the H3 loop of SPE7. The complex between phosducin (contains a C-terminal thioredoxin-like domain) and transducin βγ (Figure 7-2 D) (Gaudet *et al.* 1996) shares some similarities with the U5-15K-GYF-domain complex with respect to the thioredoxin part of the complex interface. Only one of the cysteines from the CxxC motif is conserved in phosducin and this motif is not involved in the interaction as well. The residues participating are Lys193, Glu196-Ile198, Asn200, Glu223-Gly226 and Glu229-Arg230 (corresponding to Lys82, Glu85-Ala87, Thr89 in thioredoxin, Glu223-Gly226 and Glu229-Arg230 are not present in thioredoxin). However, also the N-terminal α-helical domain participates equally in the interaction to transducin βγ.
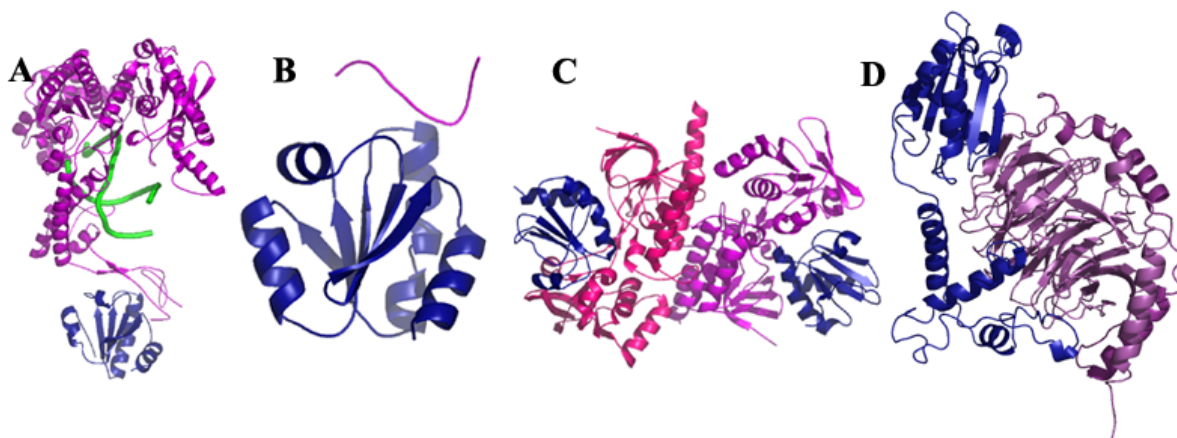
Figure 7-2 Thioredoxin and thioredoxin-like complexes. Thioredoxin or the thioredoxin-like protein is shown in blue. (A) Thioredoxin in complex with the T7 DNA polymerase (PDB code: 1T7P). (B) Thioredoxin in complex with the REF-1 peptide (PDB code: 1CQH). (C) Thioredoxin in complex with thioredoxin reductase (PDB code: 1F6M). (D) Phosducin in complex with transducin βγ (PDB code: 1F6M).

Although the last binding interface involves some of the same thioredoxin-like residues as in the U5-15K, the extension of the protein compared to thioredoxin is more comprehensive and the involved residues are also found further towards the C-terminus. Those examples however show that the different extensions of the thioredoxin protein probably have been adapted in an attempt of the proteins to fit the specific function desirable by that given protein. The U5-15K

and SnuCyp20 (a U4/U6 snRNP specific protein) (Reidt *et al.* 2000) are examples of proteins, where the spliceosome has adapted an already known fold from a homologous protein to fit its own needs and requests.

7.2. Structural basis for the bifunctionality of the GYF-domain

The GYF-domain of U5-52K interacts with proline-rich sequences (Freund *et al.* 1999), but it does not bind the U5-15K through its poly-proline binding site. A sequence database search revealed that none of the U5 snRNP specific proteins contains the proline-rich sequence (PPPPGHR) specific for interaction with the GYF-domain. Recently, it has additionally been suggested that the GYF-domain also interacts with the spliceosomal SmB and SmB' (which are common to the snRNPs U1, U2, U4, U5) by binding a less proline-rich sequence (+/GxxPPGx+) present in SmB/B' (Kofler *et al.* 2004). However, yeast two-hybrid screens (Laggerbauer *et al.* 2005) revealed that U5-52K only interacts with U5-15K and U5-102K from the U5 snRNP. If a stable interaction to Sm proteins would be present it should have been detected under the experimental conditions given and furthermore, the U5-52K protein would have been found in other snRNPs as well, since the Sm proteins are found in the U1, U2, U4 and U5 snRNPs. U5-52K interacts with U5-102K through its N-terminal part, a different region than in the interaction with U5-15K. Hence, it seems that the poly-proline binding site of the GYF-domain together with other binding sites of U5-52K could play a role in the assembly of proteins involved in the pre-mRNA splicing processes.

The alignments of the U5-15K and the U5-52K proteins from *H. sapiens*, *S. pombe* and *S. cerevisiae* (Figure 7-3) clearly show that the U5-15K is more conserved than the U5-52K between the species. This indicates the importance of these proteins for the splicing process, where the U5-15K is essential in yeast and therefore must have an important function and hence, is very well conserved. In contrast, the U5-52K has been shown to be dispensable in yeast and thus the preservation of the gene has not been of major concern for the organism. Particularly the poly-proline binding site of the GYF-domain is not fully conserved, which together with the fact that U5-52K is not essential for viability in yeast indicate that the interactions between the U5-15K and the GYF-domain are not essential.

U5-52K

```
H.sapiens    1   ................................MPKRK.VTFQGVGDEEDEDEIIVPKKK...........
S.pompe      1   MKRTLRNPGNSSTVGDVHDVFFEYDVSNVGRKRFRTKEGYYESESEEDEDQILNKEKKEGQSEDMFSDT
S.cerevisea  1   .MKYTQYPNSSKLK............RNSDRREHDEKLSDELHNQSTIYEDEELSRAEYDSDSDS.....


H.sapiens   27   ...............LVDPVAGSGGPGSRPKGKHSLDSDEEEDDDDGGSSKYDILASEDVEGQEAATLP.
S.pompe     71   SEDEKRTLPNDEAQKRRDFIENGDAERLAHKGLRNKEVLNDDSDDEDDNGKYSKLRYEDIEGQEDTNQAN
S.cerevisea 53   ............SVBDSTDNENSGKEMDEKSYEKNEDHVEDHRKRKKSKIQLLDIAEFKKEN1ADLDY


H.sapiens   81   ........SEGGVRITP....FNLQEEMEEGHFDADGNYFLN..RDAQIRDSWLDN.....IDWVKIRER
S.pompe    141   DLDADEEGSEISVPSSPKRMSFNLKEDMEEGDFDENGNFIRKNYDPESQYDAWLNGSVSNKKSIAAAREA
S.cerevisea 109  QIGNSESKVEKGVNIEP....FNIDDEIKHGVFDKDGNYIKT..ENATENDQQDNE.........EWMND


H.sapiens  132   PPGQRQASDSEEEDSLGQTSMSAQALLEGLLELLLPRETVAG.ALRRLGARGGGKGRKGPGQPSSPQRLD
S.pompe    211   EQKRKEMENRRRNQELEEFSKLPFSTVPEALSFFIARMERDESILEFLQRQSGNKKSYKKKKNNTAEGIS
S.cerevisea 164  VINTEEVNRLEKEQSVKTQNSRHYMVHEALNLLKFFLVDENETVLESLGRLNKLRKIAISKKNKSLKYVI


H.sapiens  201   RLSGLADQMVARGNLGVYQETRERLAMRLKGLGCQTLGPHNPTPPFSLDMFAEELAEEELETPTPTQRGE
S.pompe    281   PERKSADAFRKK....LIELITAGLTFLEDKIGKEDIYSETR...ESLQRIYQKLTSNSWSSFVSYDDSN
S.cerevisea 234  HG...............IELLSDLINILEKKGFSEVYEYNR......LKVQDAIEEEIFDDSSRIVNHK


H.sapiens  271   AESRGDGLVDVMWEYKWENTGDAELYGPFTSAQMQTWVSE YF PDGVYCRKLDPP..GGQFYNSKRIDFD
S.pompe    344   SSQ.........YNFKWEFD..DKTYGPYPASQIQAWSNE-YF TDAKHAAFIQLANMDEWMYPNNICFCD
S.cerevisea 282  TKL.........WGFKWLNK.LDEYHGLYANYEMSYWQKS YF KNSVIVKPHSEFDRDENWIRVSCLS..


H.sapiens  339   LYT...
S.pompe    403   VVSLKK
S.cerevisea      ..L...
```

U5-15K

```
H.sapiens    1   .MSYMLPHLHNGWQVDQAILSEEDRVVVIRFGHDWDETCMKMDEVLYSIAEKVKNFAVIYLVDITEVPDF
S.pompe      1   .MSYFLPHLHSGWHVDQAILSEQDRLVVIRFGRDHDEECIKQDEVLYRIAEKVVNMAVIYLVDIDEVPDF
S.cerevisea  1   MASVLLEQLRTGWHVDQAIVTETRRLVVIRFGRKNDRQCMIMDELLSSIAERVRNFAVIYLDIDEVSDF


H.sapiens   70   NKMYELVDPCTVMFFFRNKHIMIDLGTGNNNKINWAMEDKQEMVDIIETVYRGARKGRGLVVSPKDYSTK
S.pompe     70   NKMYELVDRTTIMFFYRNKHMMIDLGTGNNNKINWPLEDKQEMIDIIETIFRGARKGKGLVISPKDYSTR
S.cerevisea 71   DEMYELTDPMTVMFFYHNKHMMCDFGTGNNNKLNFIVDDKQEMIDILETIFRGARKNKGLVVSPYDYNHK


H.sapiens  140   YRY
S.pompe    140   HRY
S.cerevisea 141  RVS
```

Figure 7-3 Alignment of U5-52K and U5-15K from *H. sapiens* (GenBank accession no. <u>NP_006101.1</u> and <u>P83876</u>), *S. pombe* (GenBank accession no. <u>NP_595641.1</u> and <u>NP_587992.1</u>) and *S. cerevisiae* (GenBank accession no. <u>NP_012026.1</u> and <u>NP_015407.1</u>). The identical residues are marked in white with black background and the homologous residues are marked in bold black with white background. The GYF-residues are coloured in blue and the interacting residues are highlighted in green.

Other crystallographic studies on moonlighting proteins show how the proteins have evolved different surface areas for different functions and can switch between two diverse activities (Jeffery 2004). This is also found in the GYF-domain (Figure 7-4), where the interacting surface with the U5-15K is different from the interacting surface with a proline rich sequence mimicking the CD2.

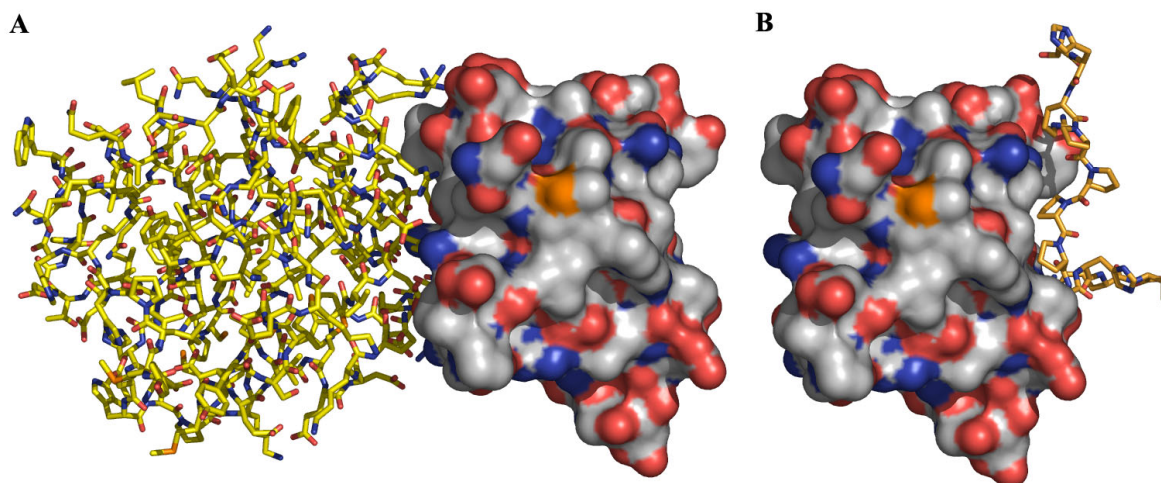**Figure 7-4 Comparison of the binding interface in the U5-15K-GYF-domain complex and the GYF-domain-CD2-peptide complex. The orientation of the two representations is the same. (A) The U5-15K is shown in yellow sticks and the GYF-domain as magenta surface representation. (B) The GYF-domain is shown as surface representation and the CD2-peptide in orange sticks (PDB code: 1L2Z).**

8. Conclusions and future perspectives

The crystal structure of the complex between the thioredoxin-like U5-15K and the GYF-domain of U5-52K has been determined. The structure reveals an unexpected interface formed by the extended C-terminus of the thioredoxin-like U5-15K and C-terminal residues of the GYF-domain of U5-52K. This interaction is a novel interaction motif for both, the GYF-domain of the U5-52K protein and a thioredoxin-like protein (U5-15K). The U5-52K is an example of a moonlighting protein, which has a function during the immune response as it interacts with an immune associated protein and furthermore plays a role in the very complicated spliceosomal machinery. The main function of the U5-52K is most likely its involvement with the splicing process, which takes place in all eukaryotes. It is however still unclear what the exact function regarding the association with the U5 snRNP is, but it could be involved in the biogenesis of the U5 snRNP or in its regeneration after spliceosome disassembly. Due to the nature of the T-cells, their receptors and binding proteins, the function of U5-52K as a CD2 binding protein is only found in mammals.

It would be interesting to determine the complex structure of other interaction partners of the two proteins, in order to investigate the binding sites in those cases, in particular the interactions to the other U5 snRNP specific protein, U5-102K. Within the U5 snRNP there are many interesting questions to address in respect to the arrangement of the snRNP and hence many interesting protein complexes and their interactions could be investigated. Also the interaction to the U4/U6 di-snRNP with U5 snRNP specific proteins is very interesting and important for the understanding of the assembly of the tri-snRNP, and thus a need for a thorough complex structure determination is present and in that way build up the snRNPs piece by piece.

The structure opens up for many interesting questions with respect to bifunctionality of proteins, which however stay unanswered until more knowledge is available about the biological processes taking place in the cell.

Crystal structure of a bacterial amidase with HDAC properties

9. Introduction

9.1. Nucleosome and post-translational modifications of histones

9.1.1. Nucleosomes

DNA is arranged with four histone proteins (H2A, H2B, H3 and H4) in a multi-protein-DNA-complex known as the nucleosome. An average of 146 base pairs (bp) of the DNA is wrapped around two copies of each protein forming an octamer. Approximately 54 bp of DNA link individual beads, i.e. for every 200 bp there is normally a nucleosome and hence the comparison to "beads on a string" (Figure 9-1 A). The linker protein histone 1 (H1) stabilises the folding of the nucleosome. The nucleosomes are then further condensed (depending of the state of the DNA) into the chromosome-structure.

The crystal structure of the nucleosome multi-protein-DNA-complex was determined in 1997 at a resolution of 2.8 Å (Luger *et al.* 1997) and it shows an arrangement of a H3-H4 tetramer and two H2A-H2B dimers with the DNA flanking the protein complex as two super helices (Figure 9-1 B+C). Besides packing the DNA, the conformational states of the nucleosomes determine the accessibility of the DNA for the different processes where DNA plays a part (transcription, replication, recombination and repair).

9.1.2. Post-translational modifications of histones

Histone proteins are small proteins. They have a flexible and basic charged N-terminus. Histones are exposed to many different post-translational modifications at the N-terminal part like phosphorylations, acetylations, methylations and ubiquitinations (Strahl and Allis 2000, Zhang and Reinberg 2001, Berger 2002). Parts of the modifications are believed to take place in order to neutralise the basic N-terminal part of histones and in that way abolish interactions with DNA, which interacts with basic proteins due to its own charge. This in turn makes the DNA more accessible to other proteins and the DNA involving processes can proceed.
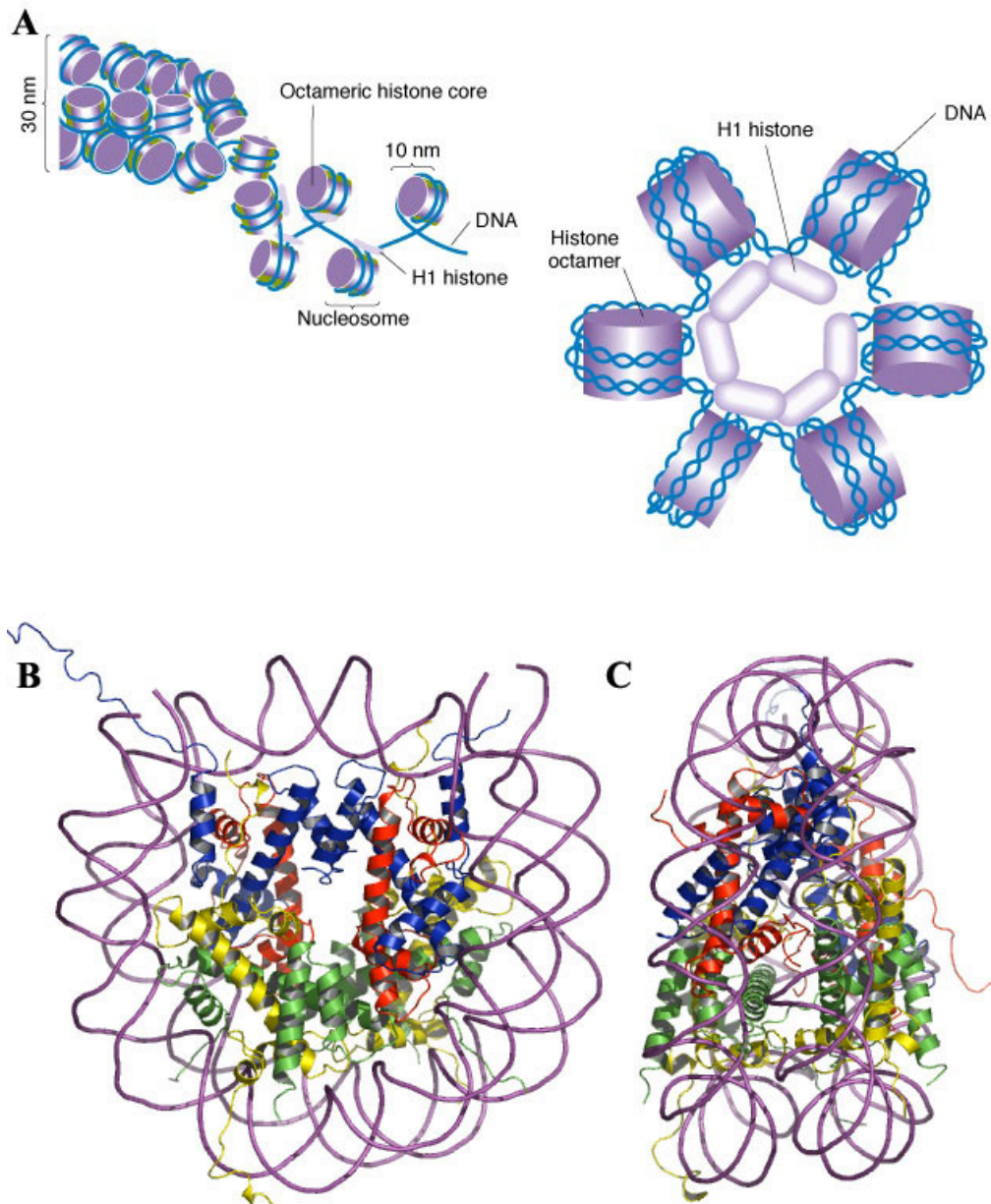
Figure 9-1 Components of the nucleosome. (A) Beads on a string. The chromatin fibre consisting of packed nucleosomes and the nucleosome in more details with the histone octamer surrounded by DNA. The figure was adapted from http://www.bio.miami.edu/dana/104/nucleosome.jpg. (B) The crystal structure determined of the histone octamer in complex with DNA. The colour codex for the four histones: H2A: yellow, H2B: green, H3: blue, H4: red and DNA: purple (PDB code: 1AOI). (C) The nucleosome core particle rotated 90°.

A hypothesis referred to as the "histone code" (Jenuwein and Allis 2001) says that the level of post-translational modifications of the histones can extend the potential information of the DNA. One of the important processes of post-translational modifications is acetylation of proteins. *In vivo*, histone acetylation is a dynamic and reversible process controlled by the opposing activities of histone acetyltransferases (HAT) and histone deacetylases (HDAC), which catalyse the addition, or removal of acetyl moieties from the ε-amino groups of lysines near the amino termini of histones. In general, increased levels of histone acetylation are associated with increased transcriptional activity whereas decreased levels of acetylation are associated with repression of gene expression (Grunstein 1997, Wade *et al.* 1997, Peterson 2002).

9.2. Class 1-4 histone deacetylases

9.2.1. **Histone deacetylases**

HDACs are members of an ancient enzyme family found in all three kingdoms (Khochbin and Wolffe 1997, Leipe and Landsman 1997, Gregoretti *et al.* 2004) and the first histone deacetylase was identified in 1996 (Taunton *et al.* 1996). HDACs are key components of chromatin remodelling complexes and major players in the regulation of gene expression in the eukaryotic cell. However, HDACs do not interact directly with DNA but function via protein-protein interactions to proteins. In cancer cells abnormal gene expression is frequently observed. It is known that HDACs associate with a number of cellular oncogenes and tumour-suppressor genes, leading to an aberrant recruitment of HDAC activity thus resulting in changes in gene expression, impaired differentiation and excessive proliferation (Grignani *et al.* 1998, Lin *et al.* 1998, Minucci *et al.* 2001). Not surprisingly, HDAC inhibitors have been shown to be efficient as anti-proliferative agents in tissue culture and, in addition, to inhibit tumour progression in model organisms. Nowadays, HDAC inhibitors are among the most promising anti-tumour therapeutics with several drug candidates currently in phase I and II clinical trials (Johnstone 2002, Rosato and Grant 2003, Villar-Garea and Esteller 2004).

The role of HDACs, however, is not only restricted to chromatin remodelling. Some HDACs have been shown to deacetylate non-histone proteins as well (Kouzarides 2000). This has in

particular been demonstrated for HDAC6, – a tubulin deacetylase (Hubbert *et al.* 2002, Matsuyama *et al.* 2002, Zhang *et al.* 2003) –, and may very well be true also for other eukaryotic enzymes, and for bacterial histone deacetylase-like amidohydrolases (HDAH; (Hildmann *et al.* 2004)).

Recent sequence analysis has divided HDACs into four groups, i.e. class 1 (HDAC1, 2, 3 and 8), class 2 (2a: HDAC4, 5, 7 and 9; 2b: HDAC6 and 10), class 3 (the sirtuins: a group of unrelated NAD (Nicotinamide Adenine Dinucleotide) – dependent deacetylase enzymes) and class 4 (with one member from human, HDAC11); the division is based on the homology with *S. cerevisiae* histone deacetylase RPD3, HDA1, SIR2 and human HDAC11 respectively (Gregoretti *et al.* 2004).

9.2.2. Class 1-2 and class 4 HDACs

Class 1 enzymes (HDAC1, 2, 3 and 8) are 350-500 amino acids in length and are in general ubiquitously expressed and primarily located in the nucleus (de Ruijter *et al.* 2003). Many class 1 enzymes are well-known transcriptional co-repressors and HDACs participate in many protein-protein complexes in the cell. For the class 1 enzymes HDAC1 and 2 complexes with the co-repressors Sin3, NuRD or N-CoR (Knoepfler and Eisenman 1999, Ng and Bird 2000, Jepsen and Rosenfeld 2002) have been characterised.

Class 2 enzymes (HDAC5, 6, 7, 9 and 10) are usually about 1000 amino acids long. Their catalytic domain contains several conserved sequence motifs that are significantly different from those of class 1 enzymes (Leipe and Landsman 1997, Grozinger and Schreiber 2002). Class 2 enzymes display a more tissue-specific expression in mammalian cells (Bertos *et al.* 2001, Fischle *et al.* 2001). Furthermore these enzymes are able to shuttle in and out of the nucleus in response to cellular signals and thus are at least in part cytoplasmic and thus acting on non-histone protein substrates. Class 2 enzymes interact with DNA binding transcription factors like: MEF2, BCL6, PLZF and TR2; and with transcriptional co-repressors: N-CoR, SMRT, BcoR and CtBP (Bertos *et al.* 2001, Fischle *et al.* 2001, Yang and Seto 2003).

In contrast to class 1 and 2 enzymes, which have been found in all fully sequenced eukaryotic organisms, class 4 proteins are so far not found in fungi. At least the most

important member of this subgroup, HDAC11, also shows tissue-specific expression (Gao *et al.* 2002).

9.2.3. Structural information of the HDACs

Structural information on HDAC enzymes, has so far only been obtained from the crystal structure determination of HDLP (histone deacetylase like protein) from the thermophilic bacterium *Aquifex aeolicus* (Finnin *et al.* 1999) with 35% identity to human HDAC1 and the crystal structure of the human HDAC8 (Somoza *et al.* 2004, Vannini *et al.* 2004), both proteins belong to the family of class I HDACs. The structures of both proteins were determined with and without different inhibitors. The overall fold is very similar with an eight stranded β-sheet surrounded by 16 α-helices and with a catalytically crucial zinc-ion in the active site (Figure 9-2 A+B).
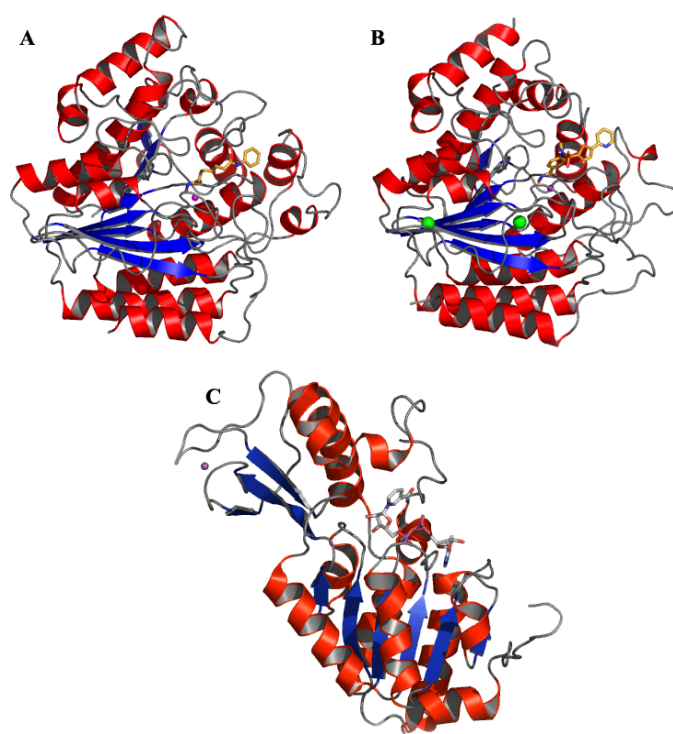
Figure 9-2 The determined structures of proteins belonging to the histone deacetylase family. The inhibitors or co-factors are shown in sticks, the bound zinc-ion in purple spheres and the two potassium ions in green spheres. (A) HDLP in complex with the inhibitor, SAHA (PDB code: 1C3S). (B) Human HDAC8 in complex with a hydroxamic acid inhibitor (PDB code: 1W22). (C) Structure of a SIR2 homologue in complex with NAD (PDB code: 1ICI).

9.2.4. Proposed catalytic mechanism

The present model of the catalytic mechanism for the deacetylation reaction contains features similar to of the one found in metallo- and serine proteases.
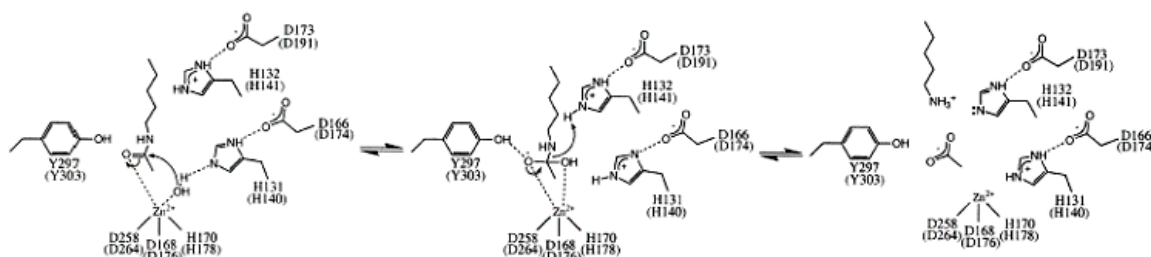

Figure 9-3 The catalytic mechanism proposed by Finnin *et al*. (1999). Amino acid numbering is according to HDLP and in brackets are the corresponding amino acids in HDAC1. See text for details.

The mechanism was proposed on the basis of the HDLP crystal structure (see Figure 9-3). The interaction of the zinc-ion in the catalytic centre with the carbonyl oxygen of the acetyl group is believed to polarize the carbonyl bond making the carbon more disposed to nucleophillic attack. Simultaneously, a water molecule is activated by the aforementioned zinc cation and a histidine-aspartate transfer relay system. This water molecule is then nucleophillic attacking the carbonyl carbon atom. The resulting tetrahedral carbon is stabilised by interactions with the zinc-ion and the active site tyrosine (Tyr297 in HDLP). Subsequently, the bond between the carbon and nitrogen breaks and the nitrogen accepts a proton from a second "outer" charge transfer relay system (His132 in contact with Asp173). Finally, the acetate moiety is thought to leave the active site through an internal channel, which opens next to the active site tyrosine. In all class 1 enzymes the two proposed charge transfer relay systems, where an aspartate interacts with a histidine, are fully conserved. However in class 2 enzymes the proposed outer charge transfer relay system is not found, since an asparagine replaces the aspartate (Figure 9-5).

9.2.5.   Class 3 Sirtuins

The class 3 HDACs are known as the sirtuins, due to the similarity to yeast SIR2 (silent information regulator-2) and are NAD-dependent enzymes (see (Marmorstein 2004) for a recent review). They deacetylate histones, but are also found in many species to mediate other processes and acting on a variety of other substrates than histones, hence they intercede a variety of biological functions (Buck *et al.* 2004). Structurally the sirtuins are completely different from the other classes of HDACs. Crystal structures of sir2 or sir2 homologues in complex with NAD have been determined and an example is shown in Figure 9-2 C, where it is obvious that the overall fold between class 1 (Figure 9-2 A+B) and class 3 enzymes is dissimilar. The sirtuins are divided into two domains, where the NAD binding domain has a typical Rossmann-fold (an opened twisted sheet with helices on both sides). The other smaller domain is placed on the NAD molecule and binds a zinc-ion, which contrary to class 1-2 enzymes, is not required for catalytic function.

9.3. HDAH from Bordetella/Alcaligenes strain FB188

9.3.1.   Characterisation of HDAH

Recently a novel enzyme (histone deacetylase-like amidohydrolase, HDAH) from *Bordetella/Alcaligenes* strain FB188 was identified and characterised (Hildmann *et al.* 2004). HDAH is a 369 amino acids protein with a molecular weight around 39 kDa. The closest sequence homology was found among the subfamily of acetylpolyamine amidohydrolases (APAH), however HDAH showed only a modest activity in deacetylating acetylpolyamines. Instead HDAH showed a significant activity in standard HDAC assays and was inhibited by known HDAC inhibitors like Trichostatin A (TSA). Furthermore HDAH showed zinc dependent activity.

9.3.2.   Homology with human HDACs

FB188 HDAH shares 35 % identity over 291 residues with the class 2 human enzyme, HDAC6, whereas it only shares 21% identity with the class 1 human HDAC8 (Figure 9-4). HDAH and HDAC6 also share a number of functional similarities. Both enzymes are zinc-dependent hydrolases that are able to efficiently deacetylate chicken histones and show the

same substrate specificity towards small peptidic substrates *in vitro*, e.g. they both process ε-N-trifluoroacetyl-lysine-containing substrates (Heltweg *et al.* 2004, Hildmann *et al.* 2004). Acetylated tubulin as well as hydroxamates such as SAHA act as competitive inhibitors. More selective class 2 inhibitors such as 3-[4-(3-(3-Chlorophenyl)-3-oxo-1-propen-1-yl)-1-methyl-1*H*-pyrrol-2-yl]-N-hydroxy-2-propen-amide (denoted 3c in the publication from Mai *et al.* (2003)) also block HDAH activity. HC-toxin and chlamydocin, which do not inhibit HDAC6 (Furumai *et al.* 2001, Koeller *et al.* 2003) also do not inhibit HDAH (Table 9-1). As HDAC6 is the most similar human homologue on both a sequence and functional basis, these functional similarities become interesting in relation to identifying a possible function of HDAH in *Bordetella/Alcaligenes* strain FB188, which is still unknown.
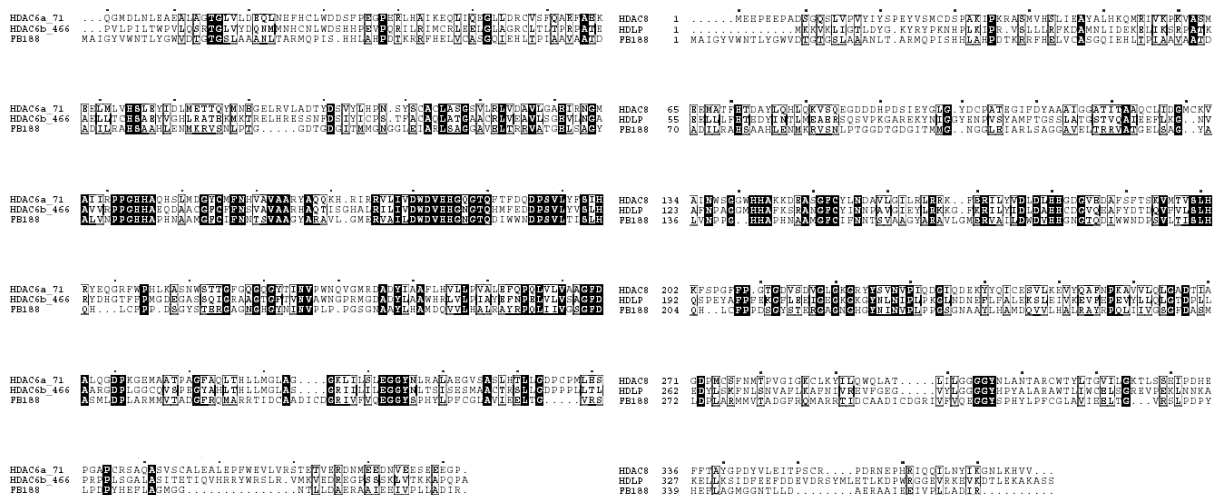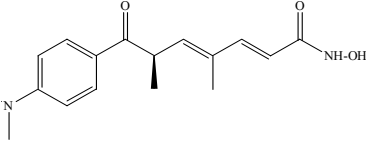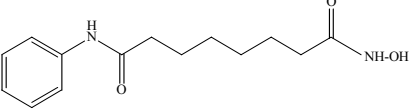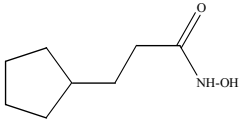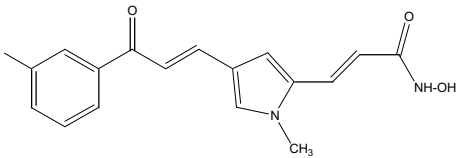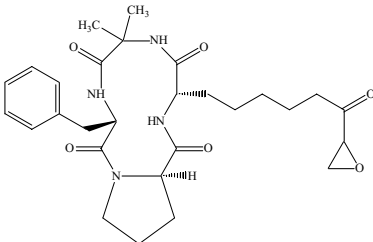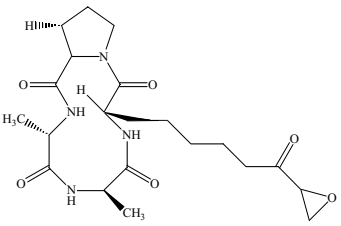

Figure 9-4 Alignments of HDAH with HDACs. HDAH is aligned with the two domains of human HDAC6 to the left (HDAC6 is the HDAC that shows the highest sequence identity to HDAH). An alignment of HDAH with human HDAC8 and *A. aeolicus* HDLP is seen to the right (HDAC8 and HDLP are the only two enzymes where the structure is known). The identical residues are marked in white with black background and the homologous residues are marked in bold black with white background. The alignment is made using CLUSTALW (Thompson *et al.* 1994) and presented with ESPript (Gouet *et al.* 1999).

HDAC6 is unique as it contains two HDAC domains (see Figure 9-4 for an alignment) and also a C-terminal zinc finger domain (Seigneurin-Berny *et al.* 2001). HDAC6 is capable of shuttling in and out of the cell nucleus, but is mostly found in the cytoplasm. HDAC6 is part of a multi-protein complex together with proteins taking part in control of protein

ubiquitination. Therefore it is suggested that HDAC6 could participate in deacetylation of proteins before their ubiquitination (Seigneurin-Berny *et al.* 2001).

Table 9-1 The effect of different inhibitors on HDAH. The class specificity of the given inhibitor is stated. The IC50 value determines the amount of inhibitor needed to inhibit 50% of the enzyme. The table is adapted from Nielsen *et al.*, (in prep.).

Inhibitor	Selectivity	IC50 [μM]
TSA	None	1.20 ± 0.12
SAHA	None	0.95 ± 0.06
CypX[§]	Class 2	0.290 ± 0.060
APHA 3c*	Class 2	2.46 ± 1.36
Chlamydocin	Class 1	> 300
HC Toxin	Class 1	No activity

*from Mai *et al.* (2003); [§] from Nielsen *et al.*, in prep.

Furthermore cytoplasmic HDAC6 co-localises with the microtubule network (Hubbert *et al.* 2002) and deacetylates α-tubulin at the Lys40 (which has a stabilising function) both *in vivo* and *in vitro*. The amino acids taking part in the catalysis are highly conserved throughout class 1 and 2 enzymes; the main exception is the already mentioned aspartate (section 9.2.4), which in class 2 enzymes is exchanged with an asparagine (Figure 9-5). Otherwise all central amino acids in the active site are fully conserved.

Figure 9-5 Alignment of HDACs from class 1 and 2. Class specific consensus residues are coloured in yellow for class 2 enzymes and in magenta for class 1 enzymes. Boxed in red are residues that constitute the zinc-ion-binding site. Boxed in blue is the inner "charge relay" system and in green is the proposed outer "charge relay" system. Black and grey boxes indicate residues within the distance of 3 Å and 3.8 Å from the bound inhibitor SAHA in the crystal structure of HDAH. Figure is adapted from Nielsen *et al.*, in prep.

9.3.3. Substrate specificity

The substrates for HDACs are not solely histones, the substrate specificity or discrimination between different substrates is therefore an interesting subject for further investigation. The preferred substrate has been shown to vary for different HDACs and HDAC bacterial homologues (Riester *et al.* 2004). In the case of HDAH the most specific substrate was found to be Boc-L-Lys(ε-methane sulfonyl)-MCA among a variety of Lys(ε-acyl) derivatives. The ideal length of the aliphatic substrate chain was examined and a clear preference was made for lysine compared to ornithine (an amino acid resembling lysine but one carbon atom shorter on the side chain) indicating the importance of the lysine side chain length. A 65 kDa protein has been identified as a possible substrate for HDAH (Hildmann *et al.* 2004), however the nature of this protein is still unknown.

9.4. Inhibitors as potential anti-cancer drugs

As mentioned in section 9.2.1 inhibitors of HDACs have turned out to be potential anticancer drugs. The immediate result of HDAC inhibition is a hyperacetylation of the target protein.

9.4.1.   Different inhibitors

Different classes of inhibitors are known. Inhibitors are divided into three different parts, the enzyme inactivating part, the spacer and the cap group. All three groups can be varied within the inhibitor in order to increase the specificity of the given inhibitor towards a certain target. A group of inhibitors contains a hydroxamate group. Well-known and characterised inhibitors as SAHA and TSA belong to this group; they have a hydroxamate group interacting in the active site, an aliphatic spacer and a benzene group (nM or µM of the inhibitor is necessary to inhibit 50% of the enzyme). Other groups are the short fatty acids and valproic acid, however they are not highly specific and the IC50 is in mM range. CHAP1 (cyclic hydroxamic acid-containing peptide) has the hydroxamic acid from TSA and the cap group (a cyclic tetrapeptide) from TPX B (Furumai *et al.* 2001) and was the first inhibitor identified with specificity for HDAC (CHAP1 has a preference for HDAC1 and 4 over HDAC6) and is specific with an IC50 in nM range. See Table 9-1 for examples of different classes of inhibitors, their specificity and the function on HDAH. For an overview of inhibitors and specificity among HDACs see the review by Johnstone (Johnstone 2002).

9.4.2.   Function of the inhibitors

The functions of the HDAC inhibitors are in general thought to be a blocking of the cell cycle and induction of apoptosis or differentiation depending on the cell-type and environmental factors (Kim *et al.* 2000, Finzer *et al.* 2001, Kim *et al.* 2001, McLaughlin *et al.* 2003). Furthermore the inhibition of histone deacetylases seems to regulate expression of genes, which are important for cell growth control and survival; finally some inhibitors can induce apoptosis. Several of the inhibitors are in phase I or II clinical trials, and new potential inhibitors are expected to become important targets in the battle against cancer. On the way to the understanding and development of those drugs, structural information of the enzymes and enzyme-inhibitor complexes are essential parts.

10. Materials and Methods

10.1. Overexpression and purification

10.1.1. Overexpression of the HDAH gene

The recombinant *Bordetella/Alcaligenes* strain FB188 HDAH was expressed and purified as described (Hildmann *et al.* 2004) with few exceptions. The gene was cloned into the pQEB-AH-NHis vector with an N–terminal His_6-tag and expressed in M15[pREP4] cells (Qiagen, Hilden, Germany). A single colony was added to 500 ml LB medium containing 100 mg/L ampicillin and 50 mg/L kanamycin and grown overnight at 37 °C. The overnight culture was 20-fold diluted in 10 L LB medium with antibiotics and the cells grew at 37 °C in a bioreactor (Applikon Biotechnology) until an OD_{600} of 0.6-0.8. The cells were induced with 0.5 mM IPTG and grew for 4 h before harvesting of the cells.

10.1.2. Purification

The cells were opened with a microfluidiser S100 (Microfluidics^{TM}) in lysis buffer (100 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 500 mM NaCl and 1 EDTA-free Protease Inhibitor Tablet (Roche Diagnostics)). The lysate was centrifuged at 100.000 xg for 30 minutes in an Avanti centrifuge (Beckman Coulter) and the supernatant and the pellet were separated. The supernatant was loaded on a 5 ml His-Trap FF affinity column (Amersham Biosciences) loaded with zinc-ions by equilibrating with a 200mM $ZnCl_2$ and eluted with a linear gradient of imidazole (0 – 0.5 M). Fractions chosen from the chromatogram were analysed on a 12.5% SDS-PAGE loaded with the BR marker (New England Biolabs, inc.). The selected fractions containing the protein were finally purified on a gel filtration column (Superdex 75 26/60 Amersham Biosciences) in order to get rid of the last impurities and to transfer the protein in a suitable crystallisation buffer (20 mM Tris-HCl pH 8.0 and 100 mM NaCl). The entire purification was performed at 4 °C. The fractions corresponding to the chromatogram peak were also analysed with SDS-PAGE. The pure protein was concentrated (Vivaspin) to 10-12 mg/ml and centrifuged at 23.000 xg for 5 minutes before crystallisation.

10.1.3. Preparation of selenomethionine substituted protein

Selenomethionine containing HDAH was prepared after the method described by Reuter and Ficner (Reuter and Ficner 1999). The plasmid was transferred into the *E. coli* strain B834DE3 (methionine deficient strain), which was added to minimal medium (containing salt solution (Na_2HPO_4, KH_2PO_4, NaCl, NH_4Cl), trace element solution (EDTA, $FeCl_3$ x $6H_2O$, $ZnCl_2$, $CuCl_2$ x $2H_2O$, $CoCl_2$ x $6H_2O$, H_3BO_3, $MnCl_2$ x $6H_2O$) 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 0.4% glucose, 1 mg/L biotin, 1 mg/L thiamin, 100 mg/L methionine and 100 mg/L ampicillin) for overnight growth. The overnight culture was 50-fold diluted in minimal medium and grown to an OD_{600} of 0.8 at 37 °C. Cells were centrifuged and the cell pellet resuspended in 1 L minimal medium deficient of methionine and left to grow for 1 h before the addition of 50 mg of selenomethionine/L culture. After 30 minutes the culture was induced with 0.5 mM IPTG and grown overnight. Purification of Se-Met HDAH was essentially similar to the purification of the native protein, with the important exception that after the His-Trap column, 5mM DTT was added to the protein and to all buffers used in order to prevent oxidation of the selenium. Se-Met HDAH was concentrated (Vivaspin) to 7 mg/ml and centrifuged at 23.000 xg for 5 minutes before setting up the crystallisation drops.

10.2. Crystallisation

10.2.1. Crystallisation conditions

Initial crystallisation screening with HDAH was performed with a variety of screens; Hampton crystal Screens 1-2, Cryo, Lite and Additive 1-3 (Hampton Research, Riverside, USA), Jena Bioscience Screen 1-10 (Jena Bioscience, Jena, Germany), Magic screen (1-92 conditions) and Structure screen (1-72 conditions). Crystals were visible at 20 °C after 3-4 days using the sitting drop vapour diffusion method in 24 wells Chryschem plates from Hampton sealed with clear tape by mixing equal volumes of HDAH (1 µl) with reservoir solution (total volume in screening: 200 µl; total volume after ideal condition found: 500 µl).

10.2.2. Different crystal forms

From the screening a few conditions were selected and refined for the final crystallisation conditions. The conditions were chosen by selecting the conditions yielding the most three-

dimensional single crystals. See Table 10-1 for the conditions used to obtain the different crystal forms.

Table 10-1 Final crystallisation conditions for the HDAH crystal forms and the corresponding determined structures.

$P2_12_12_1$	I222 (Native)	$P2_1$ (Se-Met)	$P2_1$ (Inhibitor)
1.3 M NaK-phosphate pH 5.5, 50 mM Na-Acetate, 10% 1,6-Hexanediol	1.3 M NaK-phosphate pH 5.5, 50 mM Na-Acetate, 5% 1,4-butanediol	14% PEG 8000, 100 mM CHES pH 9.5, 200 mM NaCl, 5 mM DTT	0.8 M NaCl, 200 mM Na-Cacodylate pH 6.5

10.2.3. Co-crystallisation with inhibitors

Two different inhibitors (SAHA and CypX) were co-crystallised with the inhibitor both belonging to the group of hydroxamate inhibitors (see Table 9-1 for the chemical structure of the two inhibitors). The inhibitors were incubated with the native protein for 1 h at room temperature after adding a 10-fold molar excess of the inhibitor. Afterwards the protein was crystallised as described above.

10.2.4. Cryo-cooling

Two different approaches were used for cryo-cooling: oil protection (removing the buffer from the crystal while covering the crystal with oil/other viscous solutes) and cryoprotectant addition to the crystallisation buffer. For the first approach, different viscous solutions were tested as cryoprotectant and initial tests suggested that Paratone-N would be a good cryoprotectant. There were however problems with ice rings due to difficulties removing all the reservoir solution around the crystal and the loop. The second approach was more successful using an addition of 20-25% glycerol to the reservoir solution.

10.3. *Data collection and SAD phasing*

10.3.1. Data collection

Crystals were initial tested on either a Micromax 007 rotating copper anode generator operating at 40 kV and 20 mA or a RUH3R rotating anode operating at 50 kV and 100 mA (Rigaku/MSC, The Woodlands, USA), both equipped with Mar345 dtb detectors system (MarResearch) and an X-stream system (Rigaku/MSC) for cryo-cooling. If a crystal had been soaked in a heavy atom solution and the resulting diffraction pattern showed high quality, a dataset was collected already on the rotating anode, otherwise most of the tested crystals were stored in cryo-dewars for subsequently data collection using synchrotron radiation.

X-ray data were collected from cryo-cooled crystals at beamlines BL1 at BESSY, Berlin, Germany (Initial $P2_12_12_1$ and I222 dataset and the MAD datasets), ID14-1 at ESRF, Grenoble, France (native high resolution in I222 form), BW7B at DESY, Hamburg, Germany (initial inhibitor datasets) and BM14 at ESRF, Grenoble, France (finally used inhibitor datasets). The data were processed with HKL2000 (HKL Research, Charlottesville, USA). Systematic absences along the three axes indicated the presence of screw axes, and the primitive orthorhombic space group was $P2_12_12_1$. In the monoclinic space group the systematic absences along the b axis indicated that the correct space group was $P2_1$. In the crystals of I222 symmetry, screw axes could also be present (i.e. space group $I2_12_12_1$) due to space group ambiguity, since the systematic absence conditions do not enable distinguishing between the two space groups. This meant that both space groups had to be tested[3]. In the $P2_12_12_1$ form the Matthews (Matthews 1968) coefficients were calculated to be 3.3-2.0 $Å^3$/Dalton , which equals 12 - 20 molecules in the asymmetric unit and yielding a solvent content of 62%-36.7%. For the crystals in the I222 and $P2_1$ forms values of 2.4 $Å^3$/Dalton and 2.5 $Å^3$/Dalton suggested 4 molecules in the asymmetric unit yielding a solvent content of 48.7% and 50.8% respectively.

---

[3] However, statistics speak for I222, since there are 2.4% structures determined in this space group and only 0.7% in $I2_12_12_1$ in the protein data bank (see http://www.chem.tamu.edu/xray/allspg.html).

10.3.2. Solving the phase problem

See section 5.3.2 for a more general description of the phase problem and how to solve it. MOLREP (Vagin and Teplyakov 1997), AMoRe (Navaza 1994), EPMR (Kissinger *et al.* 1999), COMO (Jogl *et al.* 2001) and PHASER (Storoni *et al.* 2004) were tested in attempts of determining the structure with the method of molecular replacement. The structure of HDLP from *A. aeolicus* (PDB code: 1C3P), HDAC8 (PDB codes: 1W22 and 1T69), different homologue models based on the sequence of HDAH and truncated versions were used as search models.

Attempts were also made to determine the structure with the method of MIR. For that purpose different heavy atom soaks were performed, the tested compounds were: ethyl mercury thiosalicylate (EMTS), $SmBr_3$[4], K_2PtCl_4, $HgCl_2$, $AuCl_3$ and Xenon. The final concentration of the heavy atom salt varied from 0.5-2 mM and the soaking period from 1-12 hours. Xenon-derivatization was performed by incubating the crystal for 15 min in a pressure chamber at 15 bar (Xcell, Oxford Cryostreams, UK).

Finally also a structure solution strategy using MAD was initiated. Initially NaI (0.5 M) was tested and the crystal was quickly soaked in the solution and mounted. Afterwards selenomethionine-substituted protein was crystallised and data were collected at the synchrotron at three different wavelengths.

10.3.3. MAD (SAD) phasing

MAD data were collected at BL1, BESSY from crystals in space group $P2_1$. Due to changes in the local environment of the selenium atoms, which influences the wavelength position of the edges of anomalous scattering an experimental fluorescence scan is always performed at the synchrotron. Figure 10-1 shows the fluorescence scan.

---

[4] Samarium has anomalous signal at the wavelength of a rotating copper anode (λ = 1.5418 Å)
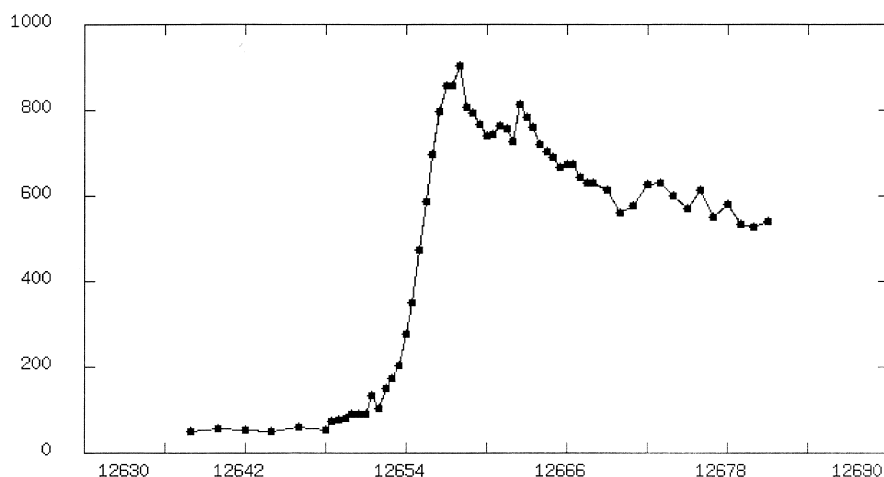
Figure 10-1 The experimental fluorescence scan of the crystal to determine the exact position of the peak and the inflection point.

From the fluorescence scan the wavelengths for the experiment were determined (see Table 10-2 for the values). The peak was slightly shifted compared to the theoretical value for the selenium K edge, which is 0.9797 Å. Two MAD experiments were performed; in the first experiment only the necessary amount of data (180°) was collected, since the effect of radiation damage on the crystal was unknown The data were collected with an oscillation of 0.5° and the detector-crystal distance was 200 mm giving a resolution on the edge of 2.54 Å. In the second experiment 250° were collected at each wavelength also with a 0.5° oscillation angle and a distance of 200 mm giving a resolution on the edge of 2.54 Å.

Table 10-2 Experimental determined values of the desired wavelengths for the MAD experiment from BESSY, Germany.

	Energy (eV)	λ (Å)	f'	f''
Peak	12657.46	0.9795	-7.9	4.7
Inflection point	12654.99	0.9797	-9.6	2.6
High energy remote	13600	0.9117		

Selenomethinonine positions were found with HKL2MAP (Pape and Schneider 2004). SAD (single wavelength anomalous dispersion) phasing in SOLVE (Terwilliger and Berendzen 1999), solvent flattening and the use of 4-fold NCS implemented in RESOLVE (Terwilliger 1999) resulted in an excellent experimental map.

10.4. Model building and refinement

10.4.1. Auto-build and building the model

The auto-build function in RESOLVE (Terwilliger 2003) was capable of placing main-and side chains corresponding to ~75% of the 4 molecules correct in the experimental map resulting from SOLVE. 80% of these residues were sequentially correct; the remaining amino acids were built as either alanines or glycines. The model was completed and corrected using the programs XTALVIEW (McRee 1999) and O (Jones *et al.* 1991).

10.4.2. Refinement of the structures

Since the overall structure of the protein:inhibitor complex was identical to the apo-structure, at this stage the model was refined against the inhibitor datasets, which were collected at a higher resolution (1.57 Å and 1.75 Å), with REFMAC5 (Murshudov *et al.* 1999) and 5% of the reflections were omitted in the refinement for the calculation of R_{free} (Brunger 1993). Water molecules were added automatically with the ARP/wARP (Perrakis *et al.* 1999) function in REFMAC5 and afterwards examined manually for reasonable hydrogen bonding possibilities. The structure of the native protein at high resolution (1.6 Å) was determined with molecular replacement using MOLREP (Vagin and Teplyakov 1997) and refined with REFMAC5. The restraints for CypX used in the refinement were created with PRODRG (Schuttelkopf and van Aalten 2004).

10.5. Structure analysis

10.5.1. Validation of the model and figures

The quality of the model was checked using PROCHECK (Laskowski *et al.* 1993). Superpositions enabling comparison of the different structures are performed in LSQMAN

(Kleywegt 1996) using the brute-force option. All figures are made in PYMOL (DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, USA) and also the electrostatic surface potential representations are made in PYMOL using APBS, which is a plug-in tool in PYMOL.

11. Results

11.1. Expression and purification

11.1.1. Overexpression of the gene

The protein was produced in LB medium in the *E. coli* strain M15[pREP4], which is especially suitable for the p-QE vector system. The production of HDAH protein was very efficient, since approximately 50 mg of purified protein was obtained from 1 L of culture.

11.1.2. Purification of the protein

Figure 11-1 shows chromatograms and SDS-PAGEs from the purification of FB188 HDAH. The protein approximately elutes with 175-200 mM imidazole from the affinity column. From the gel, fractions containing HDAH (11-12) were pooled (50 mg). Half of the amount was stored frozen using chock freezing in liquid nitrogen, and the rest of the protein was further purified on the gel filtration column. However, when thawing the protein later, it had suffered severely from the freezing and therefore precipitated. It was impossible to get the protein into solution again. A more gentle approach could be to chock freeze the protein with 15-30% glycerol, which was not tested. On the gel filtration chromatogram in Figure 11-1 the protein was eluted in a nice homogenous single peak corresponding to a monomer. Based on analysis of the purity from the SDS-PAGE, fractions corresponding to HDAH (42-50) were pooled and contained approximately 25 mg of protein, which were concentrated to 2 ml.

11.1.3. Preparation of the selenomethionine substituted protein

HDAH contains 12 methionines out of a total of 369 residues (N-terminal methionine excepted), which makes Se-Met phasing an obvious choice (30 residues per selenium). Selenomethionine-substituted protein was produced in the *E. coli* strain B834DE3, which is deficient of methionine and therefore selenomethionine, added to the medium, can be incorporated. The yield of the expression was poor (especially in comparison to the native expression), which is often seen in SeMet substituted protein production. From 3 L of culture

approximately 2.5 mg purified selenomethionine substituted protein was obtained. The protein was purified as described for the native protein with the important exception that 5 mM DTT was added to all buffers after the affinity column to prevent oxidation of the selenium atoms. The protein was only concentrated to 7 mg/ml before setting up the crystallisation trays, since SeMet containing protein is known to precipitate at lower concentrations of protein and precipitant.
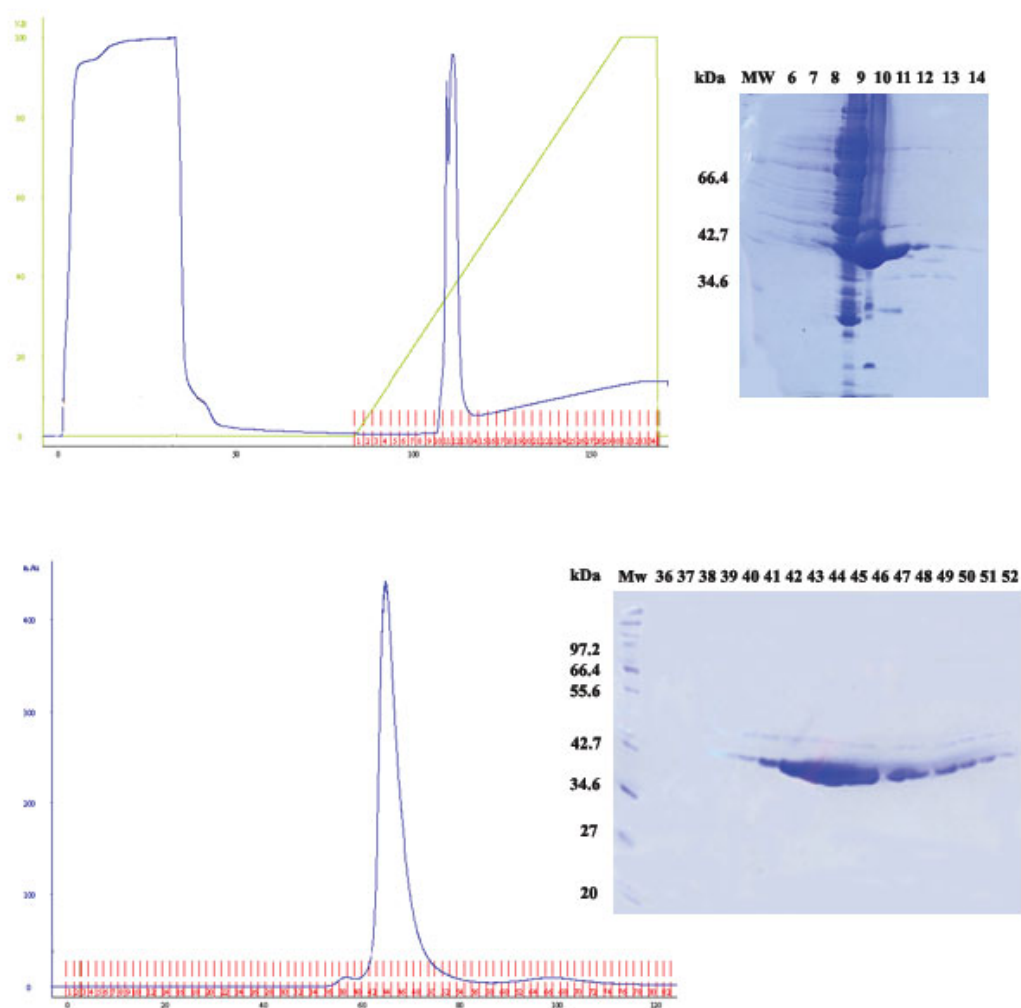

Figure 11-1 Chromatograms from the affinity and the gel filtration columns and the SDS-PAGEs of the interesting fractions. The numbering equals the fraction number seen on the corresponding chromatogram.

11.2. Crystallisation and structure determination

11.2.1. Crystallisation

HDAH turned out to crystallise very willingly (using following precipitants: NaCl, PEG 4000, PEG 8000, PEG 3350, NaK-phosphate, Imidazole, Mg-formate, Ammonium sulphate), however in a limited number of different space groups and with similar unit cell parameters. Initial crystals were obtained of the $P2_12_12_1$ and the I222 form, but due to problems obtaining phases (probably because of the crystal quality being insufficient) a second crystal form was found; $P2_1$ (see Figure 11-2 for examples of crystals in the different space groups and unit cell parameters). In this form it was possible to solve the phase problem and determine the structure. The majority of the crystals in the monoclinic space group were not single crystals, but instead they were clusters of single crystals. Before mounting the crystals they were attempted separated with an acupuncture needle, by pressing the tip of the needle at the point of common nucleation.
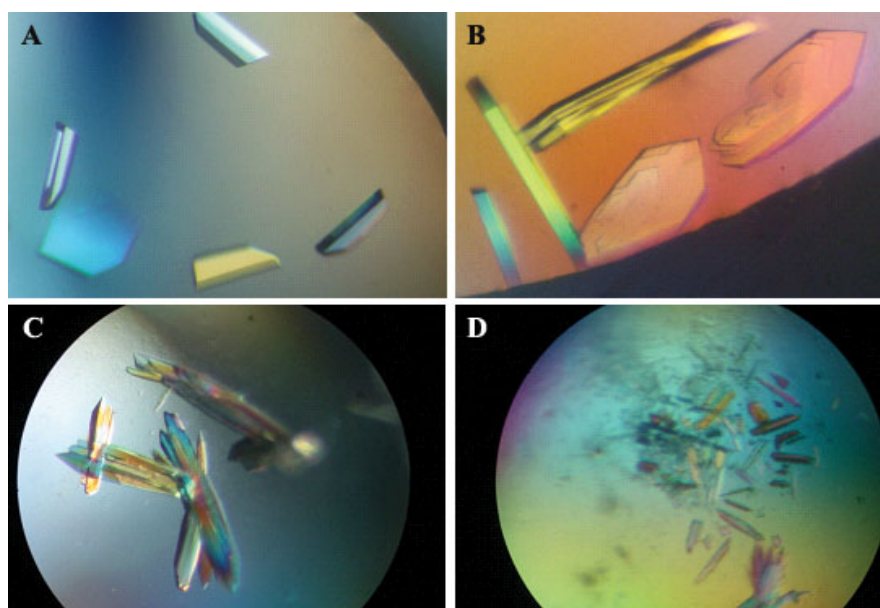
Figure 11-2 A part of the obtained HDAH crystals. (A) Crystals in the $P2_12_12_1$ form obtained with 1,6-hexanediol. (B) Crystals in the I222 form with 1,4-butanediol. (C) Crystals in the $P2_1$ form consisting of rods grown together. (D) After an attempt of separating the rods to get a "single" crystal. The small pieces were then separately tested for diffraction abilities.

This was a rather complicated manoeuvre and probably the reason why some of the tested crystals showed signs of not being single crystals, e.g. a high degree of mosaicity and/or signs of second crystal lattice on the diffraction image (Figure 11-3).
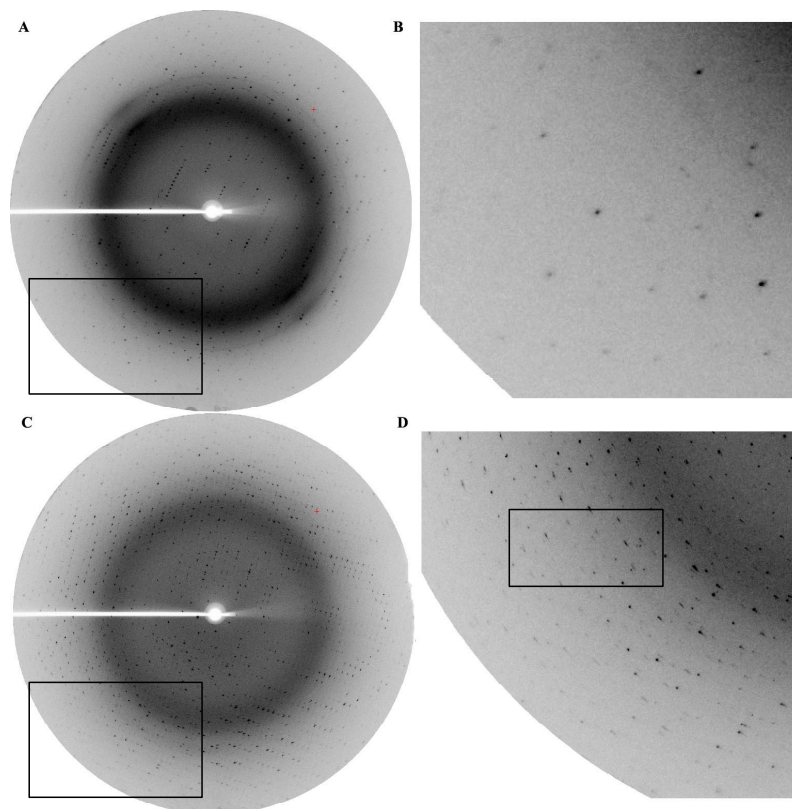
Figure 11-3 Diffraction images. (A) The diffraction pattern of the crystal where the first MAD datasets were collected. (B) Zoom of the lower left corner (rectangle shown in (A)) with the spots relatively well defined and separated. (C) The diffraction pattern of the crystal where the second MAD dataset was collected. Compared to (A) this pattern shows a very high degree of mosaicity. (D) A zoom of the lower left corner (rectangle shown in (C)). In this case the spots seemed more smeared and also signs of a second crystal lattice were found – see rectangle, where spots are very close to each other and out of periodicity.

11.2.2. Data collection

Crystals were tested and data were collected on a rotating anode and on different synchrotrons in Germany and France (see Table 11-1 for data statistics from the most important crystals).

Table 11-1 Data collection statistics.

Data set	Native (P2$_1$2$_1$2$_1$)	SeMet (peak)	Native (I222)	SAHA	CypX
Wavelength (Å)	0.91002	0.9795	0.934	0.97848	0.97848
Resolution range (Å)[1]	50-2.5 (2.59-2.5)	50-2.54 (2.63-2.54)	50-1.6 (1.66-1.6)	50-1.57 (1.63-1.57)	50-1.75 (1.81-1.75)
Space group	P2$_1$2$_1$2$_1$	P2$_1$	I222	P2$_1$	P2$_1$
Cell Dimensions (Å/°)	a = 94.0	a = 68.8	a = 93.6	a = 68.6	A = 68.1
	b = 251.8	b = 94.7	b = 127.9	b = 94.7	B = 94.5
	c = 258.0	c = 123.3	c = 251.7	c = 123.9	c = 122.6
		β = 104.2		β = 104.8	β = 104.3
Observed reflections	605957	256460	1095546	641332	600081
Unique reflections	175598	95688	198432	19884	146795
Completeness (%)[1]	82.9 (85.9)	96.7 (68.4)	99.8 (99.0)	80.9 (51.1)	94.8 (72.7)
R$_{sym}$[2] (%)[1]	12.3 (24.6)	5.4 (19.9)	8.9 (72.7)	4.3 (11.8)	5.4 (23.1)
Average (I/σ(I))[1]	13.2 (4.8)	16.9 (4.2)	14.5 (2.5)	17.4 (5.1)	15.8 (3)
Mosaicity (°)	0.515	0.615	0.734	0.275	0.527
SAD phasing					
Heavy atom sites	48 (12/molecule)				
Overall figure of merit					
Before RESOLVE	0.35				
After RESOLVE	0.72				

[1]**Numbers in parenthesis refer to the highest resolution shell. [2] R$_{sym}$ =100·$\Sigma_h\Sigma_i$|I$_i$(h)-<I(h)>|/Σ_hI(h) where I$_i$(h) is the ith measurement of the h reflection and <I(h)> is the average value of the reflection intensity.**

11.2.3. The different crystal forms

Initially the crystals of $P2_12_12_1$ and I222 seemed to appear according to the conditions given in Table 10-1, however it turned out that they grew randomly in the same drop. However, when applying the condition given in Table 10-1, the crystals tended to be in the stated space group with the given additive (1,6-hexanediol: $P2_12_12_1$ and 1,4-butanediol: I222).

Due to the large unit cell in the $P2_12_12_1$ form (see Figure 11-2) the asymmetric unit contained 12-20 molecules (most likely 16 molecules based on the post-mortem analysis of the crystal form, see section 12.1). This turned out to be a major obstacle in solving the phase problem, since molecular replacement and MAD (a search for approximately 190 selenium sites) phasing did not work. Another crystal form, I222, had already been obtained at that moment, which however turned out to be highly related to the $P2_12_12_1$ form, which again made the crystal production and data collection very difficult. A batch of selenomethionine-substituted protein was prepared with the aim of getting crystals of the I222 form. However, the crystals were very small and the resulting data were very difficult to index. At the synchrotron it was clear that usable data could not be extracted from these crystals.

Therefore the search for a new crystal form was initiated, which resulted in the discovery of the $P2_1$ monoclinic form, where the phase problem was attempted solved by all three classical methods as described below.

11.2.4. Molecular Replacement attempts

Several programs and search models were tested for a solution with molecular replacement, however no correct solution was obtained. The reason could be due to the differences between search model and the new structure mainly in the loop regions. Due to the relatively low sequence identity[5] it was difficult to predict exactly which parts of the structures that were similar, thus the creation of a good search model for molecular replacement was complicated. The presence of 4 molecules in the asymmetric unit also decreased the possibility of finding the correct solution. When looking at the structure the major

---

[5] There is approximately 21% identity of HDAH with both HDLP and HDAC8.

differences between the models and HDAH are in the areas between α1 and α3 (in HDAH) and α12 and α13 (in HDAH) (see Figure 11-11). A truncated version of HDLP was tested as a search model, since these areas also differentiate in the primary sequence alignment. This model did not produce any solution.

11.2.5. MIR phasing attempts

Attempts were also made to obtain the phases with MIR phasing and hence via isomorphous heavy atom derivatives. Many different heavy atom compounds were tested (see Materials and Methods in section 10.3.2), varying heavy atom type, concentration and soaking times. But it was not possible to identify the positions of the heavy atoms. The reason for this could be that no heavy atom had bound specifically to the protein in the crystals, however at least in the case of mercury compounds; this does not seem to be the case. HDAH contains 6 cysteines, where (as later found in the structure) Cys51, Cys207, Cys295 and Cys300 are all exposed to the surface. Cys153 is in the active site cavity and only Cys320 is integrated in the protein and possibly out of reach for a mercury atom. A more likely explanation is the lack of isomorphism between native crystals and derivative crystals. In space group $P2_1$ the crystals are separated with a needle and thus, show signs of non-singularity making the search for two isomorphous crystals even more difficult, since also the native crystals are not isomorphous.

11.2.6. SAD phasing

An initial approach using sodium iodide was attempted due to the rising numbers of structures determined with halide soaks (Dauter and Dauter 2000). However, no phases could be obtained. This was probably due to problems with the crystals resisting the soaking procedure.

Due to the presence of 12 methionines in HDAH (the N-terminal methionine not included) and 4 molecules in the asymmetric unit of the $P2_1$ crystal form, 48 selenium positions were searched for in the programs using the two datasets collected at BESSY, Berlin, Germany. MAD phasing was attempted with both HKL2MAP and SOLVE. None of the programs resulted in any interpretable electron density maps.

However when looking for the selenium sites with a SAD approach and only providing HKL2MAP with the peak dataset from the initial dataset collected in Berlin, a reasonable solution appeared (CC_{all} = 49.96 and CC_{weak} = 25.11). The phasing was done in the program SOLVE and in the final electron density map after RESOLVE (with the use of 4-fold NCS), it was possible to distinguish some secondary structure elements. Attempts were made to build the entire polypeptide chain in the electron density with the help of the known HDLP structure, this was however not possible (see Figure 11-4 A+B for the initial electron density). Different selenium sites were found with the SAD option in HKL2MAP from the second dataset collected in Berlin (CC_{all} = 53.21 and CC_{weak} = 26.84). When feeding the positions into SOLVE using SAD phasing (see Table 11-1 for statistics), the resulting electron density map from RESOLVE showed that this was indeed the correct solution (Figure 11-4 C).

MAD phasing using data collected at three wavelengths was not successful. The reason for the phasing problems could be due to radiation damage of the crystal. An overall conclusion from these experiments is that the multiplicity of the data turned out to be extremely important for the phasing. In the first dataset only 180° was collected – corresponding to a multiplicity of 1.9 (Bijvoet pairs kept separate) - where 250° were collected in the second dataset, which gives a multiplicity of 2.7 (Bivoet pairs kept separate). The reason for this is an improvement of the statistics for measuring the small anomalous differences.
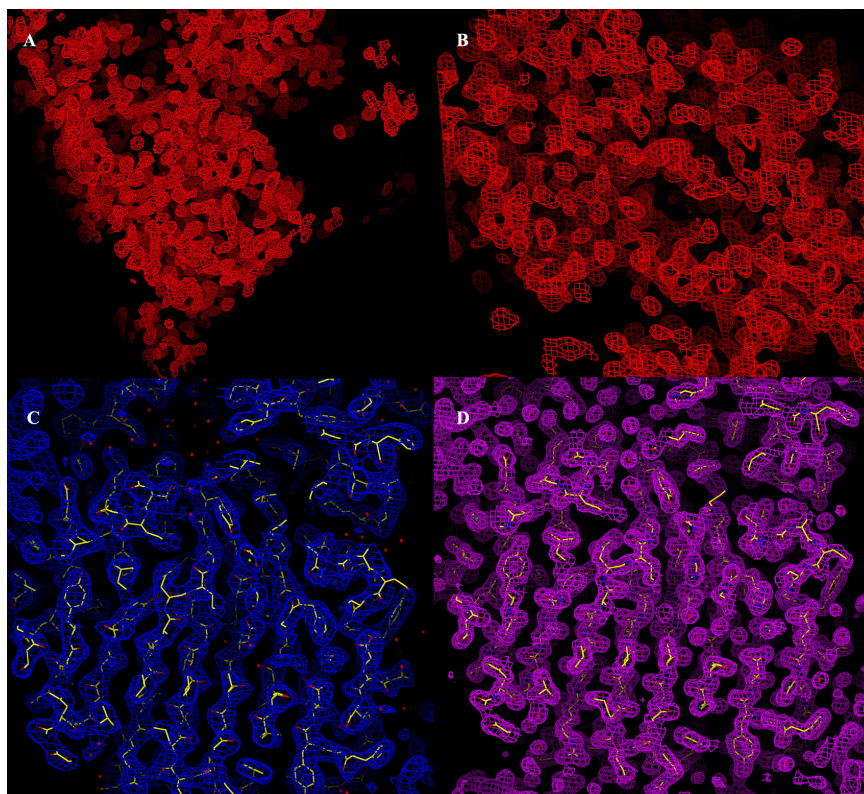
Figure 11-4 The electron density. (A) The initial electron density obtained from the first attempt of SAD phasing, solvent boundaries are detected. (B) Same initial electron density, where some secondary structure elements are found. (C) Initial electron density from the second SAD phasing at 2.6 Å resolution. Central is seen the good density for the β-sheet, the final model is depicted in sticks. (D) The final density of the CypX complex at a resolution of 1.75 Å at the same position. The figure is made directly in O.

11.2.7. Building the model

RESOLVE was able to build the majority of the model, in the final experimental SAD electron density map. However minor parts of the polypeptide chain still had to be built manually. The initial SeMet model was refined against the high-resolution inhibitor datasets, which have a resolution of 1.57 Å and 1.75 Å, respectively. The model was build and refined using cyclic repeats of refinement and model building (see Table 11-2 for refinement statistics). Water molecules were added and as the last step double conformations were build for amino acids, which showed electron density for a second conformation of the side chain. The occupancy of the different conformations was not refined due to the limited resolution.

Table 11-2 Refinement statistics.

Refinement statistics	Native (I222)	SAHA	CypX
Resolution range (Å)[1]	44.68-1.6 (1.638-1.596)	38.38-1.57 (1.611-1.57)	44.77-1.76 (1.803-1.757)
R_{cryst} (%) / R_{free} (%)[2]	16.2 (24.1)/18.6 (28.6)	17.2 (20.3)/20.1 (25.2)	15.6 (22.4)/19.1 (31.6)
Coordinate error (Å)[3]	0.051	0.053	0.068
# protein atoms:	11057	11042	11030
# ligand atoms:	28	88	56
# water molecules:	1236	1450	1288
Ramachandran plot[4]			
Most favourable regions	90.6	90.2	89.9
Additionally allowed	9.1	9.3	9.7
Generously allowed regions	0.2	0.5	0.5
r.m.s. deviations from			
Bonds (Å)/Angles (º)	0.013/1.479	0.013/1.456	0.015/1.471
Average B values (Å²)	17.306	16.498	18.739
Protein residues/ Waters	1468/1243	1468/1450	1468/1288

[1]Numbers in parenthesis refer to the highest resolution shell. [2] $R_{cryst}=100 \cdot \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the structure factor amplitudes from the data and the model, respectively. R_{free} is R_{cryst} with 5 % of test set structure factors. [3] Based on Maximum Likelihood [4] Calculated using PROCHECK (Laskowski *et al.* 1993).

11.2.8. Molecular replacement using the I222 form crystals

With the model from the $P2_1$ inhibitor containing datasets a molecular replacement with MOLREP was performed to obtain the structure of the native protein crystallised in the I222 space group. MOLREP found all four molecules in the asymmetric unit, which confirmed that the space group indeed was I222 and not $I2_12_12_1$. After a rigid body refinement the structure was manually corrected, water molecules were added and visually checked and

finally double conformations were build for amino acids showing electron density for two conformations of the side chain.

11.2.9. Crystal packing

Figure 11-5 shows the packing of the molecules in the crystals belonging to the two space groups, I222 and P2$_1$. Both forms have a solvent content of approximately 50%.
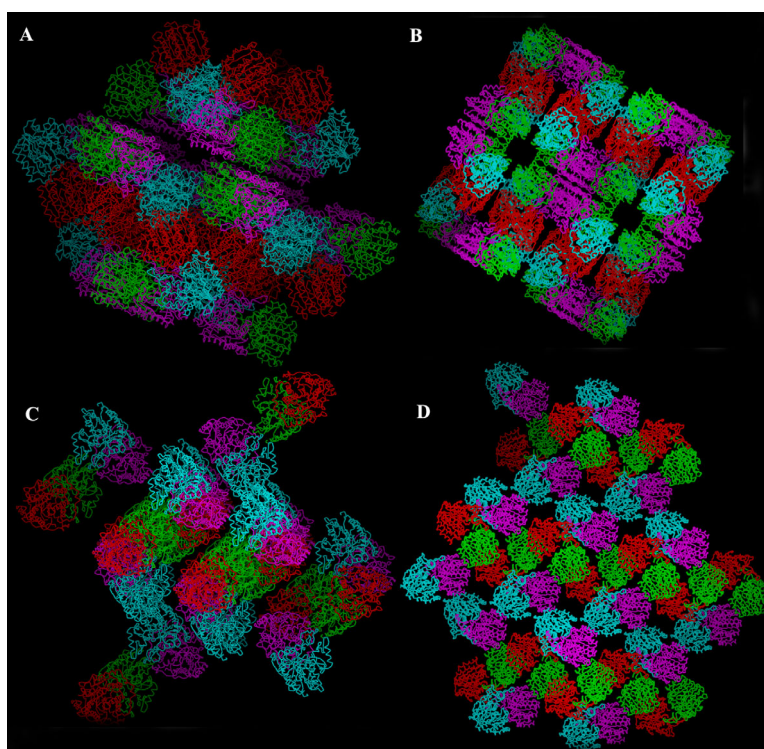
Figure 11-5 The packing of the molecules in the two different crystal forms, I222 and P2$_1$. (A)+(B) The packing in the orthorhombic I222 form. The four different molecules in the asymmetric unit are coloured in different colours. (C)+(D) The packing in the monoclinic P2$_1$ form.

11.2.10. Quality of the models

The native crystal structure of HDAH was refined at a resolution of 1.6 Å. The structures of HDAH with bound inhibitor SAHA and CypX, respectively, were refined at a resolution of 1.57 Å and 1.75 Å. All three structures comprise the protein residues 2-368 out of 369

residues and were refined to a free R-value of 18.6%, 20.1% and 19.1%, respectively, with very good stereochemistry (Table 11-2), where 90% of the residues are found in the most favourable area of the Ramachandran plots (Figure 11-6). Two cis-peptides were found in prolines (Pro139 and Pro209), however after examining the electron density, which confirmed this conformation, the conformations were kept.
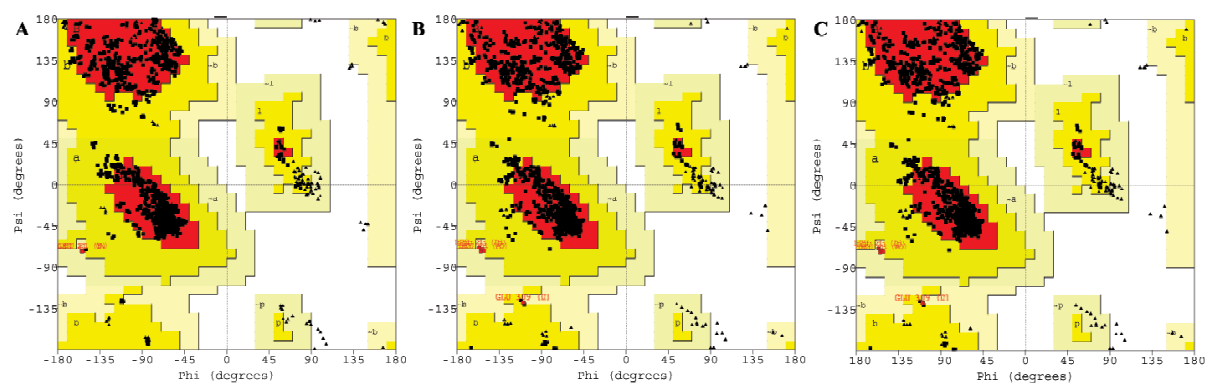
Figure 11-6 Ramachandran plot from PROCHECK, triangles denote glycines or prolines. (A) Apo-enzyme structure. (B) Structure of the complex with SAHA. (C) Structure of the complex with CypX. See Table 11-2 for values and text for description.

11.3. Overall structure

HDAH has a globular overall shape with dimensions of 50Å x 40Å x 40Å. It consists of an α/β domain with a seven stranded parallel β-sheet flanked on one side by three α-helices (α7, α8 and α9) and on the other side by four α-helices (α10, α11, α12 and α14), the remaining seven helices are shielding the top of the β-sheet (Figure 11-7). The four molecules in the asymmetric unit are very similar, upon superposition of Ca atoms the root-mean-square deviations are very small (<0.5 Å) and most of the significant differences are found in side chains on the surface of the protein due to different crystal contacts and thus these differences cannot be expected to have any biological significance.
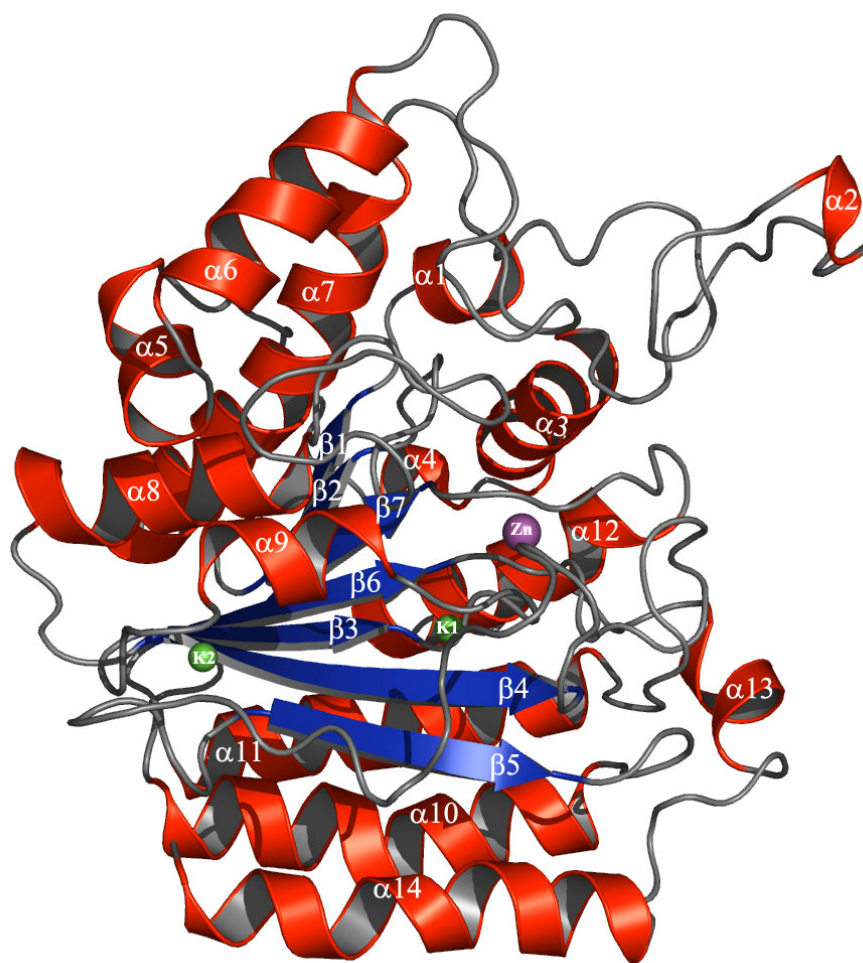
Figure 11-7 Overall structure of HDAH with bound ions. The zinc-ion is seen in purple and the two potassium ions in green.

HDAH contains a zinc-ion in its catalytic centre, but the occupancy of the zinc-ion in the crystal structures is approximately 50% as estimated from the examination of $|F_o-F_c|$ difference Fourier electron density maps calculated with different occupancies. The reason for this incomplete occupancy is most likely due to the fact that no zinc-ions were added to the buffers used in purification after the affinity column. The zinc-ion is penta-coordinated by Asp180 (Oδ1), Asp268 (Oδ2), His182 (Nδ1) and in the apo form by an acetate molecule from the crystallisation buffer. In the inhibitor complex structures the hydroxamate group of the inhibitor replaces the acetate (Figure 11-10).

There is also evidence in the electron density map for the presence of two potassium ions bound by HDAH. These ions are separated by a distance of 14 Å and are in the vicinity of the active site. The ions were found in all HDAH crystal structures, although potassium was present only in the crystallisation buffer of the apo form crystals. The observed hexa-coordination of the ions in a distorted octahedron shows typical distances for K^+-ions (Harding 2002). Furthermore, replacement of potassium by sodium (as in Somoza *et al.* (2004)) leads to a remaining positive $|F_o-F_c|$ difference electron density after refinement, indicating the presence of a heavier ion. In site 1, which is closest to the zinc-ion binding site, the K^+-ion is coordinated by Leu202 (main-chain carbonyl oxygen), Ser201 (Oγ), Asp180 (main-chain carbonyl oxygen), His182 (main-chain carbonyl oxygen), Asp178 (Oδ2) and Asp178 (main-chain carbonyl oxygen), where Asp180 and His182 also coordinated the zinc-ion. The site 2 K^+-ion is coordinated by main-chain carbonyl oxygen of Tyr226, Val197, Asp194 and Trp191 and two water molecules (Figure 11-8).
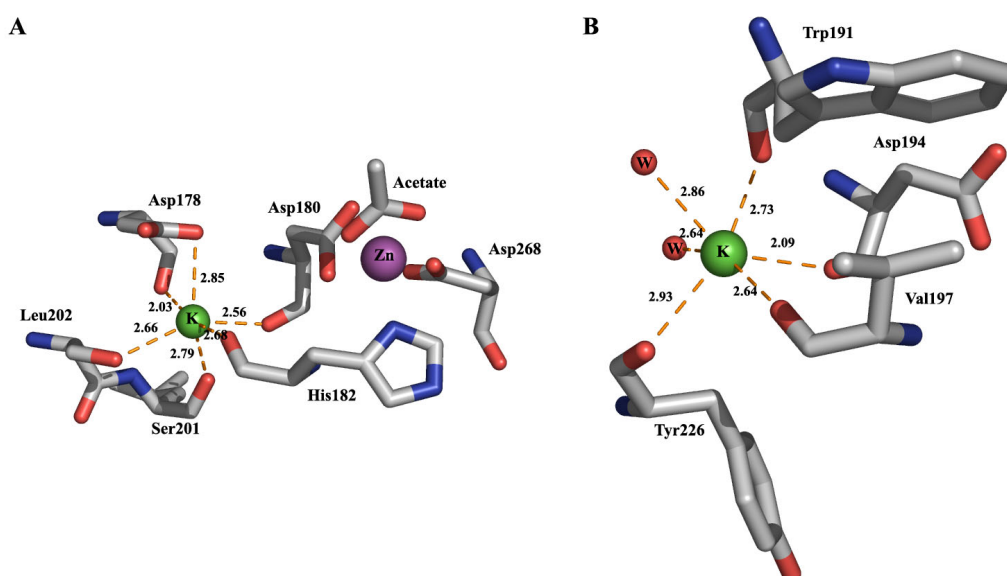
Figure 11-8 Potassium binding sites. The two potassium-ions binding sites and the zinc-ion binding site. The potassium ions are shown as green spheres, the zinc-ion as a purple sphere and red spheres are water molecules. All distances to the potassium ions are depicted. (A) Potassium binding site 1 with the zinc-ion binding site in close proximity. The amino acids participating in the coordination of the ions are also shown, Asp178, Asp180, His182, Ser201, Leu202 and Asp268 in potassium coordination and Asp180, His182, Asp268 and an acetate molecule in zinc-ion coordination. (B) Potassium binding site 2 with coordinating amino acids, Trp191, Asp194, Val197, Tyr226 and 2 water molecules.

Figure 11-9 shows the electrostatic surface representation of HDAH, HDAC8 and HDLP, all three proteins are in complex with SAHA. The position for the possible exit channel for the acetyl group is shown (see section 11.5.3). Comparison of the electrostatic surface potential reveals that HDAH is more acidic than the other histone deacetylases. HDAC8 appears to have the most basic residues exposed one, whereas HDLP exhibits an equal distribution of acidic and basic residues on the surface. However, all three proteins are rather acidic in the area around the entrance to the active site (Figure 11-9). On the figure it is also apparent that the loop area between $\alpha 1$ and $\alpha 3$ is breaking the otherwise globular form of the molecule. This area is positioned to the right of the entrance of the active site and close to the SAHA molecule. See section 12.3.1 for a discussion of the electrostatic surface potential in relation to substrate specificity.
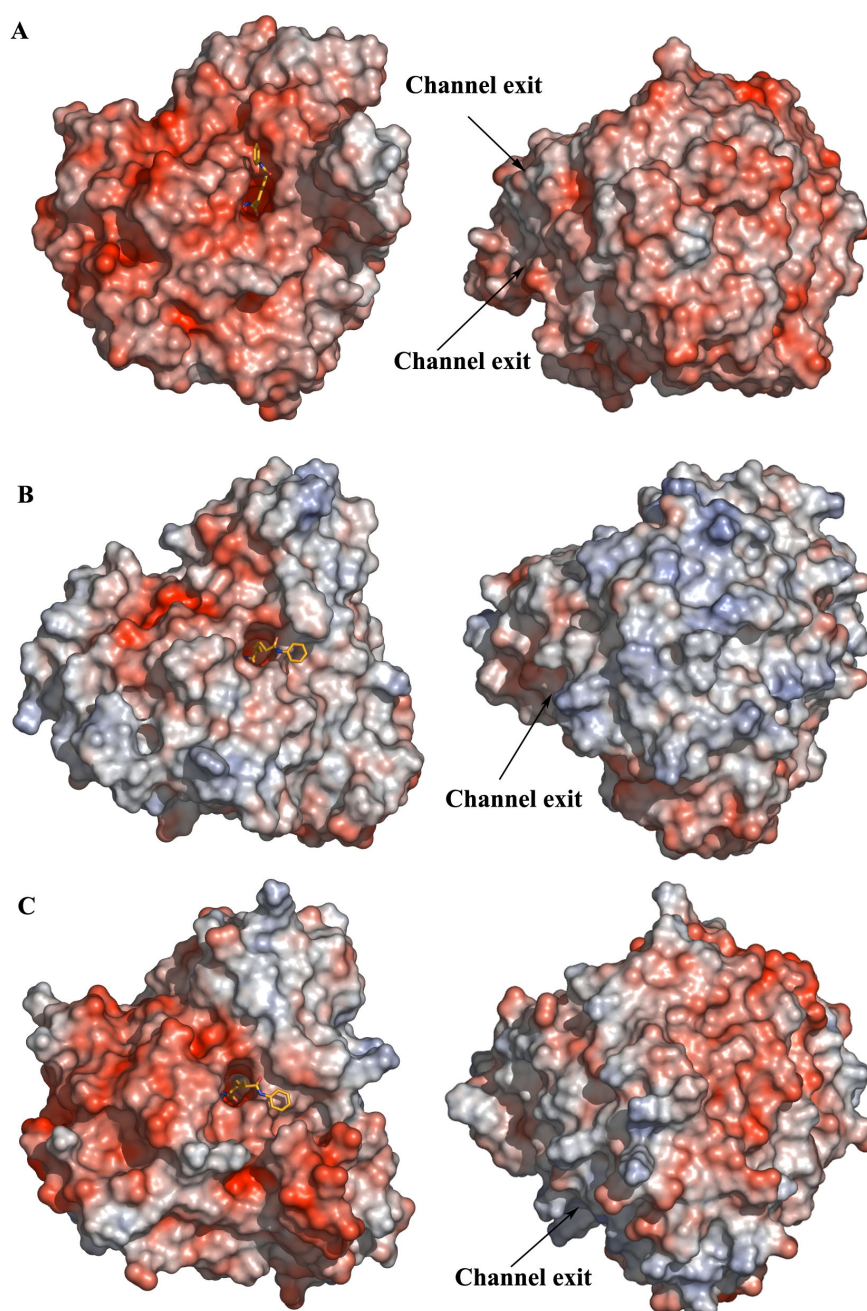
Figure 11-9 Electrostatic surface potential representation. The orientation of the proteins is the same, the left is from the front with a bound SAHA molecule and the right is turned 180°. On the figures are also depicted the approximate position for the tunnel exit in all three cases. (A) HDAH is quite acidic compared to the other proteins. (B) HDAC8 is the most basic of the proteins. (C) The electrostatic surface potential distribution of HDLP is in the middle of the other two proteins. The electrostatic surface potential is calculated with the APBS plug-in tool in PYMOL. All three proteins are shown with the same colour intensity.

11.4. Inhibitor structures

The inhibitors were co-crystallised in the presence of a 10-fold molar excess of inhibitor to the protein solution. The protein-inhibitor complex crystallised under identical conditions as the native protein and the crystals diffracted at the same resolution as the native protein. Figure 11-10 shows the two inhibitors bound, and the depicted electron density is an omit map (calculated without the inhibitor present) contoured at 0.9σ. In the figure is also shown the amino acids, which are believed to be part of the inner and outer charge transfer relay systems (Finnin *et al.* 1999); in HDAH Asp178 and Asn185 interacting with His142 and His143 (see Introduction section 9.2.4). The reason for the relatively poorly defined electron density of the inhibitor is due to the low occupancy in the position of the zinc-ion, leading to a correspondingly low occupancy of the inhibitor. This view is supported by previous findings indicating a 100-fold weaker binding of inhibitors in the absence of the hydroxamate-zinc interactions (Hildmann *et al.* 2004). Conformational flexibility of the ring moieties in both inhibitors may also contribute to the less defined electron density.
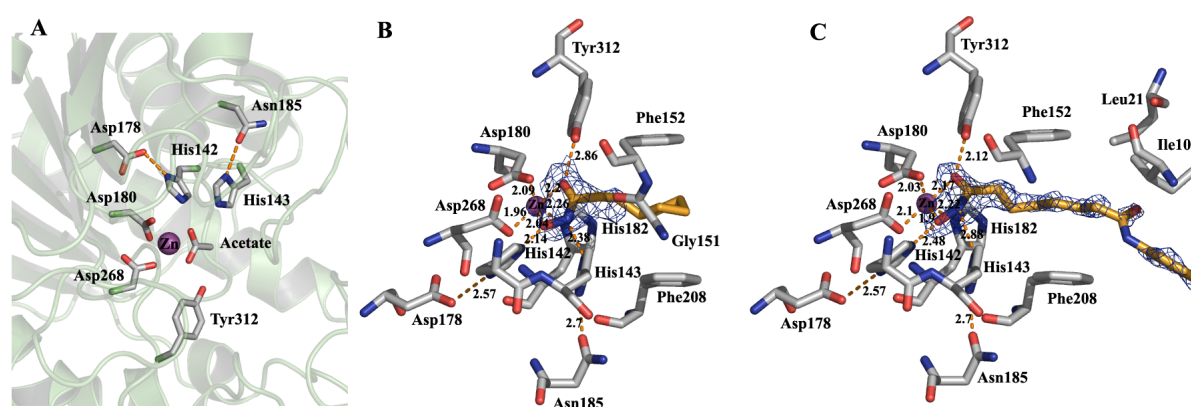
Figure 11-10 The active site without and with inhibitors bound. The inhibitors are depicted in orange. Shown are the amino acids in close vicinity of the inhibitors, the bound zinc-ion (purple sphere) and also the two amino acids proposed to take part in the charge transfer relay system. The electron density map is a 2|F$_o$-F$_c$| omit map, which is contoured at 0.9σ. (A) A view of the active site of the apo-enzyme with the zinc-ion (purple sphere) and the amino acids coordinating the ion (stick mode). (B) CypX bound to HDAH with the surrounding amino acids. The distance between the two phenylalanines is approximately 7.8 Å. (C) SAHA bound to HDAH with the surrounding amino acids. The view in (B)+(C) is turned 180° compared to (A).

The two rings in Phe152 and Phe208 (part of the active site channel) are in an ideal coplanar orientation (Figure 11-10 B+C) and could make excellent aromatic stacking with a possible inhibitor containing an aromatic ring system.

11.5. *Comparison with known structures*

11.5.1. Superposition

A structural comparison of HDAH with the known class I proteins, HDLP and HDAC8, reveals that HDAH shares ~20% identity with the HDLP in a primary structure alignment. The two structures superimpose the Ca backbone with a root-mean-square deviation of 1.47 Å over 80% of the residues. HDAH shares ~21% primary structural identity with HDAC8 and they superimpose with a root-mean-square deviation of 1.45 Å over 80% of the residues. In both cases the core of the domain superimposes very well (Figure 11-11), whereas the main structural differences are found in the loops. Ions corresponding to the two potassium ions present in HDAH are found at equivalent positions in the HDAC8 structure (Vannini *et al.* 2004) (Figure 11-11).
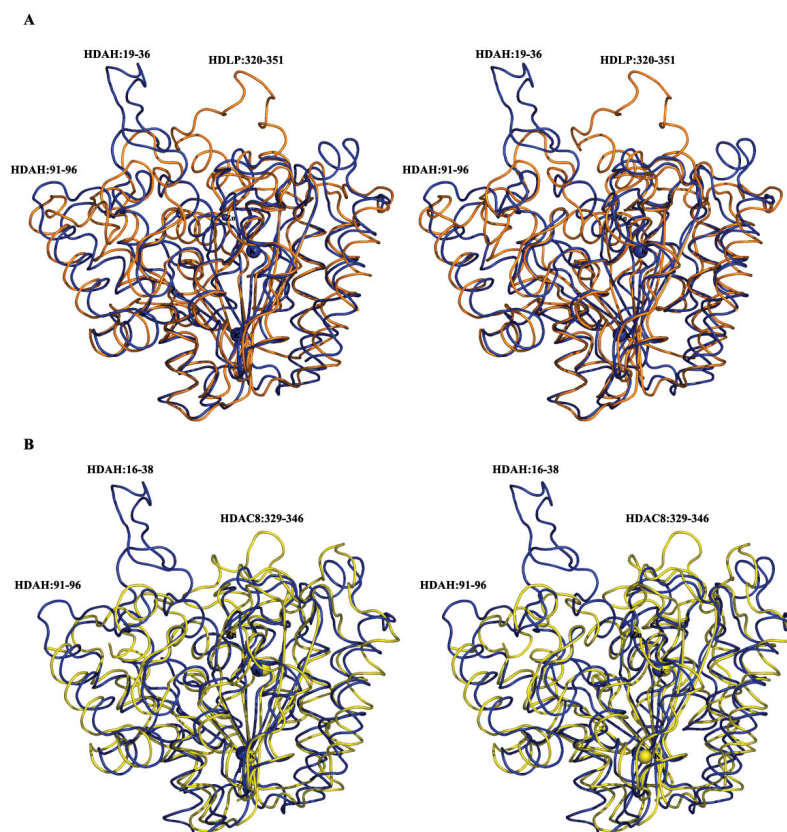
Figure 11-11 Stereo representation of the superposition of HDAH with HDLP (*A. aeolicus*) and HDAC8 (*H. sapiens*). (A) Superposition of HDAH (blue) and HDLP (orange, PDB code: 1C3P) shows that the main differences are found in the loop regions. Especially the residues 19-36 in HDAH are not present in HDLP. Furthermore HDLP has a loop extension 320-351. Loop 91-96 is also different in the two structures. (B) Superposition of HDAH (blue) and HDAC8 (yellow, PDB code: 1W22) shows that the main differences between the structures are the same as for HDLP, the class I homologue.

11.5.2. Differences in loop areas

A major difference found between HDAH and the two other proteins is amino acids 19-36 (residues between α1 and α3). In both crystal forms those residues make contacts to another molecule in the asymmetric unit stabilising the structure; in solution this part may very well be flexible and may change conformation upon binding to or interacting with another

protein. This hypothesis is consistent with the fact that the corresponding amino acids are found in close proximity of the entrance to the active site cavity.

Another difference between HDAH and HDLP is found in the region around residues 307-310 (HDLP: 292-295). Residue 309 is well conserved as a glutamine in class 2 proteins (Figure 9-5), whereas in class I enzymes it is a glycine. As a consequence of the presence of a larger side chain, i.e. to avoid stereo chemical problems, HDAH residues 135-142 (corresponding to positions 125-131 in HDLP) are shifted away (see Figure 11-12). Also in this region one of the cis-peptides, Pro139 is found.
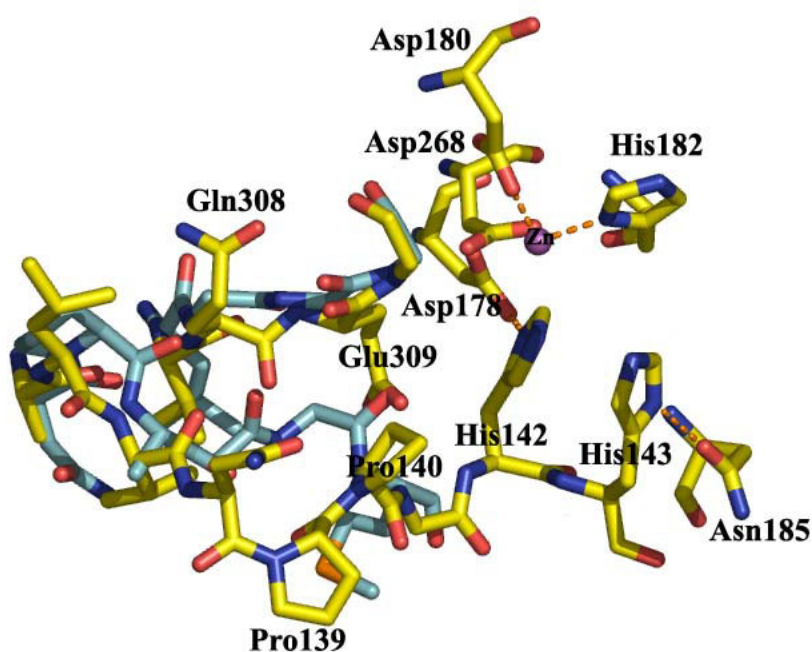
Figure 11-12 Superposition of residues around Glu309 from HDAH and HPLP. HDAH is shown in yellow, HDLP is shown in cyan and all the amino acid numbers correspond to HDAH. The superposition shows that when introducing a larger side chain, there is not enough space; hence amino acids 138-141 are displaced in HDAH in comparison to HDLP.

11.5.3. The exit tunnel

In HDAH there is an approximately 17 Å long channel leading from the bottom of the active site cavity possibly functioning as exit tunnel for the acetate product after the hydrolysis reaction. The inside of the channel is build up of the following residues: Val14, Asp15,

Thr16, His36, Ala38, Pro40, Thr42, Lys43, Ile100, Thr101, Met102, Met103, Asn138, Pro139, Pro140, Phe152, Cys153, Ile154, Gly311, Tyr312. Those are mainly hydrophobic residues, Lys43 could be an important residue and an equivalent of Arg16 and Arg27 found in the HDLP structure suggested to function as transient anchor points moving the acetate out of the channel, Lys43 is situated at the beginning of the channel. From the active site, the tunnel begins close to Gly310, Gly311 and Tyr312 and continues passing the two prolines, Pro139 and Pro140. Hereafter, the tunnel runs parallel with the active site (Figure 11-13 B) until it splits and exits to the solvent at two different areas. In FB188 HDAH, the exits of the channel are found on each side of the loop between helices $\alpha6$ and $\alpha7$ (Figure 11-7). One of the exits is then situated between the large loop ($\alpha1$-$\alpha3$) and the loop between $\alpha6$ and $\alpha7$. See Figure 11-9 for the position of the exit of the tunnel in the different proteins with a known structure. In the structures a number of ordered water molecules are found in the tunnel, although the tunnel is mainly hydrophobic (Figure 11-13 C). A channel was also proposed by Finnin *et al* (1999) and further investigated by Wang *et al* (2004) (Figure 11-13 A) in HDLP.
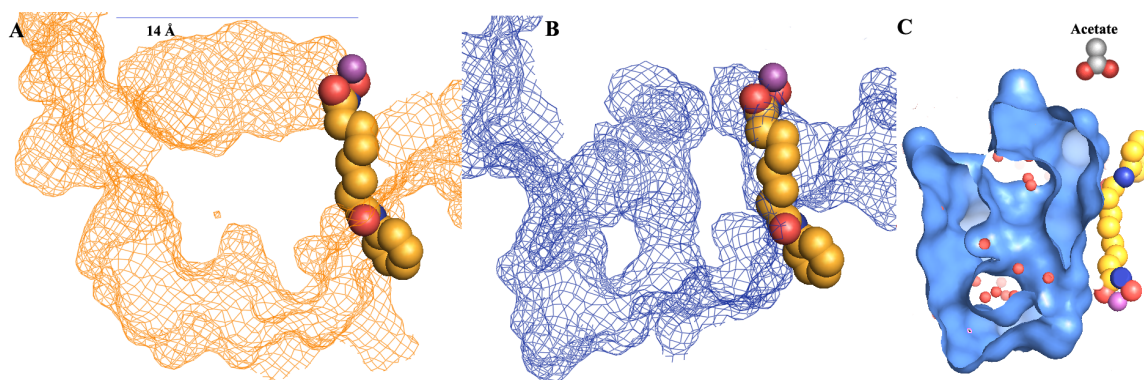
Figure 11-13 (A) HDLP (orange mesh) in the area around the active site and the adjacent proposed tunnel. The hydroxamic group of the SAHA molecule and the zinc-ion (purple) are shown in spheres. (B) HDAH (blue mesh) in the area around the active site and the adjacent possible tunnel with the two exits. The SAHA molecule and the zinc ion are shown as spheres. (C) A slab through the tunnel next to the active site. Ordered water molecules are shown in red spheres and an acetate molecule is depicted in the upper right corner. The view in (C) is turned 180° compared to (A) and (B).

12. Discussion

12.1. Different crystal forms

The first major obstacle, on the long way to determining the structures of the apo-form of HDAH and the complexes with SAHA and CypX, was the initial crystals obtained (crystallised in space groups $P2_12_12_1$ and I222). They were single crystals and diffracted well. However, the phasing was complicated by the presence of many molecules in the asymmetric unit, which made it impossible to use molecular replacement and difficult to obtain isomorphous heavy atom derivatives. Furthermore, the molecules in the $P2_12_12_1$ crystal form are most likely related by translational symmetry. This is indicated from the analysis of the average $\left|E^2\text{-}1\right|$, which is much higher than expected for a non-centrosymmetric space group (0.860 compared to the expected 0.736 in the non-centrosymmetric case). The structure in the $P2_12_12_1$ crystal form was later determined with molecular replacement and preliminary results (structure has not yet been refined) show how the four tetramers are arranged (Figure 12-1).
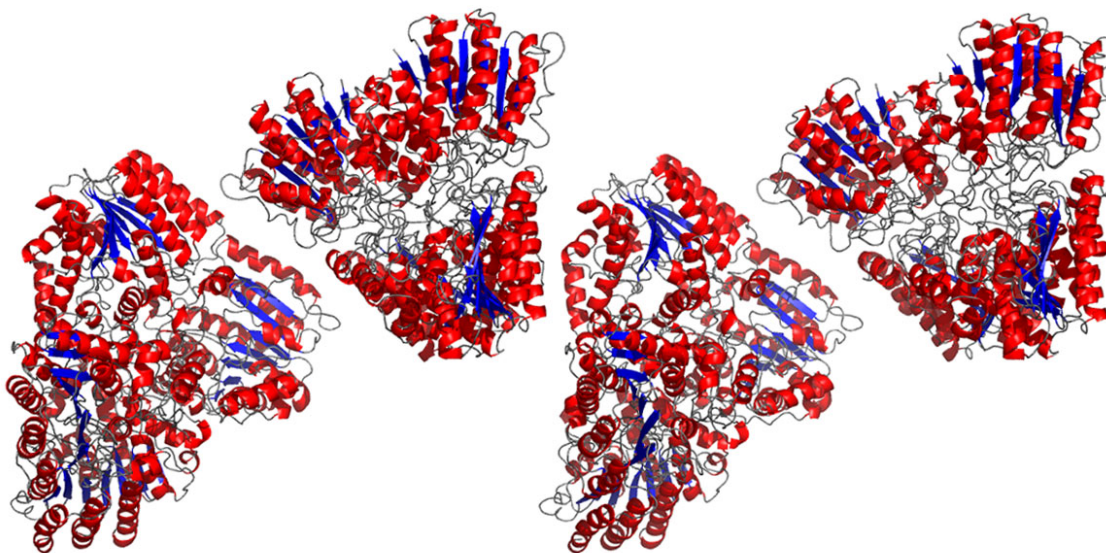

Figure 12-1 The structure in the $P2_12_12_1$ crystal form. The four tetramers in the asymmetric unit.

The two space groups are most likely connected by the fact that the P222 cell gets pseudo I-symmetry due to the translational symmetry. This is also found in the case of cellobiohydrolase I (Stahlberg *et al.* 1996).

12.2. Discussion of the catalytic mechanism in class 2 enzymes

The main difference between the sets of active site residues found in class 1 and class 2 enzymes is the presence of an asparagine in most class 2 enzymes (Introduction, section 9.2.4) in contrast to the class 1 aspartate. This aspartate and a nearby histidine (His143 in HDAH) are supposed to be part of the outer charge transfer relay system. The inner charge transfer relay system is probably the most important system for the catalysis. Preliminary analysis of both the asparagine (mutated to an aspartate) and the histidine (mutated to an asparagine) in HDAH did not abolish the activity, but instead revealed a reduction in catalytic activity of 10-fold and 100-fold, respectively (Hildmann *et al.*, in prep). Thus rather than playing an essential role in charge transfer, the asparagine in HDAH appears to have a secondary function, e.g. in orienting the histidine. This could also apply for other class 2 enzymes. Protonation of the nitrogen in the transition state may therefore very well be executed by the histidine alone or by another residue in the active site.

Mutations of the inner charge transfer relay system (His142 and Asp178) on the other hand completely eliminate catalytic activity (Hildmann *et al.*, in prep.) indicating that these two residues are essential for the catalytic mechanism. This is further supported by mutational studies on the aspartate in the His-Asp charge transfer relay systems in the serine protease, trypsin have shown that when changing the aspartate to an asparagine the catalytic activity is noticeably diminished (Craik *et al.* 1987, Sprang *et al.* 1987). Further support comes from computer-based simulations (Vanommeslaeghe *et al.* 2003), which have verified the existence of the inner His-Asp charge relay system.

Mutation of the zinc-ion ligands, either His182 or Asp268 totally abolishes or significantly reduces catalytic activity. This indicates an essential role of the enzyme-bound zinc cation, which is in agreement with earlier results (Hildmann *et al.* 2004). Mutation of active site Tyr312 decreases activity supporting the view that this residue, like His143 has an auxiliary function in the catalytic mechanism.

12.3. *Structural differences between class 1 and class 2 enzymes*

12.3.1. Electrostatic surface potential

An interesting feature in the comparison of HDAH (class 2 homologue) with known class 1 structures is the electrostatic surface potential (Figure 11-9 in section 11.3). HDAH is overall very acidic, whereas HDAC8 is very basic and HDLP somewhat intermediate with equal distributions of acidic and basic patches. The surface charge distribution gives a hint of the nature of the substrate and the differences of the substrates. HDAC8 is more basic already in the entrance of the active site cavity; the basic nature of HDAC8 does not fit very well with HDAC8 being a histone deacetylase, since histones also are very basic proteins. However, other proteins could participate in the deacetylation, since many HDACs often are found in complexes. The *in vivo* substrate of HDAC8 is so far unknown. The function of HDLP in *A. aeolicus* is also unknown. HDAH is the most acidic protein of the compared three, which could indicate a difference in substrate specificity.

12.3.2. Differences in the loops

When comparing the overall structures of enzymes from HDAC class 1 and 2 the canonical fold is very similar. In the light of this structure, with the very similar canonical fold as the class 1 enzymes, it is maybe worth to consider a revision of the classification, which originally is based on the sequence similarities. However, some of the other class 2 enzymes exhibit a more diverse sequence, which probably justifies the present classification. The differences between the two classes are mainly found in the loops. Close to the entrance region of the active site channel there are two loops, which vary in conformation between class 1 and 2 enzymes. HDAH has a big loop (amino acids 16-38 see section 11.5.2), which makes crystal contacts to a different molecule in the asymmetric unit in both crystal forms (I222 and P2$_1$). Many of the same residues from this loop take part in the contacts (Leu21, Arg29, Gln31 and Pro32) in the two different crystal forms. The loop may show a higher degree of flexibility in solution and may participate in interactions to the possible substrate. In HDLP and HDAC8 this loop is not present.

Another candidate region for substrate recognition may be loop 92-96 (between helices α6 and α7), which is poorly defined for crystals belonging to space group P2$_1$. This indicates a flexibility of the loop. The loop is in the rim area of the active site and in a region, where the sequence alignment shows a rather large diversity, both within class 2 enzymes but also in comparison to class 1 enzymes. Those features make this particular area interesting in respect of structure-based design of subtype-specific inhibitors. However, speculations about interaction of the entrance region of the active site cavity have to be made with caution. In both crystal forms of FB188 HDAH, residues from a different molecule in the asymmetric unit interact with either the opening of the active site cavity or the inhibitor (in the case of SAHA). This is also found in HDAC8 (Vannini *et al.* 2004). Therefore in Figure 11-9, where the cap region of SAHA is oriented approximately 90° different between class 1 and class 2 enzymes; this could very well be due to a crystallisation artefact and not a class specific binding.

12.4. Function of the closest human homologue, HDAC6

As mentioned earlier HDAH shows the highest sequence homology to human HDAC6 (section 9.3.2). HDAC6 is part of a multi-protein complex together with proteins taking part in control of protein ubiquitination. Consequently, it is suggested that HDAC6 could participate in deacetylation of proteins before their ubiquitination (Seigneurin-Berny *et al.* 2001). Furthermore cytoplasmic HDAC6 co-localises with the microtubule network (Hubbert *et al.* 2002) and deacetylates α-tubulin at the Lys40 (which has a stabilising function) both *in vivo* and *in vitro*. As shown tubulin acts as a competitive inhibitor of HDAH, this indicates that tubulin possibly could bind HDAH (Hildmann *et al.* 2004). Furthermore the best tripeptidic substrates for HDAH were shown to contain a basic arginine within the two neighbouring positions N-terminal to the acetylated lysine (Hildmann *et al.* 2004). This substrate specificity and the fact that HDAH is rather acidic (see Figure 11-9) indicate that the substrate of HDAH may very well contain an acetylated lysine in a sequence context comprising basic amino acids.

A possible substrate of HDAH should be found among proteins, which are acetylated and deacetylated on ε-amine of lysine residues. As mentioned tubulin inhibits the function of

HDAH *in vitro* and tubulin in Drosophila contains exactly the mentioned type of sequence context (KTK; where blue is the targeted lysine). The bacterial homologue of tubulin, ftsZ also contains a possible site KAK$_{39}$, however no acetylation or deacetylation has been reported for ftsZ so far. Another possibility could be the *E. coli* protein CheY, which is the most thoroughly studied example of an acetylatable protein in bacteria. CheY is a chemotaxis response regulator protein and the deacetylation of Lys92 affects signal generation (Ramakrishnan *et al.* 1998). In CheY, Lys92 is found in the sequence AKK. Those could all be substrate for HDAH, since also all sites are accessible on the surface of the proteins (Figure 12-2).
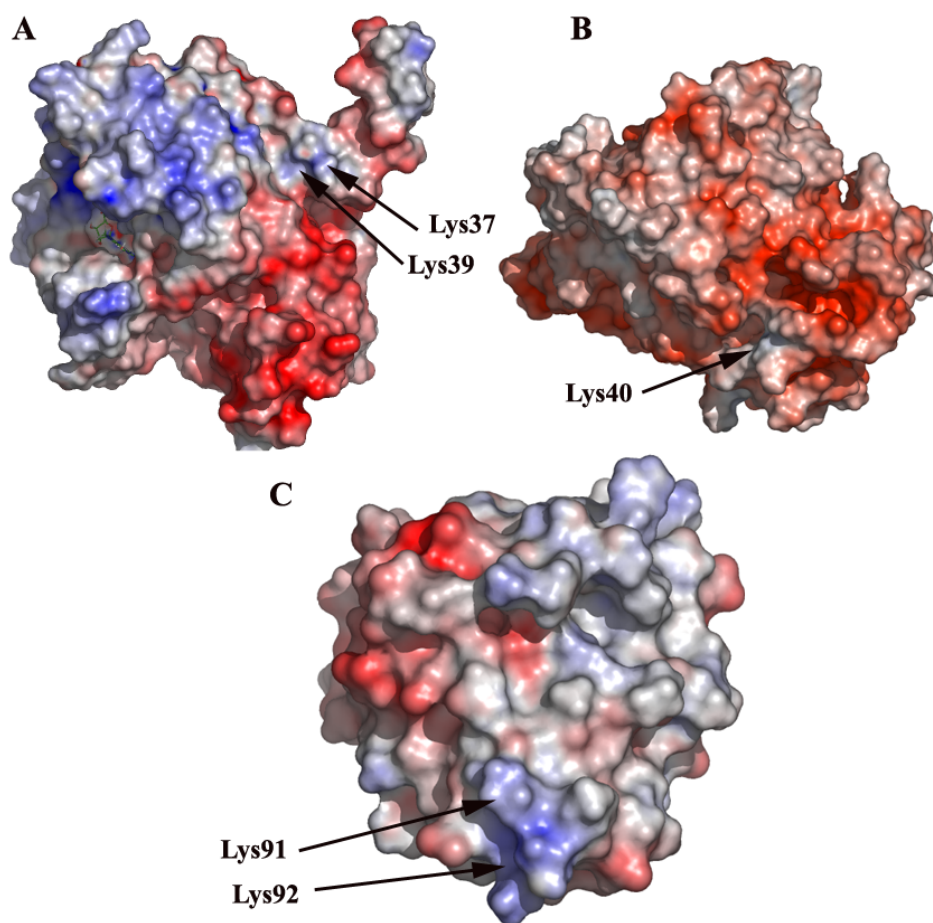

Figure 12-2 Electrostatic surface potential of ftsZ, tubulin and CheY. (A) The bacterial homologue of tubulin, ftsZ. The position of the possible acetylation site is marked with the arrows (PDB code: 1FSZ). (B) Tubulin (PDB code: 1FFX) with the acetylation site (Lys40) marked with the arrow. (C) The *E. coli* protein CheY, which can be acetylated in position Lys92 (PDB code: 1U8T).

The exact function of HDAH is still unknown, it has however been suggested that HDAH can deacetylate a 65 kDa protein from its organism, but the sequence of this protein is still unknown. More biochemical experiments are necessary to shed some light on the biological role of FB188 HDAH and the possible significance of the corresponding reversible acetylation in bacteria.

12.5. *Design of class-specific HDAC inhibitors*

With respect to designing class-specific HDAC inhibitors it is obvious to focus a bit more on the rim of the active site cavity. This area is where the different proteins with known structure, structurally diverge the most, and there is relatively low sequence homology between the individual enzymes. Another option could be a blockage of the internal channel, which may function as an exit channel for the acetate, however more biochemical experiments are needed in order to confirm that the internal channel is really the exit channel, e.g. whether mutations of Lys43 or other residues of the cavity affect the catalytic activity.

However, experiments have shown (Hildmann *et al*., in prep) that CypX is a class 2 specific inhibitor, a result difficult to explain from the crystal structure obtained. APHA 3c (Mai *et al.* 2003) is furthermore a class 2 specific inhibitor. The inhibition is probably enhanced by the presence of the pyrrole group, which makes pi-electron stacking to the two aromatic side groups (in HDAH Phe151 and Phe208) in the active site groove. However those two groups are also present in class 1 enzymes, which makes the explanation for the specificity more complex, since the reason must be found in either the active group (hydroxamic acid) or in the cap region. In the case of HC toxin and chlamydocin, which are class 1 specific inhibitors, the specificity is most probably found in the very large cap region. In HDAH residues 16-38 are in the proximity of the entrance of the active site, thus making it likely that an inhibitor with a large cap region would have stereo chemical problems to fit in the active site. More biochemical experiments are however needed in order to determine the exact role of the individual inhibitors. Also the crystal structure of HDAC6 would be helpful to see whether the area around the active site is structurally conserved.

The ultimate goal for HDAC inhibition is to find inhibitors, which inhibit different HDACs with different specificity; this can be achieved by structural based drug design in combination with biochemical tests of the given inhibitor. The structural information that is now available on the catalytic core architecture of class 2 HDACs could have a significant impact on the future design of class- or even subtype-selective inhibitors and thus will open the way to more selective novel anticancer therapeutics.

13. Conclusions and future perspectives

The structure of the apo-enzyme HDAH from *Bordetella/Alcaligenes* strain FB188 has been determined alone and in complex with two different inhibitors: SAHA and CypX. The structure was determined with the means of SAD phasing to 1.6 Å, 1.57 Å and 1.75 Å, correspondingly and had the canonical class 1 fold with a zinc-ion and two potassium-ions bound. The structure is the first structure of a HDAC class 2 homologue protein determined. Many questions like inhibitor specificity, catalytic mechanism and the possible exit tunnel are still unanswered and more biochemical experiments are necessary to reveal the answers.

Future perspectives for this project in means of crystallography could be co-crystallisation of the protein with other inhibitors in order to gain more knowledge about the specificity of the different types of inhibitors in combination with the classes of HDACs. In this respect, it is advisable for further crystallisation attempts to add zinc-ions to the buffer during the entire purification. This was discarded previous due to the very low solubility of e.g. zinc chloride. However, in the light of the current results even a small amount of zinc-ions in the buffers would probably still enhance the final occupancy of the zinc-ion and thus inhibitors in the crystal structure. Another interesting part in the design of more specific inhibitors is the interactions of the inhibitor-cap with the protein. In this case no certain statements could be made due to the interactions between the inhibitor-cap and a second molecule from the asymmetric unit. This was the case in both crystal forms achieved indicating that the area around the active site entrance is used for stabilising crystallisation. It would also be very interesting to crystallise the complex of HDAH with a possible substrate, however since the identity of the 65 kDa protein, found to interact with HDAH, is still unknown more biochemical experiments have to be performed. Furthermore, the three-dimensional structures of the many other HDACs, possibly in complex with co-repressors and regulators, will be important parts in the understanding of those processes in the cell.

14. Acknowledgements

Many people have helped me on the long and winding road, without whom this thesis would ever have finished. I wish to express my deepest thanks to the following people:

Prof. Ralf Ficner for giving me the possibility of doing my PhD in his department, for the very interesting projects and for giving me the freedom to work independently.

Prof. Oliver Einsle for being Korreferent. Oliver for being a Pymol-guru. Dr. Markus Rudolph for continuously making an effort to make me understand the wonders and principles of protein crystallography and Dr. Achim Dickmanns for discussions concerning the laboratory and for the corrections done in the manuscripts. I want to thank members of the Ficner, Einsle and Rudolph groups for pleasant working environment.

PD Andreas Schwienhorst and Dr. Christian Hildmann for being so pleasant co-workers. Christian and Dr. Julia Wittmann for the VERY thorough help with writing the German Zusammenfassung.

Dr. Jóhanna Arnórsdóttir for being a great colleague and a very good veninde. For many interesting and helpful discussions and the excellent and hyggelige time spent with Johanna Arnor and Teitur.

Jóhanna, Dr. Bente Vestergaard and Dr. Anja Strasser for reading and correcting the thesis with critical eyes. Anja for being so helpful both in the laboratory and in respect of German bureaucracy.

The Germans here in Göttingen, which have made me feel welcome and introduced me to many interesting aspects of German culture (the language, the customs, Bratwurst, Kulmbacher Bratwurst, Bier, Doppelkopf – just to mention a few things!). All my Danish friends, who despite the distance have kept in contact and visited me numerously times.

Finally I am deeply grateful to my parents for their endless support during this period and life in general and to Jose for standing by my side.

15. References

Bach, M., G. Winkelmann and R. Lührmann (1989). "20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition." Proc Natl Acad Sci U S A **86**(16): 6038-42.

Bach, M., G. Winkelmann and R. Lührmann (1989). "20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition." Proc Natl Acad Sci U S A **86**(16): 6038-42.

Bartels, C., H. Urlaub, R. Lührmann and P. Fabrizio (2003). "Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing." J Biol Chem **278**(30): 28324-34.

Bateman, O. A., A. G. Purkiss, R. van Mortfort, C. Slingsby and C. Graham (2003). "Crystal structure of η-crystallin: adaption of a class 1 aldehyde dehydrogenase for a new role in the eye lens." Biochemistry **42**: 4349-4356.

Behrens, S. E. and R. Lührmann (1991). "Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells." Genes Dev **5**(8): 1439-52.

Berger, S. L. (2002). "Histone modifications in transcriptional regulation." Curr Opin Genet Dev **12**: 142-148.

Berry, L. D. and K. L. Gould (1997). "Fission yeast dim1(+) encodes a functionally conserved polypeptide essential for mitosis." J Cell Biol **137**(6): 1337-54.

Bertos, N. R., A. H. Wang and X. J. Yang (2001). "Class II histone deacetylases: structure, function, and regulation." Biochem Cell Biol **79**(3): 243-52.

Boonacker, E. and C. J. Van Noorden (2003). "The multifunctional or moonlighting protein CD26/DPPIV." Eur J Cell Biol **82**(2): 53-73.

Brunger, A. T. (1993). "Assessment of phase accuracy by cross validation: the free R value. Methods and applications." Acta Crystallogr D Biol Crystallogr **49**(Pt 1): 24-36.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Buck, S. W., C. M. Gallo and J. S. Smith (2004). "Diversity in the Sir2 family of protein deacetylases." J Leukoc Biol **75**(6): 939-50.

Cheng, X. and T. A. Patterson (1992). "Construction and use of lambda PL promoter vectors for direct cloning and high level expression of PCR amplified DNA coding sequences." Nucleic Acids Res **20**(17): 4591-8.

Collaborative Computational Project, N. (1994). "The CCP4 suite: Programs for Protein Crystallography." Acta Crystallogr D Biol Crystallogr **50**: 760-763.

Collins, C. A. and C. Guthrie (2000). "The question remains: Is the spliceosome a ribozyme?" Nat Struct Biol **7**: 850-854.

Copley, S. D. (2003). "Enzymes with extra talents: moonlighting functions and catalytic promiscuity." Curr Opin Chem Biol **7**(2): 265-72.

Craik, C. S., S. Roczniak, C. Largman and W. J. Rutter (1987). "The catalytic role of the active site aspartic acid in serine proteases." Science **237**(4817): 909-13.

Dauter, Z. and M. Dauter (2000). "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides." Acta Crystallogr D Biol Crystallogr **56**: 232-237.

de Ruijter, A. J., A. H. van Gennip, H. N. Caron, S. Kemp and A. B. van Kuilenburg (2003). "Histone deacetylases (HDACs): characterization of the classical HDAC family." Biochem J **370**(Pt 3): 737-49.

Drenth, J. (1999). Principles of Protein X-ray Crystallography. New York, Springer-Verlag New York, Inc.

Dustin, M. L., M. W. Olszowy, A. D. Holdorf, J. Li, S. Bromley, N. Desai, P. Widder, F. Rosenberger, P. A. van der Merwe, P. M. Allen and A. S. Shaw (1998). "A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts." Cell **94**(5): 667-77.

Ejiri, S. (2002). "Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization." Biosci Biotechnol Biochem **66**(1): 1-21.

Fabrizio, P., B. Laggerbauer, J. Lauber, W. S. Lane and R. Lührmann (1997). "An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2." Embo J **16**(13): 4092-106.

Finnin, M. S., J. R. Donigian, A. Cohen, V. M. Richon, R. A. Rifkind, P. A. Marks, R. Breslow and N. P. Pavletich (1999). "Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors." Nature **401**(6749): 188-93.

Finzer, P., U. Soto, H. zur Hausen and F. Rosl (2001). "Inhibitors of histone deacetylase arrest cell cycle and induce apoptosis in cervical carcinoma cells circumventing human papillomavirus oncogene expression." Oncogene **20**: 4768-4776.

Fischle, W., V. Kiermer, F. Dequiedt and E. Verdin (2001). "The emerging role of class II histone deacetylases." Biochem Cell Biol **79**(3): 337-48.

Freund, C., V. Dotsch, K. Nishizawa, E. L. Reinherz and G. Wagner (1999). "The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences." Nat Struct Biol **6**(7): 656-60.

Freund, C., R. Kuhne, H. Yang, S. Park, E. L. Reinherz and G. Wagner (2002). "Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules." Embo J **21**(22): 5985-95.

Furumai, R., Y. Komatsu, N. Nishino, S. Khochbin, M. Yoshida and S. Horinouchi (2001). "Potent histone deacetylase inhibitors built from trichostatin A and cyclic tetrapeptide antibiotics including trapoxin." Proc Natl Acad Sci U S A **98**: 87-92.

Gao, L., M. A. Cueto, F. Asselbergs and P. Atadja (2002). "Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family." J Biol Chem **277**(28): 25748-55.

Gaudet, R., A. Bohm and P. B. Sigler (1996). "Crystal structure at 2.4 angstroms resolution of the complex of transducin betagamma and its regulator, phosducin." Cell **87**(3): 577-88.

Gottschalk, A., G. Neubauer, J. Banroques, M. Mann, R. Lührmann and P. Fabrizio (1999). "Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP." Embo J **18**(16): 4535-48.

Gottschalk, A., G. Neubauer, J. Banroques, M. Mann, R. Lührmann and P. Fabrizio (1999). "Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP." Embo J **18**(16): 4535-48.

Gouet, P., E. Courcelle, D. I. Stuart and F. Metoz (1999). "ESPript: analysis of multiple sequence alignments in PostScript." Bioinformatics **15**(4): 305-8.

Gregoretti, I. V., Y. M. Lee and H. V. Goodson (2004). "Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis." J Mol Biol **338**(1): 17-31.

Grignani, F., S. De Matteis, C. Nervi, L. Tomassoni, V. Gelmetti, M. Cioce, M. Fanelli, M. Ruthardt, F. F. Ferrara, I. Zamir, C. Seiser, F. Grignani, M. A. Lazar, S. Minucci and P. G. Pelicci (1998). "Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia." Nature **391**(6669): 815-8.

Grozinger, C. M. and S. L. Schreiber (2002). "Deacetylase enzymes: biological functions and the use of small-molecule inhibitors." Chem Biol **9**(1): 3-16.

Grunstein, M. (1997). "Histone acetylation in chromatin structure and transcription." Nature **389**(6649): 349-52.

Harding, M. M. (2002). "Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium." Acta Crystallogr D Biol Crystallogr **58**(Pt 5): 872-4.

Heltweg, B., F. Dequiedt, B. L. Marshall, C. Brauch, M. Yoshida, N. Nishino, E. Verdin and M. Jung (2004). "Subtype selective substrates for histone deacetylases." J Med Chem **47**(21): 5235-43.

Hendrickson, W. A. and C. M. Ogata (1997). "Phase Determination From Multiwavelength Anomalous Diffraction Measurements." Methods in Enzymology: 494-523.

Hildmann, C., M. Ninkovic, R. Dietrich, D. Wegener, D. Riester, T. Zimmermann, O. M. Birch, C. Bernegger, P. Loidl and A. Schwienhorst (2004). "A new amidohydrolase from Bordetella or Alcaligenes strain FB188 with similarities to histone deacetylases." J Bacteriol **186**(8): 2328-39.

Honig, B. and A. Nicholls (1995). "Classical electrostatics in biology and chemistry." Science **268**(5214): 1144-9.

Hubbert, C., A. Guardiola, R. Shao, Y. Kawaguchi, A. Ito, A. Nixon, M. Yoshida, X. F. Wang and T. P. Yao (2002). "HDAC6 is a microtubule-associated deacetylase." Nature **417**(6887): 455-8.

James, L. C., P. Roversi and D. S. Tawfik (2003). "Antibody multispecificity mediated by conformational diversity." Science **299**(5611): 1362-7.

Jeffery, C. J. (2003). "Moonlighting proteins: old proteins learning new tricks." Trends Genet **19**(8): 415-7.

Jeffery, C. J. (2004). "Molecular mechanisms for multitasking: recent crystal structure of moonlighting proteins." Curr Opin Struc Biol **14**: 663-668.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.

Jepsen, K. and M. G. Rosenfeld (2002). "Biological roles and mechanistics actions of corepressor complexes." J Cell Sci **115**: 689-698.

Jogl, G., X. Tao, Y. Xu and L. Tong (2001). "COMO: a program for combined molecular replacement." Acta Crystallogr D Biol Crystallogr **57**(Pt 8): 1127-34.

Johnstone, R. W. (2002). "Histone-deacetylase inhibitors: novel drugs for the treatment of cancer." Nat Rev Drug Discov **1**(4): 287-99.

Jones, T. A. and M. Kjeldgaard (1997). "Electron-density map interpretation." Methods in Enzymology **277**: 173-208.

Jones, T. A., J. Y. Zou, S. W. Cowan and Kjeldgaard (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr. A **47**(2): 110-9.

Jurica, M. S. and M. J. Moore (2003). "Pre-mRNA Splicing: Awash in a Sea of Proteins." Mol Cell **12**: 5-14.

Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Lührmann, J. Li and K. Nagai (1999). "Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs." Cell **96**(3): 375-87.

Kelley, L. A. and M. J. Sutcliffe (1997). "OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And DOmains." Protein Sci **6**(12): 2628-30.

Khochbin, S. and A. P. Wolffe (1997). "The origin and utility of histone deacetylases." FEBS Lett **419**(2-3): 157-60.

Kim, J. S., S. Lee, T. Lee, J. B. Trepel and Y. W. Lee (2001). "Transcriptional activation of p21WAF1/CIP1 by Apicidin, a novel histone deacetylase inhibitor." Biochem Biophys Res Commun **281**: 866-871.

Kim, Y. B., S. W. Ki, M. Yoshida and S. Horinouchi (2000). "Mechanism of cell cycle arrest caused by histone deacetylase inhibitors in human carcinoma cells." J Antibio **s53**: 1191-1200.

Kissinger, C. R., D. K. Gehlhaar and D. B. Fogel (1999). "Rapid automated molecular replacement by evolutionary search." Acta Crystallogr D Biol Crystallogr **55 (Pt 2)**: 484-91.

Kleywegt, G. J. (1996). "Use of non-crystallographic symmetry in protein structure refinement." Acta Crystallogr D Biol Crystallogr **52**(Pt 4): 842-57.

Knoepfler, P. S. and R. N. Eisenman (1999). "Sin meets NuRD and other tails of repression." Cell **99**: 447-450.

Koeller, K. M., S. J. Haggarty, B. D. Perkins, I. Leykin, J. C. Wong, M. C. Kao and S. L. Schreiber (2003). "Chemical genetic modifier screens: small molecule trichostatin suppressors as probes of intracellular histone and tubulin acetylation." Chem Biol **10**(5): 397-410.

Kofler, M. M., K. Heuer, T. Zech and C. Freund (2004). "Recognition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2." J Biol Chem.

Kouzarides, T. (2000). "Acetylation: a regulatory modification to rival phosphorylation?" Embo J **19**(6): 1176-9.

Kumar, J. K., S. Tabor and C. C. Richardson (2001). "Role of the C-terminal residue of the DNA polymerase of bacteriophage T7." J Biol Chem **276**(37): 34905-12.

Laggerbauer, B., S. Liu, E. Makarov, H. P. Vornlocher, O. Makarova, D. Ingelfinger, T. Achsel and R. Lührmann (2005). "The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation." Rna **11**(5): 598-608.

Laggerbauer, B., S. Liu, E. Makarov, H. P. Vornlocher, O. Makarova, D. Ingelfinger, T. Achsel and R. Lührmann (2005). "The human U5 snRNP 52K protein (CD2BP2) interacts with U5-102K (hPrp6), a U4/U6.U5 tri-snRNP bridging protein, but dissociates upon tri-snRNP formation." Rna **11**(5): 598-608.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Crystallogr. **26**: 283-291.

Lawrence, M. C. and P. M. Colman (1993). "Shape complementarity at protein/protein interfaces." J Mol Biol **234**(4): 946-50.

Leipe, D. D. and D. Landsman (1997). "Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily." Nucleic Acids Res **25**(18): 3693-7.

Lennon, B. W., C. H. Williams, Jr. and M. L. Ludwig (2000). "Twists in catalysis: alternating conformations of Escherichia coli thioredoxin reductase." Science **289**(5482): 1190-4.

Li, J., K. Nishizawa, W. An, R. E. Hussey, F. E. Lialios, R. Salgia, R. Sunder-Plassmann and E. L. Reinherz (1998). "A cdc15-like adaptor protein (CD2BP1) interacts with the CD2 cytoplasmic domain and regulates CD2-triggered adhesion." Embo J **17**(24): 7320-36.

Lin, R. J., L. Nagy, S. Inoue, W. Shao, W. H. Miller, Jr. and R. M. Evans (1998). "Role of the histone deacetylase complex in acute promyelocytic leukaemia." Nature **391**(6669): 811-4.

Lodish, H., D. Baltimore, A. Berk, S. L. Zipursky, P. Matsudaira and J. Darnell (1995). Molecular Cell Biology, W. H. Freeman and Company.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.

Mai, A., S. Massa, R. Pezzi, D. Rotili, P. Loidl and G. Brosch (2003). "Discovery of (aryloxopropenyl)pyrrolyl hydroxyamides as selective inhibitors of class IIa histone deacetylase homologue HD1-A." J Med Chem **46**(23): 4826-9.

Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm and R. Lührmann (2002). "Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome." Science **298**(5601): 2205-8.

Makarova, O. V., E. Makarov, S. Liu, H. P. Vornlocher and R. Lührmann (2002). "Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6*U5 tri-snRNP formation and pre-mRNA splicing." Embo J **21**: 1148-1157.

Marmorstein, R. (2004). "Structure and chemistry of the Sir2 family of NAD+-dependent histone/protein deactylases." Biochem Soc Trans **32**(Pt 6): 904-9.

Matsuyama, A., T. Shimazu, Y. Sumida, A. Saito, Y. Yoshimatsu, D. Seigneurin-Berny, H. Osada, Y. Komatsu, N. Nishino, S. Khochbin, S. Horinouchi and M. Yoshida (2002). "In vivo destabilization of dynamic microtubules by HDAC6-mediated deacetylation." Embo J **21**(24): 6820-31.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-7.

McLaughlin, F., P. Finn and N. B. La Thangue (2003). "The cell cycle, chromatin and cancer: mechanism based therapeutics come of age." Drug Disc Today **8**: 793-802.

McRee, D. E. (1999). "XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density." J Struct Biol **125**(2-3): 156-65.

Miller, R., S. M. Gallo, H. G. Khalak and C. M. Weeks (1994). "SnB: crystal structure determination via *Shake-and-Bake*." J Appl Cryst **27**: 120-124.

Minucci, S., C. Nervi, F. Lo Coco and P. G. Pelicci (2001). "Histone deacetylases: a common molecular target for differentiation treatment of acute myeloid leukemias?" Oncogene **20**(24): 3110-5.

Moore, B. (2004). "Bifunctional and moonlighting enzymes: lighting the way to regulatory control." Trends Plant Sci **9**(5): 221-8.

Mougin, A., A. Gottschalk, P. Fabrizio, R. Lührmann and C. Branlant (2002). "Direct probing of RNA structure and RNA-protein interactions in purified HeLa cell's and yeast spliceosomal U4/U6.U5 tri-snRNP particles." J Mol Biol **317**(5): 631-49.

Murshudov, G. N., A. A. Vagin, A. Lebedev, K. S. Wilson and E. J. Dodson (1999). "Efficient anisotropic refinement of macromolecular structures using FFT." Acta Crystallogr D Biol Crystallogr **55 (Pt 1)**: 247-55.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Crystallogr A **50**: 157-163.

Newman, A. J. (1997). "The role of U5 snRNP in pre-mRNA splicing." Embo J **16**: 5797-5800.

Ng, H. H. and A. Bird (2000). "Histone deacetylases: silencers for hire." Trends Biochem Sci **25**: 121-126.

Nicholls, A. and B. Honig (1991). "A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation." J. Comp. Chem **12**: 435-445.

Nilsen, T. W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Nishizawa, K., C. Freund, J. Li, G. Wagner and E. L. Reinherz (1998). "Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation." Proc Natl Acad Sci U S A **95**(25): 14897-902.

Nissen, P., J. Hansen, N. Ban, P. B. Moore and T. A. Steitz (2000). "The structural basis of ribosome activity in peptide bond synthesis." Science **289**: 920-930.

Pape, T. and T. R. Schneider (2004). "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs." J. Appl. Cryst. **37**: 843-844.

Perrakis, A., R. Morris and V. S. Lamzin (1999). "Automated protein model building combined with iterative structure refinement." Nat Struct Biol **6**(5): 458-63.

Peterson, C. L. (2002). "HDAC's at work: everyone doing their part." Mol Cell **9**(5): 921-2.

Piatigorsky, J. (2003). "Crystallin genes: specialization by changes in gene regulation may precede gene duplication." J Struct Funct Genomics **3**: 131-137.

Qin, J., G. M. Clore, W. M. Kennedy, J. R. Huth and A. M. Gronenborn (1995). "Solution structure of human thioredoxin in a mixed disulfide intermediate complex with its target peptide from the transcription factor NF kappa B." Structure **3**(3): 289-97.

Qin, J., G. M. Clore, W. P. Kennedy, J. Kuszewski and A. M. Gronenborn (1996). "The solution structure of human thioredoxin complexed with its target from Ref-1 reveals peptide chain reversal." Structure **4**(5): 613-20.

Ramakrishnan, R., M. Schuster and R. B. Bourret (1998). "Acetylation of Lys-92 enhances signaling by the chemotaxis response regulator protein CheY." Proc Natl Acad Sci U S A **95**: 4918-4923.

Reidt, U., K. Reuter, T. Achsel, D. Ingelfinger, R. Lührmann and R. Ficner (2000). "Crystal structure of the human U4/U6 small nuclear ribonucleoprotein particle-specific SnuCyp-20, a nuclear cyclophilin." J Biol Chem **275**(11): 7439-42.

Reuter, K. and R. Ficner (1999). "Overproduction, purification, crystallization and preliminary x-ray diffraction studies of the human spliceosomal protein U5-15kD." Acta Crystallogr D Biol Crystallogr **55 (Pt 4)**: 888-90.

Reuter, K., S. Nottrott, P. Fabrizio, R. Lührmann and R. Ficner (1999). "Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein." J Mol Biol **294**(2): 515-25.

Reuter, K., S. Nottrott, P. Fabrizio, R. Lührmann and R. Ficner (1999). "Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein." J Mol Biol **294**(2): 515-25.

Riester, D., D. Wegener, C. Hildmann and A. Schwienhorst (2004). "Members of the histone deacetylase superfamily differ in substrate specificity towards small synthetic substrates." Biochem Biophys Res Commun **324**(3): 1116-23.

Rosato, R. R. and S. Grant (2003). "Histone deacetylase inhibitors in cancer therapy." Cancer Biol Ther **2**(1): 30-7.

Rossman, M. G. and D. M. Blow (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Crystallogr A **15**: 24-31.

Schaffert, N., M. Hossbach, R. Heintzmann, T. Achsel and R. Lührmann (2004). "RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies." Embo J **23**: 3000-3009.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.

Schuttelkopf, A. W. and D. M. van Aalten (2004). "PRODRG: a tool for high-throughput crystallography of protein-ligand complexes." Acta Crystallogr D Biol Crystallogr **60**(Pt 8): 1355-63.

Seigneurin-Berny, D., A. Verdel, S. Curtet, C. Lemercier, J. Garin, S. Rousseaux and S. Khochbin (2001). "Identification of components of the murine histone deacetylase 6 complex: link between acetylation and ubiquitination signaling pathways." Mol Cell Biol **21**(23): 8035-44.

Somoza, J. R., R. J. Skene, B. A. Katz, C. Mol, J. D. Ho, A. J. Jennings, C. Luong, A. Arvai, J. J. Buggy, E. Chi, J. Tang, B. C. Sang, E. Verner, R. Wynands, E. M. Leahy, D. R. Dougan, G. Snell, M. Navre, M. W. Knuth, R. V. Swanson, D. E. McRee and L. W. Tari (2004). "Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases." Structure (Camb) **12**(7): 1325-34.

Sprang, S., T. Standing, R. J. Fletterick, R. M. Stroud, J. Finer-Moore, N. H. Xuong, R. Hamlin, W. J. Rutter and C. S. Craik (1987). "The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis." Science **237**(4817): 905-9.

Stahlberg, J., C. Divine, A. Koivula, K. Piens, M. Claeyssens, T. T. Teeri and T. A. Jones (1996). "Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from Thichoderma reesei." J Mol Biol **264**: 337-49.

Staley, J. P. and C. Guthrie (1998). "Mechanical devices of the spliceosome: motors, clocks, springs, and things." Cell **92**(3): 315-26.

Stevens, S. W., I. Barta, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee and J. Abelson (2001). "Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae." Rna **7**(11): 1543-53.

Storoni, L. C., A. J. McCoy and R. J. Read (2004). "Likelihood-enhanced fast rotation functions." Acta Crystallogr D Biol Crystallogr **60**(Pt 3): 432-8.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**: 41-45.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to yeast transcriptional regulator Rpd3p." Science **272**: 408-411.

Teigelkamp, S., A. J. Newman and J. D. Beggs (1995). "Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA." Embo J **14**(11): 2602-12.

Terwilliger, T. C. (1999). "Reciprocal-space solvent flattening." Acta Crystallogr D Biol Crystallogr **55**(11): 1863-71.

Terwilliger, T. C. (2003). "Automated main-chain model building by template matching and iterative fragment extension." Acta Crystallogr D Biol Crystallogr **59**(Pt 1): 38-44.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **55 (Pt 4)**: 849-61.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg (2000). "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae." Nature **403**(6770): 623-7.

15. References

Vagin, A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." J. Appl. Crystallogr. **30**: 1022-1025.

Vannini, A., C. Volpari, G. Filocamo, E. C. Casavola, M. Brunetti, D. Renzoni, P. Chakravarty, C. Paolini, R. De Francesco, P. Gallinari, C. Steinkuhler and S. Di Marco (2004). "Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor." Proc Natl Acad Sci U S A **101**(42): 15064-9.

Vanommeslaeghe, K., C. Van Alsenoy, F. De Proft, J. C. Martins, D. Tourwé and P. Geelings (2003). "*Ab initio* study of the binding of Trichostatin A in the active site of Histone Deacetylase Like Protein (HDLP)." Org Biomol Chem **1**: 2951-2957.

Villar-Garea, A. and M. Esteller (2004). "Histone deacetylase inhibitors: understanding a new wave of anticancer agents." Int J Cancer **112**(2): 171-8.

Wade, P. A., D. Pruss and A. P. Wolffe (1997). "Histone acetylation: chromatin in action." Trends Biochem Sci **22**(4): 128-32.

Wang, D. F., O. Wiest, P. Helquist, H. Y. Lan-Hargest and N. L. Wiech (2004). "On the function of the 14 A long internal cavity of histone deacetylase-like protein: implications for the design of histone deacetylase inhibitors." J Med Chem **47**(13): 3409-17.

Wang, Q., K. Hobbs, B. Lynn and B. C. Rymond (2003). "The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts." J Biol Chem **278**(10): 7875-83.

Will, C. L. and R. Lührmann (2001). "Spliceosomal UsnRNP biogenesis, structure and function." Curr Opin Cell Biol **13**(3): 290-301.

Wool, I. G. (1996). "Extraribosomal functions of ribosomal proteins." Trends Biochem Sci **21**(5): 164-5.

Yang, X. J. and E. Seto (2003). "Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression." Curr Opin Genet Dev **13**: 143-153.

Zhang, Y., N. Li, C. Caron, G. Matthias, D. Hess, S. Khochbin and P. Matthias (2003). "HDAC-6 interacts with and deacetylates tubulin and microtubules in vivo." Embo J **22**(5): 1168-79.

Zhang, Y. and D. Reinberg (2001). "Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails." Genes Dev **15**: 2343-2360.

Zhu, W., I. R. Rainville, M. Ding, M. Bolus, N. H. Heintz and D. S. Pederson (2002). "Evidence that the pre-mRNA splicing factor Clf1p plays a role in DNA replication in Saccharomyces cerevisiae." Genetics **160**(4): 1319-33.

16. Curriculum Vitae

Personal data

Name: Tine Kragh Nielsen

Address: Obere Maschstrasse 22

                    37073 Goettingen

                    Germany

Birth date: 11.03.1976

Birthplace: Pusan, Korea

Nationality: Danish

Marital status: Single

Education

1991 – 1994:   High school at Frederikshavn Gymnasium og HF

1995 – 1999:   Undergraduate work at the Department of Chemistry, University of Aarhus, Denmark

1999 – 2000:   Pre-masters laboratory work at the Centre for Structural Biology, University of Aarhus, Denmark under supervision of Asc. Prof. Søren Thirup

2000 – 2002:   Masters laboratory work at the Centre for Structural Biology, University of Aarhus, Denmark under supervision of Asc. Prof. Søren Thirup

Since 2002:   Ph.D.-student at Department of Molecular Structural Biology, Institute for Microbiology and Genetics, University of Goettingen, Germany under supervision of Prof. Ralf Ficner