Technische Universität München
Institut für Organische Chemie und Biochemie

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

Structural and biochemical characterization of cell cycle regulatory proteins and their inhibitors

Sreejesh Shanker

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation

Vorsitzender: Univ.-Prof. Dr. Dr. A. Bacher
Prüfer der Dissertation:
1. apl. Prof. Dr. Dr. h. c. R. Huber
2. Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 02.12.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 18.01.2005 angenommen.

Dedicated to mummy, papa, and sree

Acknowledgements

I would like to thank all the people who have contributed to this work.

In particular I am grateful to Professor Robert Huber for giving me the opportunity to work in his esteemed department.

My supervisors, Dr. Richard A. Engh and Dr. Tad A. Holak to whom I am indebted for their scientific guidance, inspiration, and friendly advices.

I am grateful to Roche Diagnostics for giving me the opportunity to collaborate with them and work on their projects and would like to thank all members involved in my project at Roche Diagnostics, Penzberg for providing me all the necessary materials, and logistic support for successful completion of this work.

My thanks to all the present and former members of the MPI structure biology group, in particular to the Roche and NMR group, for giving me valuable advices, help and to top it all a wonderful scientific atmosphere to work in.

My thanks to my friends circle outside the MPI for providing that non scientific buffer, so essential to survive those rainy days of research, and also providing that great Indian atmosphere which makes you feel at home, away from home.

My apologies to all others who I have not mentioned by name I am indebted to them for the many ways they helped me.

Finally, I would like to pay tribute to the constant support of my family members, whose love and sacrifices can never be repaid.

Publications

Parts of this thesis have been published or will be published in due course

Sreejesh Shanker, Mathais.Koerner, Erhard.Kopetzki, Adelbert.Grossmann, Hans Brandstetter, Norman Kairies, Robert Huber, Richard Engh.
Co-crystal structure of HDAC homolog HDLP with a novel hydroxamic group based inhibitor at 2Å resolution.
Manuscript under preparation

Madhumita Ghosh, **Sreejesh Shanker**, Igor Siwanowicz, Karlheinz Mann, Werner Machleidt, and Tad A. Holak
Proteolysis of insulin-like growth factor binding proteins (IGFBPs) by calpain.
*Biological Chemistry, **386**: 85-93, 2005*

Michael Brüggert, Till Rehm, **Sreejesh Shanker**, Julia Georgescu & Tad A. Holak. A novel medium for expression of proteins selectively labeled with 15N-amino acids in *Spodoptera frugiperda* (Sf9) insect cells.
Journal of Biomolecular NMR, **25:** 335–348, 2003.

INDEX

1 Introduction

Cancers which are the result of successive accumulation of genetic lesions with subsequent deregulation of cell cycle control are now righteously called “cell-cycle disease”. Since its prevention and cure is of prime importance, the aim of this thesis was to provide some structural and functional insights into some of the important regulatory proteins of the cell cycle involved in cancer and to study small molecule inhibitors as potential drugs for the future.

Scope of work

The first Part of this thesis involves structural investigations into the histone deacetylase (HDACs) family of proteins and studies with small molecule inhibitors against HDACs. Histone deacetylases (HDACs) stabilize local DNA condensation leading to silencing of tumour suppressor genes like $p21^{WAF/CIP1}$. The deregulation of HDAC recruitment to promoters appears to be one of the mechanisms by which these enzymes contribute to tumorigenesis. Histone deacetylase inhibitors (HDACI) bind to the active site of HDACs and block the gene silencing effects of HDACs. The HDACI by the fundamental mechanism of blocking HDAC activity mediate diverse role in the cells like inducing hyperacetylation, reactivativating suppressed genes and have pleiotropic cellular effects that inhibit tumour-cell growth and survival. HDACI have been shown to be selectively cytotoxic against tumor cells while leaving normal cells apparently unharmed. HDACI inhibitors thus represent a new wave of anticancer drugs and are exciting prospects for anticancer therapies. In this study we have identified a novel thiophene containing hydroxamic group based HDACI and further shown by crystallographic studies as to how it binds to the protein.

The second part of the thesis deals with formulation of two insect cell media IML 406 and IML 455 for ^{15}N labeling studies of mammalian proteins. For protein-ligand binding studies or for structure determination with NMR spectroscopy on proteins expressed in insect cells the use of selectively labeled protein samples is essential. Here

we present the complete formulations of the two insect media, IML 406 and IML 455, for the high-yield production of selectively ^{15}N-labeled proteins in insect cells. The selective labeling using the baculovirus expression vector system represents a complement or even an alternative to the bacterial expression system. Based on findings a first simple overview of the network of the amino acid metabolism in *E. coli* and insect cells focused on nitrogen could be provided. For some amino acids the expression of labeled proteins in insect cells can replace the cell-free protein expression. Preliminary ^{15}N labeling studies were also carried out on the mammalian cyclin dependent kinase (CDKs) family of proteins. CDKs are principle regulators of cell cycle progression, CDKs in addition to coordinating the progression through cell cycle receive and integrate the growth regulatory signals, and also play important role in differentiation and apoptosis.

The final part of the thesis involves proteolytic studies on insulin like growth factor binding proteins (IGFBPs) by the endopeptidase calpain. IGFBPs modulate bioavailability of IGFs both in circulation and in cellular microenvironment, besides they have some important IGF independent actions that are not yet completely understood. Normal cells are kept under a strict vigil by the cell cycle checkpoints. Extracellular growth factors like insulin like growth factors (IGFs) play a key role in cellular division and proliferation. Deregulation of these survival factors result in uncontrolled cell growth and suppressed apoptosis, which constitute the first step in neoplastic development. Calpains are Ca^{2+}-dependent cysteine proteases, which play a role in the cell cycle, specifically in the G_1 to S phase transition. Rules that govern calpain cleavage specificity are poorly understood. The study was carried out with the purpose of understanding the cleavage pattern of calpains, we performed calpain induced proteolytic studies on the insulin-like growth factor binding proteins (IGFBP-4 and 5). IGFBPs were been found to be cleaved in the unstructured solvent exposed regions of the protein.

The work presented here is original and shows preliminary results in the field of cancer therapeutics. Laboratory work was based on the biochemistry and molecular biology techniques. Further biophysical methods such as X-ray crystallography, NMR and isothermal titration calorimetry were employed for structural and biochemical studies.

1.1 Cell cycle and chromatin

Cells perpetuate by duplicating their contents and then dividing into two and this cycle of cell division forms the basis of propagation of all living things. The fundamental task of the cell cycle is to ensure that DNA is faithfully replicated and identical chromosomal copies are distributed equally among the daughter cells (Sherr 1996). In normal cells the cell cycle is a tightly regulated and carefully balanced process through which one cell divides into two. The cell cycle can be divided into four phases: G_1 (gap 1), S (synthesis), G_2 (gap 2) and M (mitosis) phase (Fig.1.1). The S and M phases reflect stages in cell cycle progression where DNA synthesis and replication and mitosis occur in a temporally regulated fashion, separated by the two gap phases (McLaughlin et al. 2003).

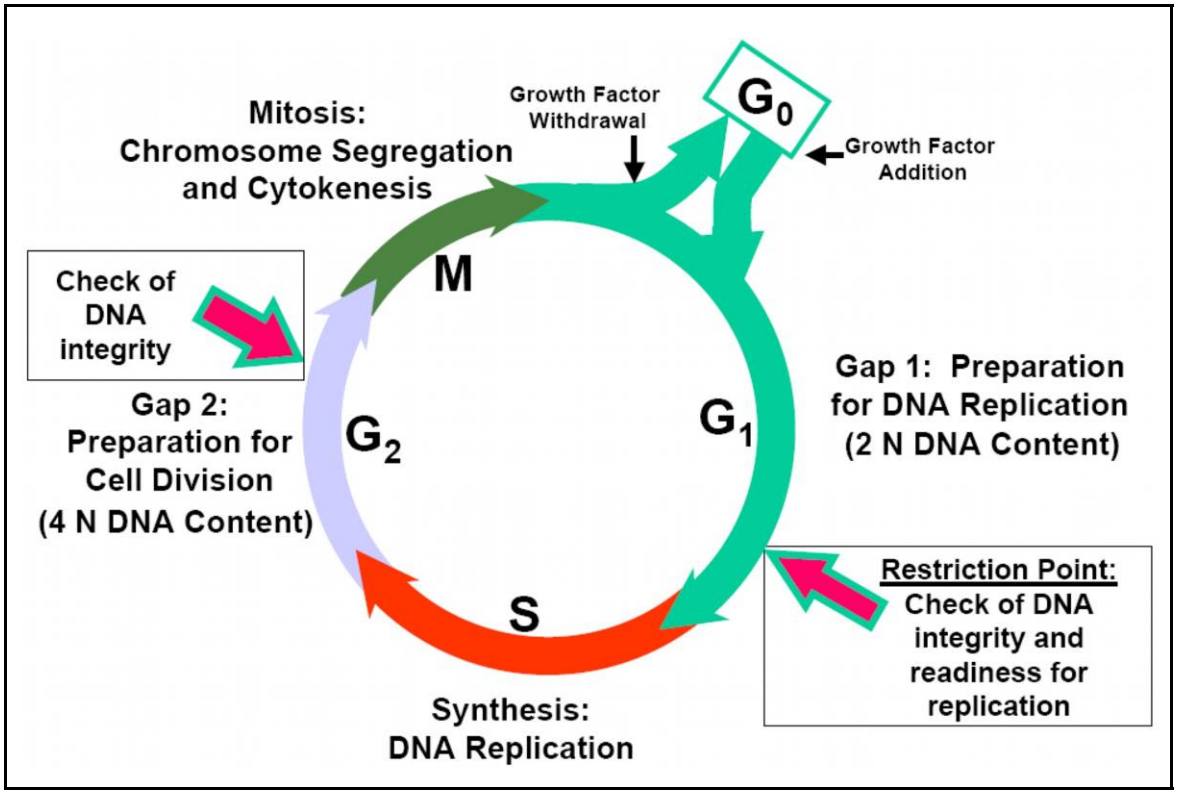


Figure 1.1 Phases of the cell cycle. Schematic sketch of the cell cycle, showing the various phases and checkpoints and also outlining the activities of each stage.

During G_1 phase, cells are responsive to extracellular mitogenic signals and respond by either advancing towards another division or withdrawing into a resting state (G_0). In G_1 the histones necessary for the formation of new chromatin are synthesized, and

the cell is metabolically active and continually grows but does not replicate its DNA. G_1 phase is followed by the S phase, during which chromosomes are faithfully duplicated. The completion of chromosome duplication is followed by the G_2 phase, during which the cell growth continues and proteins are synthesized in preparation for mitosis (Sherr 1994a; 1996; Grana et al. 1998). The fidelity of cell cycle progression is maintained by an array of regulatory decision points such as the restriction point (R), which is operationally defined within G_1, as a point of control that determines when a cell becomes committed to cell cycle progression (Carpentier et al. 1987; Dyson 1994; Sherr and Roberts 1999) and this control point is often deregulated in tumor cells. Other control points act in either a positive or negative fashion, to delay or allow completion of the cycle (Sherr and Roberts 1999). In addition, a series of surveillance mechanisms, known as checkpoints, assess the integrity of cell cycle progression (McLaughlin et al. 2003). Passage through the restriction point and entry into S phase is controlled partly by cyclin-dependent protein kinases (CDKs) that are sequentially regulated by cyclins D, E, and A. In general, CDK activity requires cyclin binding, which depends on both positive and negative regulatory phosphorylations (Norbury et al. 1992) and can be constrained by at least two families of CDK inhibitory proteins (Elledge and Harper 1994). D-type cyclins act as growth factor sensors, with their expression depending more on extra cellular cues than on the cell's position in the cycle (Sherr 1994b). The regulation of the cell cycle is integrated with gene expression and chromatin control. Transcription of certain genes, such as those involved in DNA synthesis and replication, is activated as cells approach S phase. This provides a point of control in regulating the cell cycle, and is achieved through pathways that connect the cell cycle machinery with processes, such as chromatin control (McLaughlin et al. 2003). Transcription in eukaryotic cells is influenced by the manner in which DNA is packaged. In quiescent cells, DNA is tightly compacted to prevent accessibility of transcription factors. DNA is packaged into chromatin, a highly organized and dynamic protein–DNA complex. The fundamental subunit of chromatin, the nucleosome, is composed of an octamer of four core histones, an H3/H4 tetramer and two H2A/H2B dimers, surrounded by 146bp of DNA (Ito 2000; Strahl-Bolsinger et al. 1997) (Fig.1.2 A). These proteins are basic and highly conserved throughout evolution. Histone H1 is important in determining the level of DNA condensation (Marks et al. 2001).

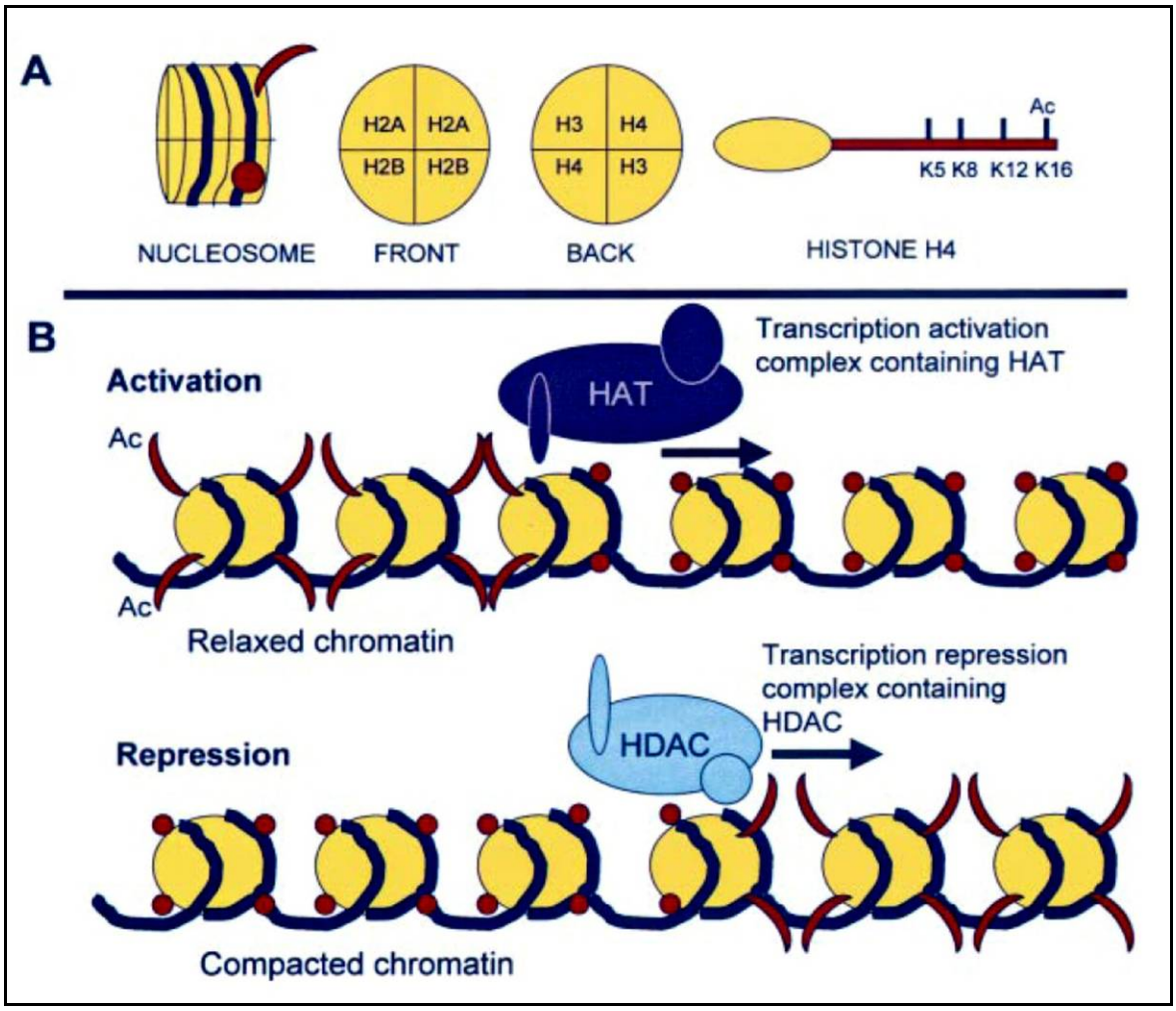


Figure 1.2 (A) Schematic representation of a nucleosome. Yellow represents the histones. Dark red depicts the histone tail that can be modified to loosen DNA (purple) winding. The dark red circle represents a tail without an acetyl (Ac) group. The dark red ` banana shape ' represents a histone tail with an acetyl group, relieving the tight packing of the DNA. **(B) Transcriptional repression and activation in chromatin**. Yellow circles represent core histone octamers; in the upper panel, acetylated histone tails (dark red) are depicted emerging from the octamer. DNA is purple, and the solid black arrow represents complex movement. Both histone acetyltransferase (HAT; activation) and HDAC (repression) require several cofactors (for DNA binding, for recruitment of the complex, for remodeling of the DNA helix to reduce the accessibility of transcription factors) for their activity (de Ruijter et al. 2003).

Local chromatin architecture plays an important role in the regulation of gene expression. During activation of gene transcription, this compact, inaccessible DNA is made available to DNA binding proteins via modification of the nucleosome (Ito et al.

2000). This architecture of chromatin is strongly influenced by post-translational modifications of the histones. All four-core histones have an amino-terminal tail, which is lysine rich and contains about half of the positively charged residues and most of the post-translational modification sites. The N-terminal tail of the histones passes through and around the enveloping DNA double helix (Marks et al. 2001). The modification of the structure of these N-terminal tails by methylation, phosphorylation, and acetylation/deacetylation, is crucial in modulating gene expression, as it affects the interaction of DNA with transcription-regulatory non-nucleosomal protein complexes. The acetylation of core histones is probably the best understood type of modification. Histone acetylation occurs at the ε amino groups of evolutionarily conserved lysine residues located at the N-termini of core histones. All core histones are seen to be acetylated *in vivo* (de Ruijter et al. 2003). Modifications of histones H3 and H4 are, however, much more extensively characterized than those of H2A and H2B (Wade, 2001). Important positions for acetylation are Lys^{9} and Lys^{14} on histone H3, and Lys^{5}, Lys^{8}, Lys^{12} and Lys^{16} on histone H4 (Bjerling et al. 2002). Steady-state levels of acetylation of the core histones result from the balance between the opposing activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs). In general, increased levels of histone acetylation (hyperacetylation) are associated with increased transcriptional activity, whereas decreased levels of acetylation (hypoacetylation) are associated with transcriptinal scilencing (Ito et al. 2000; Forsberg and Bresnick 2001; Wade 2001) (Fig.1.2 B).

1.2 Histone deacetylases

1.2.1 Classification of the HDAC family

Histone deacetylases (HDACs) catalyse the removal of the acetyl moiety from acetyl-lysine within core histones, thereby stabilizing local condensation of the DNA with concomitant downregulation of gene expression. The deregulation of HDAC recruitment to promoters appears to be one of the mechanisms by which these enzymes contribute to tumorigenesis (Finnin et al. 1999). There are two protein families with HDAC activity: the recently discovered SIR2 family of NAD^{+} dependent HDACs and the classical HDAC

family. Members of the classical HDAC family fall into two different phylogenetic classes, namely class I and class II (Fig. 1.3). The class I HDACs (HDAC -1, 2, 3 and 8) are most closely related to the yeast (*Saccharomyces cerevisiae*) transcriptional regulator RPD 3. Class II HDACs (HDAC -4, 5, 6, 7, 9 and 10) share domains with similarity to HDA1, another deacetylase found in yeast (Bjerling et al. 2002). Recently a new member of the HDAC family has been identified, HDAC11. This protein contains all the necessary features to be designated as a HDAC, but it remains to be classified (Gao et al. 2002). Members of both classes share a conserved 390 amino acid region of homology known as the deacetylase core. The deacetylase core belongs to a super family of genes that includes the *A. aeolicus* HDAC homologue HDLP (35.2% identity with HDAC 1) (Finnin et al. 1999) HDACs of class I are expressed in most cell types, but the expression pattern of class II HDACs is more restricted suggesting that they might be involved in cellular differentiation and developmental processes. Class I HDACs are found almost exclusively in the nucleus. The localization of HDAC1 and HDAC2 is exclusively nuclear, due to the lack of a nuclear export signal (NES). HDAC3, however, has both a nuclear import signal and a NES, suggesting that HDAC 3 can also localize to the cytoplasm. Class II HDACs are actively maintained in the cytoplasm and transported to nucleus only when required. The Class II HDACs also possess a more extensive N-terminus. Both classes interact with different set of co-repressors (Fischle et al. 2002; Yang et al. 2002).

1.2.2 HDAC complexes

HDAC1 and HDAC2 are the best characterized of the HDAC proteins and are generally found in stable, multi-component complexes of proteins, which are then recruited by DNA binding proteins. Three complexes containing HDAC1 and HDAC2 have been characterized so far, the Sin3, NuRD, and CoREST complexes (Fig. 1.4). A "core complex," consisting of HDAC -1/ 2 and the histone binding proteins RbAp46/48, this complex stably associates both *in vivo* and *in vitro.* The Sin3 complex consists of this core complex, along with SAP18 and SAP30, which appear to aid in stabilizing the protein associations and mSin3A, which serves as a scaffold for the assembly of the complex and its interaction with various DNA binding proteins. The NuRD complex (nucleosomal

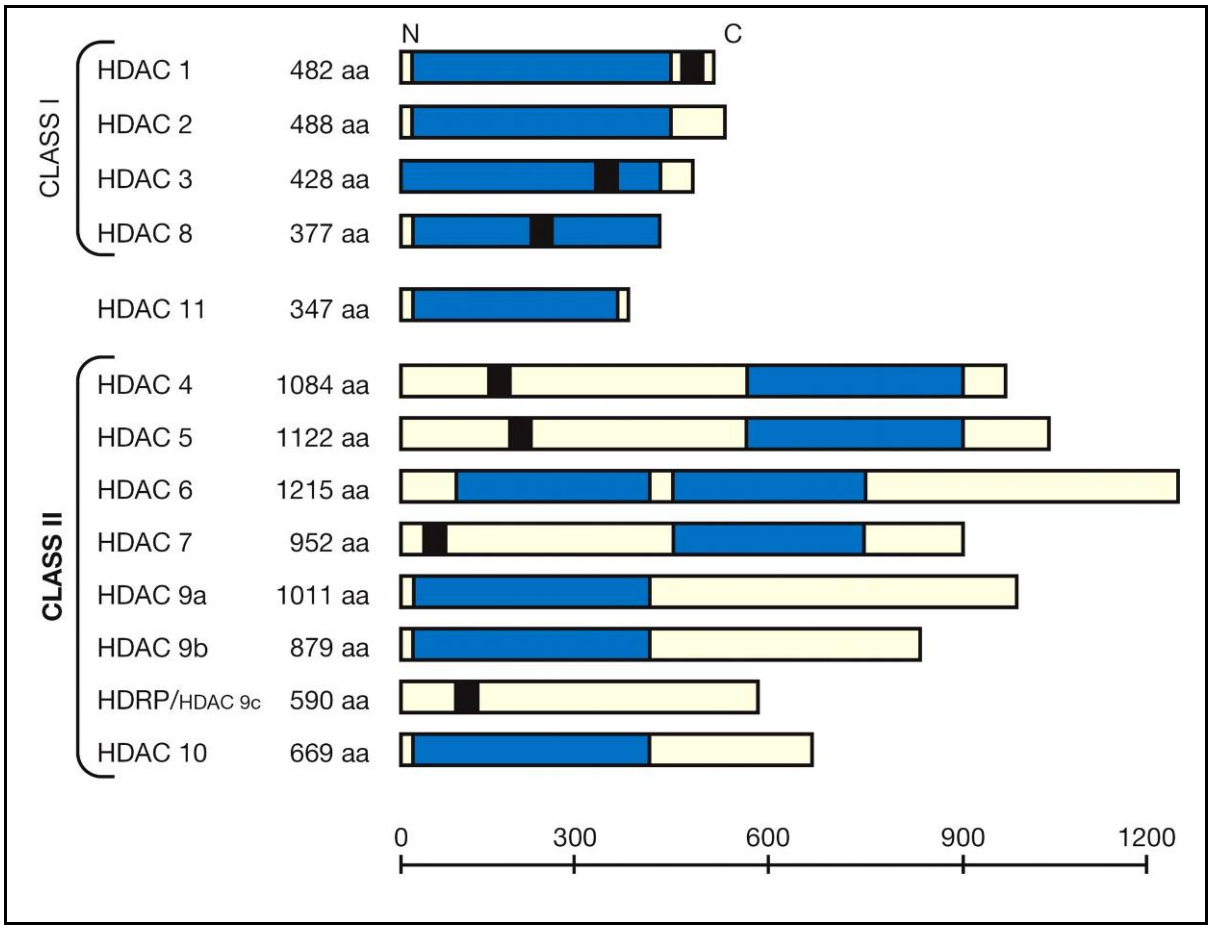


Figure 1.3 Schematic representation of the different isoforms of HDAC. Bars depict the length of the protein. The catalytic domain is shown in blue. Note that HDRP does not possess any deacetylase activity. Black depicts a NLS. N, N-terminus, C, C-terminus.

remodeling and deacetylation) also contain the core complex, as well as MTA2, which is related to a protein (MTA1) found to be over expressed in metastatic tumor cells, and CHD3 and CHD4 (also called Mi-2^{α} and Mi-2^{β}), which possess the DNA helicase/ATPase domains found in the SWI/SNF family of chromatin remodeling proteins (Grozinger and Schreiber 2002). The coupling of chromatin remodeling and deacetylase activity may play an important role in the deacetylation of chromatinized histones in the cell. Although histones can be deacetylated by HDAC1 in the absence of ATP, but ATP hydrolysis, and presumably chromatin remodeling, is required for the deacetylation of oligonucleosomes in vitro (Tong et al. 1998; Zhang and Raveche 1998). Unlike the other two complexes, the

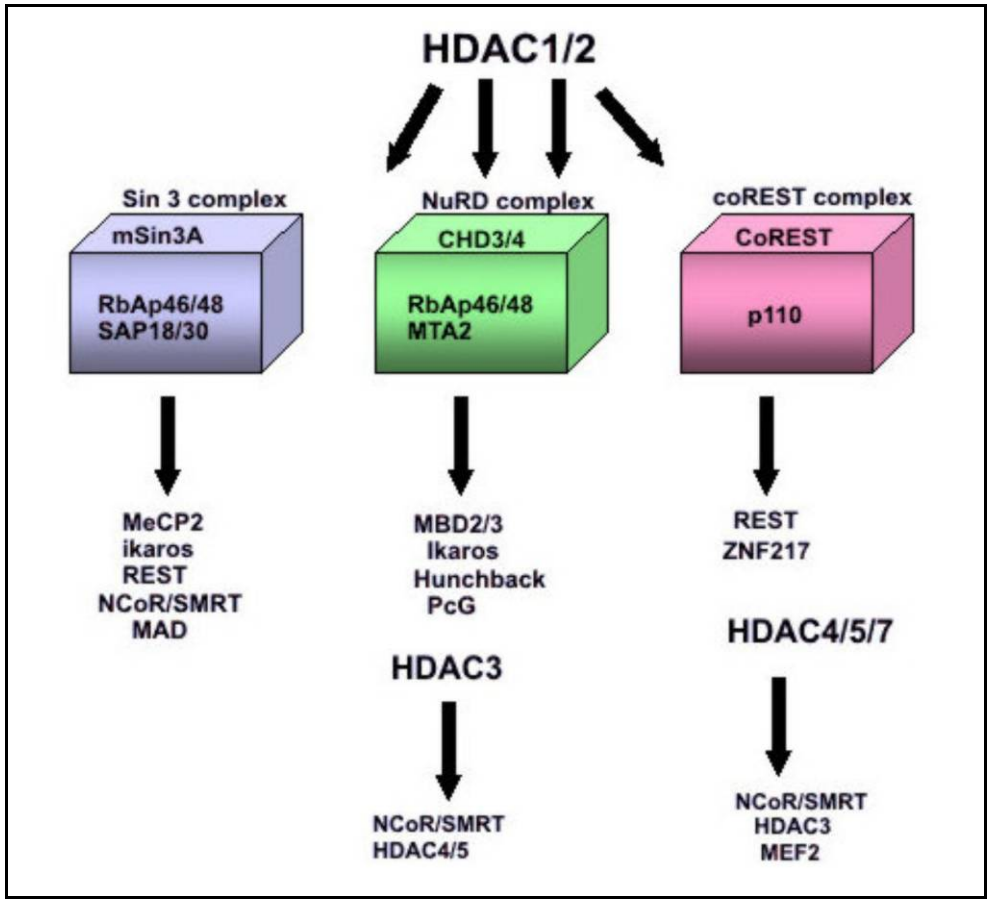


Figure 1.4 Protein Complexes of HDAC Deacetylases. HDAC1 and 2 are found in three main complexes, the Sin3, NuRD, and CoREST complexes, which are recruited to promoters or DNA binding domains by various proteins. HDAC1 and 2 can also be recruited directly by certain transcription factors. HDAC3, 4, 5, and 7 directly associate with the nuclear corepressors NCoR and SMRT, and with each other. HDAC4, 5, and 7 can also interact directly with members of the MEF2 family of transcription factors.

CoREST complex contains HDAC1 and HDAC2 but, does not appear to include RbAp46 or RbAp48. It also contains a protein homologous to MTA1 and MTA2, termed CoREST which is competent to recruit HDAC 1/ 2 and silence expression from a reporter gene. CoREST contains two 50 amino acid SANT domains, which resemble the DNA binding domains of Myb-related DNA binding proteins. MTA1 and MTA2 as do several other proteins involved in transcriptional regulation. Such proteins include SWI3, ADA3, NCoR and TFIIIB, for which SANT is named (Pijnappel et al. 2001). These SANT domains may

mediate critical protein-protein interactions. Indeed, the N-terminal SANT domain of CoREST is required for the recruitment of HDAC 1 to promoters (You et al. 2001).

1.2.3 Histone deacetylases and cancer

Histone deacetylases are involved in various forms of tumorogenisis. In general, there are three separate known levels that are dependent on histone deacetylases and are likely linked to tumorigenesis. On the first level are the cell cycle restraining transcriptional repressors including Mad and Rb. On a second level are transcriptional repressors that normally block the process of differentiation in certain cellular lineages, for example, the nuclear hormone receptors such as RARα in hematopoietic differentiation. A final level controlled by histone deacetylase containing complexes is the strong correlation between genomic methylation and the transcriptional silencing of various tumor-suppressor genes, such as $p21^{WAF/CIP1}$ (Cress and Seto 2000).

1.2.3.1 Disruption of the Mad/Sin3/HDAC complex

In growing cells, Myc-Max heterodimers activate transcription of growth stimulatory genes, such as E2F, that are regulated by E-box elements in their promoters (Sears et al., 1997; McArthur et al., 1998). However, upon cellular differentiation Myc is replaced by Mad and the Mad/Max heterodimer represses transcription of growth stimulatory genes. It is known that Mad recruits Sin3/HDAC complex to carry out this repression. Disruption of this Mad /Sin3/HDAC complex by an oncogene v-Ski (Li et al. 1996) releases the transcriptional repression by Max. The normal c-Ski is found as a part of the Sin3a complex and helps bind Sin3 to Ncor/HDAC subunit, v-Ski disrupts this complex association and thus releases the transcriptional repression by Max**.**

1.2.3.2 Histone deacetylase interact with retinoblastoma tumour suppressor protein (Rb)

The Rb protein is a critical regulator of cell growth it acts by repressing the transcription of those genes whose products are needed for cell cycle progression. The means by which pRb exerts control over cell cycle progression is thought to be its

association with the E2F family of transcription factors (Dyson, 1998; Nevins, 1998). This regulation occurs by one of two mechanisms:

- pRb can bind to the transcriptional activation domain of E2F, thereby blocking its activation and
- transcription from promoters containing E2F sites can be actively repressed by complexes containing E2F, Rb, and HDAC, SWI/SNF family members.

E2F regulated promoters are involved in two activities. Expression of enzymes required for DNA synthesis and expression of cell cycle regulatory proteins like cyclin A and E. The RB/E2F transcription regulatory pathway is found disrupted in nearly all human tumors (Sellers et al.1997; Weintraub 1995**;** Luo et al. 1998).

1.2.3.3 Histone deacetylases and interaction with fusion proteins in leaukemia

Acute leukemia results when the differentiation of immature hematopoietic cells is blocked and the cells proliferate without restraint. Acute promyelocytic leukemia (APL) is most often associated with chromosomal translocation t(15;17), which fuses a large portion of the retinoic acid receptor α (RARα) to the coding sequence of a second gene, PML (promyelocytic leukemia). In a smaller percentage of APL cases, a similar translocation t(11;17) occurs in which the same coding region of RARα is fused to a different protein, PLZF (promyelocytic leukemia zinc finger). PLZF normally plays a role in central nervous system development and in hematopoiesis. RARα also plays a role in hematopoiesis. Specifically, RARα binds to DNA as a heterodimer with an RXR (retinoid-X receptor) protein. In the absence of retinoic acid, which normally induces differentiation of promyelocytic cells, the RARα/RXR heterodimer binds to a transcriptional corepressor that contains NCoR, Sin3A, and histone deacetylase (Grignani et al. 1998). In the presence of retinoic acid, the co-repressor complex is displaced and is replaced by a co-activator complex that contains the histone acetyltransferases p300/CBP and PCAF. However, when the RARα protein is present in either of these fusion proteins it is no longer responsive to physiological levels of retinoic acid. Thus, under these conditions, RARα

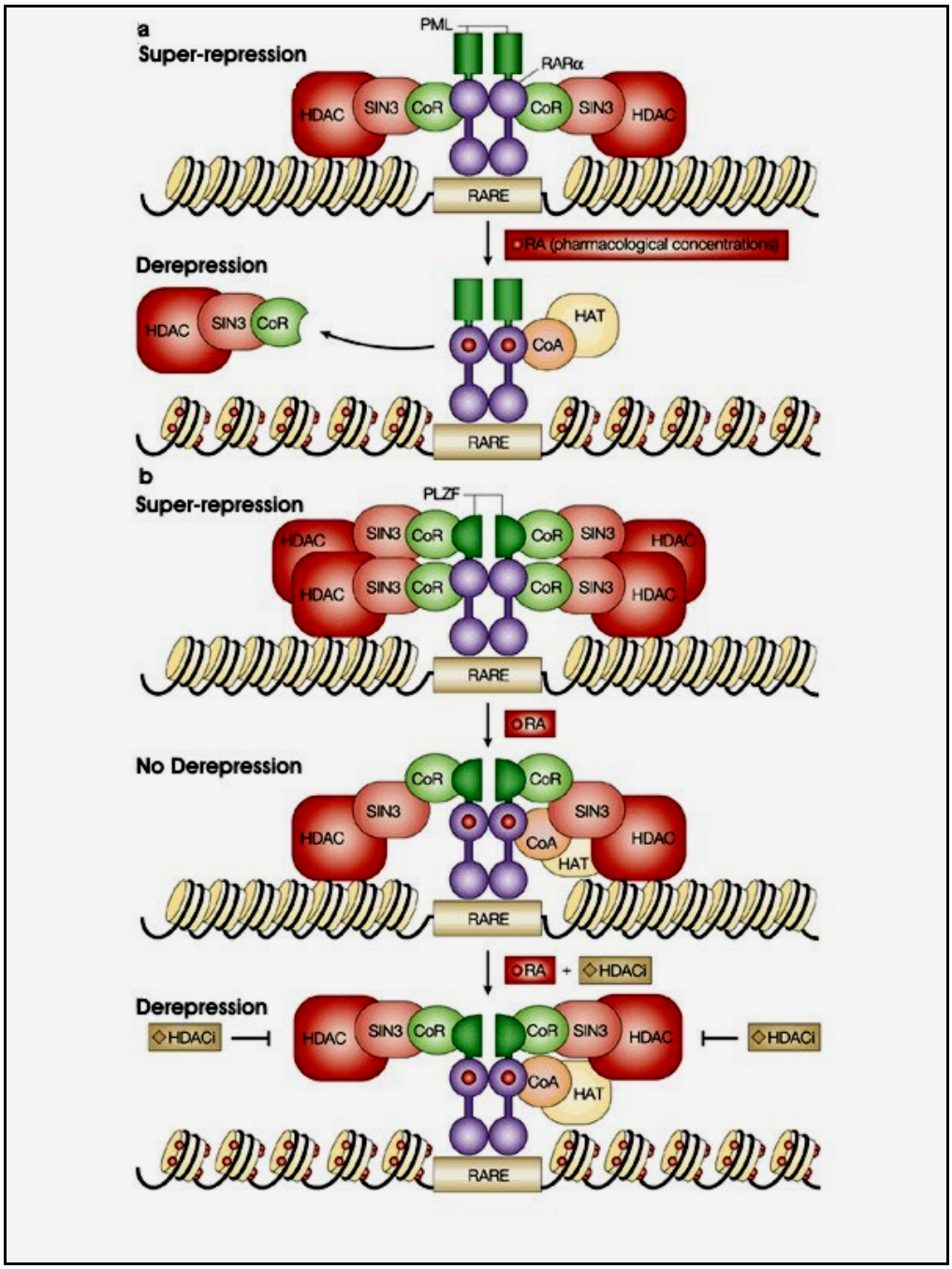


Figure 1.5 Molecular basis of HDAC repression in APL cells. (a) The t(15;17) chromosomal translocation generates the PML–RARα fusion protein that, recruits multiple HDAC complexes, leading to 'super-repression' of target genes. To disrupt this complex, pharmacological concentrations of retinoids are required; this leads to recruitment of co-activator (CoA)–HAT complexes and gene expression. (**b)** The t(11;17)translocation generates the PLZF–RARα fusion protein, which contains two CoR interaction surfaces. The one in the PLZF portion of the molecule cannot be dissociated by retinoids, and additional HDAC inhibitor (HDACI) is needed to relive the repression.

becomes a constitutive transcriptional repressor, which blocks normal differentiation and leads to leukemia.The PML-RARα fusion protein is responsive to pharmacological levels of retinoic acid but PLZF-RARα fusion protein is completely insensitive to retinoic acid as PLZF protein itself contains an interaction domain that can bind to the NCoR/Sin3A/HDAC complex. Thus, the fusion protein interacts with corepressors through two domains, one of which is not sensitive to retinoic acid and in such a case treatment is possible only by giving both retinoic acid and histone deacetylase inhibitor (described below) (Fig. 1.5).

1.2.4 Histone deacetylase inhibitors (HDACI)

HDACI represent a structurally diverse class of HDAC inhibitors produced naturally or as synthetic products which have an ability to block HDAC mediated gene silencing. These inhibitors act by binding to the active site of HDACs and thus disrupting the ability of HDACs to bind to their natural acetyl lysine substrate. HDACI by the fundamental mechanism of blocking HDAC activity mediate diverse role in the cells like inducing hyperacetylation, reactivativating suppressed genes and have pleiotropic cellular effects that inhibit tumour-cell growth and survival. HDAC inhibitors thus represent a new wave of anticancer drugs and are exciting prospects for for anticancer therapies. In this study we have identified a novel hydroxamic group based HDACI which has the potential to become a future drug.

1.2.4.1 Classes of HDACI

HDAC inhibitors have been subdivided into 5 classes:

Small molecule hydroxamic inhibitors: The general structure of these substances consists of a hydrophobic linker that allows the hydroxamic acid moiety to chelate the cation at the bottom of the HDAC catalytic pocket, while the bulky part of the molecule acts as a cap for the tube. Most of the chemicals in this group are very potent (functioning at nanomolar to micromolar concentrations *in vitro*) and reversible inhibitors of class I/II HDACs (Villar-Garea and Esteller 2004), some of the best known HDACI come under this class e.g. Trichostatin A (TSA), Suberoyl anilide hydroxamic acid (SAHA) (Fig. 1.6 [1, 2]). The design of many synthetic drugs has been inspired by the TSA structure.

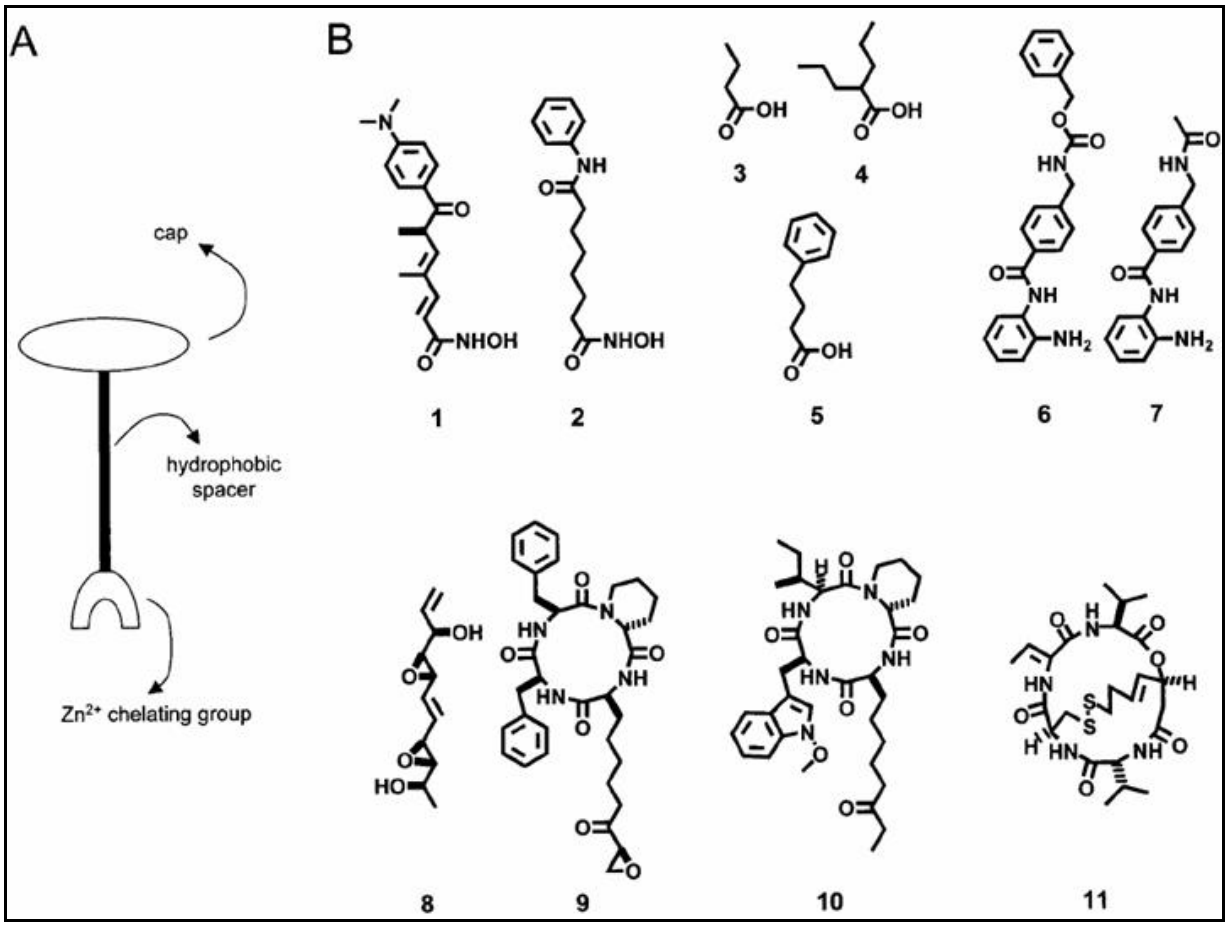


Figure 1.6 Compounds with proven HDAC-inhibitory activity. (a) Scheme showing the general structure of most of these substances. **(b)**Examples of chemicals included in the different HDAC inhibitor groups (see text): TSA (1) and SAHA (2) are hydroxamic acids; butanoic acid(3), valproic acid (4) and 4-phenylbutanoic acid (5) are carboxylic acids; MS-275 (6) and N-acetyldinaline (7) are benzamides; and depeudecin(8) and trapoxin A (9) are epoxides; also shown are apicidin (10) and depsipeptide FK228 (11).

Carboxylic acids: There are only a few drugs in this group: butanoic, valproic and 4-phenylbutanoic acids (Fig. 1.6 [3, 4, and 5]). Despite being much less potent than the hydroxamic acids (inhibition occurs *in vitro* at millimolar concentrations and at very high doses *in vivo*) and their pleiotropic effects, these are currently among the best studied HDAC inhibitors: valproic acid and phenylbutyrate have already been approved for use in treating epilepsy.

Benzamides: The benzamide class is generally less potent than the previous two classes mentioned (Miller et al. 2003). The exact mechanism by which benzamides eg. MS-275, N-acetyldinaline (Fig. 1.6 [6, 7]), exert their antiproliferative activity is yet to be elucidated. It is believed that the diaminophenyl group is very important for the inhibitory behavior; probably, both amino functionalities chelate the metallic ion in the catalytic site. (Marks et al. 2004).

Epoxides: These chemicals are supposed to trap HDACs through the reaction of the epoxide moiety with the zinc cation or an amino acid (forming a covalent attachment) in the binding pocket. However, the lability of the epoxide functionality prevents significant *in vivo* activity, which makes them of little pharmacologic interest. The only HDAC inhibitors in this set of compounds are a number of natural products with significant *in vitro* activity, such as depeudecin and trapoxin A (Fig. 1.6 [8, 9]). and some cancers, respectively, whereas butanoic acid (or its prodrug forms, such as pivaloyloxymethylbutyrate) is undergoing clinical trials (Gore et al. 2002).

Electrophilic ketones: Electrophilic ketones have been shown to inhibit HDACs. The first member of this new class of inhibitors was trifluromethyl ketone. They show inhibition at low micromolar to high nanomolar range. These agents also possess antiproliferative activities in vitro and induce histone hyper acetylation and p21 gene expression (Miller et al. 2003).

1.2.4.2 Mechanisms of anticancer action of HDACI

Depending on the cell type and the concentration of the compound used HDACI can mediate a diverse range of intrinsic effects on cell growth and survival.

HDACI mediate anticancer effect through altered gene expression: HDACI are thought to induce their effects through the activation and/or repression of genes following hyperacetylation of histones and chromatin remodeling (Lindemann et al. 2004). The total number of genes affected by HDACI ranges from 2 to17% depending on the HDACI used in the study. A significant number of apoptotic (i.e., CD95, CD95 ligand, gesolin, Bak,

Bcl-X1) and cell cycle regulatory genes (cyclin D; cyclin E; cyclin A; $p21^{waf1/cip1}$) have been identified as being activated or repressed by HDACI (Lindemann et al. 2004).

Cell cycle arrest by HDACI : Nearly all HDACI studied are capable of inducing cell cycle arrest at the G_1/S transition (Johnstone 2002; Marks et al. 2000; Gabrielli et al. 2002). The CDK inhibitor $p21^{waf1/cip1}$ was one of the first target genes identified (Richon et al. 2000). The activation of $p21^{waf1/cip1}$occurs in a p53 independent manner (Huang et al. 2000) and is necessary for the induction of G_1 arrest by HDACI (Sandor et al. 2000) (Burgess et al. 2001). The up regulation of $p21^{waf1/cip1}$ may not be solely responsible for the G_1 arrest, as repression of cyclin D and cyclin A is also likely to contribute to the loss of CDK 2 and CDK 4 kinase activities and presence of hypophosphorylated Rb (Sandor et al. 2000). Two genes involved in DNA synthesis, CTP synthetase and thymidylate kinase are also transcriptionally repressed by HDACI (Glaser et al. 2003) and also lead to G_I/S arrest.

- **Induction of apoptosis by HDACI:** HDACI results in induction of a large number of proapoptotic genes and repression of anti-apoptotic genes although the exact molecular pathways, which are essential for induced apoptosis, have not been elucidated (Lindemann et al., 2004). Some candidate genes, which may contribute to HDACI mediated cell toxicity, are CD95, Bcl-2 and $p21^{waf1/cip1}$
- **HDACI target cell cycle checkpoint:** $p21^{waf1/cip1}$ is responsible for the HDACI mediated G_1/S arrest and as stated above occur at much lower doses. HDACI mediated G_2 phase arrest has been detected in comparatively restricted number of cell lines and requires higher HDACI doses. Unlike cells which undergo apoptosis at higher doses of HDACI these cells seem to be resistant to the cytotoxic affects of these drugs: This G_2 arrest is due to the HDACI sensitive G_2 checkpoint that operates through the caffeine sensitive ATM/ATR checkpoint signaling pathway possibly via Chk2 pathway (Qiu et al. 2000). The cells that have lost the G_2 checkpoint result in their entering to mitosis with normal kinetics but undergo aberrant mitosis with the condensed

chromosomes unable to separate at metaphase (Qiu et al. 2000; Krauer et al. 2004).

HDACI target important nonhistone proteins*:* HDACI also target non-histone protein including transcription factors (e.g. p53, GATA1, RelA) nuclear import proteins (e.g. importin-α7), signal transduction molecules (e.g., β-catenin), cytoskeletal proteins (e.g. α-tubulin) and DNA repair enzymes (e.g. Ku70, WRN) can be regulated by acetylation (Chen et al. 2001; Cohen et al. 2004; Krauer et al. 2004; Kouzarides 1999; Wolffe et al. 2000). HDACI-induced hyperacetylation of Ku70, results in release of Bax and the sensitization of the cells to apoptosis may prove to be one of the most important non-histone effects of HDACI (Cohen et al., 2004). In its hypoacetylated form Ku79 binds to Bax and sequesters it from the mitochondria thereby raising the cellular apoptotic threshold. Upon addition of HDACI, Ku70 releases Bax thus lowering the apoptotic threshold. The significance of these results maybe related to the ability of HDACI to synergise with a host of pro-apoptotic anti cancer compounds that have diverse molecular targets (Lindemann et al. 2004).

1.2.4.3 Tumor–selectivity of HDACI

HDACI have the remarkable property of being selectively cytotoxic against tumor cells while leaving normal cells apparently unharmed, Although HDACI induce equivalent histone hyperacetylation (Kitazono et al. 2002; Brinkmann et al. 2001; Qiu et al. 2000) and induction of $p21^{waf1/cip1}$ in both normal and tumor cells; while normal cells undergo cell cycle arrest the tumour cells undergo apoptosis.in response to HDACI. (Lindemann et al. 2004). However it is known that there are fundamental epigenetic differences between normal and tumor cells that could alter the transcriptional response to HDACI. Few hypotheses have been given to understand this differential control. First while HDACI potentially activate a number of genes, but those genes that contain hypermethylated promoter regions are unresponsive to HDACI due to the dominant gene silencing effect of CpG methylation (El-Osta et al. 2002). A number of important antiproliferative and proapoptotic genes such as those for p14ARF, $p15^{INK4b}$, $p16^{INK4A}$, RB, caspase-8 and Apaf1 proteins are hypermethylated (Rountree et al. 2001; Esteller 2003) and hence

unresponsive to HDACI, but malignant cells have genomes which are hypomethylated and these cells therefore would be hyper responsive to inhibition by HDACI. Secondly the functional status of control points and checkpoints may also underlie the tumor selective cytotoxicity of HDACI. As functional loss of such points is fundamental requirement for tumorigenesis and cells that lose their ability to up regulate $p21^{waf1/cip1}$ and arrest in G_1/S are clearly sensitized to HDACI-induced apoptosis (Lindemann et al. 2004).Finally recent studies have shown that multiple genes with specific molecular pathways such as IGF/IGF-IR/AKT, TNF/TNF-R7NF-κB and CD95R/caspase-8/FLIP signal transduction pathways., the Apaf-1/caspase-9/surviving apoptotic pathway, and the ubiquitin /proteosome protein degradation pathway (Mitsiades et al. 2004) are coordinately regulated by the HDACI. As many of these pathways function aberrantly in a range of tumors but are effectively regulated in normal cells, this provides some insight into the tumor specificity of these drugs and why they are so effective against a diverse set of solid and hematological tumors.

1.3 Cyclins and cyclin dependent kinases

Cyclins and cyclin dependent kinases are the principle regulators of cell cycle progression, through activation of cell cycle checkpoints. Orderly transition between cell cycle phases requires the scheduled activity of CDKs, controlled in part by their association with cyclins and endogenous CDK inhibitors and also partly governed by their phosphorylation state. CDKs in addition to coordinating the progression through cell cycle receive and integrate the growth regulatory signals that are targeted to the cell cycle. Here these signals control the cell cycle by turning the CDK switches on or off (Pavletich 1999). In cancerous cells, aberrant expression of CDKs, cyclins and endogenous CDK inhibitors, results in deregulated CDK activity, thus providing a selective growth advantage. Because of their critical role in cell cycle progression, as well as the association of their activities with the processes of differentiation and apoptosis, the CDKs comprise an attractive set of targets for novel antineoplastics (Shapiro et al. 1998).

1.3.1 Structural features of CDKs

The structure of monomeric CDK2 has the same overall fold as other eukaryotic protein kinases (PKA; Knighton et al. 1991). The structure consists of a small N-terminal lobe dominated by β-sheet and the large PSTAIRE helix, a larger C-terminal lobe which is primarily helical, and a deep cleft at the junction of the two lobes that is the site of ATP binding residue (Fig.1.7). In the monomeric CDK 2 structure, two regions differed from the canonical kinase structure and were predicted to function as regulatory elements (De Bond et al. 1993). One is an α-helix present in other protein kinases, but having the unique sequence PSTAIRE (one letter amino acid code) found only in cyclin-dependent kinases; the other is a regulatory loop that, like most other eukaryotic kinases has the activating phosphorylation site.

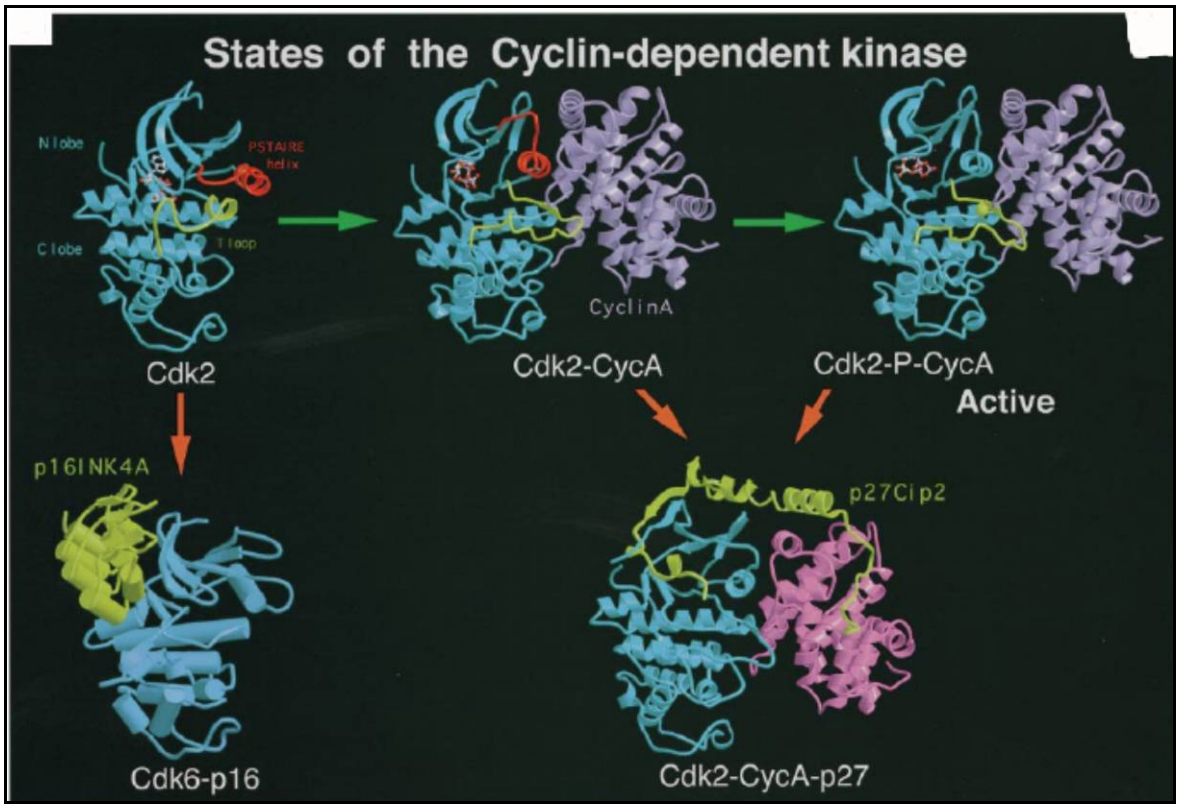


Figure 1.7 Cyclin-dependent kinase structures corresponding to non-activated (monomeric Cdk2), partially active (Cdk2-cyclinA complex), fully active (phosphorylated Cdk2-cyclinA complex), inhibited (p27-Cdk2-cyclinA complex), and non-activatable (p16-Cdk6 complex) states. In the monomeric Cdk2 and cyclinA-bound Cdk2 structures, the PSTAIRE helix is highlighted in red and the T loop in yellow. In the

phosphorylated Cdk2-cyclinA complex, theT loop is in yellow and the phosphate group is indicated by a yellow sphere. Where present, ATP is shown in ball and stick representation (Jeffrey et al. 1995) (Russo et al. 1996; Russo et al. 1998).

1.3.2 Mechanism of activation

Monomeric CDKs are found in an inactive form and undergo a sequential activation process that involves binding to cyclins and phosphorylation. Binding to cyclins is the first step in activation the cyclin binds to one side of the catalytic cleft and forms a continuous protein-protein interface by contacting key residues in both the lobes. Of special significance are the contacts which cyclins make with the PSTAIRE helix as this brings about conformational changes in the CDKs, thus leading to activation of CDKs. Other key contacts are made to the T loop, and parts of the N and C lobes (Jeffrey et al. 1995). The cyclin moves the PSTAIRE helix into the catalytic cleft and rotates it by a 90º angle. It also changes the T loop structure and position, moving parts of it by over 20 Å.

The functional significance of the PSTAIRE helix is that it carries on it a catalytic site residue, Glu51, conserved among eukaryotic protein kinases. In the absence of bound cyclin this glutamic acid side-chain lies outside the catalytic cleft, but is moved into the catalytic site when the cyclin binds. Here together with a lysine residue, an aspartic acid residue and a magnesium ion it coordinates the ATP phosphate atoms and correctly orients them for catalysis. The functional significance of rotation of the T loop is that prior to cyclin binding the T loop is positioned in front of the catalytic site thus blocking the access of polypeptide substrate from the ATP site. Once cyclin binds it rotates the loop by a 20º angle, thus placing this loop outside the catalytic site, which relives the blocking of the catalytic site. This positional change also exposes the phosphorylation site on the T loop whose phosphorylation is necessary for complete activity. The T loop contains the residue Thr160 (Cdk2) that is a site for phosphorylation by CDK activating kinases (CAKs). When the T loop becomes phosphorylated, it undergoes a further conformational change. This change is brought about by the phosphate group which acts as an organizing center and is bound by three arginine side-chains, each coming from a different part of the structure (one from the N lobe, one from the C lobe, and one from the T loop). The arginine residues, in turn, hydrogen bond to other CDK and cyclin groups, and extend the

organizing influence of the phosphate group. The significance of this CDK region is that it is part of the catalytic cleft, and in particular the putative polypeptide substrate interaction site. Phosphorylation thus completes the reorganization of the substrate binding site that was started by the cyclin and brings about complete activation (Pavletich 1999).

1.3.3 Mechanism of catalysis by CDK-Cyclin complex

Once the CDK-cyclin complex is formed and the phoshorylation of the T loop complete, the CDKs are fully active. The CDK catalytic core provides highly specific binding sites that allow the two substrates, ATP and protein to nestle against one another in the desired orientation allowing the γ-phosphate of ATP to face the hydroxylated side chain on the protein substrate. Catalytic residues then promote the otherwise unlikely transfer of the γ-phosphate to the oxygen of the hydroxyl group (Morgan 1997).

1.3.4 CDKs and cell cycle regulation

The principle CDKs responsible for G_1 progression and entrance into S phase are CDK2, CDK4, and CDK6. The D-type cyclins (D1, D2, and D3) are the regulatory subunits for CDK -4/ 6 (Fig.1.8). Cyclin E, which activates CDK2, is expressed in late G_1 and early S phase (Dulic et al. 1994; Koff et al. 1992; Ohtsubo et al. 1998). Cyclin A can also activate CDK2, but is essential later, at the G_1/S boundary and throughout S phase (Girard et al., 1991; Pagano 1997; Zindy et al. 1997).

A substrate for CDK 4 and CDK 6: The major substrate of D-type cyclin CDK 4/ 6 complexes is the retinoblastoma protein, RB which was discussed in the HDAC section. The cyclin-CDK 4/ 6 induce phosphorylation of RB thus rendering it partially inactive ie it can no longer recruit the repressor complexes containing HDAC and hence leading to activation of E2F family members which ensure that the cell progresses to the S phase. The importance of pRb phosphorylation in G1 progression is underscored by two observations. First, loss of CDK 4/6 activity only arrests cells when pRb is functional. Second, ectopic expression of the CDK 4/6 specific inhibitor $p16^{INK4a}$ requires functional pRb to impose a G_1 arrest. Thus, the first molecular event involving the actions of a CDK

required for G_1 progression is phosphorylation of RB, by D-type cyclin–CDK 4/6 complexes.

1.3.5 Sequential action of CDK4 and CDK6 and CDK2

Cyclin E lies downstream of E2F and herein lies the link between the regulation of RB–E2F complexes by D-type cyclins and the activation of CDK2 (Ohtani et al. 2004) (Zerfass-Thome et al. 1996). The initial phosphorylation of RB by CDK 4/6 partially inactivates its negative regulatory functions. This inactivating event, however, is sufficient only to facilitate cyclin E expression. This in turn allows for subsequent phosphorylation of Rb by cyclin E–CDK 2 complexes leading to its full inactivation. This results in further increased levels of E2F and cyclin E and onset to the S-phase. Soon thereafter increasing levels of CDK 2-cyclin A complexes phosphorylate E2F and inhibits its binding to DNA (Morgan 1997).

1.3.6 Endogenous inhibitors of CDKs

The activity of many CDK-Cyclin complexes is controlled by endogenous cyclin dependent kinase inhibitors (CKI). In mammalian cells, two classes of CKI are found, the Cip/Kip and Ink4 families.

Cip/Kip family: It is represented by p21, p27 and p57 which inhibit a broad range of CDKs especially the ones involved in G_1/S transition ie. CDK2 and CDK 4/6. The mechanism of inhibition by p27 is shown in Fig. 1.7. It interacts with both cyclin and CDK and inhibits by inserting a small 3_{10}-helix inside the catalytic cleft, where it mimics the ATP structure both in its position and contacts it makes. A tyrosine side chain conserved among this family members mimics the van der Waals interaction made by the ATP purine group (Pavletich 1999). p21 also plays a important role in CDK inhibition and is especially important in cell cycle, as it lies downstream of tumor suppressor p53 thus connecting the p53 and Rb pathway also HDACI are thought to upregulate p21 and bring about G_1/S arrest.

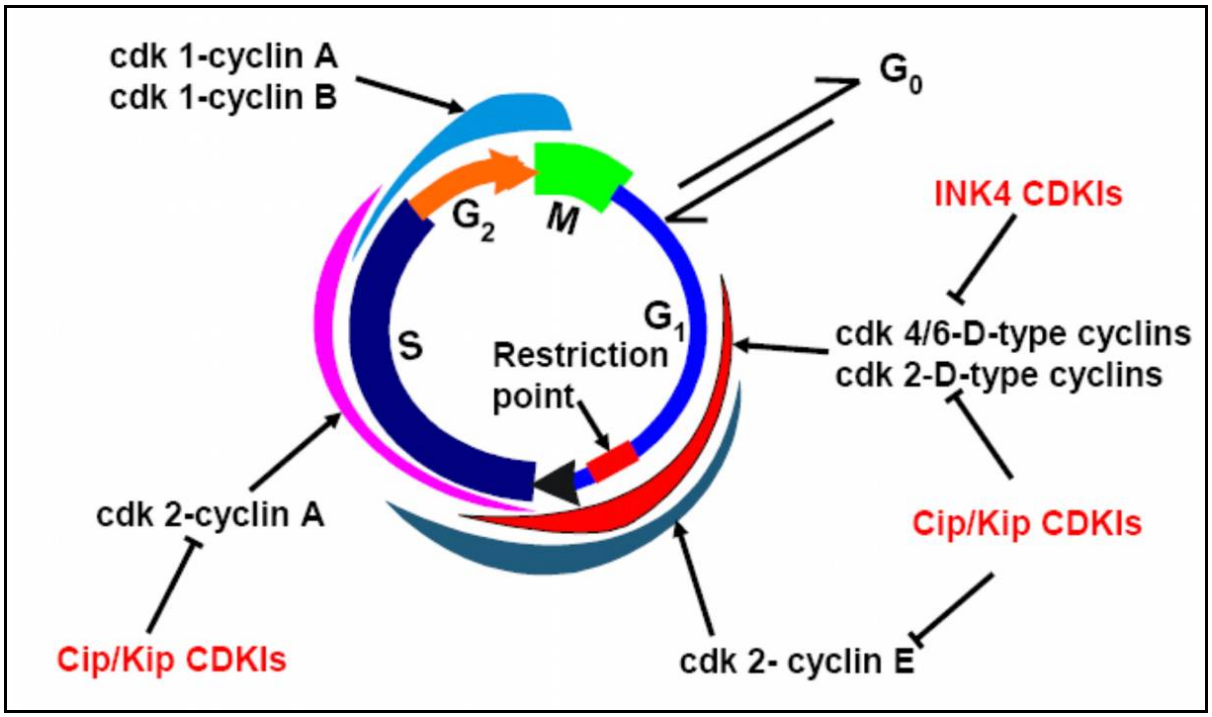


Figure 1.8 Schematic depiction of the mammalian cell cycle. Cell cycle progression through different phases of the cycle are accomplished in part by different CDK-cyclin complexes, indicated here and described in the text. Cyclin dependent kinase inhibitors (CDKIs) exert negative regulatory effects on cell cycle progression and are represented in red.

Ink4 family: Comprises of four members p15, p16, p18, and p19. They have relatively narrow specificity for CDK 4/6-cyclin D complexes. Most observations suggest that these inhibitors act by blocking the association of CDK 4/6 with cyclin D thus preventing the activation of the kinase. Structural studies have shown that the two members' p16 and p19 inhibit this association by binding to one side of the catalytic cleft, opposite to the cyclin binding site and brings about structural distortion of CDK4/6. The N and C lobes aretwisted away from each other thereby blocking translocation of PSTAIRE helix into catalytic site. It also disrupts the ATP binding site and pushes the T loop further away from desired catalytic site. The binding and mechanism of action are thus distinct from Cip/Kip family.

1.4 The insulin-like growth factor IGF system

As a group of essential cell growth modulators, insulin like growth factors (IGFs) and their cognate tyrosine kinase, transmembrane receptors and play a critical role in regulating somatic cell growth and proliferation. IGF I induces cell proliferation by regulating the cell cycle machinery, primarily in the G_1 phase of the cell cycle.

1.4.1 The insulin like growth factor binding proteins (IGFBP)

The insulin like growth factor binding proteins (IGFBPs) are a family of six proteins that bind to IGF I and II with very high affinity and antagonize the binding of IGFs with the type I IGF receptor thereby controlling their bioavailability and distribution in the serum (Fig.1.9). The six IGFBPs posses a very strong structural sequence homology though their biochemical effects are different. Both the amino and carboxy terminals of the IGFBPs are cysteine rich and the disulfide bond pattern is essentially conserved across all of the IGFBPs. In contrast the central domain of IGFBPs has no conserved cysteine residues and this region has very poor sequence homology amongst the IGFBPs (Baxter 2000; Clemmons 2001; Payet et al. 2004). IGFBPs not only regulate IGF action but also stimulate several biological effects that are independent of their ability to bind to the IGFs. These being inhibition or enhancement of cell growth and induction of apoptosis. IGFBPs can also enhance the IGFs availability by mechanisms, which are yet to be elucidated. IGFBPs were also shown to bind to important viral oncoproteins like HPV oncoprotein E7. This implies additional roles for IGFBPs in the pathways of cell proliferation, apoptosis, and malignant transformation.

1.4.2 The IGFBP degrading proteases

The effects of IGFBPs on IGFs are regulated in part by IGFBP proteases, which cleave the proteins at the flexible regions and thereby release fragments with low affinity for IGF binding (Fig. 1.9). This would result in an increase in IGF bioavailability and their bioactivity at the cell surface. Since IGFBP proteases are emerging as key regulators of IGF action at the cell surface they are now considered to be a member of the IGF family.

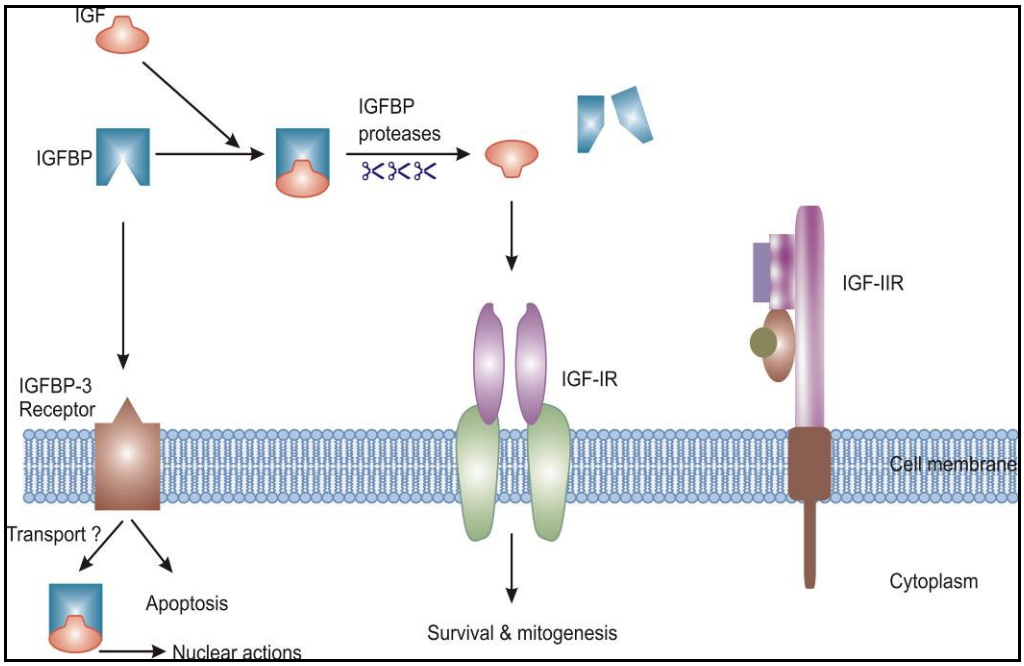


Figure 1.9 The IGF-IGFBP axis. IGFBPs bind IGFs with high affinity, thereby regulating the bioavailability of free IGFs in the serum. Specific IGFBP proteases cleave IGFBPs and thus regulate the levels of free IGFs. Free IGFBP 3 can transverse the cellmembrane and mediate IGF independent nucleur actions.

1.5 The calpain system

Calpains are non-lysosomal, Ca^{2+}-dependent cysteine proteases, which are ubiquitously distributed across cell types and vertebrate species. They play a role in the cell cycle regulation but physiological function of calpains in the cell cycle and rules that govern calpain cleavage specificity are poorly understood. The two ubiquitously expressed calpain forms, with proteolytic activities requiring μM and mM calcium, were identified and named as μ and m (also referred to as I and II) calpains, respectively (Goll et al. 2003). Calpains are heterodimeric proteins, composed of a large 80-kDa catalytic subunit organized in four domains (I–IV) and a common small, 30-kDa regulatory subunit organized in two domains (V-VI). Domain II is a cysteine protease domain and contains the catalytic cysteine, histidine and asparagine residues. Domain IV is a Ca^{2+} binding domain in which five EF-hand motifs are present. The small regulatory subunit is composed of an N-terminal glycine-clustering hydrophobic region (domain V) and a C-

terminal Ca^{2+} binding domain (domain VI). The hydrophobicity of the N-terminal domain (domain V) has been taken as an indication for its role in membrane anchoring (Moldoveanu et al. 2003). Calpains have potential biological functions in apoptosis, pathology of degenerative diseases, in mediating intracellular calcium signals (Glading et al. 2002). A number of studies indicated that the calpains have a role in the cell cycle, specifically in the G_1 to S transition.

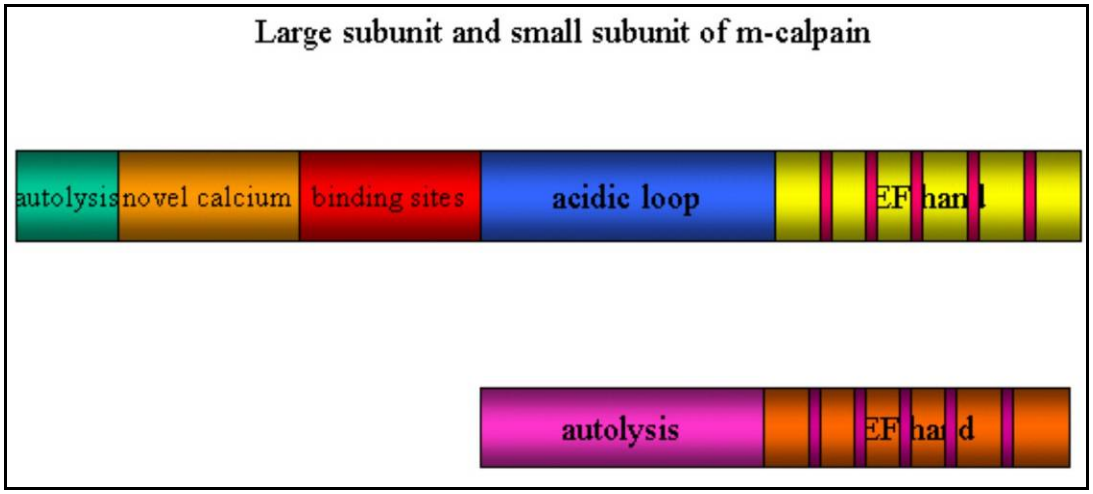


Figure 1.10 Schematic diagram of the large and small subunit of human m-calpain. structure. highlighting the various domains

The rules that govern calpain specificity have not yet been determined. Experimental reports published so far indicate that proteolysis by calpains are limited and does not lead to small peptides (Croall and DeMartino 1991; Goll et al. 2003). A number of studies indicated that the calpains have a role in the cell cycle, specifically in the G_1 to S transition (Goll et al. 2003). Calpain mediated degradation of $p21^{KIP1}$, which is a member of CIP1/KIP1 family of CDKIs, had been reported in preadipocyte cell cycle progression and differentiation (Patel and Lane 2000). However, biological role of calpain in the cell cycle regulation is still poorly understood. In this thesis an attempt has been made to characterize calpain preferred cleavage positions in the insulin-like growth factor binding proteins (IGFBPs). The reasons for using IGFBPs were the following:

- Proteolysis of IGFBPs is extensively characterized by numerous reports this is because limited proteolysis of IGFBPs is the major mechanism for the release of IGFs from IGFBP/IGF complexes, generating IGFBP fragments with reduced affinity for IGFs (Binox et al. 1999; Khandwala et al. 2000) and thereby increasing the concentration of IGFs at the cell surface and allows IGFs to bind and activate the IGF receptor (Bunn and Fowlkes 2003).

- IGFBPs are structurally well characterized, they are composed of globular domains linked by large flexible fragments, thus providing model proteins that contain diverse structural elements (Kalus et al. 1998).

- Recently, the presence of specific cell-surface IGFBP receptors has been reported. IGFBP-3 and –5 can enter the cell and bind to several different targets in the cytoplasm. Further more it has been shown that both the proteins can localize in the nucleus and posses nuclear localization sequence (NLS) in their mid-region (Jaques et al. 1997). This raises the possibility that nuclear IGFBP may directly control gene expression.

1.6 Baculovirus expression system

Baculovirus are a diverse group of viruses found mostly in insects which represent their natural host and they are not known to have any non arthropod hosts (O'Reilly et al., 1992). The baculovirus based expression systems are one of the most powerful expression systems available. The most commonly used baculovirus in the BEVS is the "*Autographa californica* multiple nuclear polyhedrosis virus" (AcMNPV) (Nobiron et al. 2003), which belongs to the family baculoviridae. It belongs to the group of double stranded DNA viruses and has a genome size of approximately 130 kbp with ability to incorporate hetrologous genes. The properties of insect cell lines for baculovirus propagation include growth rate, ability to support virus replication and plaque formation, ability to grow in both adhesion and suspension cultures. Most commonly used cell line is the *Sf 9* cell line, which originated from IPLB-SF-21 cells derived from *Spodoptera frugiperda* pupal ovarian tissue. One of the major advantages of BEVS over bacterial expression systems is

represented by the ability to allow easy production of functional heterologous proteins like, enzymes (Kumar et al. 2001) antibodies, or receptors (Cascio et al. 1993; Zhu et al. 2001). The insect baculovirus expression system provides a eukaryotic environment that is conducive to the proper folding, disulphide bond formation, oligomerisation and posttranslational modifications of the protein essential for the biological functionality of a large number of foreign proteins. The most important parameters that influence protein yields in the baculovirus system are: cell density at the time of infection the multiplicity of infection (MOI) (Chan et al. 1998) the age and the composition of the medium (Doverskog et al. 2000; Chiou and Wu 1990). To obtain high cell densities, and therefore high protein yields, several media and feeding strategies (Kim and Ziegler 2000) have been developed to enhance cell growth. The identification of essential and non-essential amino acids for cell growth was one of the most important factors. So far, alanine, cysteine, glutamic acid, glutamine, aspartic acid and asparagine were found to be non-essential amino acids. Further agents influencing cell growth are growth factors, vitamins and other compounds which are provided by chemically ill defined substances, like yeastolate or fetal calf serum. In general the serum can be replaced by lipid mixtures but the yeastolate contains additional nutrients which, although not completely identified, are beneficial for growth. The use of media without yeastolate leads to a dramatic decrease of growth rate and maximal cell density (Kim et al., 1999). Serum and yeastolate are also sources of free amino acids (Mitsuhashi, 1982) at different extent.

NMR-based structural studies and ligand binding studies require selectively ^{15}N labelled and/or uniformly labeled proteins. In *E. coli* ^{15}N-ammonium chloride is used as a sole source for nitrogen, whereas commercially available media for uniform labeling in insect cells contain all ^{15}N-amino acids, which increase the costs dramatically. Unfortunately, only few reports on labeling proteins in insect cells can be found in the literature (Creemers et al. 1999). In addition the total composition of the used media has not been revealed. In this study two novel media IML 406 and IML 455 were formulated for selective labeling of proteins expressed in *Sf 9* insect cells and comparison studies were done with similar labelling trials in *E. coli*

1.7 Introduction to methodology

This section gives an overview of the significant methods used for structural and functional characterization employed in the study; including X-ray crystallography and NMR for structural studies and ITC used for thermodynamic characterization of interactions.

1.7.1 X-ray crystallography

X-ray crystallography exploits the fact that X-rays are diffracted by crystals. Crystals can be described as an infinite array in which the building blocks (the symmetric units) are arranged according to well-defined symmetries into unit cell which are translationally repeated in 3-dimensional space. Growth of good quality single well defined crystals form the basic prerequisite for X-ray crystallography. This requires a homogeneous and pure protein (~ 95%), which is then concentrated to a supersaturated state to obtain crystals. Crystal growth basically involves three steps: nucleation, growth and cessation of growth. Different techniques are employed for setting crystallization trials which include sitting drop vapor diffusion, hanging drop vapor diffusion, sandwich drop and batch, micro batch under oil, micro dialysis, and free interface diffusion. Sitting and hanging drop methods are most commonly used, as they are easy to set up, require a small amount of sample, and allow large amount of flexibility during screening and optimization. The sitting drop method was used in this study in this method a small amount of sample is mixed with reservoir in a particular protein to reservoir ratio (i.e. 1:1, 2:1 etc) and placed on a platform in vapor equilibration with the reservoir that usually consists of one or more precipitants and a buffering agent (Fig 1.11).

Once good crystals are obtained they are harvested either in a capillary or loop and mounted on a X-ray source and X-ray detector, when a beam of X-ray passes through the crystal they are diffracted in all directions and recorded on the X-ray detector in some particular pattern, which is known as the diffraction pattern. A number of such diffraction patterns from different angles of the same crystal are recorded to give a data set from which the structure of an individual molecule can be deduced.

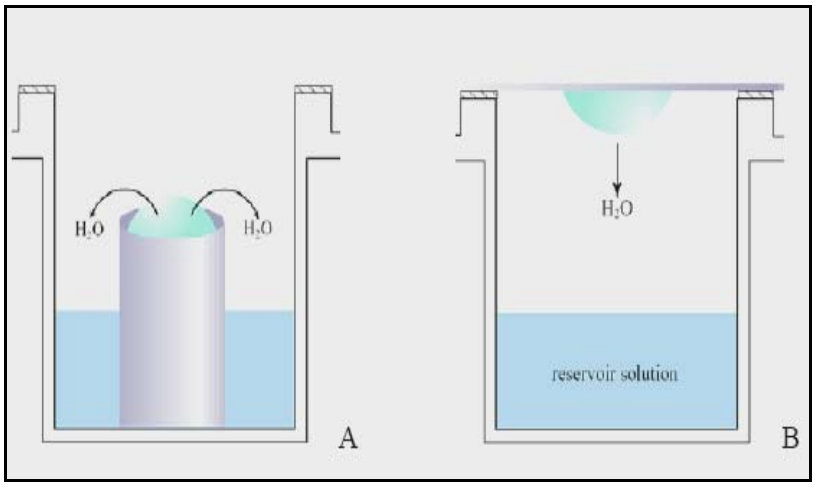


Figure 1.11 (A) Sitting drop and (B) hanging drop crystallization techniques

At times in sitting drop method crystals adhere to the bottom and are difficult to harvest. In such cases hanging drop technique is used wherein the sample drop is applied on a thin siliconized coverslip and placed in an inverted position (see Fig B) over the reservoir. In both cases the chambers are sealed from outside.

The diffraction or scattering of X-rays is based on the interference phenomenon.which is explained by the Braggs law.Once a good diffraction data is obtained the next step is to calculate the electron density map so as to ascertain the exact position of each atom in the asymmetric unit of the crystal. Calculation of electron density is dependent on two main factors the amplitude and the phase of the diffracted beam, which are then Fourier, transformed to give the electron density map. The amplitude of the beam can be derived from the measured intensity of reflection, but information about the phase angle is lost and without correct phase angles the calculation of an interpretable electron density map is not possible. This leads to the so-called phase problem in X-ray crystallography. There are a few methods, to date, which can help to circumvent the phase problem. They are

- Molecular replacement (MR)
- Single isomorphous replacement (SIR)
- Multiple isomorphous replacement (MIR)
- Multiple-wavelength anomalous dispersion (MAD) and
- Direct methods

The method of *Molecular replacement* (MR) was used in this study and is briefly introduced here it depends on the availability of a sufficiently homologous model structure. That is, if the structure of a protein (search molecule) is known that is homologous to the crystallized protein (target molecule) with unknown structure, the former can be used as a model to calculate a starting set of phases that can be interactively refined (Hoppe, 1957; Huber, 1965). For this, the search molecule must be oriented and positioned in the unit cell of the target molecule in such a way that maximum overlap of the models is achieved. This is done by a rotational and translational function. In the rotation step, the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the next step the translation needed to superimpose the now correctly oriented molecule onto the other is calculated. Molecular replacement has become an extremely useful technique for protein phase angle determination.

Once the electron density is obtained its further refined and fitted to give the final structure of the molecule under study.The introduction presented here was adapted from Blow (2002). For more details about crystallization procedures refer to McPherson (1999), and for the basis of X-ray crystallography refer to Drenth (1994) or Giacovazzo (1992).

1.7.2 NMR Spectroscopy

Nuclear magnetic resonance spectroscopy complements X-ray crystallography for small and medium size proteins (below 30 kDa) (Montelione et al., 2000; Prestegard et al., 2001). Its advantages are that it does not require crystals and allows you to study the protein in solution state; it does not destroy the sample and has many uses apart from solving structures. It can be used to determine structure-function relationships (Shuker et al. 1996), to do comparative studies on metabolic pathways from different organisms (Bruggert et al. 2003), to find binding partners and study their binding sites (Stoll et al. 2001), A series of spectra taken under different conditions may be used to monitor aggregation and formation of amyloid fibrils (Zurdo et al. 2001) to investigate dynamics of proteins to distinguish multiple conformations (Muhlhahn et al. 1998) and its ability to detect weak ligand binding to target molecules have made it increasingly important in drug discovery.

1D-NMR: A simple one-dimensional proton experiment is the most basic spectrum in NMR spectroscopy that can be acquired in a short time (usually not longer than a few minutes) for samples as dilute as 0.01 mM and contains great amount of information. It tells us about the folding status of the proteins as to whether a protein is folded or unfolded which is important for any further functional or structural studies on the protein as only folded proteins retain their functional activity and three dimensional structure (Rehm et al 2002) (Fig.1.12). One-dimensional spectra of protein molecules contain overlapping signals from many hydrogen atoms because the differences in chemical shifts are often smaller than the resolving power of the experiment. This problem has been bypassed by designing experimental conditions that yield a two-dimensional NMR spectrum.

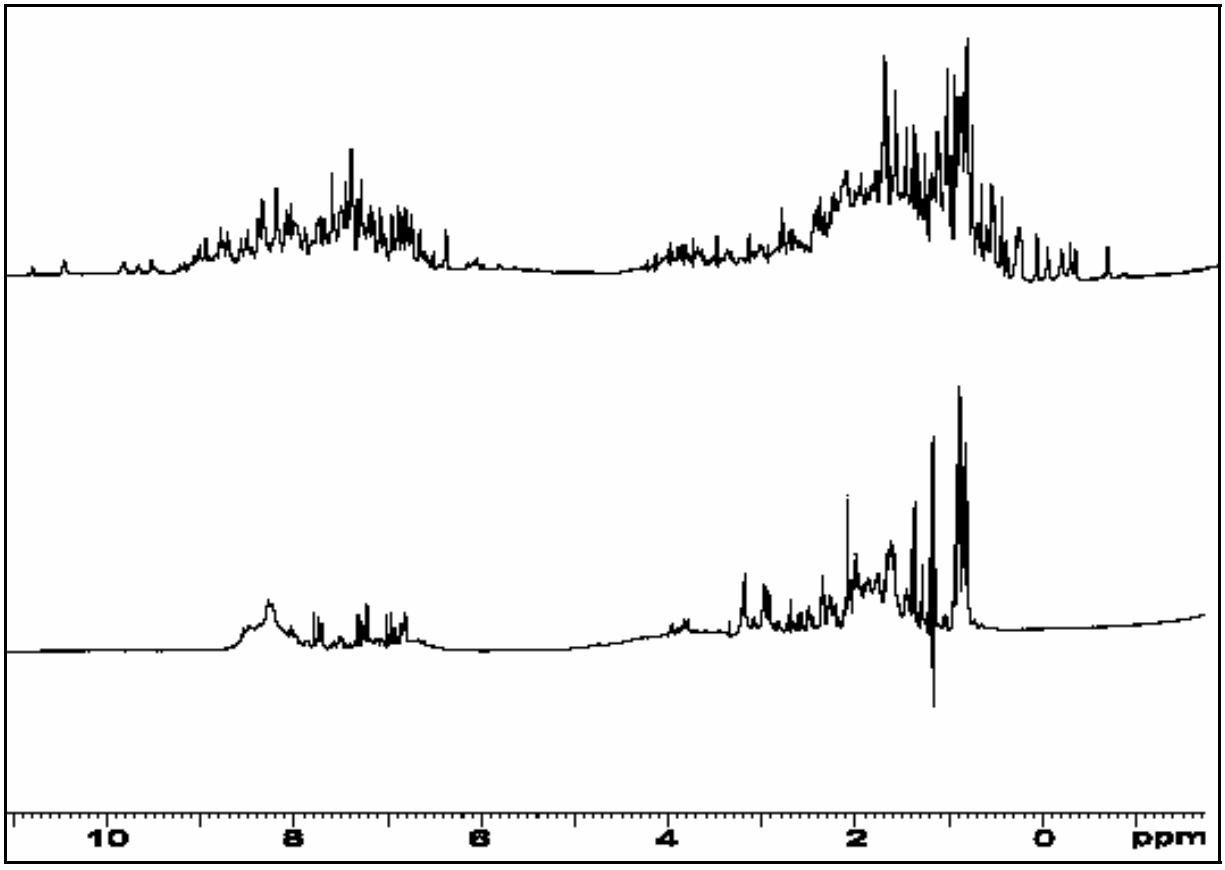


Figure 1.12 Characterization of protein structures by one-dimensional NMR spectroscopy: Upper panel: A typical one-dimensional proton NMR spectrum of a folded protein with signal dispersion downfield (left) of 8.5 ppm and upfield (to the right) of 1 ppm. Lower panel: An unfolded protein sample. Strong

signals appear around 8.3 ppm, the region characteristic for amide groups in random coil conformation. No signal dispersion is visible below approximately 8.5 ppm. Also to the right of the strong methyl peak at 0.8 ppm no further signals show up.

2D-NMR: Due to the greatly improved resolution of two-dimensional experiments, these are frequently used for screening and binding studies. The simplest and most powerful among them is the heteronuclear single-quantum coherence (HSQC) experiment. This method has been employed in this study to investigate the different metabolic pathways employed by insect cells and bacteria based on nitrogen. In a large scale approach (Yee et al., 2002) have recently investigated more than 500 proteins from five different organisms, using ^{15}N-HSQC experiments to screen for those proteins amenable for NMR structure analysis. For this kind of spectrum ^{15}N-labeled protein samples are required. The HSQC shows one peak for every proton bound directly to a nitrogen atom and thus exactly one signal per residue in the protein (apart from proline which is devoid of proton bound nitrogen and some additional side chain signals appear which can easily be identified).

1.7.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a thermodynamic technique, which monitors the bioenergetics of a reaction. The use of ITC to measure the binding of a macromolecule to its ligand relies on the fact that such an interaction is accompanied by a heat effect. This heat change is then used to obtain information about the binding constant, K_a, and the enthalpy of binding ΔH. Moreover, the temperature dependence of the enthalpy of binding allows one to calculate the entropy of the system. ITC is of great use in structure based drug design. This method was used in this study to characterize HDAC inhibitors. ITC is also finding a niche as a tool in kinetic studies with potential to assess inhibition characteristics. In ITC, a syringe containing a “ligand” solution is titrated into a cell containing a solution of the “macromolecule” (Fig 1.13) at constant temperature and heat released or absorbed is measured which in turn is directly proportion to the amount of binding. As the macromolecule in the cell becomes saturated with ligand, the heat signal diminishes until only the background heat of dilution is observed.

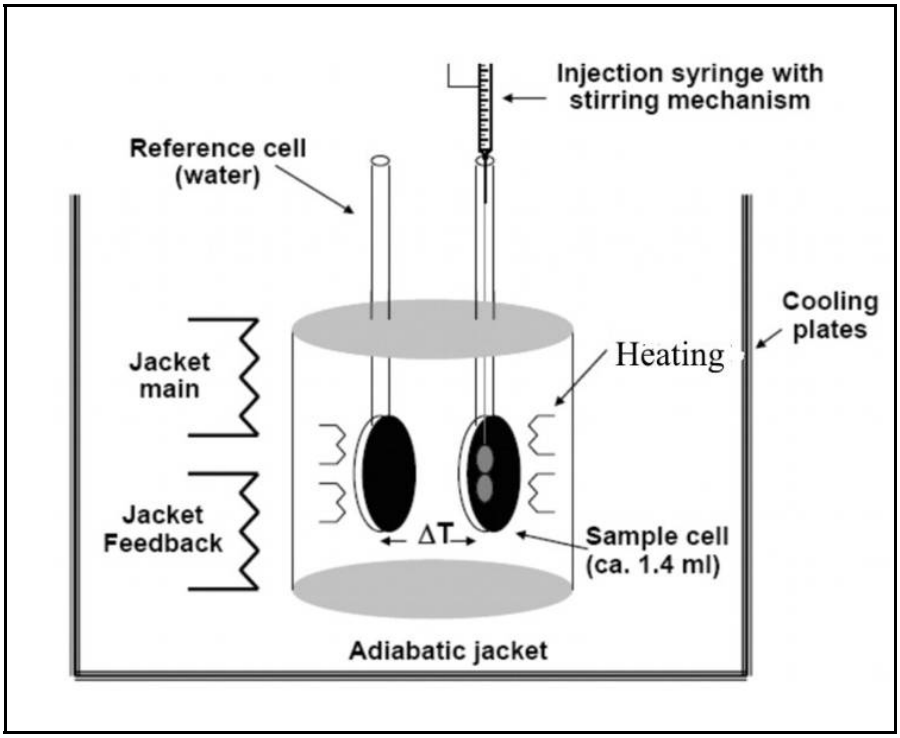


Figure 1.13 Schematic diagram of the VP isothermal titration The syringe rotates in place during the ITC experiment. The end of the syringe has been adapted to provide continuous mixing in the ITC cell. The plunger is computer-controlled and injects precise volumes of ligand.

2 Materials and Methods

2.1 Materials

All chemicals were of analytical grade and were purchased from Sigma, Fluka and Merck laboratories if not stated otherwise.

2.1.1 Enzymes and Other Proteins

Restriction endonucleases, NEB buffers and T4-DNA ligase	New England Biolabs.
pfuturbo DNA polymerase, pfu buffer and Dpn I.	Stratagene.

Taq-DNApolymerase, Pwo-DNA-polymerase.	Peq Labs.
Hen egg white lysozyme.	
RNaseA.	
DNaseI	
Thrombin.	
Factor Xa	
HAS	
CIP	
Complete Protease Inhibitors Cocktail.	Roche Applied Science.
Phenylmethylsulfonyl fluorid (PMSF).	Roche Applied Science.

Table 1 Molecular Weight Marker for SDS-PAGE Electrophoresis (NEB, FRG)

Protein	**Source**	**Apparent MW (Da)**
Maltose-binding protein- β-galactosidase	*E. coli*	175000
Maltose-binding protein-paramyosin	*E. coli*	83000
Glutamic dehydrogenase	Bovine liver	62000
Aldolase	Rabbit muscle	47500
Triosephosphate isomerase	Rabbit muscle	32500
B-lactoglobulin A	Bovine milk	25000
Lysozyme	Chicken egg white	16500
Aprotinin	Bovine lung	6500

2.1.2 Kits and reagents

Table 2 List of various kits and materials used

BD Advantage 2 PCR kit	BD Biosciences.
QIAquick PCR purification kit	Qiagen.

QIAprep spin miniprep kit	Qiagen
Rapid DNA ligation kit	Roche Applied Science
Quick change site-directed mutagenesis kit	Stratagene
Transfection buffer A and B	Pharmingen
IPTG, dATP, dGTP, dTTP, dCTP	Peq Labs

2.1.3 Isotopically enriched chemicals

Table 3 List of isotopically enriched chemicals

Deuterium oxide, D2O 99 %, 99.99 %	Campro Scientific, Berlin, FRG
Deuterium oxide, D2O 95 %, 95 %	Campro Scientific, Berlin, FRG
^{15}N-ammonium chloride, NH4Cl 99.9 %	Campro Scientific, Berlin, FRG
^{15}N-L-alanine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-arginine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-aspartic acid, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-glutamic acid, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-glycine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-isoleucine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-leucine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-Lysine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-Phenyloalanine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-Serine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-Tyrosine, 99,9 %	Campro Scientific, Berlin, FRG
^{15}N-L-Valine, 99,9 %	Campro Scientific, Berlin, FRG

2.1.4 Media composition

All media, if not mentioned here, were prepared exactly as described in (Sambrook and Russell 2001).

Minimal medium for uniform ^{15}N labeling of amino acids in *E. coli*

The composition for 1L media used for uniform labeling is given below

Table 4 Minimal Medium (MM) for Uniform Enrichment with $^{15}NH_4Cl$

	Stock solutions	
1	Thiamin	1 %
2	Antibiotics	
3	$MgSO_4$	1M
4	Zn-EDTA solution:	
	EDTA	5 mg/ ml
	Zn (Ac)	28.4 mg/ ml
5	Trace elements solution:	
	H_3BO_3	2.5 g/ l
	$CoCl_2*H_2O$	2.0 g/ l
	$CuCl_2*H_2O$	1.13 g/ l
	$MnCl_2*H_2O$	9.8 g/ l
	$Na_2MoO_4*2H_2O$	2.0 g/ l
	If difficult to dissolve, pH was lowered with citric acid or HCl.	
6	Glucose, separately autoclaved	5 g/ 25 ml
	Working solution	
1	Composition for 1L media is given below	
	NaCl	0.5 g
	Trace elements solution	1.3 ml

	Citric acid monohydrate	1 g
	Ferrous citrate (dissolved in 120µl conc. HCl, heated)	36 mg
	KH_2PO_4	4.02 g
	K_2HPO_4 $3H_2O$	7.82 g
	Zn-EDTA solution	2 ml
	NH_4Cl or $^{15}NH_4Cl$	1 g
2	pH was adjusted to 7.0 with NaOH	
3	The mixture was autoclaved	
4	25 ml separately autoclaved glucose was added	
5	Other compounds were added (previously sterile filtered):	
	Thiamin	560 µl
	Antibiotics (half of the usual amount for the LB-medium)	

Medium for ^{15}N selective enrichment of amino acids in *E. coli*

The composition for 1L media used for selective enrichment is given below and the amino acid to be enriched was added in its ^{15}N isotope form

Table 5 Composition of 1L media for selective enrichment of amino acids in *E. coli*

1	Ala, Glu, Gln, Arg, Gly	400 mg
	Asp, Met	255 mg
	Cytosine, Guanosine, Uracil	125 mg
	Asn, Ile, Val, Leu, His, Lys, Pro, Thr, Tyr	100 mg
	Phe, Thymidine, Thymine	50 mg
	$CaCl_2$	10 mg
	NaAc	2 g

	KH_2PO_4	10 g
	Citric acid	1 g
	Trace elements solution	1 ml
	Ferrous citrate	36 mg
	Zn-EDTA	1 ml
	Ser	1.6 g
	NH_4Cl	1 g
2	The pH was adjusted to 7.4 with 1M NaOH,	
3	The mixture was autoclaved,	
4	After autoclaving the following sterile filtered solutions were added:	
	Cys, Trp, nicotinic acid	50 mg
	Biotin	0.1 mg
	$MgSO_4$	2 ml
	Thiamin	560 µl
	Glucose	20 ml
	Antibiotics (2x of the usual amount for the LB-medium) Concentrations of stock solutions' were like for MM.	

Medium for ^{15}N selective enrichment of amino acids in insect cells

Table 6 Composition of 1L media formulated in this study for selective enrichment of amino acids in Insect cells. KBM10 media is shown for comparison.

Contents g/ l	KBM10	IML405	IML455	IML406
NH_4Cl	0	0	0,265	0
Glucose	5	5	5	5
KCl	2,87	2,87	2,87	2,87
$CaCl_2 \times 2H_2O$	1	1,4	1,4	1,4
$MgCl_2 \times 6H_2O$	2,14	2,14	2,14	2,14
$Mg_2SO_4 \times 7H_2O$	2,72	2,72	2,72	2,72
$NaHCO_3$	0,35	0,35	0,35	0,35
NaH_2PO_4	0,88	0,9	0,9	0,9
L-Glutamine	1	1	0	1
L-Arginine-HCl	0,7	0,7	0,7	0,7

L-Asparagine	0,35	0,35	0,35	0,35
L-Aspartic acid	0,35	0,35	0	0,35
L-Glutamic acid	0,6	0,6	0,6	0,6
Glycine	0,65	0	0	0
Glycine-HCl	0	0,65	1	1
L-Histidine	0,2	0,2	0,2	0,2
L-Isoleucine	0,5	0,175	0,175	0,175
L-Leucine	0,075	0,4	0,4	0,4
L-Lysine-HCl	0,625	0,625	0,625	0,625
L-Methionine	0,05	0,05	0,05	0,05
L-Phenylalanine	0,15	0,15	0,15	0,15
L-Proline	0,35	0,35	0,35	0,35
L-Serine	0,55	0,55	0,55	0,55
L-Threonine	0,175	0,175	0,175	0,175
L-Valine	0,1	0,1	0,1	0,2
L-Cystine	0,05	0,05	0,05	0,05
L-Tryptophan	0,1	0,1	0,1	0,1
L-Tyrosine	0,072	0,072	0,072	0,08
ml/ l				
Yeastolat-ultafiltrate	20	20	20	20
FCS	50	50	50	50
Lipid-mixture	10	10	10	10
Trace elements	1	1,5	1,5	1,5
MEM-vitamin	10	10	10	10
µg/ l				
Vitamin B-12	240	240	240	240
pH	6,2-6,3	6,2-6,3	6,2-6,3	6,2-6,3
Osmolality mOsm/ kg	320-340	320-340	320-340	320-340

LB Medium

10 g/ l Bacto-tryptone

5 g/ l Bacto-yeast extract

10 g/ l NaCl

pH of the medium was adjusted to 7.0 with 5 N NaOH and then made up the final volume to 1 lit with ddH2O. Later medium was sterilized by autoclaving (For plates, medium was supplemented with 15 g/ l agar and recommended amount of appropriate antibiotics).

Terrific broth (TB) medium

12 g/ l Bacto-tryptone
23.9 g/ l Bacto-yeast extract
8 ml/ l glycerol
2.2 g/ l $KH_2 PO_4$
9.4 g/ l K_2HPO_4

SF 900 II for insect cells

Prepared according to manufacturers protocol.

2.1.5 Stock solutions and buffers

All stock solutions and buffers, if not mentioned here, were prepared exactly as described in (Sambrook and Russell 2001).

IPTG stock solution

IPTG was dissolved in water (2.38 g/ 10 ml) to the end concentration of 1 M. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium, unless otherwise indicated.

Ampicillin stock solution

Ampicillin was dissolved in water (1 g/ 10 ml) to the end concentration of 100 mg/ ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Kanamycin stock solution

Kanamycin was dissolved in water (0.5 g/ 10 ml) to the end concentration of 50 mg/ ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Chloramphenicol stock solution

Chloramphenicol was dissolved in ethanol (0.5 g/ 10 ml) to the end concentration of 50 mg/ ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Buffers for Ni-NTA purification under native conditions

- **Lysis buffer** 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole, EDTA-free protease inhibitor cocktail.
- **Ni-NTA wash buffer** 50 mM Tris, 300 mM NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole, pH 8.0.
- **Ni-NTA elution buffer** 50 mM Tris, 300 mM NaCl, 10 mM β-mercaptoethanol, 200 mM imidazole, pH 8.0.

Buffers for Ni-NTA purification under denaturing conditions

- **Lysis buffer A** 100mM NaH_2PO_4, 10mM Tris pH: 8.0, 6M Guanidinium chloride, 10mM β-mercaptoethanol.
- **Ni-NTA wash buffer B** 100mM NaH_2PO_4, 10mM Tris pH: 6.5, 6M Guanidinium chloride, 10mM β-mercaptoethanol.
- **Ni-NTA elution buffer** 100mM sodium acetate pH: 4.0, 6M Guanidinium chloride, 10mM β-mercaptoethanol.
- **Refolding buffer** 100mM Tris pH 8.0 200mM Arginine**,** 1mM EDTA**,** 2mM Reduced glutathione, 2mM Oxidised glutathione, 0.05 % NaN_3.

Buffers for GST purification

- **Lysis buffer** 50 mM Na_2HPO_4 pH 8.0, 250 mM NaCl, 10 mM β-mercaptoethanol, EDTA-free protease inhibitor cocktail.
- **GST-sepharose wash buffer** 50 mM Na_2HPO_4 pH 8.0, 250 mM NaCl, 10 mM β-mercaptoethanol.
- **GST-sepharose elution buffer** 50 mM Na_2HPO_4 pH 8.0, 250 mM NaCl, 50Mm glutathione 10 mM β-mercaptoethanol.

- **Crystallization buffer:** 25 mm Tris pH 8.0, 250mm NaCl, 2.5 % Glycerin, and 10mM β-mercaptoethanol.
- **Phosphate-buffered saline (PBS)**

2.1.6 *E. coli* strains used in this study

Table 7 Plasmids and experimental organisms

#	Strain type	Product supplier	Reference
1	BL21(DE3)	Invitrogen	Philips, T.A. et al., 1989
2	BL21(DE3)pLysE	Invitrogen	Philips, T.A. et al., 1989
3	BL21(DE3)pLysS	Invitrogen	Philips et al., 1989
4	BL21 Star(DE3)	Invitrogen	Makrirides et al., 1995; Lopez et al.,1999
5	BL21 Codon plus (DE3)-RP	Invitrogen	
6	TOP10	Invitrogen	

2.1.7 Plasmids for protein expression

pET28a, pET15b, pET30/ LIC/ Xa, pGEX-4T-2, pGEX-3X

2.1.8 Insect cell line

Spodoptera frugiperda (*Sf9*) cells, (Gibco Germany)

2.2 Methods

2.2.1 Molecular Biology Techniques

2.2.1.1 Molecular cloning

Plasmid constructs produced during this study were cloned with the aid of polymerase chain reaction (PCR), followed by restriction digestion and ligation. The standard PCR protocol followed is summarized in table 7.

Table 7 Standard protocol for polymerase chain reaction.

Reaction composition	Amount (µl)	Thermal cycling parameters			
Sense primer (10 pm)	1.5	Segment	Cycles	Temperature (ºC)	Time
Anti sense primer. (10 pm)	1.5	1	1	95	1 min
Template DNA (100-150 ng)	2.0	2	20-30	95	1 min
				50-55	30 sec
				68-72	1 min/ kb
dNTPs (10 mM)	1 0	3	1	72	4 min
pfu DNA polymerase	1.0	4	1	4	Hold
10 x pfu buffer	5.0				
Sterile ddH2O	38				

All PCR reaction mixtures were prepared in sterile thin wall PCR tubes and PCR was performed using the master cycler (Eppendorf). The amplified PCR products were analyzed by agarose gel electrophoresis (1 % agarose dissolved in TBE buffer containing ethidiumbromide [10 µg/ ml]). Desired PCR amplified products were purified by QIAquick PCR purification kit following the manufacturer's instructions. Restriction digestion of purified PCR products and desired vector was then carried out using the specific restriction endonucleases in amounts recommended by manufacturers (NEB) at 37°C in a 50 µl reaction for a couple of hours. The restriction digested vectors were then dephosphorylated by treating with 5 units of calf intestinal phosphatase at 37°C for one hour. All reactions were stopped by heating at 65/ 85°C for 20 min. Desired digested products were then agarose gel purified using QIA quick gel extraction kit following the manufacturers instructions. Ligation reactions were then set up using T4 DNA ligase,

following the standard protocols at vector to insert ratio of 1:3 to 1:5 and were usually kept overnight. 1 to 2 µl of this mixture was then used for transformation into TOP10 *E. coli* cells using electroporation.

2.2.1.2 Site directed mutagenesis

Quick-Change® site-directed mutagenesis kit was used to introduce single amino acid substitutions, deletion and addition mutations at desired sites in the plasmid containing gene of interest. Gene encoding a supercoiled double stranded DNA carrying a gene of interest was used as a template. Complementary oligonucleotide primers containing the desired mutations were used for PCR following manufacturer's instructions. The extension of mutagenic primers, generated mutated covalently closed and nicked plasmids. Following PCR, the product was treated with 1 µl of Dpn I at 37°C for few hrs. The Dpn I endonuclease was used as it is specific to methylated and hemimethylated DNA which is not found in the newly synthesized daughter strands and hence helps to digest the non mutated parental strands and select mutated DNA. After digestion 1-2 µl of this DNA was used for transformation.

2.2.1.3 Transformation

Electrocompetent *E. coli* cells were used to amplify hetrologous genes and overexpress hetrologous proteins in this study 50 to 100 ng of plasmid DNA was gently mixed together with the 40 µl aliquot of electrocompetent cells and placed between the electrodes of a 0.1 cm electroporation cuvette (Biorad, FRG). The cuvette was then put into the electroporator (Stratagene, FRG), and a pulse of 1660 V was applied. The value of the time constant was observed (usually 4-5 ms). The mixture was then washed out from between the electrodes with 1ml of the sterile prewarmed (37°C) LB medium (without antibiotics), and transferred to a sterile 1.5 ml tube for shaking (800rpm) at 37°C. After 1 hr, 50 to 100µl of culture were streaked on a LB agar plate containing the appropriate antibiotic and were placed in an 37°C .incubator for 12 to 16 hrs. The colonies were then analysed for positive clones using either colony PCR or restriction digestion and further confirmation was carried out by gene sequencing.

2.2.2 Protein expression in *E. coli*

2.2.2.1 Small scale test expression of proteins

Once the clones were obtained and transformed into expression strains, small scale test expression were done to identify the optimal condition for hetrologous protein production. 5 ml LB containing the appropriate antibiotic was inoculated with a fresh single bacterial colony and incubated overnight at 37°C with vigorous shaking (280 rpm) in a 100 ml flask. Pre-warmed 50 ml LB medium in 250-500ml flask was inoculated with 1ml of the overnight culture, supplemented with appropriate antibiotic, and incubated at 37°C with shaking (180 rpm) until the OD_{600} reached 0.6 to 0.8 values. Induction by IPTG addition followed upto 1 mM IPTG end concentration was usually used. The cells were then grown for another 2-6 hrs or overnight depending on temperature (15-37°C). 1ml of sample was then taken at regular intervals, the cells were pelleted and lysed by sonification and loaded on a SDS-PAGE gel and visualized by coomasie blue staining to see levels of protein induction.

2.2.2.2 Large scale over-expression of proteins

Once the conditions were optimized for protein expression, large-scale production of recombinant proteins from *E. coli* was carried out by culturing cells at previously optimized conditions. 50 ml LB/ enriched media were inoculated with a fresh single bacterial colony and incubated overnight at 37°C with shaking (280 rpm) in a 100 ml flask. For minimal medium 50 µl culture grown in LB media was used as an inoculum, instead of single bacterial colony. 10-15 ml of the overnight was then pelleted and cells were resuspended in fresh media containing antibiotic and this culture was then used to inoculate 1l of the desired media containing appropriate antibiotics, which was then incubated at 37°C with shaking (180 rpm) until the $O.D_{600}$ reached 0.7 to 0.8 values (O.D. was measured using spectrophotometer). Once cells reached the desired OD they were induced with upto 1 mM IPTG end concentration. Cells were harvested 3-4 hrs post-induction and frozen at -80˚C till further use.

2.2.3 Cloning and protein expression in baculovirus/ insect cell system

2.2.3.1 Maintenance and adaptation of *Sf 9* insect cells

The *Sf 9* cell lines, which double in every 18 to 24 h were grown in Sf-900 II SFM media supplemented with 1 % antibiotic-antimycotic liquid and 5 % fetal bovine serum and also in various formulated media (Table 7). The cells were maintained in spinner flasks as suspension culture or in 75 cm^2 tissue culture flasks as adhesion culture. The cells cultured in adhesion cultures were passaged at 80 to 90 % confluency and finally diluted to 30 % confluency. The suspension cultures were routinely passaged at cell densities of 1.8 to 2.0 × 10^6 cells per ml and diluted to a final cell density of 0.3 to 0.4 × 10^6 cells/ ml. All cultures were incubated at 27°C and the suspension cultures were stirred at 90 rpm. The cells had to be adapted to new media conditions in a stepwise manner. Therefore after seeding at 30 % confluency in a 25 cm2 tissue culture flask they were exposed to a medium containing 50 % of the new and 50 % of the initial medium. At confluency the *Sf 9* cells were diluted in the same mixture. After reaching a constant growth rate the cells were exposed to a 75 % new medium and passaged until a constant growth rate was reached again. In the final step 100 % of the new medium was used. After adaptation to the new conditions the cells were maintained in suspension culture for further growth and expression studies.

2.2.3.2 Cloning into baculovirus

Cloning into baculovirus was done in a two step process which involved cloning of the desired gene into a transfer vector followed by homologus recombination between the transfer vector and viral DNA. The transfer vector is designed for amplification of DNA in bacteria following the normal protocols of cloning described above. The Col E1-origin makes the replication of transfer vector possible in *E. coli* and leads to high plasmid copy number. The other important feature of the transfer vector is that the cloning site is surrounded at both ends by DNA sequences, which contain parts of ORF603 and ORF1629 from the baculovirus DNA. These two ORFs flank the locus of the polyhedrin gene in the AcNPV DNA. The integration of the desired gene into viral DNA.is then

accomplished via homologous recombination of transfer vector and virus DNA following cotransfection.

2.2.3.3 Cotransfection

2 x 10^6 *S f9* cells were seeded onto 60 mm tissue culture plates and allowed to adhere to the plate to form an adhesion culture. Separately 0.5µg of BaculoGold DNA (Pharmingen) and 5 µg of recombinant baculovirus transfer vector containing the insert were mixed well by vortexing and then the mixture was incubated for 5 min at 27°C. Thereafter, 1 ml of transfection buffer B was added to the mixture. Medium was aspirated off from cells that were previously allowed to attach to the plate and replaced with 1 ml of the transfection buffer A. The previously prepared transfection buffer B/ DNA mix was then added drop wise to the experimental cotransfection plate. After every 3 to 5 drops, the plate was gently rocked to mix the drops with the medium and then the plates were incubated at 27°C for 4 hrs. After 4 hrs the medium was aspirated and 3 ml fresh pre warmed medium was added. The plates were then further incubated at 27°C for 4-5 days, following which the plates were checked for signs of infection under an inverse microscope. Infected cells appeared larger with enlarged nuclei., the supernatant was then collected and assayed for cotransfection efficiencies by the end-point dilution assay.

2.2.3.4 Endpoint dilution assay (EPDA)

1 × 10^6 *Sf 9* cells per well were seeded and allowed to attach firmly on a 12-well EPDA plate. 0, 1, 10, and 100 µl of the recombinant virus supernatant was then added to separate wells. All plates were incubated at 27°C for three days. After incubation, infected cells were differentiated from uninfected once based on the enlargement of nucleus of infected cells.

2.2.3.5 Preparation of high titer viral stock

2 × 10^6 *Sf 9* cells were seeded on a 15 cm plate and allowed to adhere. 25 µl of a low titer recombinant viral stock was then added and then incubated at 27°C for 3 days. After 3 days of incubation, signs of infection were checked as before and, the supernatant

from the plates were harvested by spinning down the cellular debris in a tabletop centrifuge at 10,000 x g. The virus supernatant was stored at 4°C in a sterile tube and covered with foil to protect against light. Later, large-scale virus amplification was carried out by adding 100 µl of this supernatant to 500 ml suspension cultures at a cell density of 1×10^6 *Sf 9* cells. Followed by incubation for 4-5 days after which the the culture was centrifuged and supernatant collected and stored at 4°C shielded from light. These high titer viral stocks were then used for protein expression studies. The stocks were stored at 4°C for further applications. The presence of the gene in the amplified virus titer was checked by isolating DNA from the baculovirus.

2.2.3.6 Isolation of DNA from baculovirus

750 µl of an occlusion negative cell suspension was transferred to a micro centrifuge tube and centrifuged at 5,000 rpm for 3 min at room temperature. To the supernatant 750 µl of ice cold 20 % PEG in 1 M NaCl was added, mixed and incubated on ice for 30 min. The viral particles were pelletted by centrifuging at maximum speed for 10 min at 4°C. Pelleted viral particles were centrifuged again at maximum speed for 2 min and the residual supernatant was aspirated off. Thereafter, the pellets of viral particles were resuspended in 100 µl of TE buffer. Plasmid DNA was isolated from viral particles using Easy-DNA kit (invitrogen) following the manufacturers instructions. Presence of the gene of interest was identified by PCR amplifying the isolated plasmid DNA with forward and reverse primers designed against up and downstream sequences of the gene of interest.

2.2.3.7 Growth and expression studies

For growth studies adapted *Sf 9* cells were maintained for 10 passages in adhesion culture in the desired medium. After this period the insect cells were transferred to suspension culture and grown in 50 ml of the medium at an initial density of 3×10^5 cells per millilit. The density and vitality of the cells was measured after 24, 48, 72, 96, 120 and 144 h of culturing. The test expressions were performed in 100 ml suspension cultures. The cells were infected with a MOI of 8 at a cell density of $2.0–2.2 \times 10^6$ cells per millilit. After 60 hrs the cells were harvested by centrifugation at 900 g and 4°C. The GST was

purified and the quantity was determined using Bradford assay. To investigate the influence of the amount of a certain amino acid provided by yeastolate or serum on cell growth and protein yield, the particular amino acid was not added separately to the medium.

2.2.3.8 Expression of recombinant proteins

Test expressions were set up to determine the optimal condition for recombinant protein expression. The MOI and expression time are crucial parameters for optimum protein yield. Test expressions were accomplished with adhesion cultures. For each MOI tested 2×10^6 *Sf 9* cells were plated on a 15cm plate. Cells were infected when they were 50 to 60 % confluent with different MOIs between 1 and 8 and grown at 27°C. After 24, 36, 48, 60 and 72 hrs the cells were harvested and the protein yield evaluated by means of SDS PAGE and Western blot. After the determination of optimal conditions, large scale protein expression was carried out. The cells were grown in suspension culture in spinner flasks with continuous agitation at 90 rpm and 27°C. Spinner flasks of 100, 250, 500 and 1000 ml capacity were used and cultures were grown at 50-60 % maximal flask capacity. Once the cells reached a density of 1.8×10^6 cells/ ml the cells were diluted by addition of fresh media at a ratio of 1:1 and upgraded to the next higher flask capacity.

2.2.3.9 Production of unlabelled proteins

After reaching a cell density of 2.0 to 2.2 x 10^6 cells per millilitre in around 2 l culture the culture was infected with the respective recombinant virus at an optimal MOI (usually 3-8). After the optimal expression time (50-60 h) the cells were pelleted and frozen at -80°C until further use. For obtaining larger protein yields cells were also grown in a bioreactor which could take a working volume of 4 lits. The bioreactor was sterelized by autoclaving before each run and was fed with 1 lit suspension culture. Further cells are grown in controlled conditions of 27°C temperature, 6.3 pH, increased agitation of 100 rpm and controlled oxygen levels which helps in sustaining higher cell densities. The cells were then infected with recombinant virus at cell densities of 3.0-3.2x10^6 and were harvested 60 hrs post infection.

2.2.3.10 Production ^{15}N-labelled proteins

The basic methodology remains the same but cells should be adapted to the media before protein expression, the media should contain a ^{15}N enriched sample of the desired amino acid. The cell density at the time of the infection differs as a function of the used medium. Expression volumes, expression time and MOI remain the same as in unlabeled samples.

2.2.3.11 Determination of the osmolality

Osmolality is defined as the concentration of all osmotically active particles (millimoles) in a kilogram of a liquid. Insect cells can exist only within a narrow range of osmolality (290-350 mOsm/ kg). Fluctuations in the osmolality severely effects growth and protein yield of insect cells. Since in this study developed media were used, whose individual components were varied, control of the osmolality was absolutely necessary. The osmolality was determined using a freezing point osmometer Osmomat 010 (Gonotec). For measuring osmolality first the instrument was calibrated with distilled water and secondly with an aqueous solution of NaCl with an osmolality of 300 mOsm/ kg. For calibration 50µl of solution is used, and the freezing point for both solutions is determined. With the help of these two values osmolality of the media can be tested. Since during the culturing of the cells the osmolality changes, this value must be closely monitored.

2.2.3.12 Determination of the cell density and cell viability

The cell density can be determined with the help of heamocytometers or a particle size-measuring instrument. With the heamocytometer 100 µl of the cell culture is pipetted into the counting chamber and covered with a cover glass. Under an inverse microscope the cells in the central square of the chamber are counted and after multiplication of this number with the dilution factor one receives the number of cells per millilitre performed with a particle. The determination of the cell number with the particle size measuring instrument has the advantage that one receives a plot with the particle size distribution and the exact cell density. This distribution also informs about vitality of the cells and possible

contamination. Cultures with vital cells have a very close particle size distribution between 13 and 19 µm, with an average of 16-17 µm. below this range the dead cells are observed and at a range of 1-4 any bacterial or yeast contamination is observed. Cultures with vital cells above 95 % were chosen for passaging and protein expression. The Vitality of the cells was also determined with trypan blue exclusion test wherein 100µl of 0.4 % trypan blue is added to 900µl of cell culture, only dead cells take up the dye and these are then counted. The percentage of the dead cells to the entire cell number is then determined and for healthy cultures should be below 5 %.

2.2.3.13 Quantitative amino acid assay

The quantitative analyses of free amino acids were performed with an amino acid analyzer LC3000 (Eppendorf). The serum was deproteinated by addition of 200 µl 10 % sulfosalycylic acid to 800 µl serum. Before loading of samples yeastolate was diluted 1:80 and 1:160, the serum 1:1 and 1:2. The amino acids were separated by a strong cation exchange column PEEK and detected after post-column derivatisation with ninhydrin. For the viral expression of GST in insect cells, the plasmid pAcG2T was chosen. The plasmid was used for co-transfection with linearized BaculoGold virus DNA (Pharmingen).

2.3 Protein purification

Recombinant proteins were purified to homogeneity using basic chromatographic techniques like affinity chromatography, ion exchange chromatography and gel filtration chromatography. Native proteins were obtained after removal of the affinity tag by proteolytic digestion and subsequent separation of an untagged protein from the affinity tag. Folding status and the secondary structural content of proteins were monitored by NMR and CD spectroscopy and then further crystallographic trials were done to obtain the three dimensional structure of proteins.

2.3.1 Affinity chromatography

Affinity chromatography separates proteins on the basis of a reversible interaction between a protein and a specific ligand coupled to a chromatographic matrix. Both His and

GST tagged proteins were used for affinity chromatography. Cellular lysates of one-litre *E. coli* cultures were incubated with 2-5 ml of 50 % Ni-NTA slurry (Qiagen)/ glutathione sepharose fast flow (Pharmacia) which are pre-equilibriated with 5 column volume of respective lysis buffer. Incubation time varies from 3 hrs to overnight. The protein bound resin was then washed and protein eluted with their respective buffers (see materials).

2.3.2 Ion exchange chromatography (IEC)

Ion exchange chromatography separates proteins according to their net charge. mono S (cation exchange) as well as mono Q (anion exchange) column purifications was performed during this study. Purification parameters for each column were chosen based on the extent of purity of protein to be fractionated and the isoelectric point of the protein to be purified from other contaminant proteins.

2.3.3 Gel filtration chromatography

Further purification was done with gel filtration chromatography, which separates molecules according to their hydrodynamic radius. Unlike ion exchange or affinity chromatography, molecules do not bind to the chromatography medium so buffer composition does not directly affect resolution. A significant advantage of gel filtration is that conditions can be varied to suit the type of protein. Superdex™75 prepgrade, Superdex™200 prepgrade columns (Pharmacia) were used for purifying proteins during this study.

2.3.4 SDS-polyacrylamide gel electrophoresis

SDS-PAGE under denaturing and reducing conditions were used to separate proteins based on mass. Mass of desired proteins worked on during this study and the effectiveness of each protein purification scheme was analyzed by this method. Protein samples to be resolved on SDS-PAGE were prepared by addition of 5x sample buffer (10 % SDS, 1.5 M β-mercaptoethanol, 250 mM tris-HCl pH 6.8, 50 % glycerol and 0.5 % bromophenol blue) to the final concentration of 1x followed by incubation for 3 min at 98°C. 10-20 µl of this sample was then loaded on the gel. For the present study, 10-15 %

tricine gels were used and electrophoretic separations were always performed by applying a constant current of 125 V.

2.4 Methods to detect and estimate proteins in gels

All proteins purified during this study were separated on SDS-PAGE gels and detected by western blot and coomassie blue staining.

2.4.1 Western blot

Proteins were resolved on SDS-PAGE and then transferred onto the nitrocellulose membrane with the aid of semi-dry electroblotting apparatus. Protein transfer from the gel onto the nitrocellulose membrane was carried out at 110 mV of constant current for one hour. Non-specific binding sites for immunoglobulins on the membrane were blocked with skimmed milk powder. The membrane was initially incubated with primary antibodies (1/ 1000 dilution in 1 % skim milk) for few hrs at 4˚Cand then with alkaline phosphatase conjugated secondary antibodies (1/ 2000 dilution in tris buffer). Membrane was thoroughly washed with Tris buffer and then developed by incubating with BCIP/ NBT alkaline phosphatase substrate for 15 to 30 min.

2.4.2 Coomassie blue staining

Proteins resolved on polyacrylamide gels were simultaneously fixed and stained in the coomassie staining solution (Sambrook et al.1989). After rinsing the gel briefly in water, the gel was destained in the destaining solution, which eliminates the blue background while the proteins retain the blue colour. Gels were dried and/ or scanned for further analysis.

2.4.3 Bradford assay

The concentration of proteins in solutions was estimated with the assistance of the Bradford reagent (BioRad; Bradford, 1976). 10 µl of the protein solution to be measured was added to 990 µl of 5X diluted BioRad-reagent working solution. As control similar

mixture was prepared with 10 µl water instead of the protein solution. OD was measured at 595 Å and subsequently converted into the protein concentration on the basis of a BSA calibration curve.

2.4.4 UV determination

Precise protein concentrations were determined by finding out the absorbtion maxima of the protein by performing a UV scan from 340 Å to 240 Å, and then the concentration was calculated using the Beer-Lambert law. Extinction coefficients were theoretically calculated.

2.5 Crystallization methods

Purified proteins were concentrated in crystallization buffer to high concentrations (5-50 mg) using membranes with appropriate molecular weight cutoffs (Amicon, Viva science), crystallization trials were then set up using all the available screens (Hampton, Emerald, Jena and in house) at protein to reservoir ratio of 1:1 using the sitting drop method. Reservoir volume varied from 500 to 100 µl depending on whether 24 well or 96 well plates were used and protein amounts varied from few µl to 100 nl. In the latter case crystallization robot (Cartesian) was used which gives an enormous advantage of being able to screen large number of conditions with very little protein (~30 µl/ 100 wells) and time (~ 10 min/ 96 well plate) and is the basic need for high throughput crystallography. Once preliminary crystallization conditions were found, further optimization was carried out by putting up grid screens, which were set around the preliminary conditions with varying parameters like pH precipitant concentration, protein to reservoir ratio and temperature, until good diffracting crystals were obtained. The crystals were then harvested using appropriate loops/ capillary and then flash frozen using liquid nitrogen and measured at the synchrotron facilities at DESY (Hamburg) and SLS (Switzerland).

2.6 NMR sample preparation

For the NMR experiments the protein solutions were concentrated with a Centricon10 (Amicon) to the volume of 450 µl: 50 µl D_2O (99.9 %) was then added to the

sample. The sample concentration ranged from 0.1 to 0.8 mM. All NMR spectra were acquired at 300 K on a Bruker DRX-600 spectrometer. 1H-^{15}N-HSQC spectra (Mori et al., 1995) were recorded with 128 increments in the indirect ^{15}N dimension with scans varying from 4 to 1024 depending on the concentration of individual samples. Measurement times ranged thus from 2 to 24 hrs. Processing and analysis of the spectra was performed using the programs Xwinnmr (Bruker) and Sparky (Goddard & Kneller, 2001), respectively.

2.7 ITC experiments

For the ITC experiments the proteins were dialysed in the ITC buffer (25mM Tris, pH 8.0, 250 mM, NaCl, 10mM β-mercaptoethanol, 2.5 % glycerol and 1 % DMSO) and used at a concentration of 20 to 60 μM. All Inhibitors were dissolved in 1 % DMSO and made to a final concentration of 1 mM to 500 μM. Control experiments were set up with just the buffer and inhibitor to observe any heat change due to solute dilution. Typical ITC run conditions are given below

Run parameters

Injections:	30-40
Volume:	5 μl
Duration:	10 sec
Interval:	420 sec
Temprature:	25˚C

All runs were set up with a protein to inhibitor ratio of 1:10 to 1:20. ITC results were analysed with the program Microcal Origin.

2.8 CD spectroscopy

CD spectrum was collected on a jasco J-715 spectropolarimeter using 0.1-cm path length cuvette. The protein concentration was estimated according to Beer-Lambert Law. The spectrum was converted to mean residue ellipticity after correcting with the corresponding buffer (25 mM Tris, 250 mM NaCl, 3 mM DTT, 0.25 % Triton, pH 7.6)

3 Results and Discussion

3.1 Structural studies on histone deacetylases

3.1.1 Structural and biochemical studies on HDAC homolog HDLP

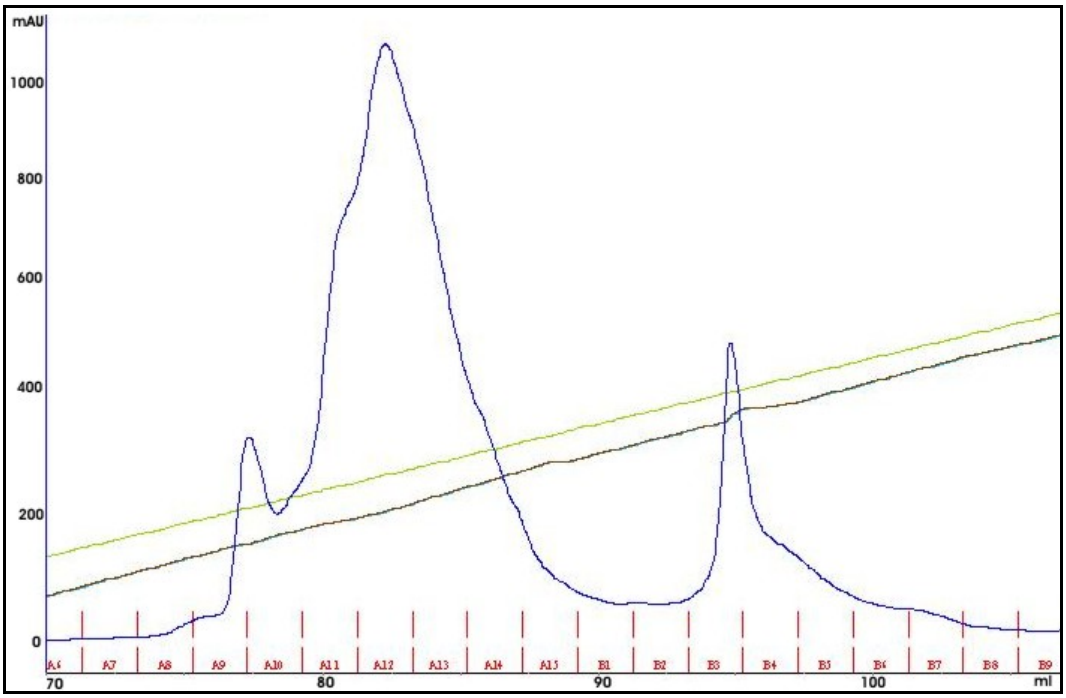
Structural and biochemical studies were done with HDLP protein from the thermophilic archaebacteria *Aquifex aeolicus.* The deacetylase core of HDACs includes the *A. aeolicus* homolog, which has a 35.2% identity with the mammalian HDAC1.

3.1.1.1 Cloning, expression and crystallization of HDLP

The initial construct having a synthetic HDLP gene was provided from Roche Diagnostics in a modified pGEX vector (Construct A) containing a modified N-terminal with a thrombin cleavage site followed by a linker and factor Xa cleavage site. It also contained specific site mutations Cys75Ser and Cys77Ser. The gene was subcloned into a His tag containing pET 15b vector between BamH1 and Hind III restriction sites and the modified N-terminal linker region was removed using site directed mutagenesis (Construct B). Active site mutants Tyr297Phe/Tyr297Leu and F141Y/F141L were generated by site directed mutagenesis kit (Stratagene).

Construct A: Protein expression trials were first conducted on construct A, which was expressed as a GST tagged HDLP protein in *E. coli.* The protein was found to be best expressed after induction with 1mM IPTG and incubation at 37˚C for 3 hr followed by harvesting. Protein purification was done under native conditions. Cells were lysed by sonification in GST lysis buffer, followed by high-speed centrifugation (65,000g, 1 hr) after which the supernatant was applied to GST bed and eluted following the principles of affinity chromatography. This was followed by proteolysis of the GST tag by thrombin protease (10units/mg). Further purification was accomplished by ion exchange chromatography using the Mono Q anion exchanger. The protein was initially dialyzed with Mono Q buffer A (25 mM Tris, pH 8.0, 0 mM NaCl, 10 mM β mercaptoethanol and 5% glycerol) and then bound to the column. Elution was done by salt gradient with buffer B which was similar to buffer A but had 1 M NaCl (Fig 3.1).

A



B

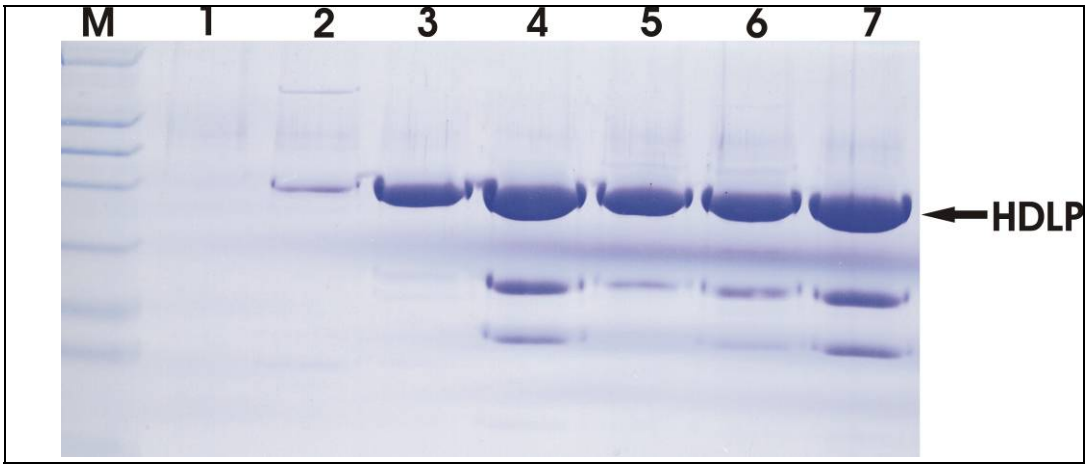
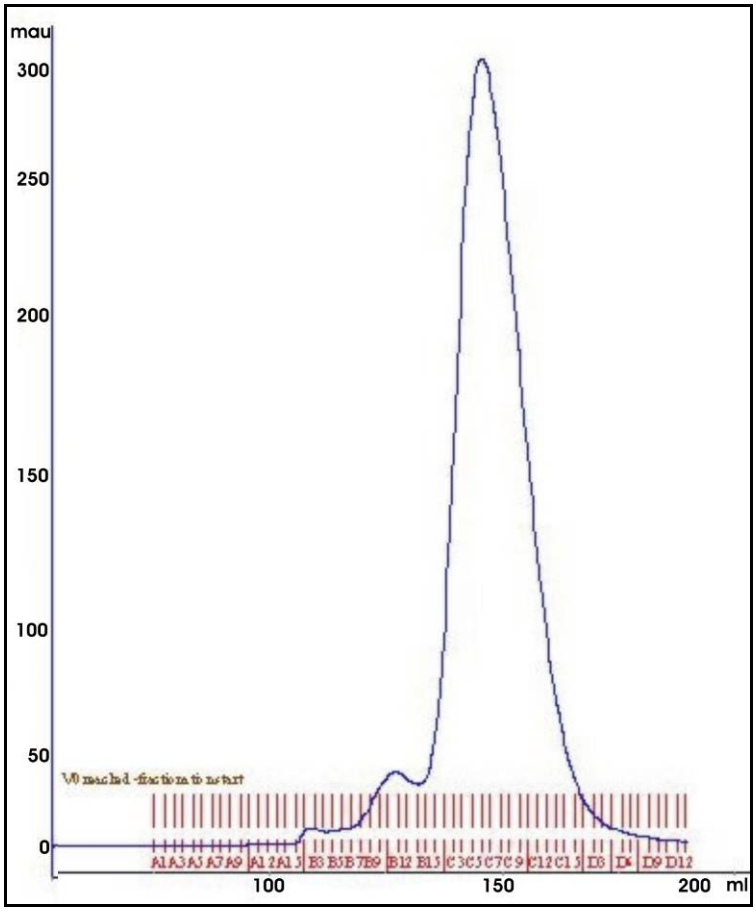


Figure 3.1 HDLP purification by ion exchange chromatography. (a) Elution profile of HDLP monitored at UV 280 (blue line), the concentration(yellow line), conductivity (brown line) and the fractions (red sticks) are also shown. (b) SDS PAGE showing fractions from under the observed mono Q peak showing HDLP.

A



B

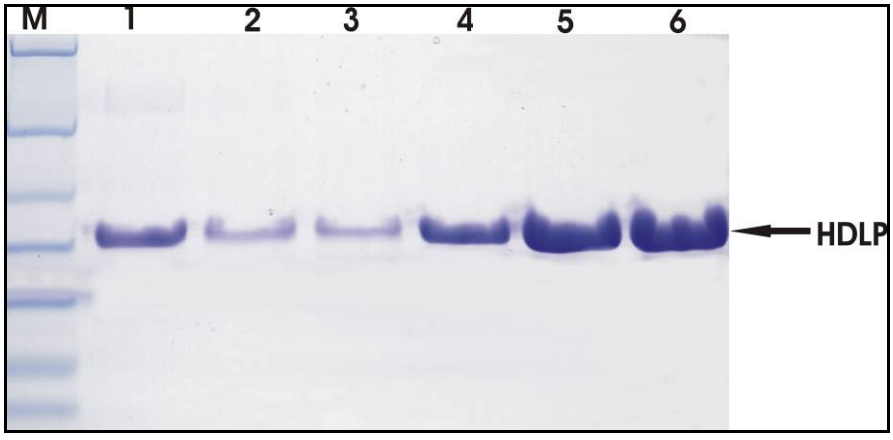


Figure 3.2 HDLP purification by gel filtration chromatography. (a) Elution profile of HDLP monitored at UV 280 (blue line) and the collected fractions (red sticks) are shown. (b) SDS PAGE showing fractions from under the observed gel filtration peak peak showing HDLP.

Gel filtration chromatography was then used as a final step purification using a Superdex 75 column. The running buffer used was 25 mM Tris pH 8.0, 250 mM NaCl, 2.5 % glycerol and 10 mM β mercaptoethanol (Fig.3.2). The purified protein was then reconstituted with Zn, which is essential for HDAC activity; by incubating the protein with 5 fold molar excess of Zn followed by desalting. The protein was then concentrated to ~10 mg/ml and co-crystallization trials were set up with different inhibitors provided by Roche Diagnostics.

Construct B: was used to express a His-tagged HDLP protein. Expression and purification conditions were similar to construct A the only change was that the affinity chromatography step was done with native Ni-NTA buffers instead of native GST buffers.

3.1.1.2 1D-NMR studies

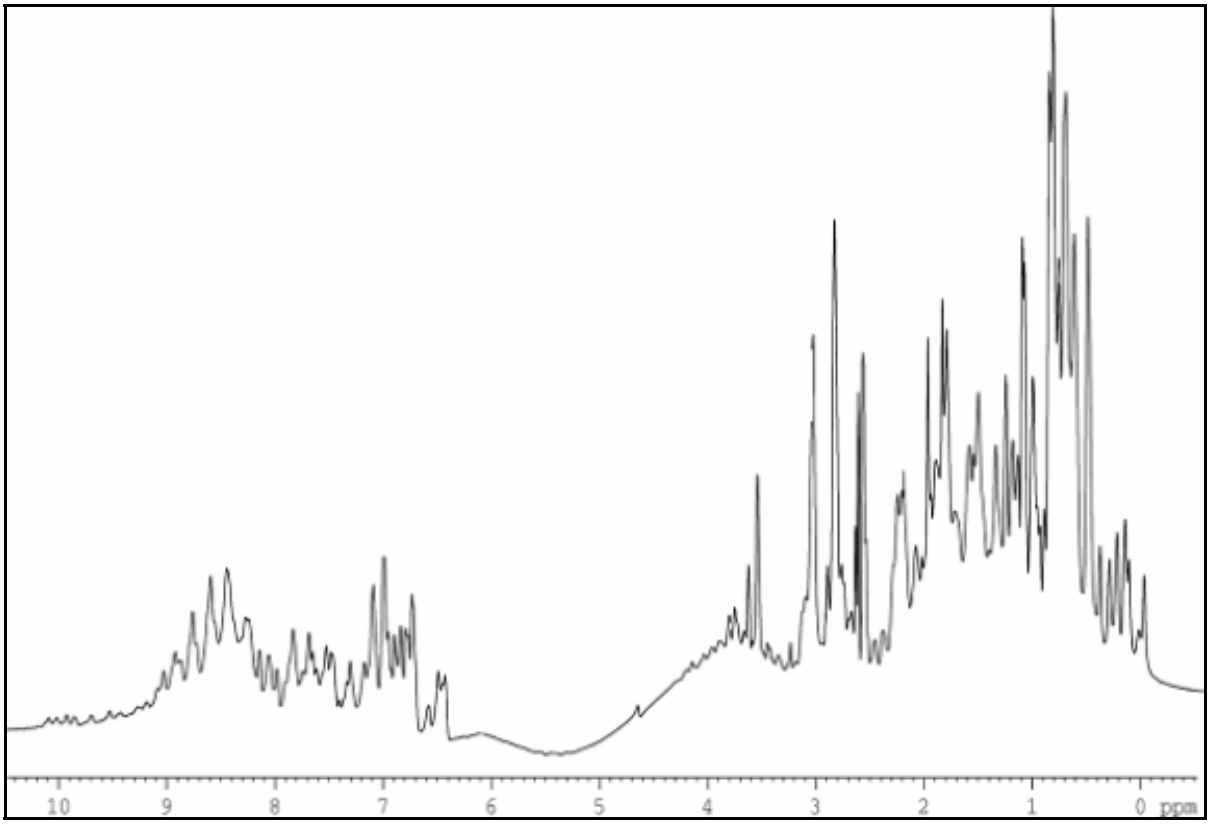


Figure 3.3 Characterisation of HDLP by 1D-NMR spectroscopy. Signal dispersion downfield of 8.3ppm and upfield of 1ppm show a well folded protein.

1D-NMR gives valuable insights to protein folding. Only folded proteins tend to retain their complete activity and 3-D structure (Rehm et al. 2002). Hence knowing the folding status of HDLP was important for both ITC and crystallization experiments. The NMR spectra (Fig.3.3) shows a folded HDLP protein with signal dispersion downfield (left) of 8.3 ppm and upfield (right) of 1 ppm. These peaks are characteristics of a folded protein. The protein was further concentrated for crystallization trials.

3.1.1.3 Screening for potential HDACI

We screened an array of structurally diverse inhibitors by isothermal titration calorimetry (ITC).

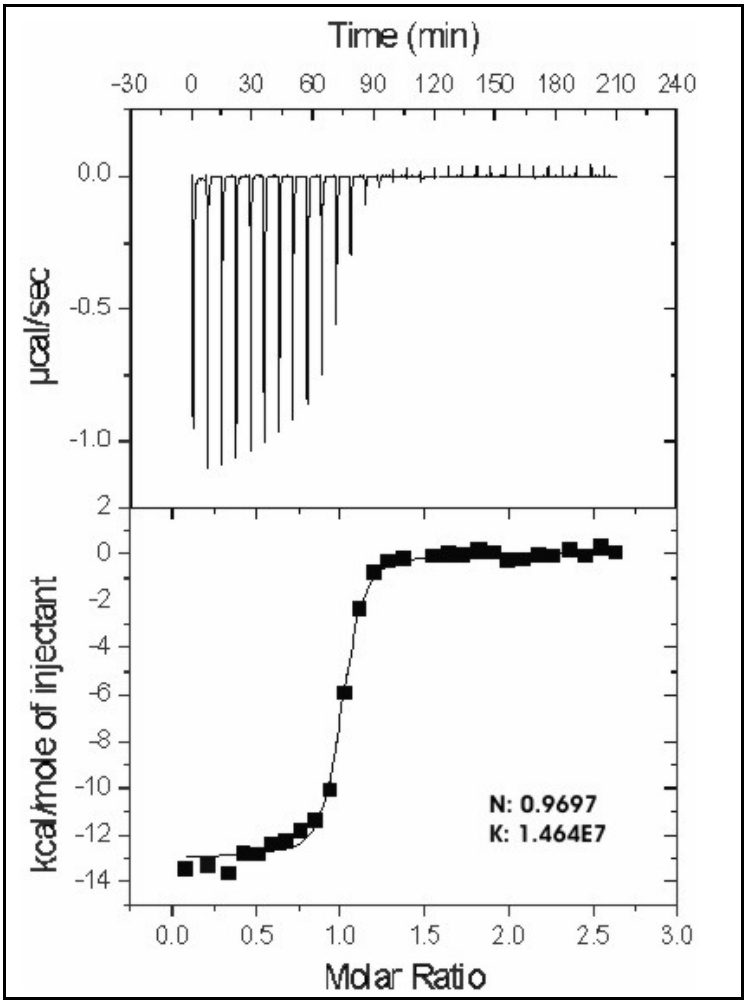


Figure 3.4 ITC showing the binding curves of HDACI to HDLP. The curve was obtained from the thiophene based hydroxamic acid group inhibitor and shows tight binding.

The inhibitors were designed to bind to the active site of the protein based on the previous known protein structure: Most of the inhibitors showed differential binding affinities ranging from micromolar to millimolar concentrations and thus were not good for crystallization trials. One class of inhibitors, the thiophene group containing hydroxamic acid based inhibitors, however showed binding in low nanomolar range 55-75 nM (Fig. 3.4) and thus was chosen as the lead compound for crystallization trials. The inhibitor consists of a metal binding domain which comprises the hydroxamic group, a linker domain consisting of the thiophene ring and a capping domain, which is denoted by the R group (Fig. 3.5).

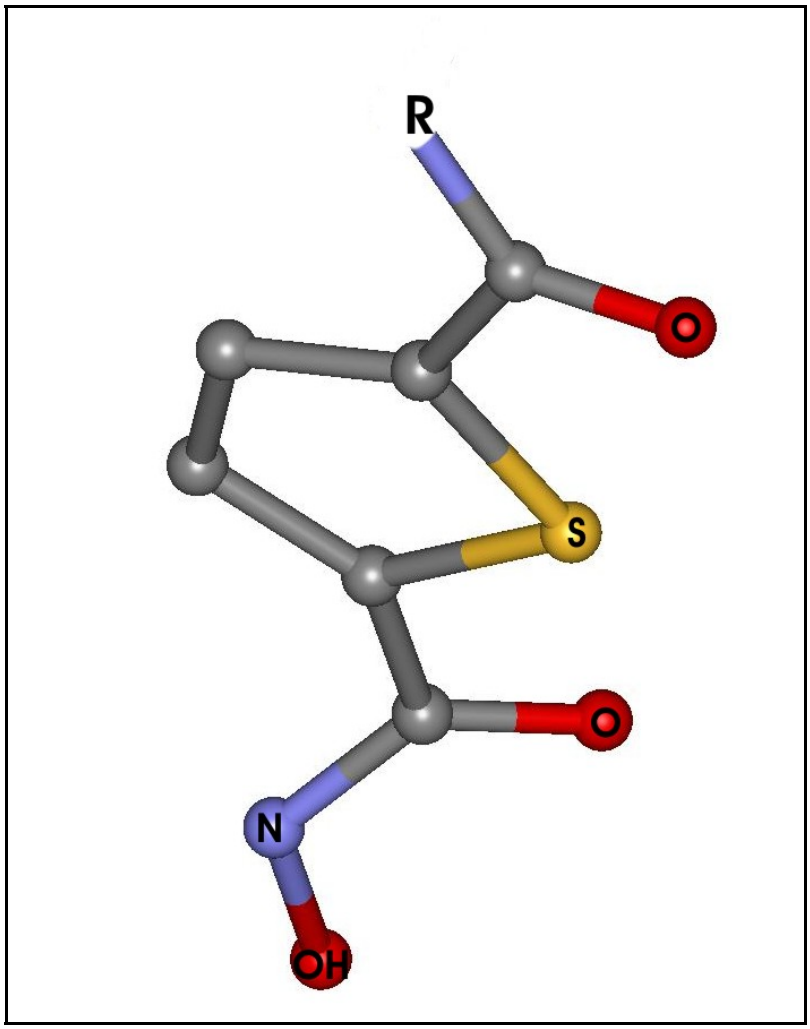


Figure 3.5 Ball and stick representation of the inhibitor highlighting major groups. R stands for various substituents, which were used at that position.

3.1.1.4 Crystallization set up

Crystallization trials were set up using classical sitting drop vapor diffusion by mixing 5 fold molar excess of inhibitor (dissolved in 1% DMSO) to protein and reservoir at a ratio of 1:1:1. All available screens were tried. The crystals were measured at the Swiss light source and on the BW6 beam line at Doris (DESY, Hamburg, Germany) using the 156-mm MarCCD detector (Mar-USA, Evanston, IL) at 100K and were processed with MOSFLM.

Construct A: Crystallization trials succeeded in conditions of 10% PEG 8000, 35% MPD, and 100 mM Tricine pH: 9.0. The conditions were further optimized to give single well defined crystals. Needle shaped crystals appeared overnight and grew to a size of ~ 0.2 x 0.1 x 0.2 mm within 2 weeks. The crystals diffracted to a resolution of 2.4 Å and analysis showed they belonged to space group P1. The crystals were found to contain protein-Zn complex but there was no inhibitor bound to the active site. Many attempts were then tried to obtain a co-crystal with the inhibitor including micro seeding, soaking, and adding inhibitor prior to gel filtration.

Construct B: Crystallization trials succeeded in a buffer containing 34% PEG 1200, and 25mM HEPES pH: 8:0. Flat plate shaped crystals appeared within hours of plating and grew to maximum size of 0.3x0.2x0.1 mm within a week (Fig. 3.6). Crystals diffracted to 2 Å and were found to contain inhibitor bound to the active site and the structure of the co-crystal structure was solved**. .**

3.1.1.5 Crystal packing

We solved the co-crystal structure of the HDLP-HDACI complex to 2 Å resolution. The structure reveals a totally new crystal packing then previously observed (Finnin et al. 1999) consisting of 4 monomers in the asymmetric unit and belonging to space group P1. The crystal diffracted to a high resolution and the diffraction pattern from of the set of images is shown in Fig. 3.7. The electron density map around the inhibitor and the surface representation of the 4 monomers indicating their respective positions in the asymmetric unit is shown in Fig. 3.8.

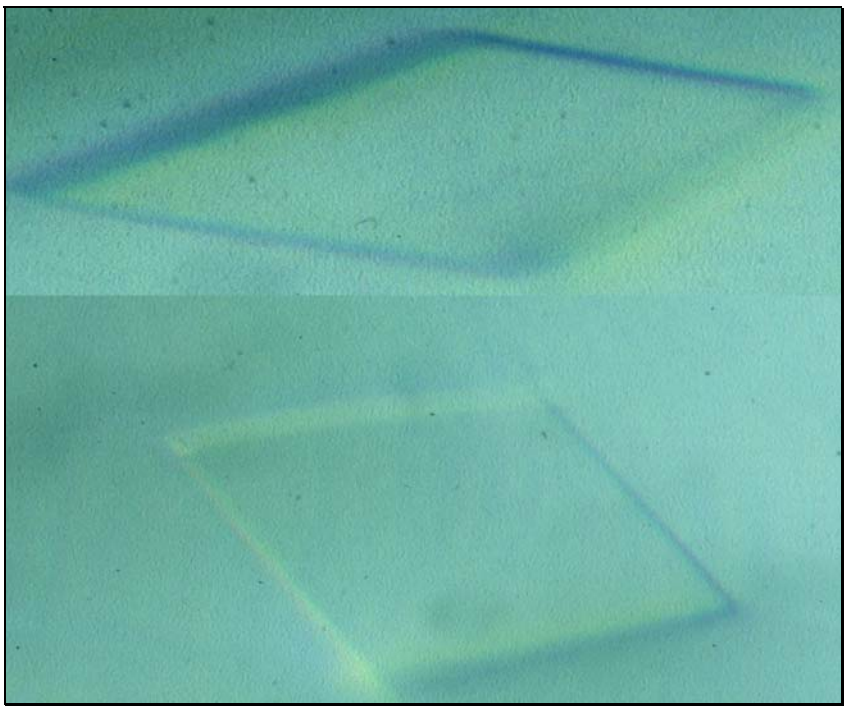


Figure. 3.6 Photograph showing the characteristic shape of the HDLP crystals obtained in the study.

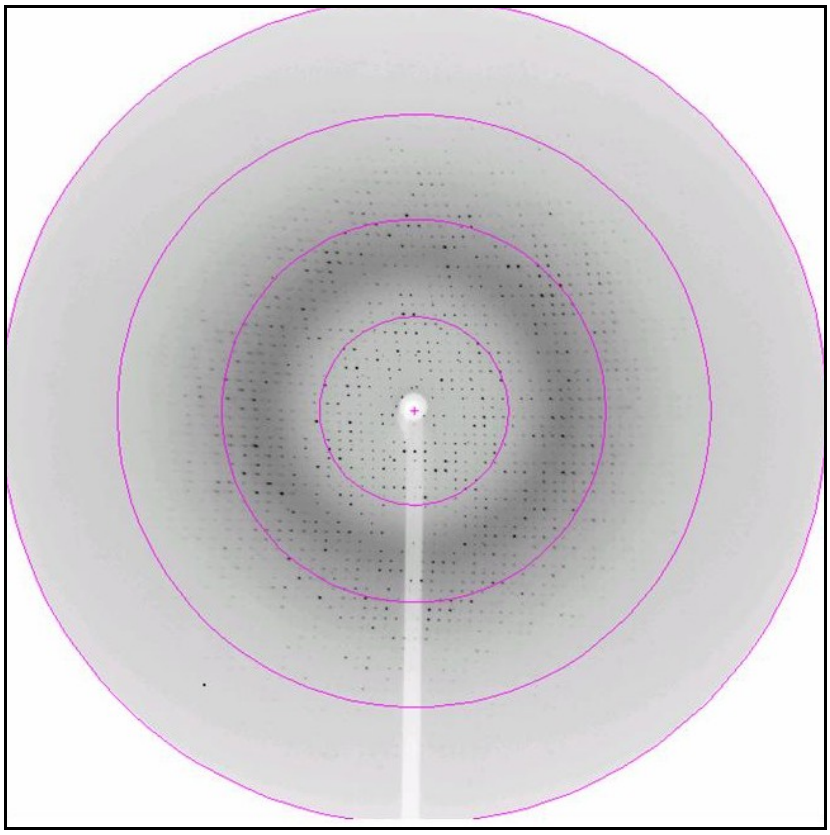


Figure 3.7 Diffraction pattern Represents one typical diffraction pattern image from the entire data set.

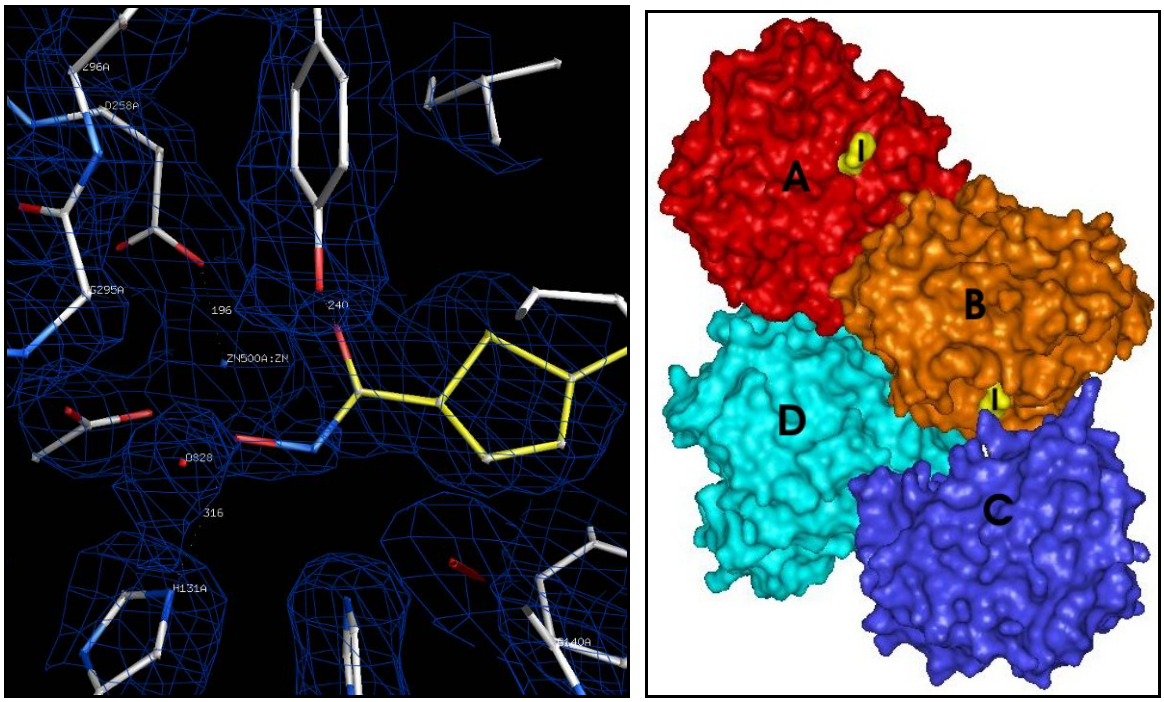


Figure 3.8 A 2Fo-Fc Electron density map. showing a close up view of the HDLP-Zn-HDACi interaction. Inhibitor is shown in yellow and is seen hydrogen bonding with Zn and active site residues. B. Surface representation of the HDLP inhibitor complex: yellow surface represents the inhibitor (I), and the remaining colours show the surface of the four monomers (A, B, C, D) of HDLP in the asymmetric unit.

It is unlikely that this tetramer is found in solution or is caused by inhibitor binding as gel filtration, dynamic light scattering and mass spectroscopy both prior to and after adding inhibitor show only the molecular weight of monomeric HDLP monomer. The monomers interact at their interfaces by the formation of salt bridges in the unit cell of a crystal**.** Each monomer binds one inhibitor molecule; one Zn ion and two sodium ions the crystallographic data and refinement statistics are shown in Table 8.

3.1.1.6 Overall structure

The overall protein structure shows that HDLP consists of a single α/β domain which includes a central 8 stranded parallel β sheet surrounded by 11 α helices. (Fig.3.9). Four helices pack on either face of the β sheet and the rest are clustered to one side of the β sheet. From the carboxyl end of the β strands emerge large well defined loops and these

loops along with the α helices give rise to two important structurally prominent features: a deep, tube shaped pocket and an internal cavity next to the pocket.

Table 8. Shows the crystallographic and refinement data of our inhibitor and compares it to the previously published HDLP:TSA and HDAC8:TSA structures.

	HDLP:TSA	HDLP:HDACI	HDAC8:TSA
Data			
Space group	P212121	P1	P43
Cellparameters(a, b, c, α, β and γ)		76.6, 80.1, 85.9 90, 107, 108	81.0, 81.0, 114
Resolution range	2.1	2	1.9
Completeness (%)	93.8	98	99
Total reflections	180,427		241,421
Unique reflections	50,796	114,204	54,547
Rsym(%)	7.1	7	5.8
Model			
Molecules in assymetric unit	2	4	2
Other Ligands	2Zn	4Zn, 10Na	2Zn, 3Ca, 4 Na
R factor	22.4	22	17.2
R free	25.8	28	22.1
Rms deviations from ideal geometry			
Bonds(Å)	0.008	0.015	0.01
Angles(°)	1.78	1.6	1.27

The pocket has a depth of 11 Å, it constricts in the middle to 4.5 Å by 5.5 Å and then again widens out at the bottom. At the bottom of the pocket lies the the catalytic site of the enzyme.The pocket walls are mainly covered with hydrophobic and aromatic residues which are highly conserved among the family of HDACs (Pro 22, Gly 140, Phe 141, Phe 198, Leu 265 and Tyr 297). Of particular importance are Phe 141 and Phe 198, whose phenyl groups face each other in parallel at a distance of 7.5 Å, marking the most slender portion of the pocket. The internal cavity is also formed by the loops and it measures approximately 13 Å by 5 Å and reaches a depth of 15 Å. It is primarily lined with glycine residues the the aromatic residues Phe 141 and Tyr 297 which line the catalytic pocket. The physiological role of this cavity might be to act as a sink for the diffused acetate product

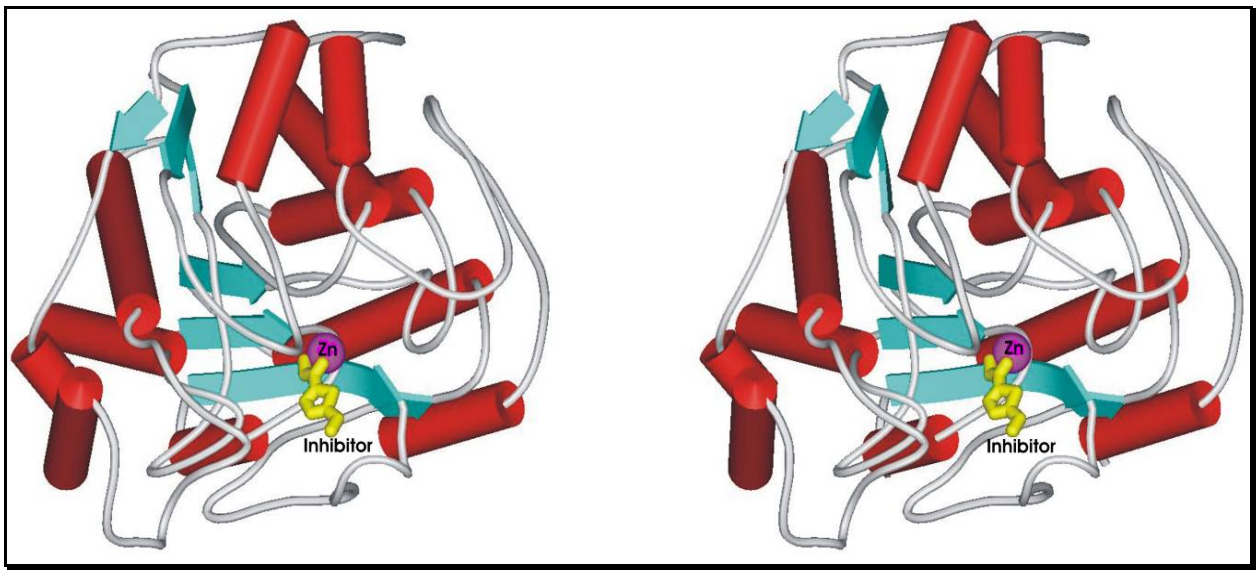


Figure 3.9 The secondary structure of HDLP showing the inhibitor and Zn ion binding site.

3.1.1.7 Catalytic site and inhibitor binding

The catalytic machinery lies at the bottom of the pocket, at the center of this region lies a Zn ion, which is coordinated by the carboxylate group of Asp 166 and Asp 258 and by the N δ1 atom of His 170. Other key residues at the base of the pocket include two histidines (His 131 and His 132), two aspartic acids (Asp 166 and Asp 173) and a tyrosine (Tyr 297). Each of the histidines hydrogen bonds through its N δ1 atom to the carboxylate oxygen atom of the respective aspartic acid residues thus forming a His-Asp charge relay system characteristic of the active site of serine proteases, where it polarizes the imidazole Nε and increases its basicity. The His 131-Asp 166 charge ralay pair is more deeply buried in the pocket than the His 132-Asp 173 charge relay which is partially exposed to solvent. The Tyr 297 residue is positioned next to the zinc, opposite to the two charge-relay systems. Its hydroxyl group lies 4.4 Å from the zinc atom. In the crystal structure of HDLP presented here the inhibitor is seen to bind the protein by inserting its long hydrophobic chain into the catalytic pocket. The hydroxamate moiety then coordinates the Zn ion at the bottom of the pocket in a bidentate fashion (Fig. 3.11) via its carbonyl and hydroxyl groups forming a pentacoordinated Zn ion. It also simultaneously coordinates with Y297 (2.3 Å) and both the charge relay histidines 131 and 132 (2.8 Å, 2.7 Å).

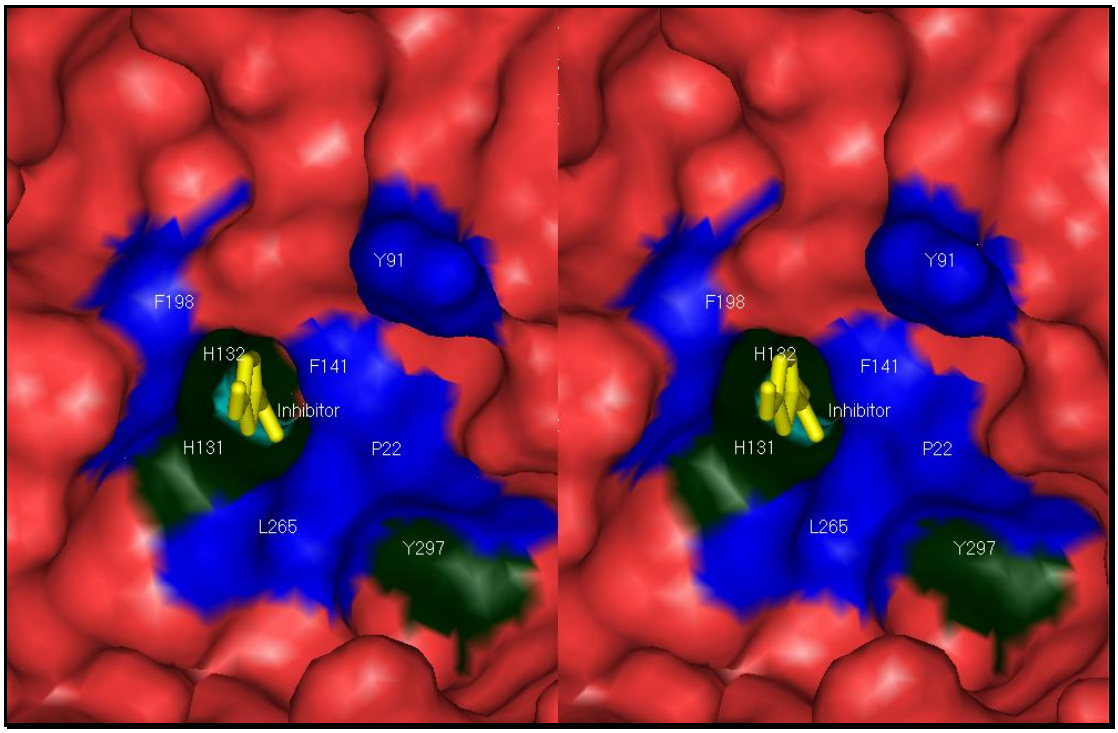


Figure 3.10 Surface representation of HDLP-inhibitor contacts

Blue represents residues lying on the walls of the catalytic pocket; green color represents residues interacting with inhibitor at the bottom of the catalytic site

This is followed by the linker region this region fits in the hydrophobic channel of the protein with its thiophene ring sandwiching itself between the the side chains of Phe 141 and Phe 198 and is involved in aromatic interactions with the two aromatic residues the linker region is further involved in van der Waals interactions with the side chain of leu 265 and main chain of Gly 140 of the protein, the rest of the inhibitor has been shown by the R-which is primarily involved in capping the entry of the pocket and interacts mainly with residues Pro 22 and Tyr 91 of the protein. The different R groups tested did not change the binding affinity of the inhibitor: protein interaction and hence are not discussed individually.

3.1.1.8 Novel sodium binding sites

Our HDLP protein structure reveals 2 metal binding sites occupied by sodium ion in our structure. The sodium ions Na1 and Na11 are located at a distance of 7 Å and 10 Å

respectively from the Zn^{2+} ion at the active site. Both sodium ions are bound in a octahedral geometry by six oxygen atoms of the protein, Na1 by the carboxylate of D 166, the side chain hydroxyl group of S 189, and the backbone carbonyl group of residues D 166, D 168, H 170 and L 190. Na11 is coordinated by the side chain hydroxyl group of S 263 and the backbone carbonyl group of residues 260, 261, 263, 266 and 268 (Fig. 3.11). The coordinating distances vary from 2.5-3.0 Å, which is commonly observed for coordination geometry of the sodium ion.

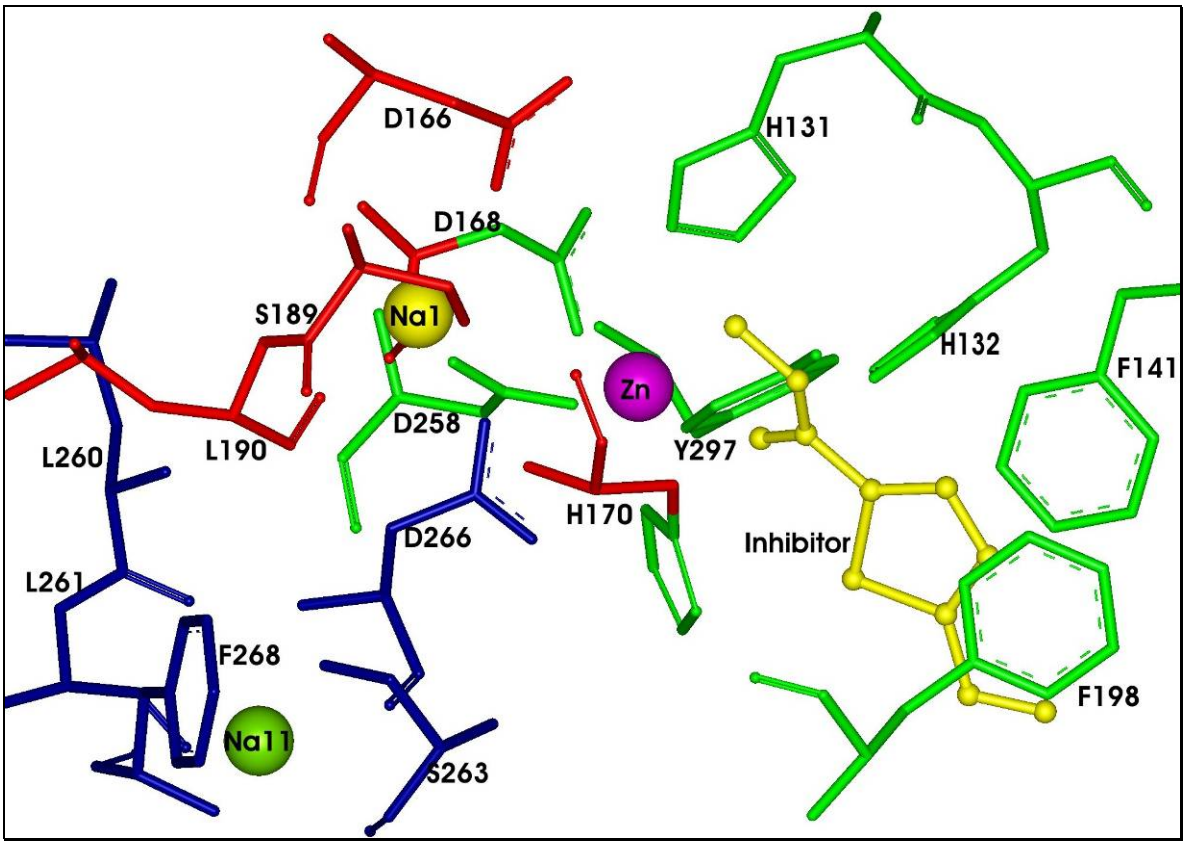


Figure 3.11 Close up view of HDLP: showing residues (green) coordinating with the Zn (pink sphere) and inhibitor (yellow) at the active site and residues (green and blue) coordinating with the proposed Na binding sites: Na1 (yellow sphere) which lies at a distance 7 Å from the active site and Na11 (green sphere) which lies at a distance of 10 Å.

It is also seen that 2 of the residues D168 and H170, which interact with Na1, also interact with the Zn ion at the active site. Another weak metal binding site, which was partially occupied by sodium was found 15 Å from the active site and was coordinated by carbonyl

oxygen of residues 179, 182, 185 and 216, and by two water residues. But since only two monomers contained sodium at this position it was not considered a major metal binding site.

3.1.1.9 Analysis of active site mutants

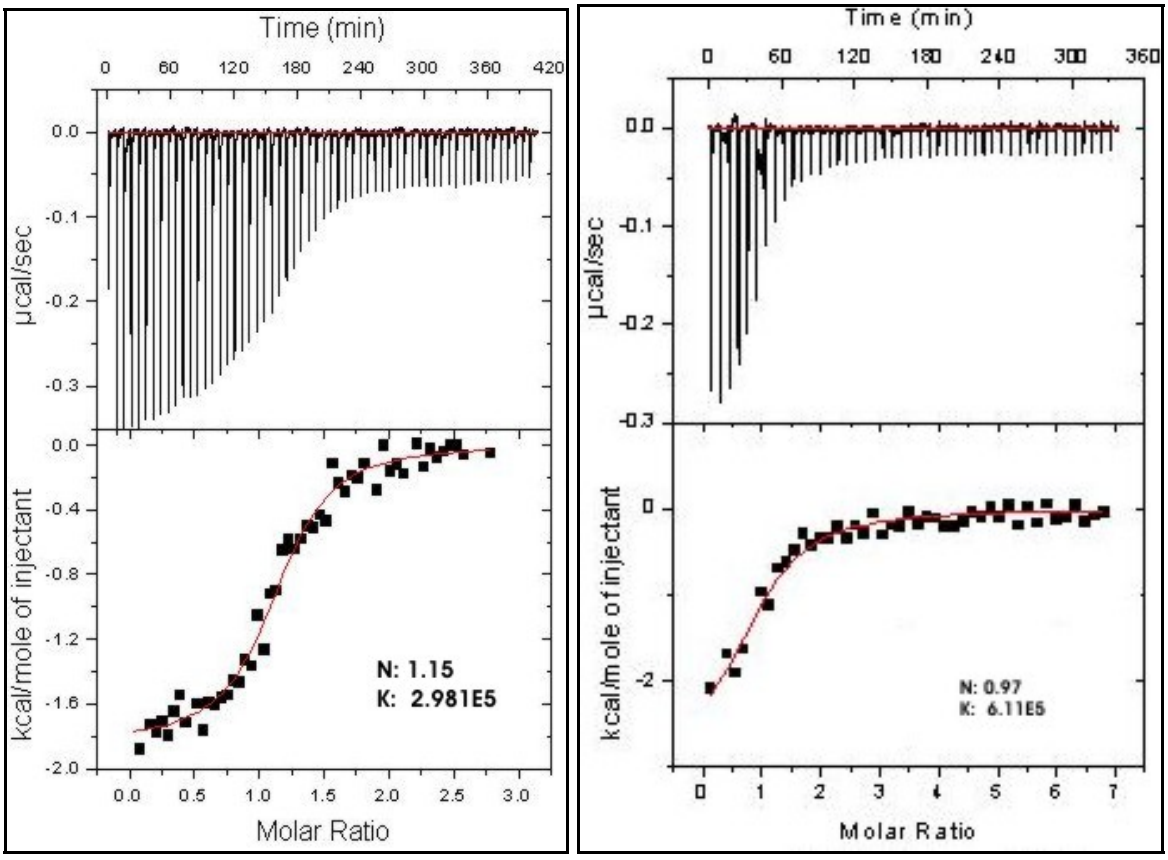


Figure 3.12 ITC analysis of mutants. (a) Mutant F141Y shows weaker binding then wild type, HDLP (b) Mutant Y297F shows further weakening of binding affinity.

To determine which residues were crucial for inhibitor binding, active site mutants Tyr 297/Phe 297, Tyr 297/L 297 and mutants Phe 141/Tyr 141 and Phe 141/Leu 141 were generated and their interactions with inhibitor were characterized by ITC (Fig. 3.12). Results show that the active site mutant Tyr297 /Phe 297 mutation reduces the binding affinity of the inhibitor by 100 folds whereas Tyr 297/Leu 297 abolishes the interaction. This is consistent with the fact that tyrosine 297 makes a hydrogen bond with the hydroxamic group of the inhibitor and this interaction is lost when the tyrosine residue is mutated. The other mutations at the Phe 141 site were done to test the hypothesis that the

thiophene moiety of the inhibitor is involved in aromatic interactions with the aromatic residues Phe 141 and Phe 198. Results showed that the Phe 141/ Tyr 141 leads to a 50 fold decrease in the inhibitor binding affinity and the Phe 141/ Leu 141 mutant abolishes the inhibitor binding. Thus showing that there exists, specific aromatic interactions between the thiophene residue and the aromatic phenylalanine residues, which lead to increased binding affinity of the inhibitor. An NMR study of the mutants showed that these proteins were properly folded and weakening or abolishment of inhibitor binding was not due to any unfolding or aggregation of the mutant proteins.

3.1.2 Preliminary investigations on mammalian HDAC family

Due to lack of any structural information on any of the mammalian HDAC family members (at the beginning of the study) the study was initiated to obtain the structural information of class 1 mammalian HDACs (HDAC 1, 2, 3 and 8). HDAC8 was chosen as it lacks the extensive C-terminal domain found in the other family members, which may be responsible for binding to other co-repressor complexes. It has been shown earlier that HDACs 1 and 2 are inactive when produced by recombinant methods and in-vivo is active only in large protein complexes. HDAC8 shows 34% similarity with HDAC3 but is not known to be part of any complexes unlike HDAC3.

3.1.2.1 Cloning of HDAC8

The full length ~1.1 kb HDAC gene was fished out from a pool of cDNA (Stratgene) and cloned into a pET 30 LIC vector in a two step fashion, first a set of primers were designed against the full length HDAC8 and against residues 15 to 324 lacking any tags and after a successful PCR primers with tags for pET 30 LIC were designed and cloning into the vector was done according to manufacturer's instruction. Several HDAC 8 constructs were made and protein expression checked (Table 9). Also for better expression of protein the pET 30 LIC was modified by deletion mutation which shortened its N-terminal and removed the region between the His tag and start of HDAC8 coding region. The mutation was carried out using site directed mutagenesis kit (Stratagene).

3.1.2.2 Expression and Solubility Tests of HDAC8 Constructs in *E. coli*

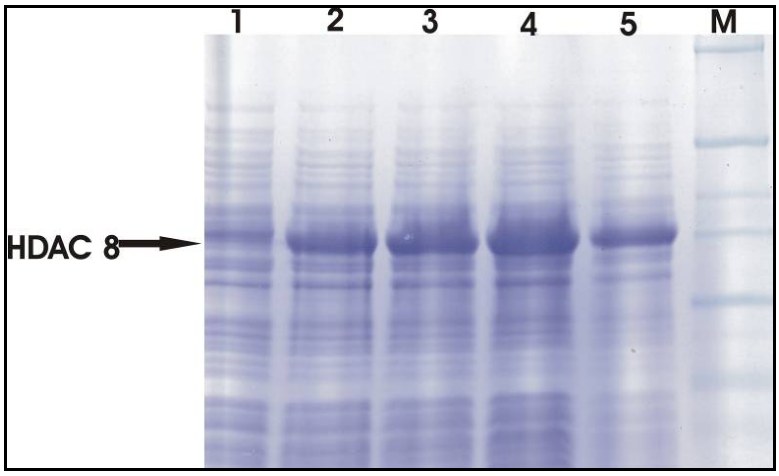


Fig. 3.13 Coomassie stained SDS PAGE.showing induction assay of the mutated His-HDAC8 (1-377) construct. C stands for sample before IPTG induction and 1, 2, 3 and 4 stand for (time intervals [hr]: 1, 2, 3 and 4)

Time course expression test: Protein induction was tested on various constructs by growing cultures in 50 ml LB-medium, containing appropriate antibiotics; OD_{600} was monitored until the value 0.6-0.7 was reached. At that time (t = 0) the culture was induced with 1mM IPTG (end concentration). The culture was grown overnight. 1ml samples for electrophoresis were taken before induction (C), and after 1, 2, 3 and 4, hours after induction (t = 1, 2, 3, 4). The samples were centrifuged and pellet was dissolved in 50 µl of the 2x SDS PAGE loading buffer (Sambrook and Russell 2001) and 10 µl was loaded on PAGE gel and observed with coomasie blue staining (Fig.3.13). The results show clear induction of the full length HDAC8 protein which is seen to increase to a maximum after the first 3 hrs of induction. The expression profile of the various constructs is summarized in (Table 9).

Solubility test Once the constructs were examined the best two constructs i.e. number 5 and 6 which contained the full length HDAC8 and the catalytic domain of HDAC8 respectively in a modified pET 30 LIC vector, were expressed for solubility test to find out whether protein was produced in the soluble fraction or in pellet as inclusion bodies. The bacterial cultures were grown and harvested for 3 hrs post induction. The pellet was re-

suspended in PBS buffer, the suspension was sonicated to disrupt the cells and then the suspension was centrifuged for 60 min at 4°C with an rpm of 20,000. The pellet was dissolved in 5 ml buffer containing 8 M urea, 100 mM Tris-HCl pH 8.0. Both supernatant and dissolved pellet were separately incubated with Ni-NTA resin equilibrated with lysis buffer and then were separately eluted in their respective elution buffers and further 20 µl samples were taken for PAGE and Western blot analysis. Results showed that most of the protein was produced as inclusion bodies and only a small fraction of protein was found in the soluble fraction. Trials were done to refold the protein from inclusion bodies but met with little success. The next approach was to try and get more fraction of the protein in the soluble fraction and this was achieved by growing the protein at 20 °C for 12 hrs after induction.

Table 9 Expression test of the His-tagged constructs of HDAC8

S.N0.	Construct	Vector	Expression
1	HIS-HDAC8 (1-377)	pET 30 LIC	low
2	HIS-HDAC8 (15-377)	pET 30 LIC	poor
3	HIS-HDAC8 (15-324)	pET 30 LIC	low
4	HIS-HDAC8 (1-324)	pET 30 LIC	poor
5	HIS-HDAC8 (1-377)	Muated pET 30 LIC	higher
6	HIS-HDAC8 (15-324)	Muated pET 30 LIC	higher

The protein was further purified by ion exchange chromatography using anion exchange column and gel filtration chromatography .Gel filtration showed that most of the protein was found in an aggregated fraction around >200 kDa and only a small part of the protein was found as a monomer, but quantity of the purified protein was not high enough to set crystallization trials as protein aggregated at higher concentrations so instead we studied the secondary structure of the protein by CD spectroscopy.

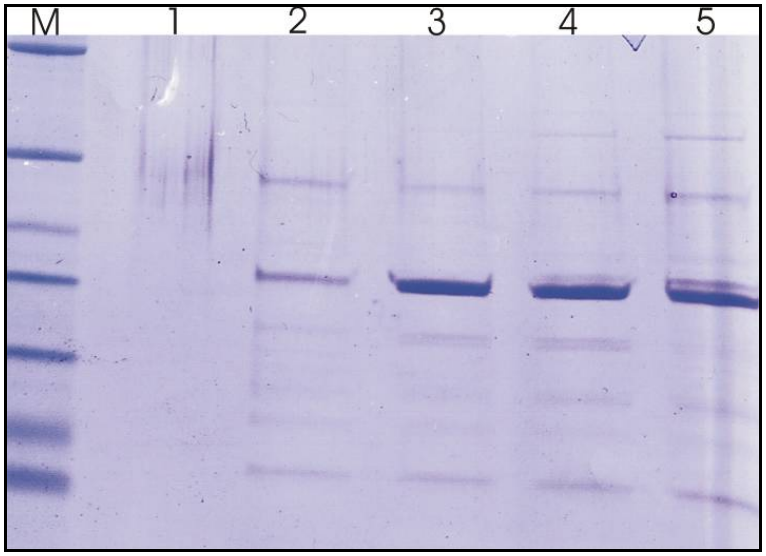


Fig. 3.13 .SDS PAGE showing purified protein fractions (lane 2-5) after ion exchange chromatography.

The CD spectra showed two minima at 208 and 222 nm which resemble signal seen with a protein containing both α-helices and β sheets thus the protein was shown to be folded. During the course of the study the first crystal structure of HDAC 8 was published complexed with inhibitors and it was shown that half of the protein structure comprised of canonical secondary structures elements and the other half is comprised of loops that link the regular secondary elements thus complementing our CD data.

3.1.3 Discussion

Analysis of HDLP mutants

The active site mutant Y297/ F297and Y297/ L297 studies showed interesting features about the inhibitor protein interactions, the conservative mutation Y297/ F297 showed a 100 fold reduction in the binding affinity, and this is consistent with the fact that the hydroxamic group makes hydrogen bond with the OH group of tyrosine and in the absence of OH is unable to do so thus highlighting the importance of Y 297 at the active site, mutation of Y297/ L297 completely abolished the interaction and this may be due to change in the integrity of the active site due to replacement of an aromatic ring by a hydrophobic aliphatic chain thus also highlighting the importance of the aromatic ring to

maintain structural integrity of the active site, the other mutation site was phenylalanine 141 and the objective was to study the aromatic interaction of the thiophene inhibitor with the aromatic rings of F141 and F198. The F141/ Y141 conservative mutation showed a decrease of binding affinity by 50 fold and the F141/ L141 mutant totally abolished the activity of the inhibitor thus highlighting the importance of the thiophene ring of the inhibitor. The reduction in activity associated with the F141/ Y141 mutation could be due to the orientation of the hydroxyl group which could lead to steric hindarance with the incoming cap group of the inhibitor.To further test as to whether the reduction in in binding affinity was indeed due to disruption of aromatic interactions we also tested another inhibitor SAHA which has been shown to be a potent HDAC inhibitor and is currently undergoing clinical trials. In our assay the binding affinity of SAHA to wild type HDLP was found to be approx 50 fold lesser to our inhibitor and was found in the low micromolar range and when titrated against the mutant F141/ L141 SAHA did not show any significant decrease in binding affinity this could be explained by the fact that SAHA does not possess the thiophene ring and instead as a aliphatic region at the same position and thus no additional aromatic interactions are observed. A similar inhibitor with a benzene ring instead of the thiophene ring was shown bound to the HDAC 8 crystal structure (Vannini et al. 2004). Thus our study highlights the importance of the aromatic interactions at the F141 and F198 containing region of the HDAC catalytic pocket in inhibitor binding.

Comparison of HDLP and HDAC 8 structures

Overall structure

The crystal structure of our HDLP structure was compared to the recently published HDAC8 structures (Somoza et al. 2004; Vannini et al. 2004). It shows that the structures are quite similar but do have some differences. The overall structure is similar and comprises of the 8 stranded parallel β sheets flanked by α helices. The architecture of the active site and its immediate surroundings are conserved between the two proteins and, based on sequence homology, are likely to be common to all zinc-dependent HDACs. Similar to HDLP, the active site Zn ion of HDAC8 is pentacoordinated, having Asp-178, His-180 and Asp-267 as ligands in addition to the hydroxamic acid group of the inhibitors.

The major difference is seen in the loop regions especially the N-terminal loop which differs (residues 30-36 of HDAC8 and residues 18-26 in HDLP) in both size and composition. In both the proteins this loop region lines a large portion of one side of the active site pocket and extends to the surface. This loop in HDLP has been shown to have some inherent flexibility (Finnin et al. 1999). In HDAC 8 this loop is shorter by two amino acids and the composition is also varied then other class 1 HDAC members. The result of these changes is that HDAC8 has a wider pocket opening and the loop is highly flexible. Inhibitor studies have shown (Somoza et al. 2004) that this region is highly flexible and can accommodate various ligands which are physiologically important. The loop in some of the HDAC8 structures has peeled away from the protein exposing the adjacent cavity, which was then found to be bound by another inhibitor molecule. Such a case was not observed with the HDLP structure in which the cavity was occluded by the loop. A sequence alignment of HDLP with other class 1 HDACs shows that the other enzymes of this family possess longer loops structurally similar to the one observed in HDLP and in such a case their catalytic center would be more rigid then HDAC8, structure. But further structural insights into these enzymes are needed to establish these results.

Metal binding sites

The sodium site Na1 in our HDLP structure is highly conserved among the HDAC family and with the exception of L 190 which is a phenylalanine in HDAC1, 2 and 3 all other residues are identical. The HDAC8 structures also show that the metal binding is coordinated by the same conserved residues (D 176, D 178, H 180, S 199 and L 200) although one group has a sodium ion bound at the position (Somoza et al. 2004) and another has a potassium ion bound (Vannini et al. 2004) which could be due to different buffers used for crystallization. The possible functions for this metal binding site in the mechanism of the deacetylation could be to provide a structural role by stabilizing the geometry of the active site. It also contributes to the positive electrostatic potential and could help stabilize the transition state. It may also play a role in the stabilization of the negatively charged acetate product and could also modulate the basicity of the histidine 131 through its interaction with aspartic acid 166 of the buried Asp166-His131 charge relay system. The Na11 binding site is only partially conserved with only residues 260,

263 and 268 being conserved among class 1 HDACs. The role of this site could be to provide structural stability to the active site. This site was not reported in either of the HDAC8 structures, but there was a second potassium site reported by Vannini et al which is identical to the partially occupied sodium site in our structure. The residues making this site are also highly conserved but their function is yet to be understood.

Mechanism of deacetylation

The HDLP and HDAC8 structures place a constraint on the possible catalytic mechanisms employed by the class 1 and 2 HDACs. The deacetylase active site has features of both metallo and serine proteases (Finnin et al. 2004). The probable first step in the reaction is a nucleophilic attack on the carbonyl carbon of the substrate, the carbonyl oxygen of the of the N-acetyl amide bond could bind to the zinc thus placing the carbonyl carbon in close proximity to the Zn bound water molecule, the Zn ion in coordination with tyrosine 297 could polarize the carbonyl group so that carbon is better electrophile, the Zn bound water molecule acts as a nucleophile as its pK_a is lowered by the Zn ion itself and also by the buried charge relay system of His 131-Asp166 to which it is bounded. In the HDLP- HDACI structure shown here the hydroxyl group of the hydroxamic acid mimics the Zn bound water. The resulting tetrahedral intermediate could be stabilized by the interactions with Zn and Tyr 297 and as mentioned previously the Na1 site could also play a role here. The collapse of the tetrahedral intermediate would cause the release of acetic acid and lysine products.

In conclusion the structural studies carried out in the thesis on HDAC family members have shown the mechanism by which the thiophene based inhibitor binds to the catalytic site of HDACs and highlighted the importance of aromatic-aromatic interactions in the linker region which can be exploited in future drug designs and thusof a basis for further design of such aromatic group based inhibitors as antitumour agents. In future it would be interesting to study the structure of the C-terminal of class 1 HDACS and members of the class 2 HDACs so as to be able to develop specific inhibitors for a particular member of the family

3.2 A novel medium for expression of proteins selectively labeled with ^{15}N-amino acids in *Spodoptera frugiperda* (*Sf 9*) insect cells

3.2.1 Formulation of insect cell culture media

For protein-ligand binding studies or for structure determination with NMR spectroscopy on proteins expressed in insect cells the use of selectively labeled protein samples is essential. While bacterial expression systems are widely used for production of uniformly or selectively ^{15}N-labeled proteins the usage of the baculovirus expression system for labeling is limited due to lack of a suitable labeling media. A medium for the selective ^{15}N-amino acid labeling in insect cells has to fulfill several criteria:

- The quantities of amino acids should be reduced to a minimum for high cell densities and protein yield.
- If possible, the use of serum or yeastolate should be decreased without negative influence on protein yields.
- The formulation should be known and readily available. In this study two insect cell media were formulated namely: IML 406 and 455, for the high-yield production of selectively ^{15}N-labeled proteins in insect cells. The protein glutathione S transferase (GST) was used for the labeling studies as it can be easily produced in both systems. The quantities of ^{15}N-amino acids utilized in the production of labeled GST were similar in the case of bacterial and viral expression systems.

3.2.2 Selection of the medium

In order to find a suitable medium for selective labeling in insect cells, several media were tested for growth rate and maximal cell density. Table 10 summarizes the characteristics of different widely used media for insect cells, named SF 900II, IPL 41 and KBM 10. The SF 900II media was used as a reference for comparison of the growth rate and maximal reachable cell density. The high contents of amino acids, (Radford et al.

1997) and glucose in medium enabled the highest reachable cell density in all examined media.

The comparison of the published formulations of IPL 41 and KBM 10 showed that the total mass of amino acids in IPL 41 is about 100% higher than in KBM 10. There are also great differences in the proportion among certain amino acids. Nevertheless nearly identical maximal cell densities were detected for IPL 41 and KBM 10. Based on these results a reduced content of amino acids was used for the developed media IML 406 and IML 455.

Table 10. Comparison of the maximal cell density (Nmax), the doubling time, the time of exponential growth (texp.) and protein yield (Y) in the media SF 900II, KBM 10, IPL 41 (serum-free and with 5%FCS), and IML 406 and IML 455. The yield of GST obtained in SF 900II was set to 100%.

Medium	Nmax.	Doubling	t_{exp}	Y
	X 10^6 ml^{-1}	time, h		
SF 900 11	5.7	20-22	100	100%
KBM 10	3.4	22-26	80	60-63%
IPL 41	3.1	22-26	80	58-63%
(5% FCS)				
IPL 41	3.2	22-26	80	26-31%
IML 406	4.2	22-26	80	65-74%
IML 455	2.2	24-28	70	43-48%

The initial media contained yeastolate and 5% serum and a second medium with 10% serum but without yeastolate. Whereas the medium with yeastolate enabled healthy cell growth after six weeks of adaptation, the medium without yeastolate led to slow growth rates after this time. Based on the result the medium with yeastolate was taken as a starting point for studying the influence of serum and yeastolate on the amino acid composition. Different amounts of amino acids are provided by yeastolate and serum.

Table 11 gives a quantitative analysis of the total amount of each amino acid supplied by serum and yeastolate. While the content of free amino acids in serum is very low the addition of some amino acid species via yeastolate is very high. The yeast autolysate contains very large amounts of glutamic acid, isoleucine, leucine, phenylalanine and valine. For rest of the amino acids the portion provided by yeastolate is below 15%. Because the developed medium was to be used for selective labeling of proteins with ^{15}N-amino acids, the separate addition of these amino acids was their main source. The content of amino acids was modified in a way that at least two third of each amino acid, except alanine, was supplied not from yeastolate or serum. The resulting medium was termed IML 406.

Table 11 Overview of the proportion of free ^{14}N-amino acids provided by yeastolate (YF) and serum (FCS)

Amino acid	Portion of ^{14}N-amino acid provided by		Amino acid	Portion of ^{14}N-amino acid provided by	
	YE	FCS		YE	FCS
L-alanine	94%	6%	L-leucine	25.70%	<1%
L-arginine	8.30%	0%	L-lysine	12.90%	<1%
L-asparagine	17.40%	0%	L-methionine	20.90%	<1%
L-aspartic acid	14.90%	<1%	L-phenylalanine	24.50%	<1%
L-cystine	14.10%	<1%	L-proline	5.60%	0%
L-glutamic acid	29.90%	2.6%	L-serine	12.50%	<1%
L-glutamine	0.40%	<1%	L-threonine	30%	<1%
Glycine	7.70%	<1%	L-tryptophan	16.80%	0%
L-histidine	8.80%	<1%	L-tyrosine	12.40%	<1%
L-isoleucine	31.80%	<1%	L-valine	32.20%	<1%

To investigate the possibility of using ammonium chloride for selective ^{15}N-labeling in insect cells, we used the medium IML 455, which contained 5 mM ammonium

chloride instead of glutamine and aspartic acid to enhance the formation of glutamine from glutamic acid. In a further step the cell growth and the protein yield of these two media were compared with those of other widely used insect cell media and are listed in Table 10. For IML 406 the growth studies show a slightly higher maximal cell density when compared to IPL 41 and KBM 10. The cell density in IML 455 is 50% lower than for IML 406 and the period of exponential growth is shortened in comparison to the residual media. Except for the slightly prolonged doubling time in medium IML 455, the doubling times of the other media are nearly identical. The determination of protein yields with test expressions of GST in different media was also performed. On the one hand no significant differences in yields could be observed between KBM 10, IPL 41 with 5% serum and IML 406. On the other hand the yield in medium IML 455 was more than halved when compared to reference SF 900II, whereas in serum-free IPL 41 the amount of expressed GST was reduced by 70%. Based on these data the media IML 406 and IML 455, showing sufficient protein yields, were used for ^{15}N-labeling studies using NMR spectroscopy. To ascertain if the quantities of certain amino acids provided by serum and yeastolate are sufficient for cell growth to high densities, these compounds were not added separately to the IML 406 medium with the consequence that serum and yeastolate were the sole sources for these compounds. A selection of amino acids essential and non-essential for *Sf 9* insect cells is given in Fig.3.14. In adhesion cultures no influence on cell growth for all amino acids was seen in the modified IML 406 over a period of ten passages, whereas in suspension cultures significant differences were detected.

In the case of cystine no high cell densities and protein yields were obtained. Reduction of arginine, serine and tyrosine led to a decrease of the maximal cell density by 50%. In the medium without the separate addition of histidine the cell density reached 75% of the complete IML 406.The reduction of the other examined amino acids had no negative influence on the insect cells in suspension cultures. Furthermore we performed expression studies with GST in the above-mentioned modified media (Fig.3.15). In the case of tyrosine and serine the yield dropped in a range of 10–20% of the complete IML 406. In IML 406 containing only half the amount of separately added serine and tyrosine the cell growth and protein yield was equal to the normal IML 406. The yields for arginine, glycine and lysine were decreased to 40–65% of the normal values, whereas in

the medium without separate addition of histidine the yield was reduced to 80%. For the other investigated amino acids no significant reduction in yields were detected.

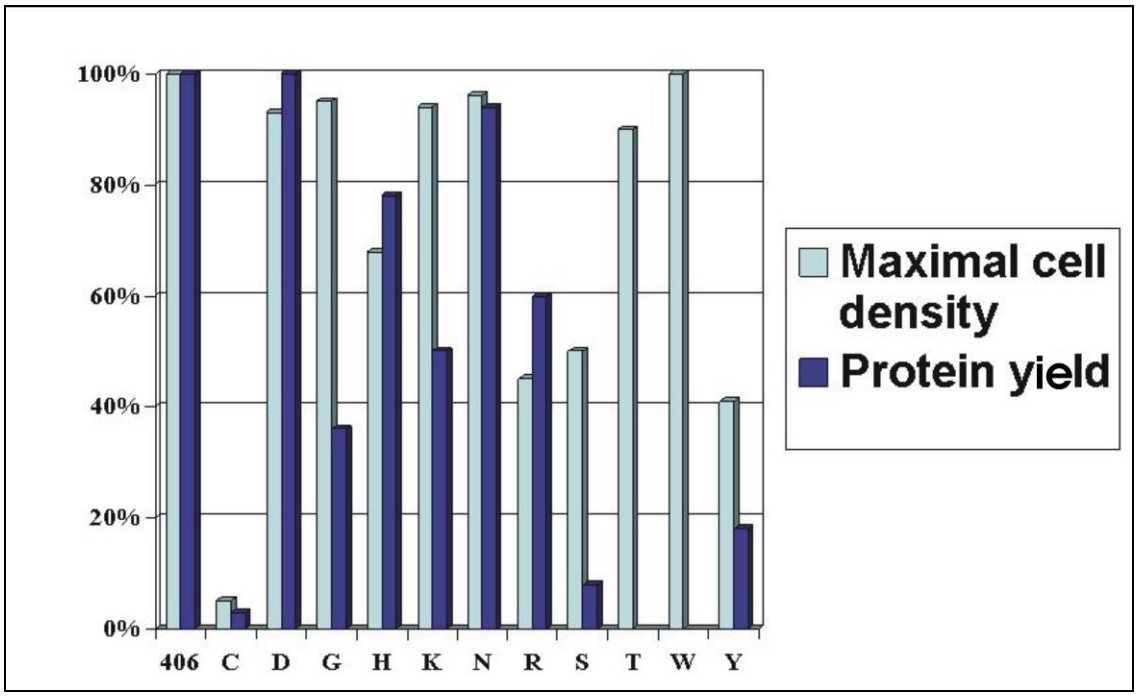


Figure. 3.14 Overview of the maximal cell density and protein yield in IML 406 without separate addition of single amino acids. The values for complete IML 406 were taken as 100%. In case of tryptophan and threonine the protein yields were not determined.

3.2.3 ^{15}N-labeling studies

For different labeling studies on GST, the protein was expressed in *Sf 9* and *E. coli* using single ^{15}N-amino acids. In both expression systems the yields of GST ranged between 8 and 12 mg of the protein per liter of culture. In most cases the quantities of ^{15}N-amino acids utilized in IML 406 and the medium for bacterial expression were similar. One exception was lysine that was added to IML 406 in an amount three times larger than that in bacterial media. Since the assignment of GST is not available, a comparison of the complete number of resonances, their shape and intensities of the signals should be sufficient to provide qualitative information about the efficient use of the new media. The starting points for the interpretation were the known metabolism pathways in *E. coli*.

Table 12 presents a comparison between the signals seen in the HSQC spectra of selectively labeled GST expressed in *Sf 9* and *E. coli*. The results of selective labeling of GST with ^{15}N-glycine, ^{15}N-phenylalanine, ^{15}N-valine and ^{15}N-leucine are presented below. Selected examples for ^{15}N amino acids (showing cross labeling) are also depicted. Fig. 3.15 shows the ^{15}N-HSQC spectrum of the purified GST labeled with ^{15}N-glycine in *Sf 9* and *E. coli*.

Table 12 Results of selective labeling of GST in *E. coli* and *SF 9* cells

^{15}N-aminoacid	Number of amino acids in		Number of signals		Number of same signals
	GST expressed in		in GST spectra		
			(strong/weak) from		
	E. coli	BEVS	*E. coli*	BEVS	
GLY	17	16	16	17	15
LYS	21	21	19	19/2	17
VAL	10	10	14/7	11/6.	10
PHE	9	9	18/6	9/3.	8/2.
LEU	28	28	24/4	34/10	22
GLU	16	16	50/18	40/4	31
ASP	18	19	49/18	0/0	

The spectrum exhibits 17 strong and several weak peaks in case of expression in insect cells. In the spectrum obtained from the bacterial expression system 17 intensive peaks can be identified. 15 identical signals can be found in both spectra. GST expressed in *Sf 9* contains 16 glycines, GST from *E. coli* 17 glycines. The number of strong peaks in both organisms is close to the number that is expected for glycine. 550 mg l^{-1} serine in IML 406 was used, and in the medium for *E. coli* 1.6 g l−1. HSQC spectra of ^{15}N-lysine labeled GST in insect cells and *E. coli* are very similar. The protein possesses 21 lysines and the spectra of both systems show 17 identical resonances. In both spectra the total number of intensive signals is smaller than the number of lysines in GST. For two peaks a difference in chemical shift can be found depending on the host of expression. The differences were not caused by posttranslational modifications since a MALDI-TOF

analysis of GST expressed in *Sf 9* revealed only the mass of unmodified GST. The higher amount of lysine in IML 406 had no influence on the total number of intensive peaks. For labeling of GST with ^{15}N-valine 200 mg l^{-1} of the amino acid was used for bacterial and viral expressions (Fig.3.16). The HSQC spectrum shows 11 strong peaks and additionally some weak peaks in the case of expression in *Sf 9*.

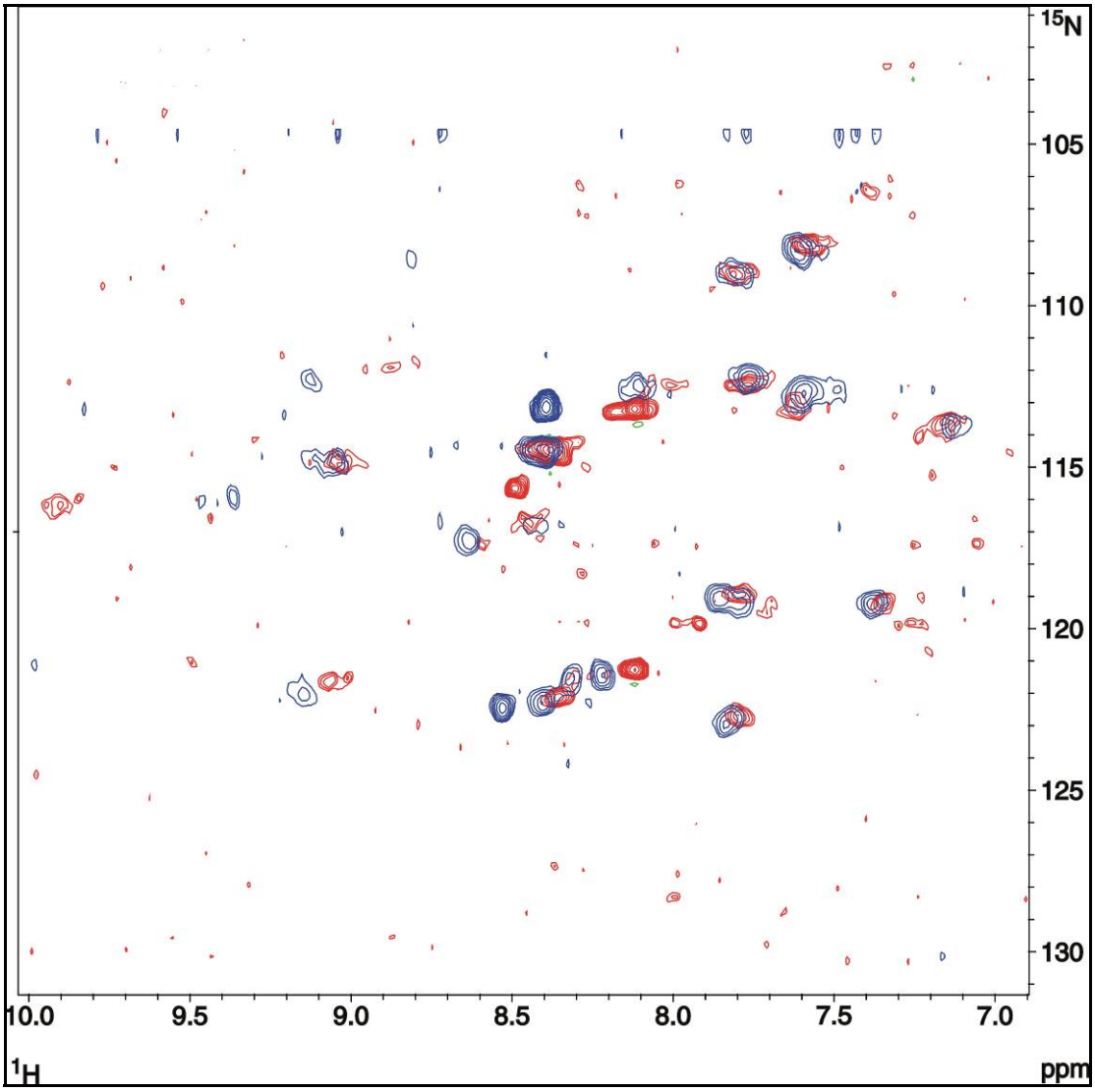


Figure 3.15 Overlay of ^{15}N-HSQC spectra, recorded with GST expressed in *E. coli* (red) and *Sf 9* (blue). The protein was selectively labeled with ^{15}N-glycine.

The spectrum of bacterially expressed ^{15}N-valine GST contains 14 strong and 7 weak peaks. Several strong signals can be assigned to peaks in the ^{15}N-glutamic acid

HSQC spectrum (vide infra). Ten resonances are identical in both spectra, which correspond to the number of valines in GST. For the eleventh signal in the spectrum from *Sf 9* no corresponding signal from other spectra can be found.

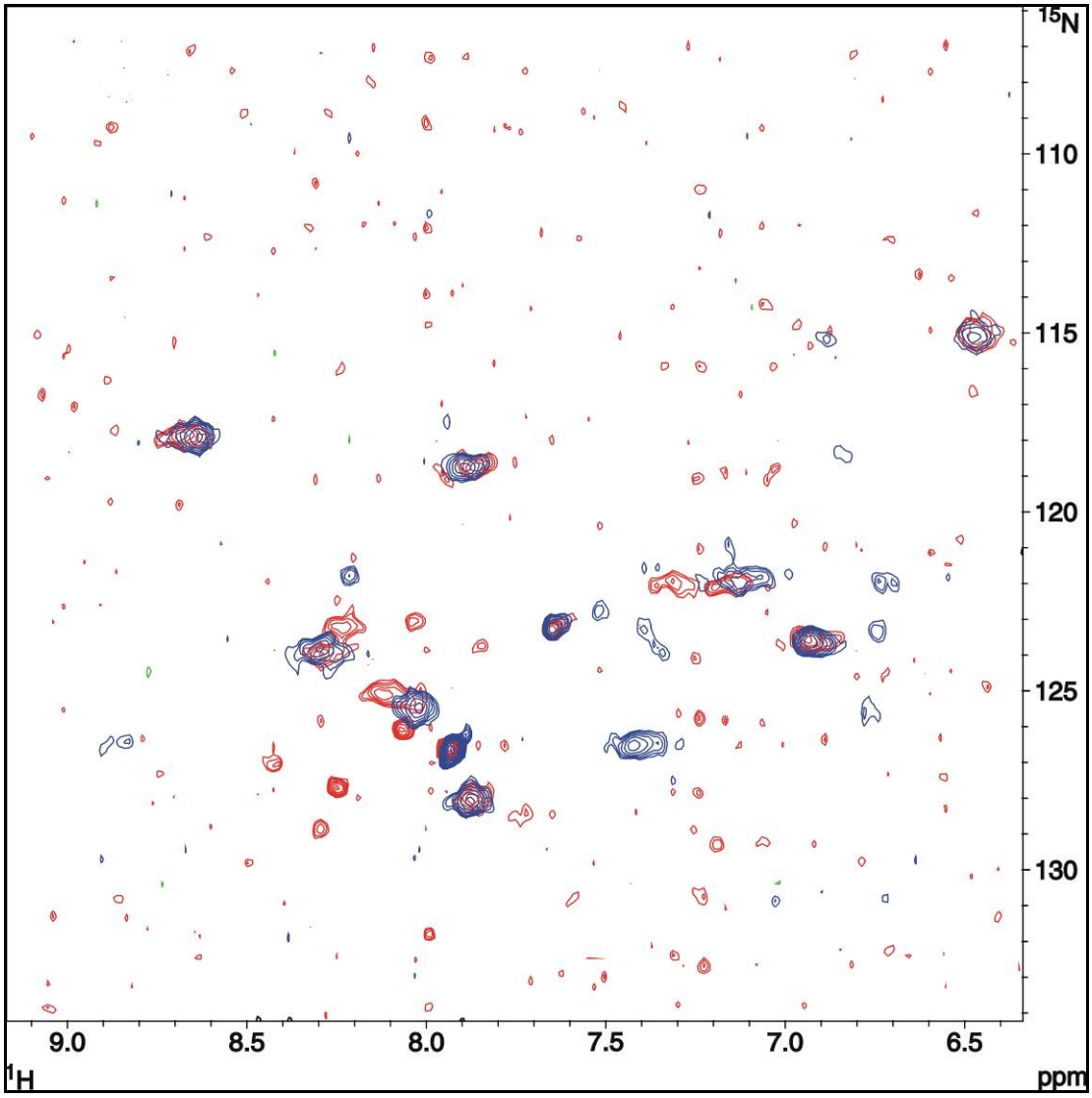


Figure. 3.16 Overlay of ^{15}N-HSQC spectra, recorded with GST expressed in *E. coli* (red) and *Sf 9* (blue). The protein was selectively labeled with ^{15}N-valine.

For the identification of cross labeling we labeled GST with ^{15}N-alanine in *E. coli*. The HSQC spectrum contains 10 peaks, three less than expected. Seven signals could be

assigned to weak peaks in the ^{15}N-valine spectrum recorded with GST expressed in bacteria (Fig. 3.17). In the case of viral expressed GST these peaks are not visible.

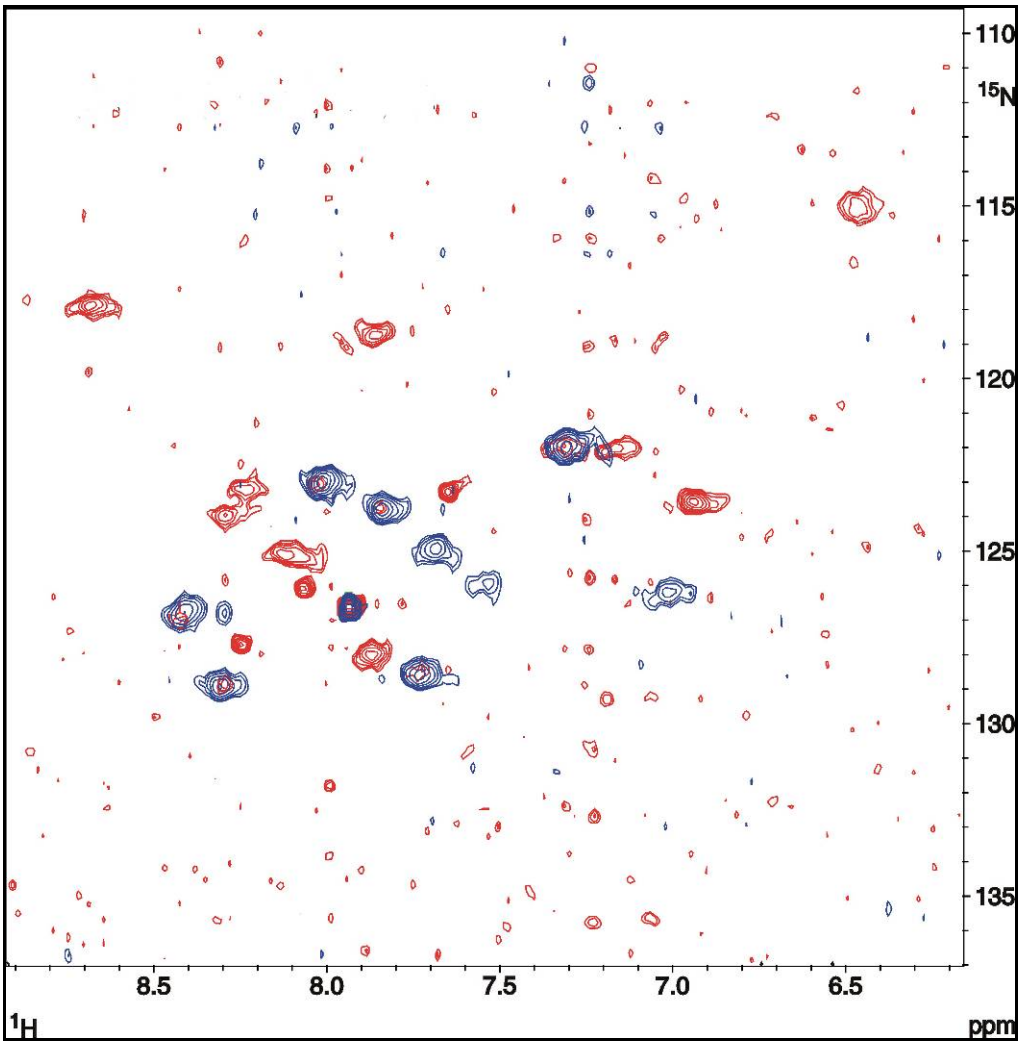


Figure. 3.17 Superposition of ^{15}N-HSQC spectra, of GST labeled with ^{15}N valine (red) and ^{15}N alanine (blue) in *E. coli*.

Fig. 3.18 depicts GST labeled with ^{15}N phenylalanine in *E. coli* and insect cells. GST holds nine phenylalanines. Comparison of both spectra reveals significant differences in the number of signals. Whereas the viral expression of GST leads to nine strong signals, the HSQC spectrum of GST from *E. coli* contains more than 15 strong signals. One strong signal of the HSQC spectrum from *Sf 9* has no counterpart in the spectrum from *E. coli*. More than ten peaks of the HSQC spectrum derived form bacterially expressed protein are

identical with signals in the spectrum of GST labeled with ^{15}N-tyrosine in *E. coli* (Fig.3.19).

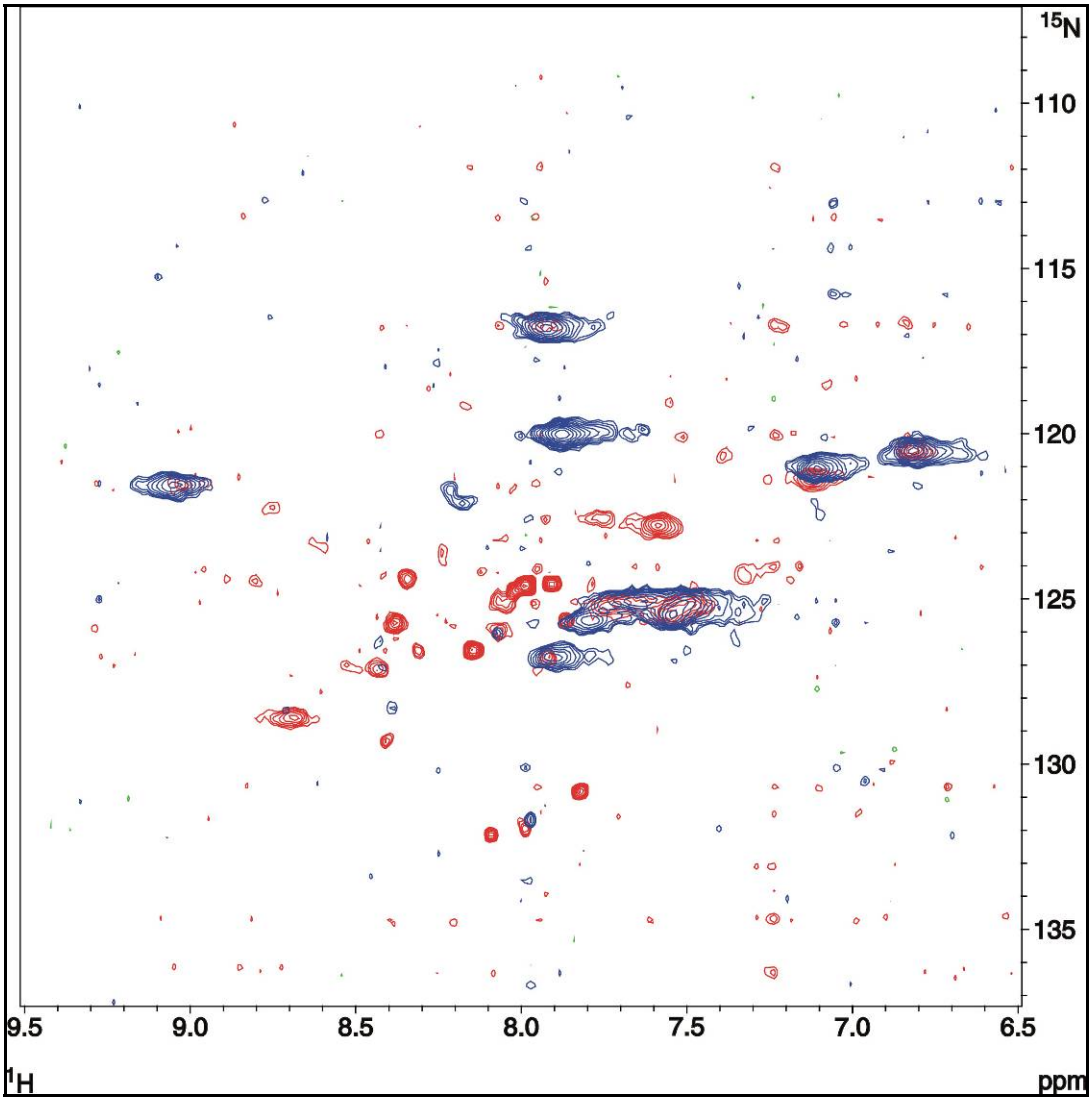


Figure 3.18 Overlay of ^{15}N-HSQC spectra, recorded with GST expressed in *E. coli* (red) and *Sf 9* (blue). The protein was selectively labeled with ^{15}N-phenylalanine.

Strong signals from one spectrum correspond to weak peaks in the other spectrum. In IML 406 250 mg l^{-1} ^{15}N-phenylalanine were used, in the medium for bacterial expression 100 mg l^{-1}. Despite the higher amount of phenylalanine for *Sf 9* no conversion to other amino acids was detectable since phenylalanine and tyrosine are essential for insect cells. Also the possibility of using reduced amounts of certain amino acids for

efficient labeling was investigated. For this purpose ^{15}N-glycine labeled GST was expressed in IPL 41 and we used KBM 10 for expression of the ^{15}N-valine or ^{15}N-phenylalanine labeled GST. In the case of the^{15}N-glycine labeling in IPL 41 the spectrum showed the same result as in IML 406. The usage of KBM 10 with ^{15}N-valine or ^{15}N-phenylalanine led to spectra of poor quality. The peaks were at the same protein concentration less intensive. For ^{15}N-leucine (Fig.3.20) differences in the number of peaks are visible depending on the expression system.

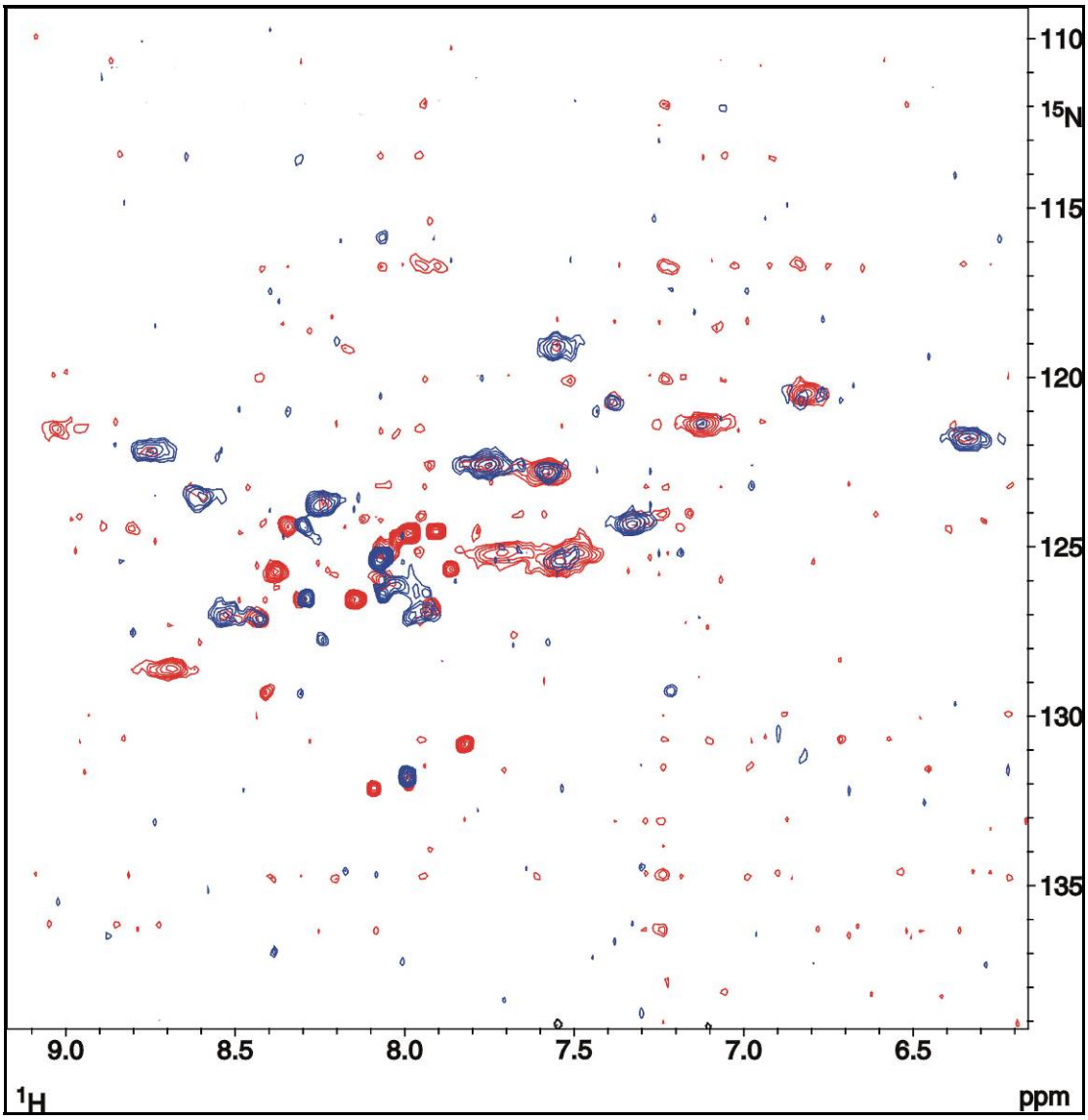


Figure 3.19 Superposition of ^{15}N-HSQC spectra, of GST labeled with ^{15}N phenylalanine (red) and ^{15}N tyrosine (blue) in *E. coli*.

Whereas the bacterial expression of GST led to a spectrum of about 24 strong and several weak signals the expression in *Sf 9* gave an HSQC spectrum with more than 30 strong peaks. GST contains 28 leucines and 22 peaks can be found in both spectra. We also labeled GST with ^{15}N-isoleucine in *E. coli* and compared the HSQC spectrum with the spectrum of ^{15}N-leucine labeled GST in *E. coli*. Several strong peaks in the spectrum of leucine correspond to weak peaks in the spectrum of isoleucine. At least in *E. coli* there exists an efficient conversion from leucine to isoleucine. A comparison of the respective spectra revealed an efficient conversion of isoleucine to valine in *E. coli*. In insect cells these reactions are not visible. But the spectrum of GST labeled with ^{15}N-leucine in *Sf 9* contains still a large number of signals, which could not be assigned to amino acids investigated so far.

3.2.3.1 ^{15}N-labelling with ammonium chloride, glutamic acid or aspartic acid

It was of major interest to investigate the possibility of using ammonium chloride for ^{15}N-labeling of proteins in *Sf 9*. The spectrum contained two groups of signals in a region that is typical for amides in the side chains of glutamine or asparagines. In other regions of the HSQC spectrum no peaks were visible. Especially the incorporation of ^{15}N-nitrogen in alanine reported by Drews et al (Drews et al. 2000) was not detectable even after recording the HSQC spectrum with 1024 scans. For completeness ^{15}N-glutamic acid (Fig.3.21) and ^{15}N-aspartic acid was used for labeling of GST in *Sf 9* and *E. coli*. In the bacterial expression system both amino acids led to HSQC spectra with more than 50 signals. GST contains 18 aspartic acids and 16 glutamic acids. The superposition of both spectra revealed that most of the strong signals were identical. An overlay of the HSQC spectra derived from ^{15}N-isoleucine, ^{15}N-leucine, ^{15}N-phenylalanine, ^{15}N-tyrosine and ^{15}N-valine with the ^{15}N-glutamic acid HSQC spectrum from *E. coli* showed that a large number of strong signals in these spectra correspond to weak or strong signals in the ^{15}N-glutamic acid spectrum.

Furthermore, the HSQC spectrum of GST labeled with ^{15}N-glutamic acid contained a number of resonances, which could be identified as cross labeling to alanine by superposition of respective HSQC spectra. In contrast to these results the labeling of GST

with ^{15}N-glutamic acid in insect cells led to a spectrum with only 40 strong and 4 weak peaks.

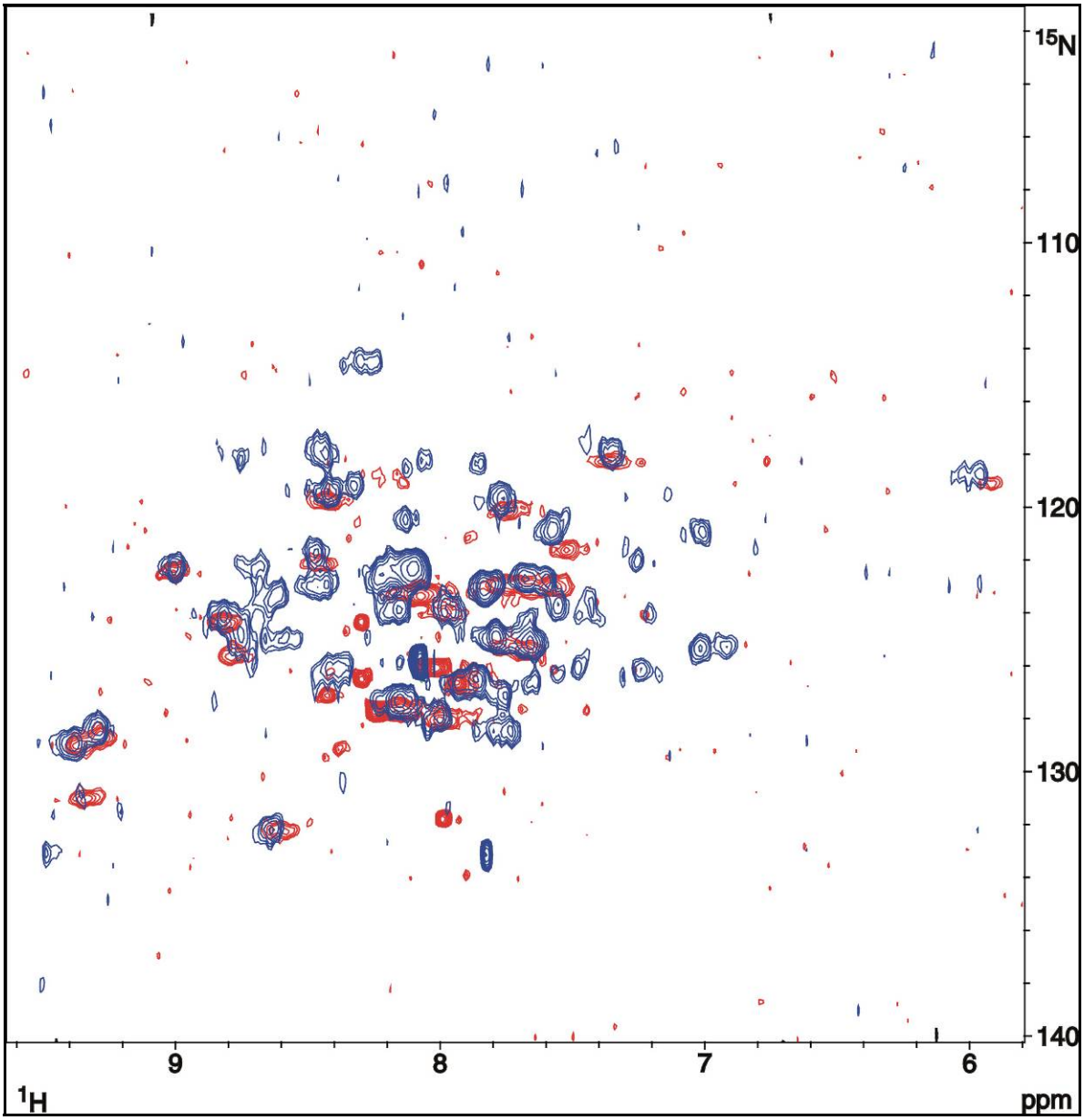


Figure 3.20 Overlay of ^{15}N-HSQC spectra, recorded with GST expressed in *E. Coli* (red) and *Sf 9* (blue). The protein was selectively labeled with ^{15}N Leucine.

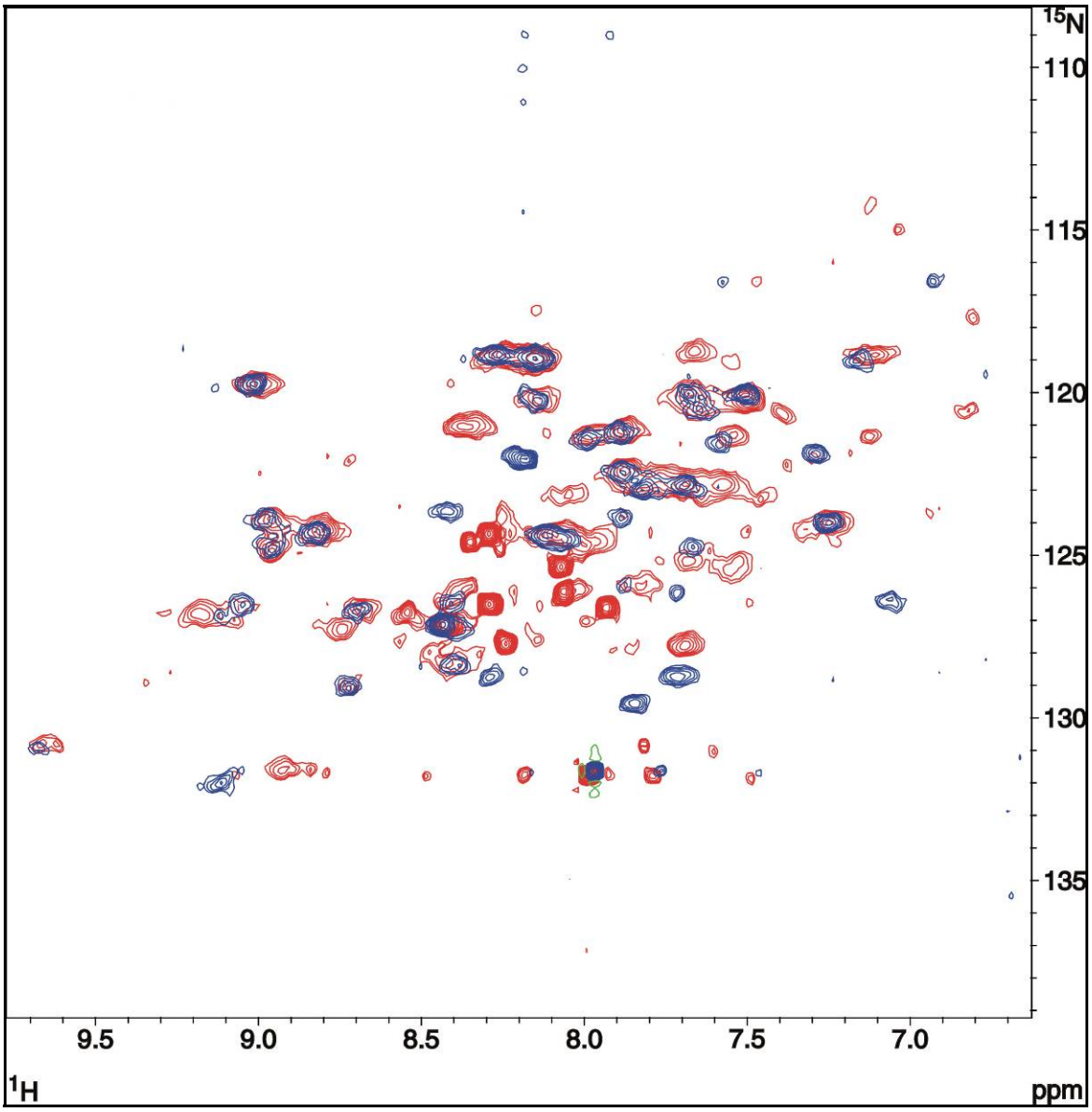


Figure 3.21 Superposition of ^{15}N-HSQC spectra, of GST labeled with ^{15}N glutamic acid (red) in *E. coli* and *Sf 9* (blue).

The comparison with the HSQC spectrum of GST derived from *E. coli* exhibited a high conformity of the strong signals in both spectra. Seven signals in the spectrum of viral expressed GST were identified as alanines. In contrast to the expression in bacteria, cross labeling to isoleucine, leucine (Fig.3.22), phenylalanine, tyrosine or valine is not detectable. In the case of labeling with ^{15}N-aspartic acid in *Sf 9* no interpretable spectrum could be obtained.

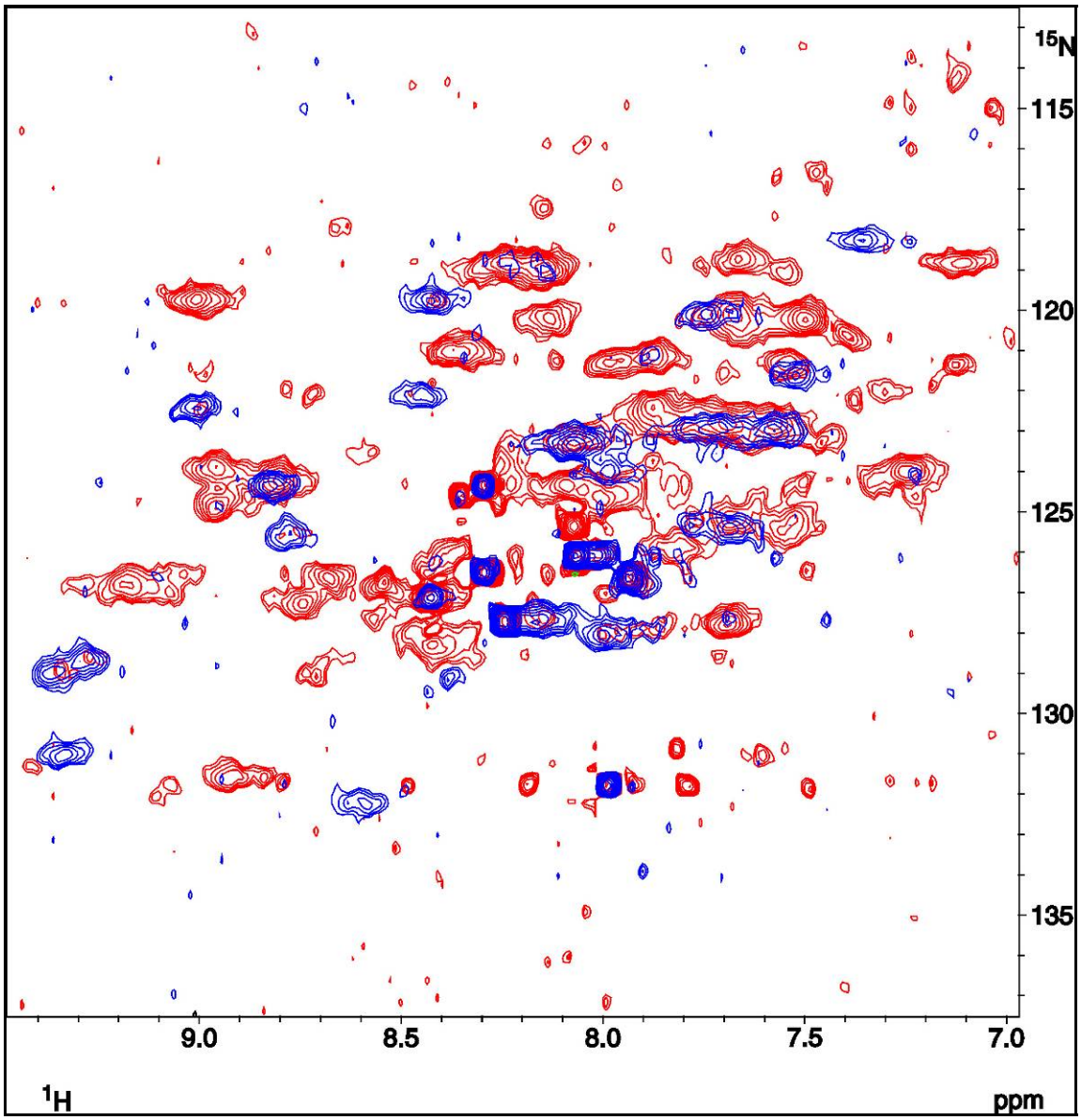


Figure 3.22 Superposition of ^{15}N-HSQC spectra, of GST labeled with ^{15}N glutamic acid (red) and ^{15}N Leucine (blue) in *E. coli*.

Finally we compared the HSQC spectra of selectively labeled GST expressed in insect cells with the spectrum of the uniformly ^{15}N-labeled GST expressed in *E. coli* (Fig.3.23) using ^{15}N-ammonium chloride as sole nitrogen source. Almost all strong peaks are identical in both spectra. Only four signals of the viral expressed GST have no equivalent resonances in the uniformly labeled GST.

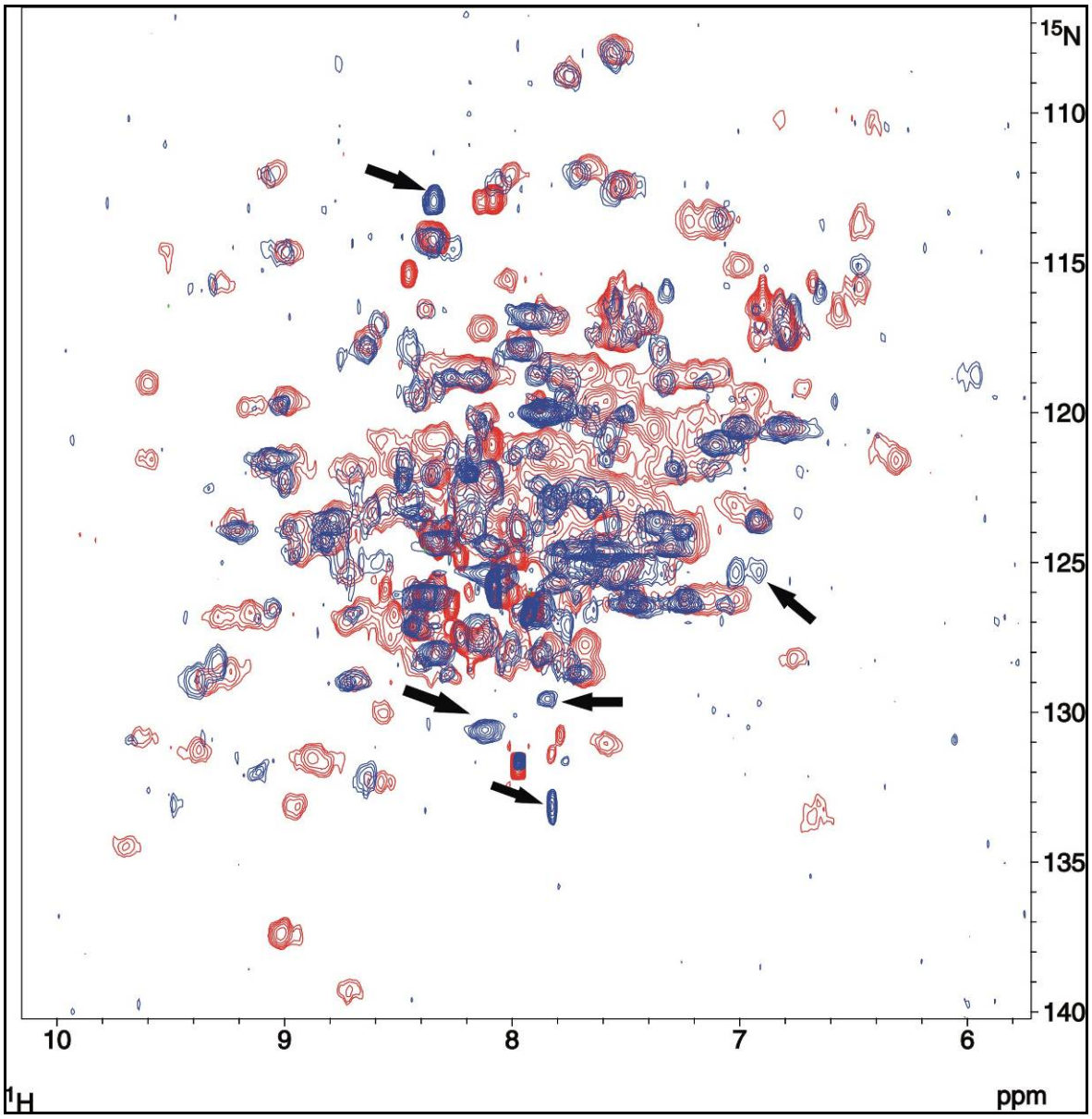


Figure. 3.23 Comparison of ^{15}N-HSQC spectra, of selectively labeled expressed in *Sf 9* insect cells (blue) with the HSQC spectra of uniformly ^{15}N labeled GST expressed in *E. coli* (red). Strong signals which are not found in the spectra of ^{15}N GST from *E. coli* are marked with arrows.

Some of them are present in the spectra of selectively labeled GST expressed in *E. coli*. Also, most of the weak signals are expected to originate from the protein. A large fraction of the peaks in ^{15}N-GST can be assigned to a certain type of amino acid using the HSQC spectra of selectively labeled GST expressed in insect cells.

3.2.4 Preliminary investigations on cyclin dependent kinases

The kinase CDK6 was expressed and purified as a GST fusion protein using the baculovirus expression vector system. The fusion protein obtained had high purity and possessed high stability. The cleavage of GST-CDK6 with thrombin resulted in near homogenous CDK6 with reduced stability. Preliminary labeling trials were then performed using selectively labeled media, co-crystallisation trials were also performed.

3.2.4.1 Protein Expression and purification

The GST-CDK6 construct was cloned into the baculovirus vector as mentioned earlier (Materials and methods). For the expression trials 2.0 to 2.2x 10^6 cells were grown in 2 L agitating bottles in SF 900 11 medium. The cells were infected with the appropriate baculovirus containing the GST-CDK6 gene with a MOI of 5. Cells were grown at 27˚C for 60 hrs post infection before harvesting. GST-CDK6 expression was also carried out in a biofermentor, for larger production of the protein. In case of fermenter, 4 L suspension culture was grown to a cell density of 3 million cells per milliliter and then infected with the appropriate virus (MOI=5). The express ion time amounted to 65 hrs. The yield was 25-30 % more highly compared with the agitating cells. Cells were stored overnight at -80˚C, The cells were lysed using native GST buffers and centrifuged at 25,000 rpm for 1hr. The supernatant was then applied to the GST coloumn and eluted with high glutathione gradient. SDS PAGE showed a band corresponding to the weight of the GST-CDK6 fusion protein. The yield of the expressed protein was estimated by Bradford assay to approximately 25mg/l of which 40% was GST. The eluted protein was then put for thrombin digestion by adding 10unit enzyme/mg fusion protein in thrombin cleavage buffer, complete cleavage of the fusion protein was achieved within 6 hrs at 4˚C. The cleaved mixture was then again run through the GST column and the flow through containing only CDK 6 was collected.

The protein was found to be stable only in conditions of greater then 250 mM NaCl and 7.2 pH, below these concentrations the protein tends to aggregate. Further purification was done by gel filtration chromatography (Fig. 3.24) in a buffer containing 400 mM NaCl, 15 mM Tris pH 7.8 the protein was purified to high purity and 6-7 mg pure protein

was obtained with 1litre culture. The protein was stable to upto 3 weeks at 4°C after which it tends to precipitate.

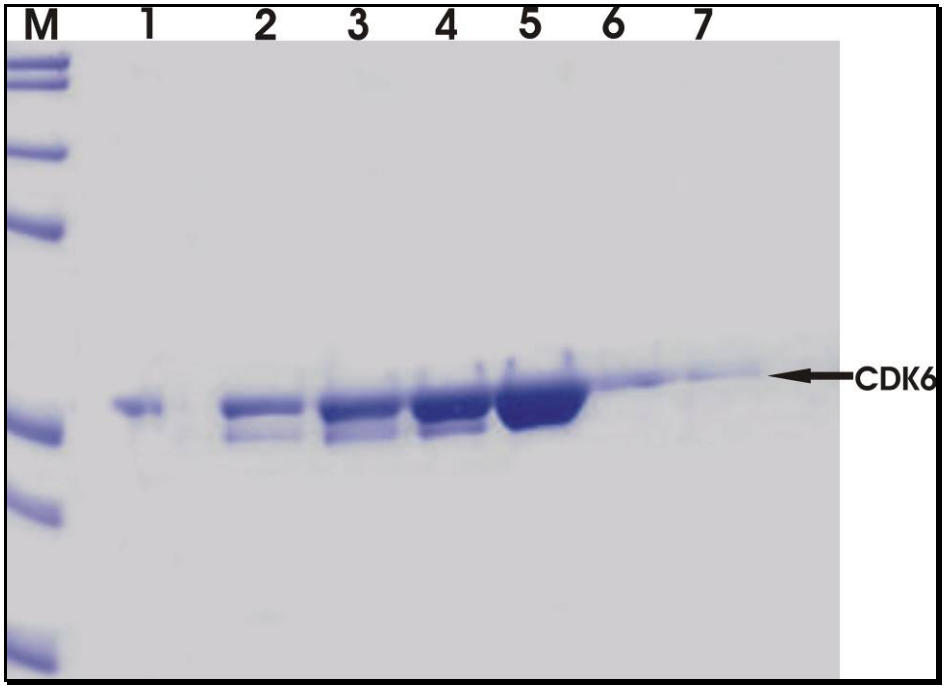


Figure 3.24 CDK 6 purification by gel filtration chromatography. SDS PAGE showing fractions (lane 1-7) containing CDK 6from under the observed gel filtration peak

3.2.4.2 1D-NMR studies

The protein was checked for proper folding by 1D-NMR and it was seen that the protein was nicely folded (Fig.3.25). The very intensive and sharp signals with 1.2 ppm and 3.6 ppm came from small organic compounds like DTT, isopropanol, in which the PMSF was solved. During the course of the experiment some of the protein precipitated which indicated that the protein was unstable.

3.2.4.3 Preliminary ^{15}N selective labeling trials on CDK6

Initial attempts at labeling GST-CDK6 fusion protein to investigate the possibility of selective labeling in insect cells had given poor results using the IPL 41 serum free

medium (which was chosen as its composition was known), which had initiated the need for formulation of the above said IML 455, and IML 406 media. After successful labeling

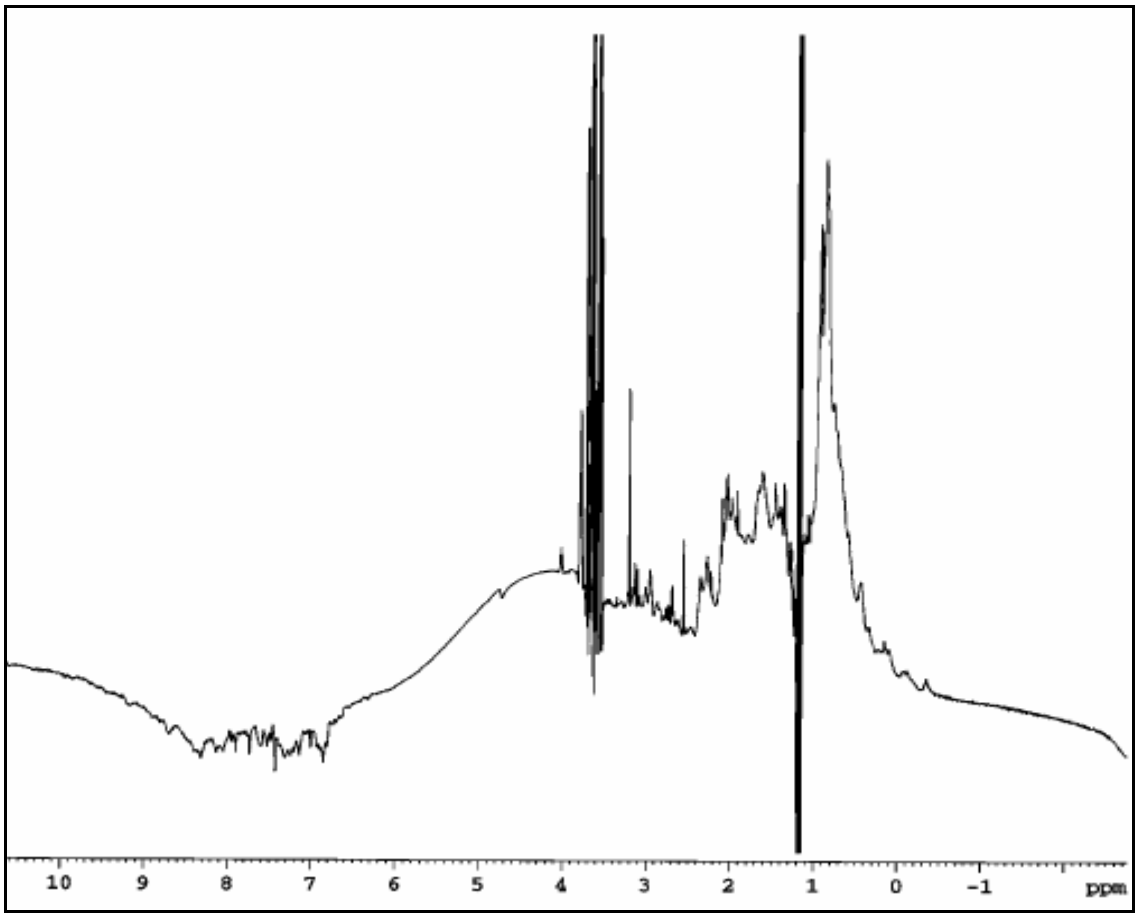


Figure 3.25 1D-NMR spectra of GST-CDK6 fusion protein, Spectra shows a folded protein.

trials with GST, the possibility of multiple selective labeling of GST-CDK6 using lycine, ^{15}N-leucine, ^{15}N-phenylalanine and ^{15}N-valine was carried out (Fig.3.26). The spectrum shows a multiplicity of intensive signals within the range between 105 and 118 ppm in the nitrogen dimension, which typically corresponds to the range in which glycine signals are observed; over 30 peaks can be identified. The number of glycines in combined GST-CDK6 fusion proteins is 34. Thus most of the glycines could be observed. In the remaining part of the spectrum can be identified over 80 peaks. A high portion of the

desired amino acids in GST-CDK6 is thus labeled proving the efficiency of the formulated media in labeling proteins expressed in insect cells.

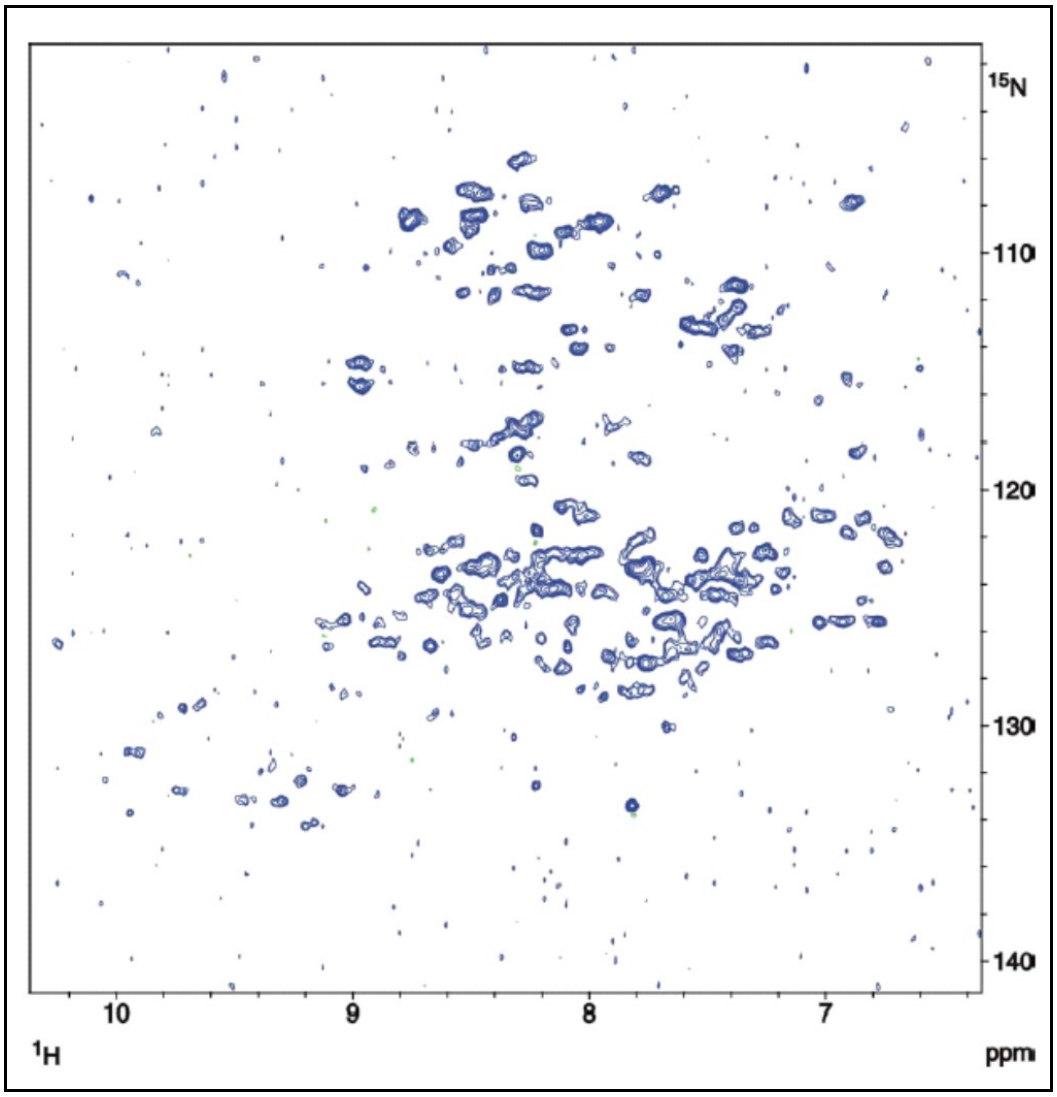


Figure 3.26 HSQC spectra of GST-CDK6 showing selective labeling with ^{15}N-glycine, ^{15}N-leucine, ^{15}N-phenylalanine and ^{15}N-valine. The spectrum shows a multiplicity of signals. In the region characteristic of glycine around 30 peaks can be observed

3.2.4.4.Crystallization trials

Crystallization trials were set up with the protein in an attempt to get crystals of monomeric CDK 6. Co-crystallization trials were also set up with inhibitors designed

against the ATP domain of the protein, the protein was concentrated to approximately 20mg/ml and 5-fold excess of inhibitor was added. But even after many trials and trying various conditions good diffracting crystals could not be obtained. The possible reason was that the protein tends to aggregate at higher concentrations.

3.2.5 Discussion

The medium formulated here, IML 406, has similar behavior with respect to growth rate and protein yield as other commonly used insect cell media like, for example, IPL 41. The reduced amount of amino acids compared to IPL 41 or SF 900II has no negative influence on protein yield and no cost-intensive feeding strategies were used. The formulation of a medium, which can be flexibly utilized for selectively labeling, was highly desired as compositions of commercial media are not fully known. Also a defined number of amino acid species not essential for *Sf 9* cells were successfully labeled. Therefore the selective labeling using the baculovirus expression vector system represents a complement or even an alternative to the bacterial expression system.

Influence of free amino acids provided by serum and yeastolate

Some amino acids (e.g., valine or leucine), which are often used for selective ^{15}N-labeling in *E. coli*, are present in high amounts in yeastolate. Hence increased amounts of these amino acids were separately added to the medium. The consequence is an increase of costs for labeling experiments with these amino acids. One solution to prevent this is the reduction of yeastolate. In the case of nonessential amino acids the cells are able to produce these compounds in sufficient amounts. In media without separate addition of arginine, serine or tyrosine the maximal cell density was one half to that of a complete IML 406. This indicates that the quantities of these amino acids provided by serum or yeastolate are low (8%, 12% and 12% of the total mass for arginine, serine or tyrosine, respectively). For serine and tyrosine a high degree of ^{15}N-labeling is therefore possible. In the case of arginine we obtained about 60% of the yield of the complete IML 406. The halving of arginine, serine or tyrosine in the medium did not reduce the maximal cell density and the protein yield. Therefore the labeling experiments can be performed with

reduced amounts of these amino acids. Histidine is provided in small amounts by serum and yeastolate (9% of total mass) and consequently indicates only a small consumption of histidine by *Sf 9* insect cells. For efficient ^{15}N-labeling the added quantities of ^{15}N-histidine used here are therefore sufficient. In the case of glycine, lysine, threonine and tryptophan the maximal cell density was not influenced but protein yield reduced by 40 to 50%. The separate addition of these amino acids was necessary to obtain high protein yields and a high degree of ^{15}N-incorporation. Aspartic acid, asparagines and glutamine are not essential for *Sf 9* insect cells. The growth and expression studies show that the cells are able to synthesize these amino acids in sufficient quantities. In the case of the nonessential amino acid cystine, it was not possible to reach high cell densities and protein yields. The addition of 80 mg l^{-1} ^{15}N-cystine should be sufficient to obtain a high degree of isotopic labeling. In general, the amount of free amino acids provided by yeast extract can be eliminated, since insect cells can grow in media without the yeast extract but they have to be supplemented with 10% (v/v) dialyzed serum (Gibbs et al. 1993)

Selective labeling with ^{15}N-amino acids

The selective labeling with ^{15}N-amino acids in *Sf 9* resulted in most of the cases in an equal or better quality of the NMR spectra when these were compared to the spectra obtained from GST expressed in bacteria. Even in the cases were the amount of the ^{15}N-amino acid was only two third of the total mass we obtained interpretable HSQC spectra. With one exception the used amount of ^{15}N-amino acids was sufficient for selective ^{15}N-labeling. The number of scans used for acquisition of HSQCs was similar for most of the investigated amino acids in both expression systems and therefore, the ^{15}N-incorporation is similar using the bacterial medium and IML 406. In some cases of bacterially expressed GST we were forced to use a lower number of scans to be able to distinguish between expected and cross labeled peaks. This demonstrates the high efficiency of amino acid conversion in *E. coli*. In several HSQC spectra of GST derived from bacterial expression very sharp signals were identified. They stem probably from impurities or flexible residues. Since most of the amino acids are essential for insect cells, a conversion between these amino acids is therefore not possible. Since the assignment of GST is not available, we compared the number of total peaks and determined the number of identical peaks in

the spectra obtained for one ^{15}N-amino acid. In a first approach only strong peaks as provided by this amino acid in its ^{15}N-version were considered. Weak peaks were considered as formed from ^{15}N-amino acids by cross labeling like transaminations. These signals were compared with strong signals in HSQC spectra of GST labeled with other ^{15}N-amino acids. In the first step we focused on metabolism pathways known in *E. coli* (Duffy et al. 1998). If a high conformity of the signal pattern is visible, we concluded that the weak signals in one spectrum represented cross labeling to an amino acid causing strong signals in the other spectrum. Signals that do not correspond to peaks in other spectra cannot be assigned to a certain type of an amino acid. In a last step we compared the signals of selectively labeled GST with the peaks of ^{15}N-GST obtained using $^{15}NH_4Cl$ as sole nitrogen source. The findings confirmed the incorporation $^{15}NH_4$ in the amide groups of arginine and glutamine but the formation of ^{15}N-alanine was not detected. In both expression systems we detected a powerful conversion between glutamic acid and aspartic acid.

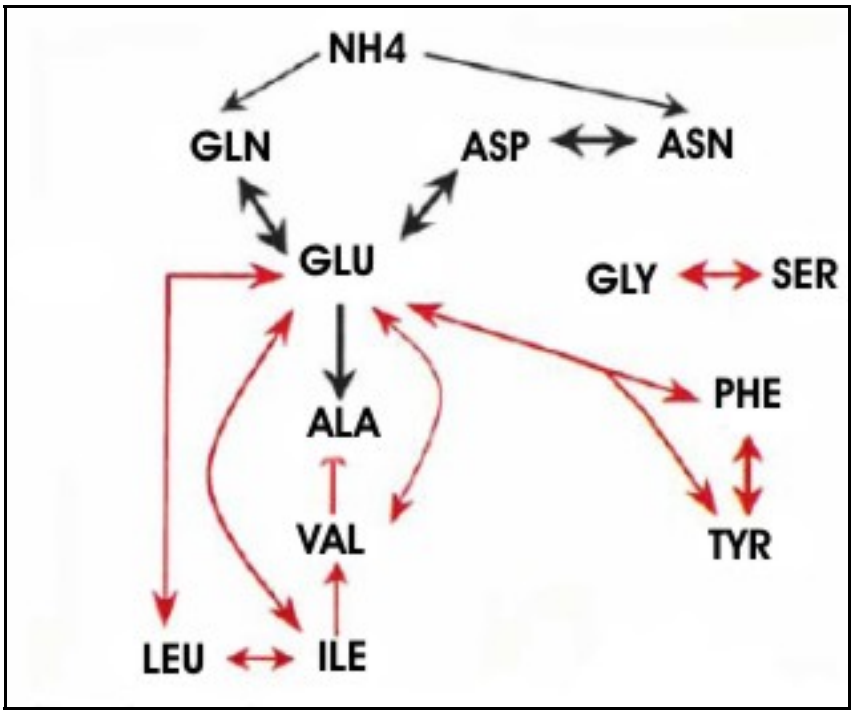


Figure 3.27. A simplified presentation of the amino acid metabolism in *E. coli* and *Sf 9* in respect to ^{15}N: The black arrows symbolize pathways, which are present in both organisms; the pathways that only exist in *E. coli* are shown in red. The strength of the arrows reflects the intensity of the conversion.

The transamination between these amino acids plays a central role in bacteria since they harbor many aminotransferases, which use glutamic acid as donor for amino groups. In *Sf 9* this reaction is limited to amino acids involved in the citric acid cycle and alanine. The final comparison of the HSQC spectra of GST selectively labeled in *Sf 9* with uniformly ^{15}N-labeled GST expressed in *E. coli* showed that most of the very weak signals in the spectra of selectively labeled protein can be considered as impurities.

Based on our findings we propose in Fig. 3.27 a simple and provisional overview of the network of the amino acid metabolism in *E. coli* and insect cells focused on nitrogen. Especially the central role of glutamic acid for synthesis of amino acids in *E. coli* is clearly visible. Fig. 3.27 also shows that the expression in *Sf 9* is better suited for selective labeling of amino acids. The expression system offers the possibility to selectively label amino acids to high degrees and without cross labeling for tyrosine, phenylalanine, glycine, serine, cysteine, arginine and valine. Even amino acids can be labeled that are normally not used in *E. coli* due to extensive cross labeling. After elimination of yeastolate from IML 406 the degree of ^{15}N-incorporation in *Sf 9* will be higher than in *E. coli*. The expression in insect cells is then in many cases a potential alternative to the in-vitro-translation. This is important for proteins that are not available in high yields or in a functional form in bacteria.

3.3 Pattern of proteolytic cleavage of IGFBPs by μ-calpain

To elucidate the cleavage pattern of calpains, we performed calpain induced proteolytic studies on the insulin-like growth factor binding proteins (IGFBP-4 and -5). Our results show that calpain cleavage sites are in the non-conserved, unstructured regions of the IGFBPs. Compilation of the calpain-induced proteolytic cleavage sites in several proteins reported in the literature, together with our present study, has not revealed clear preferences for amino acid sequences. We therefore conclude that calpains seem not to recognize amino acid sequences but instead cleave with low sequence specificity at unstructured or solvent exposed fragments that connect folded, stable domains of target proteins.

3.3.1 In-vitro cleavage analysis

The SDS-PAGE analysis of the in vitro assay clearly showed that there were calpain cleavage sites in both IGFBP-4 and IGFBP-5 and both proteins were rapidly cleaved into smaller fragments (Fig. 3.28 A and B). For 1000 nM of IGFBPs as little as 20 nM of μ-calpain was sufficient to induce cleavage and further increase in concentrations of the protease up to 250 nM did not result in any change in the cleavage pattern. Also there was no change observed with increasing time: cleavage patterns obtained within the first 30 min. were the same as those obtained after 4 hr. This was expected as calpain is an autolytic enzyme (Li et al. 2002) and autolyses within the first 30 min. which results in the protease activation but increased time interval does not result in any extra cleavage sites which shows calpain mediated proteolysis is limited.

The protein stability assays were conducted to ascertain that the protein cleavage pattern was indeed due to calpain cleavage and not to random degradation at room temperature. SDS-PAGE reveals that the proteins were stable and did not degrade at room temperature for up to 48 h, as shown in Fig. 3.29.

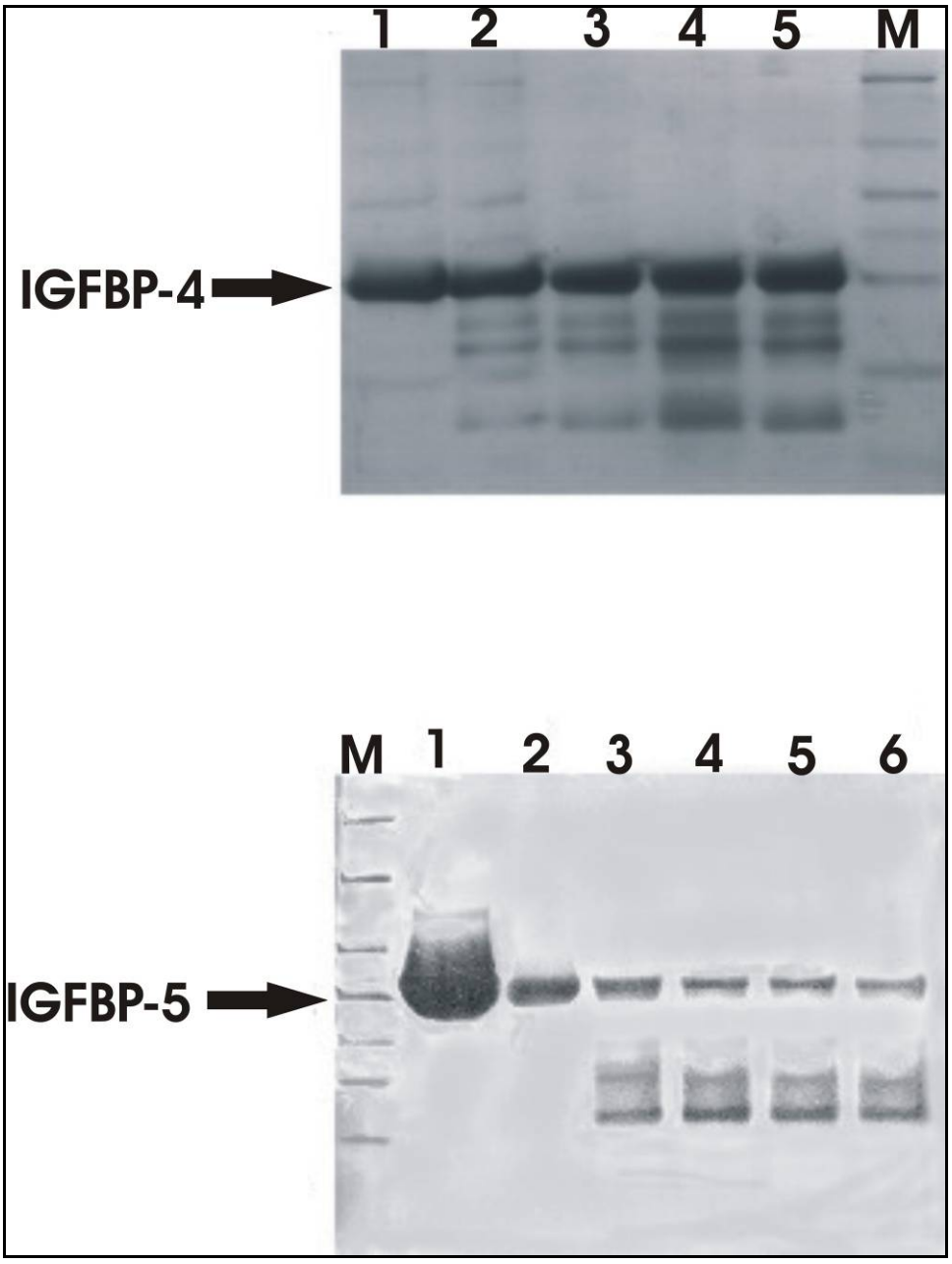


Figure 3.28 A. SDS-PAGE analysis of IGFBP-4 proteolysis by μ-calpain.

Lane 1 contains aliquots of the undigested protein as the control. Lanes 2-5 contain aliquots of the digestion reaction of IGFBP-4 cleavage at different calpain concentrations (20 nM, 50 nM, 100 nM, and 250 nM) in the calpain reaction buffer. M, shows the prestained molecular weight protein marker (kDa).**B. SDS-PAGE analysis of IGFBP-5 digested by μ-calpain.** M, shows the prestained marker. Lane 2 consists of the undigested IGFBP-5 protein. Lanes 3-6 show IGFBP-5 cleavage at different time intervals of 0 min, 30 min, 1 hr, 2 hr and 4hr, respectively, at a constant calpain concentration of 20 nM in the calpain reaction buffer. IGFBP-4 is a 28-kDa protein and IGFBP-5 is 28.5 kDa in molecular weight.

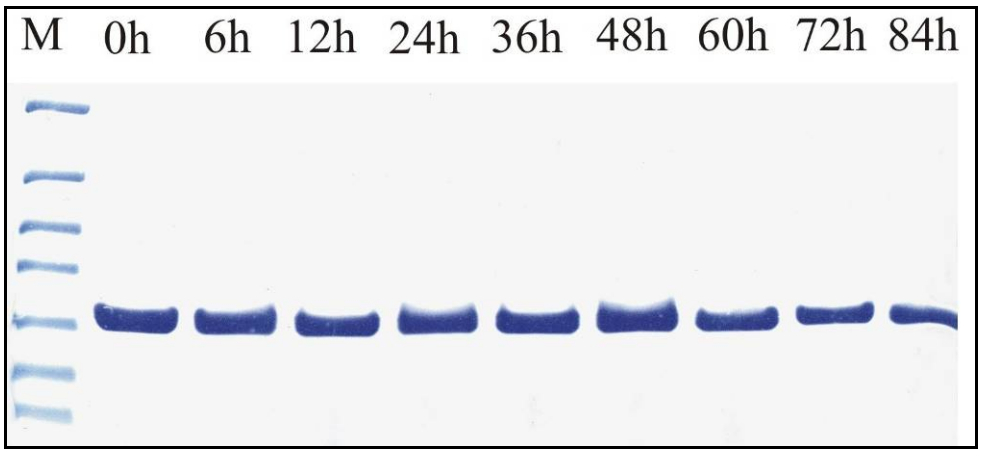


Figure 3.29 IGFBP stability gel showing the protein at different time intervals.

Lane M shows the prestained molecular weight protein marker (kDa) and the remaining lanes contain aliquots of the protein at different time intervals. The protein remains stable even after 3 days.

3.3.2 Calpastatin inhibitory assay

This assay was undertaken to ensure that the cleavage pattern observed was due to calpain and not due to any contaminant protease in the reaction mixture. Fig. 3.30 shows the results of the inhibitory assay of the endogenous proteinous inhibitor of calpains, calpastatin. Calpastatin efficiently blocked the cleavage of both IGFBP-4 and IGFBP-5 as no cleavage fragments of the proteins were observed. Domain I of calpastatin was chosen for the purpose as it posses the best ability to inhibit μ or m-calpain as compared to other domains. This result clearly indicated that the cleavage pattern observed was entirely due to calpain and not to any other contaminating protease. Also no cleavage was observed in the absence of calcium, in agreement with the fact that calcium is necessary to induce the autocatalytic activity of calpain, which results in the formation of the functional catalytic domain

3.3.3 NH_2-terminal sequencing

Characterization of the cleavage sites was accomplished by subjecting the proteolytic products to the NH_2-terminal sequencing by Edman degradation. The amino acid sequences obtained for IGFBP-4 were: GSSHH, A_3IHC, M_{73}HGQG, A_{87}IQESL,

A_{123}KIR, and G_{139}APR. Exact molecular weights of the isolated fragments were determined by MALDI-MS. Similarly, for IGFBP-5, the amino acid sequences obtained from the N-terminal sequencing were S_3FVH, C_8EPCD, Q_{142}SKFV and A_{153}HPRI.

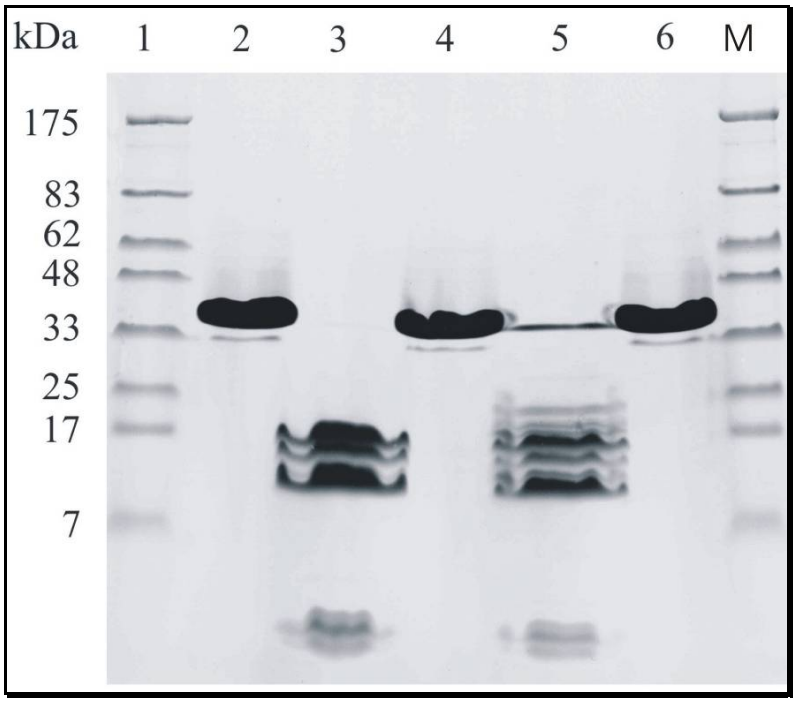


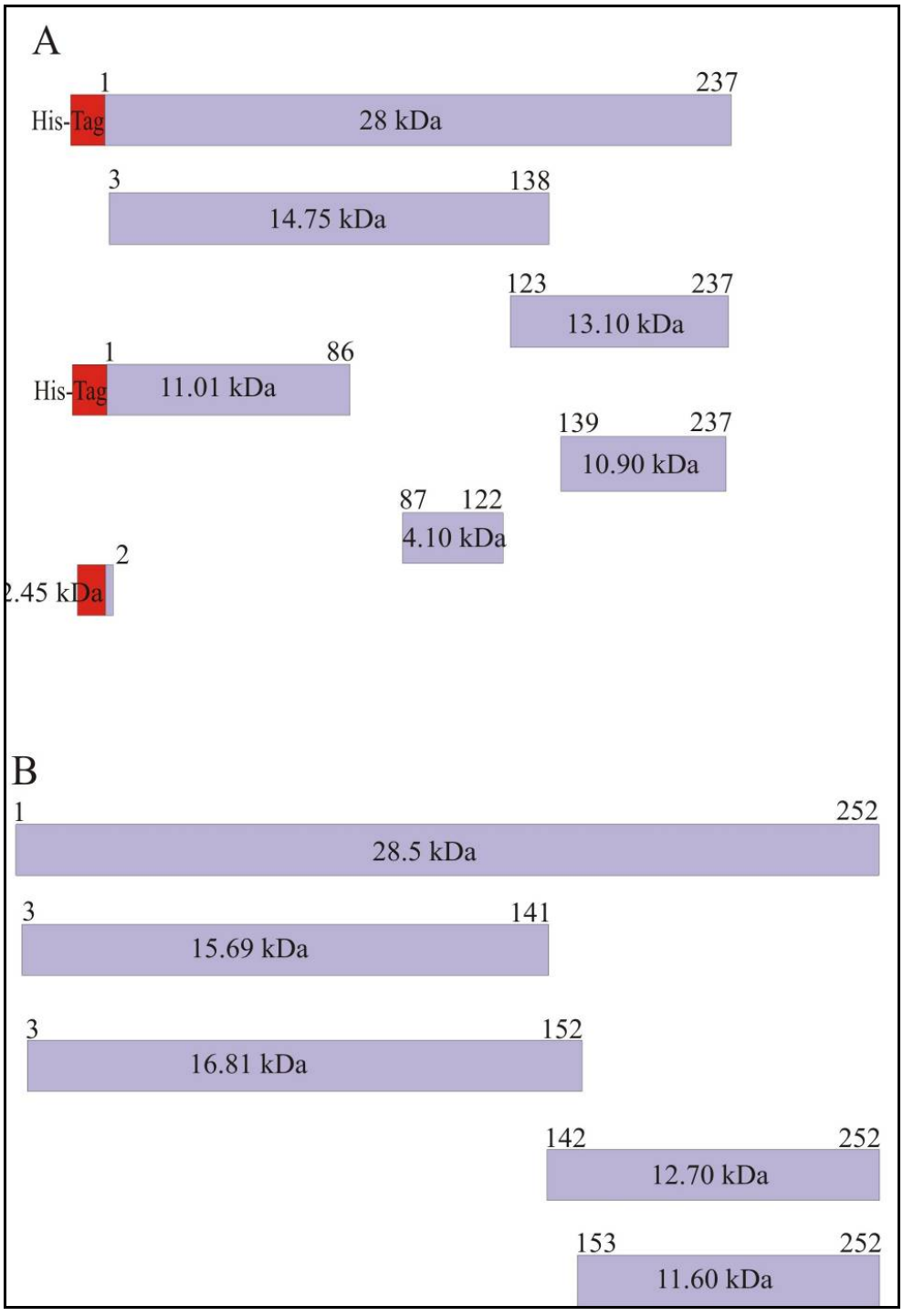
Figure 3.30 Calpastatin inhibitory and Ca^{2+}-dependent protease assay. Lane 1 and 7 show the molecular weight protein marker. Lane 2 shows IGFBP protein the protease, μ-calpain and calpastatin in the absence of Ca^{2+} where no cleavage occurs. Lanes 3 and 5 contain aliquots of IGFBP-4 and IGFBP-5 respectively, digested with μ-calpain in the absence of calpastatin and supplemented with 1mM of Ca^{2+}. Lanes 4 and 6 show the inhibition of the IGFBP-4 and 5 cleavages by μ-calpain in the presence of calpastatin at molar concentrations of 1: 1 with calpain.

The results of the generated fragments from the PVDF analysis and the reverse phase HPLC were mapped onto the IGFBP-4 and IGFBP-5 sequences and all the theoretically possible cleavage fragments for both IGFBPs were simulated.

3.3.4 Mass spectrometry

Mass spectrometry showed that fragments of the following sizes were generated for IGFBP-4: 14.75 kDa, 13.10 kDa, 11.01 kDa, and 10.90 kDa, 4.10 kDa, 2.45 kDa (Fig.

3.31 A). For IGFBP-5, fragments of 15.69 kDa, 16.81 kDa, 12.70 kDa and 11.60 kDa were obtained (Fig. 3.31 B).



FI. 3.31 Schematic representation of the fragments generated by calpain cleavage.

A, Fragments for IGFBP-4 and, **B**, IGFBP-5.

These mass spectrometry results, together with the results from N-terminal sequencing, identified cleavage sites at amino-acid positions 2, 86, 122, and 138 in IGFBP-4 (Fig. 3.31). Hence in IGFBP-4 calpain appears to act on the bond between Glu 2 and Ala 3 generating fragment of approximately 2.45 kDa, which corresponds to the 20 His-tag residues along with the first three residues of the amino terminus of the protein. The other three major cleavage sites were found in the weakly conserved regions of the protein at Glu 86 and Ala 87 generating a fragment of approximately 11.01 kDa and 16.0 kDa, but another small fragment of 4.10 kDa was observed in mass spectrometry instead of the large 16.0 kDa fragment. This discrepancy arose because this larger fragment was further cleaved at another major internal cleavage site, Phe 122 and Ala 123 which generated fragments of 4.10 kDa and 13.10 kDa. The last major site was the bond between Asp 138 and Gly 139, which resulted in fragment of 10.90 kDa. The NH_2-terminal sequencing of IGFBP-4 yielded another weak cleavage site at the bond between Leu 72 and Met 73, which corresponds to the N-terminal region of the protein but mass spectroscopy, did not show any such fragment with the corresponding molecular weights. Hence this site was not considered as a major cleavage site.

In IGFBP-5, we identified cleavage sites at residues 2, 141, and 152, as shown in Fig. 3.32. Three major cleavage sites were observed: the first site is between the residues Gly 2 and Ser 3 located at the very beginning of the protein at the amino terminus. NH_2-terminal sequencing of IGFBP-5 also showed another very close site between His 7 and Cys 8, but this was probably due to degradation of the N-terminus following initial cleavage, as no corresponding fragment was seen in mass spectra. The other two cleavage sites are located in the weakly conserved domain of the protein similar to IGFBP-4 but not at the same residues. The first, located between Thr 141 and Gln 142, results in fragment of 15.69 kDa and 12.70 kDa, and the second between Thr 152 and Ala 153, thus generating fragments of 16.81 kDa and 11.61 kDa.

3.3.5 Discussion

The IGFBP family comprises six proteins (IGFBP-1 to 6) that bind to IGFs with high affinity and a group of IGFBP-related proteins (IGFBP-rP 1-9), which bind IGFs with lower affinity (Clemmons et al. 2001; Khandwala et al. 2000). A key conserved

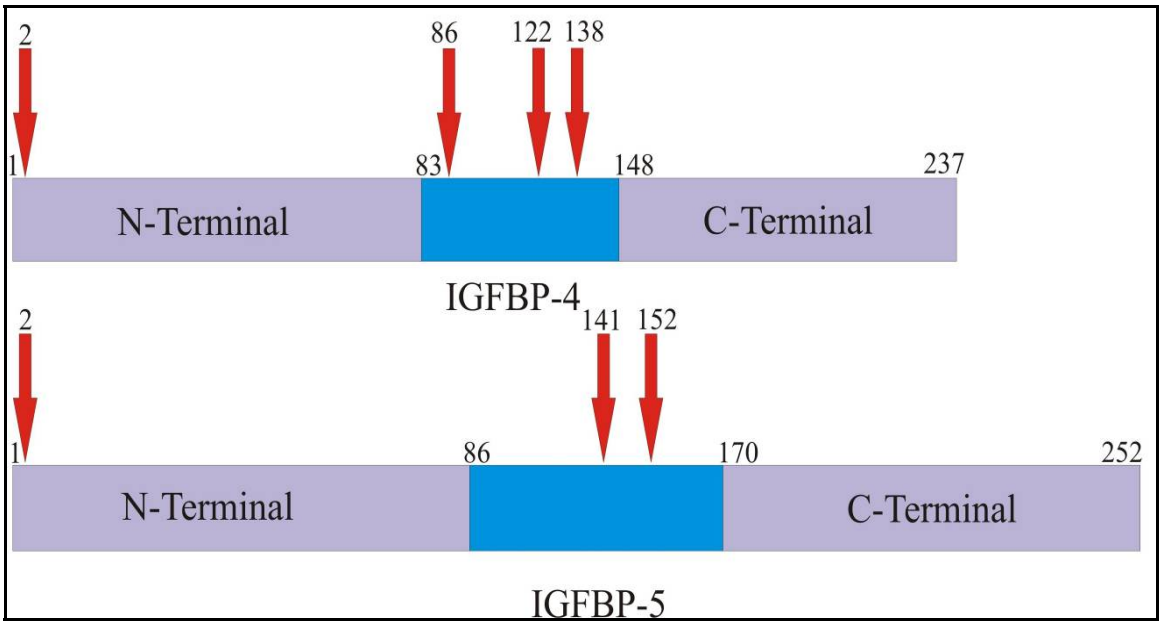


Figure 3.32 Summary of the location of the principal calpain cleavage sites in IGFBP-4 and IGFBP-5.

structural feature among the six IGFBPs is a high number of cysteines (16-20 cysteines), clustered at the N-terminus (12 cysteines) and also, but to a lesser extent, at the C-terminus. IGFBP-1 to 6 share a high degree of similarity in their primary protein structure (identities around 30-40%), with highest conservation at the N- and C-terminal regions.

It has been shown that these regions participate in the binding to IGFs (Baxter 2000; Payet et al. 2003; Shand et al. 2003). The structure of the N-terminal domain of IGFBP-5 discloses a rigid, globular structure that consists of a centrally located three-stranded antiparallel β-sheet (Zeslawski et al. 2001). This is in contrast to the central, weakly conserved domain in IGFBPs for which the NMR spectra indicates a highly flexible, unstructured conformation This domain is also the site of proteolytic activity in IGFBPs. Our data indicate that there are four major calpain cleavage sites in IGFBP-4 and three in IGFBP-5 all located in this region of the protein except for the first cleavage sites in the N-terminal domain of both the proteins. Generally for all IGFBPs two to three major proteolytic cleavage sites always occur at the beginning and towards the end of the variable domain. Calpain follows this rule too though we also observe cleavage sites at the second residue of the N-terminal domain of both IGFBP-4 and -5. The conserved N-terminal regions of IGFBPs are resistant to proteolytic attack, and degradation of C-

terminal regions has been observed only for IGFBP-3 (for plasmin, PSA, cathepsin D, and thrombin) and IGFBP-5 (thrombin) in *in-vitro* studies.

Since IGFBP-3 and IGFBP-5 show structural similarities and both are synthesized by cell types derived from the mesenchyma, calpain is expected to cleave IGFBP-3 in the same fashion as it cleaves IGFBP-5. There is evidence that a calpain-like, cysteine protease is a key regulatory component required for passage through G_1 phase of the cell cycle in eukaryotes (Mellgren 1997) and as IGFBP-3 can traverse into the nucleus calpain might also regulate the IGF-independent functions of IGFBP-3 (Baxter et al. 2000).

There are a limited number of reports on subsite specificity of calpains and Croall and Demartino, 1991and more recently Goll et al. 2003, have reviewed these. In general the literature data indicates that the calpains cleave target proteins at a limited number of sites and produce large polypeptide fragments rather than small peptides or amino acids Our present study has not revealed clear preferences for amino acid sequences, although in many cases a hydrophobic residue is present in the P_2 position In conclusion calpains seem not to recognize amino acid sequences in target proteins but cleave with low sequence specificity at unstructured or solvent exposed fragments that connect folded, stable domains of target proteins suggesting that calpains may modulate functions of substrate proteins by cutting their interdomain regions.

Summary

The focus of this work was on the structural and functional characterization of important cell cycle proteins that are involved in regulation of gene expression, cell cycle progression and differentiation, in particular those whose deregulation is associated with cancers, including the HDACs, CDKs and IGFBP family of proteins. In order to facilitate selective and uniform ^{15}N labeling of these proteins by NMR two insect cell media IML406 and IML455 were formulated.

Histone deacetylases (HDACs) catalyse the removal of the acetyl moiety from acetyl-lysine within core histones, thereby stabilizing local condensation of the DNA with concomitant downregulation of gene expression. HDAC deregulation plays a key role in cancer cell growth and HDAC inhibition has been shown to enhance transcription of a set of genes (2-5 %) that includes tumour suppressor genes. HDAC inhibitors therefore represent potential anticancer drugs and are the focus of much research. In this study the co-crystal structure of HDAC homolog HDLP with a thiophene based hydroxamic group inhibitor was solved. The bioenergetics of the HDLP-inibitor interaction was characterized by using isothermal calorimetry (ITC) and the inhibitor was shown to be a tight binding inhibitor with binding affinities in the low nanomolar range. Active site mutations were also carried out to determine important residues involved in protein inhibitor interactions.

For ^{15}N labeling studies by NMR, with proteins expressed from the baculovirus system, a well defined ^{15}N labeling media was required. Two insect cell media, IML 406 and IML 455, were thus formulated for high yield protein production of selectively ^{15}N labeled and uniformly ^{15}N labeled proteins in insect cells. The protein glutathione S transferase (GST) was used for expression and labeling trials to optimize the medium. Also GST labeling in insect cells was compared to GST labeling in *E.coli*. For the most studied amino acids essential for insect cells the ^{15}N-HSQC spectra, recorded with GST labeled in insect cells, showed no cross labeling and provided therefore spectra of better quality compared to NMR spectra of GST expressed in *E. coli*. Based on these findings a first simple overview of the network of the amino acid metabolism in *E. coli* and insect cells focused on nitrogen could be provided. The second part of this study involved cloning and expression studies on human CDK 6 purified from insect cells. Cyclin

dependent kinases are the principle regulators of cell cycle progression, through activation of cell cycle checkpoints and their activities are also associated with the processes like differentiation and apoptosis. In cancerous cells, deregulated CDK activity is often observed. CDK 6 was expressed and purified to high purity and stability. Crystallization trials were conducted to obtain uncomplexed monomeric CDK 6 and also an attempt was made to obtain its crystal structure with inhibitors targeting the ATP domain, but crystals which gave good diffraction pattern could not be obtained. Preliminary ^{15}N labeling studies with CDK 6 could be successfully carried out using the formulated IML 406 media.

In the final part of the study the proteolytic cleavage pattern of IGFBPs (IGFBP-4 and-5) with the endopeptidase calpain was carried out. The calpains form a growing family of structurally related intracellular multidomain cysteine protease. It plays an important role in cell cycle progression from G1 to S phase of cell cycle and its cleavage of IGFBPs could be a means of regulating the IGF independent action of IGFBPs which are not completely understood. The study was performed to elucidate rules that govern calpain cleavage specificity. The study has not revealed clear preferences for amino acid sequences, although in many cases a hydrophobic residue is present in the P_2 position. In conclusion, we propose that calpains seem not to recognize amino acid sequences in target proteins but cleave with low sequence specificity at unstructured or solvent exposed fragments that connect folded, stable domains of target proteins suggesting that calpains may modulate functions of substrate proteins by cutting their interdomain regions.

Zusammenfassung

Im Zentrum dieser Doktorarbeit stand die strukturelle und funktionelle Charakterisierung wichtiger Proteine des Zellzyklus, welche Funktionen in der Regulation der Genexpression, sowie im geregelten Ablauf von Zellzyklus-Progression und Zellteilung erfüllen. Zu nennen sind hier vor allem jene Proteine, deren Deregulation mit der Krankheit Krebs in Verbindung steht, wie zum Beispiel die Histon-Deacetylasen (HDACs), die Cyclin-abhängigen Kinasen (CDKs) und die Familie der insulinähnlichen Wachstumshormone (IGFBPs). Um die selektive ^{15}N-Markierung dieser Proteine für NMR-Experimente zu ermöglichen, wurden zwei Insektenzellmedien, IML406 und IML455, entwickelt.

Histon-Deacetylasen (HDACs) katalysieren die Entfernung der Acetylgruppen von Acetyl-Lysin innerhalb der Core-Histone und stabilisieren dadurch die lokale Kondensation der DNA mit nachfolgender Verminderung der Genexpression. Die HDAC-Deregulation spielt eine Schlüsselrolle im Wachstum von Krebszellen und es wurde gezeigt, dass die HDAC-Inhibierung die Transkriptionsrate einer Reihe von Genen (2-5% des gesamten Genoms), darunter auch Tumor-Suppressor-Gene, steigert. HDAC-Inhibitoren stellen daher potentielle Krebsmedikamente dar und werden intensiv erforscht. In dieser Arbeit wurde die Kristallstruktur des HDAC-Homologs HDLP im Komplex mit einem auf Thiophen basierenden Inhibitor mit Hydroxamgruppe gelöst. Die bioenergetischen Eigenschaften der HDLP-Inhibitor-Interaktion wurden durch isotherme Titrations-Kaloriemetrie (ITC) charakterisiert und es wurde gezeigt, dass der Inhibitor mit Affinitäten im nanomolaren Bereich an sein Substrat bindet. Des weiteren wurden Mutationen im aktiven Zentrum durchgefuhrt, um wichtige, an Protein-Inhibitor-Interaktionen beteiligte Reste zu ermitteln. Auch wurden mehrere Konstrukte von Säuger-HDAC8 kloniert und gereinigt, um deren Sekundärstruktur anschließend durch CD-Spektroskopie zu analysieren.

Für NMR-Versuche mit ^{15}N-markiertem Protein, das mit dem Baculovirus/Insektenzell-System exprimiert wurde, war ein spezielles Markierungsmedium nötig. Dafür wurden zwei Insektenzell-Medien entwickelt, IML 406 und IML 455, welche eine hohe Proteinausbeute in Verbindung mit selektiv ^{15}N-

markiertem (IML 406) und einheitlich ^{15}N-markiertem (IML 455) Protein ermöglichen. Das Protein Glutathion-S-Transferase (GST) wurde für die zur Optimierung des Mediums nötigen Expressions- und Markierungsexperimente benutzt. Ein Vergleich der ^{15}N-Markierung des GST in Insektenzellen mit der Markierung in *E.coli* wurde ebenfalls durchgefuhrt. Die aufgenommenen ^{15}N-HSQC-Spektren der meisten Aminosäuren des in Insektenzellen exprimierten GST ließen dabei kein Cross-labeling erkennen und lieferten Spektren höherer Qualität als die ermittelten NMR-Spektren von in *E. coli* exprimiertem GST. Basierend auf diesen Erkenntnissen konnte ein erster einfacher Überblick über das Netzwerk des Aminosäurestoffwechsels in *E. coli* und Insektenzellen mit Fokus auf den Stickstoff erstellt werden. Der zweite Teil dieser Studie umfasste Untersuchungen des humanen CDK 6, welches kloniert, in Insektenzellen exprimiert und aufgereinigt wurde. Cyclin-abhängige Kinasen agieren als die Hauptregulatoren der Zellzyklus-Progression, indem sie spezielle Kontrollpunkte des Zellzyklus aktivieren. Die Aktivitäten der CDKs haben auch Einfluss auf Prozesse wie die Zelldifferenzierung und die Apoptose. In Krebszellen wird häufig eine abnorme Aktivität der CDKs beobachtet. In dieser Arbeit wurde CDK 6 exprimiert und als stabiles und reines Protein aufgereinigt. Damit wurden Kristallisationsversuche durchgeführt, mit dem Ziel, die Struktur von nicht komplexiertem monomerem CDK 6 zu ermitteln. Es wurden auch Kristallisationsansätze mit Inhibitoren, die gegen die ATP-Domäne gerichtet waren pipettiert. Allerdings konnten keine Kristalle mit ausreichend guter Rontgenstreuung Beugungsverhalten für eine Strukturbestimmung erhalten werden. Anfangliche ^{15}N-Markierungsversuche mit CDK 6 konnten erfolgreich mit Medium IML406 durchgeführt werden.

Im letzten Teil dieser Doktorarbeit wurde eine Analyse der proteolytischen Spaltung von IGFBPs (IGFBP-4 und-5) mit der Endopeptidase Calpain durchgeführt. Die Calpaine formen eine wachsende Familie von strukturell verwandten intrazellulären Multidomänen-Cystein-Proteasen. Sie spielen eine wichtige Rolle im Verlauf des Zellzyklus beim Übergang von der G1 zur S-Phase und die Funktion der Spaltung der IGFBPs könnte in der Regulation der IGF- unabhängigen Aktivität der IGFBPs liegen. Die hier vorliegenden Untersuchungen wurden angestellt um grundlegende Regeln der Kalpain-Spaltungs-Spezifität zu ergründen. Dabei konnten zwar keine klaren Präferenzen für spezifische Aminosäuresequenzen festgestellt werden, allerdings war in vielen Fällen

ein hydrophober Rest in der P_2 – Position betaligt. Calpaine scheinen also keine spezifischen Aminosäuresequenzen im Zielprotein zu erkennen, sondern spalten mit nur geringer Sequenzspezifität, und zwar bevorzugt in unstrukturierten oder dem Lösungsmittel exponierten Bereichen des Proteins, welche gefaltete und stabile Domänen desselben verbinden. Die Aufgabe des Calpains könnte also in einer Modulation von Funktionen des Substratproteins liegen, gesteuert durch die proteolytische Spaltung in dessen Zwischendomänenregionen.

References

Baxter, R.C. 2000. Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities. *Am J Physiol Endocrinol Metab* **278:** E967-976.

Bjerling, P., Silverstein, R.A., Thon, G., Caudy, A., Grewal, S., and Ekwall, K. 2002. Functional divergence between histone deacetylases in fission yeast by distinct cellular localization and in vivo specificity. *Mol Cell Biol* **22:** 2170-2181.

Brinkmann, H., Dahler, A.L., Popa, C., Serewko, M.M., Parsons, P.G., Gabrielli, B.G., Burgess, A.J., and Saunders, N.A. 2001. Histone hyperacetylation induced by histone deacetylase inhibitors is not sufficient to cause growth inhibition in human dermal fibroblasts. *J Biol Chem* **276:** 22491-22499.

Bunn, R.C., and Fowlkes, J.L. 2003. Insulin-like growth factor binding protein proteolysis. *Trends in Endocrinology and Metabolism* **14:** 176-181.

Burgess, A.J., Pavey, S., Warrener, R., Hunter, L.J., Piva, T.J., Musgrove, E.A., Saunders, N., Parsons, P.G., and Gabrielli, B.G. 2001. Up-regulation of p21(WAF1/CIP1) by histone deacetylase inhibitors reduces their cytotoxicity. *Mol Pharmacol* **60:** 828-837.

Carpentier, Y., Liautaud-Roger, F., Labbe, F., Loirette, M., Collery, P., and Coninx, P. 1987. Effect of gallium at two phases of the CA 755 tumour growth. *Anticancer Res* **7:** 745-748.

Cascio, M., Schoppa, N.E., Grodzicki, R.L., Sigworth, F.J., and Fox, R.O. 1993. Functional expression and purification of a homomeric human alpha 1 glycine receptor in baculovirus-infected insect cells. *J Biol Chem* **268:** 22135-22142.

Chan, L.C., Greenfield, P.F., and Reid, S. 1998. Optimising fed-batch production of recombinant proteins using the baculovirus expression vector system. *Biotechnol Bioeng* **59:** 178-188.

Chen, L., Fischle, W., Verdin, E., and Greene, W.C. 2001. Duration of nuclear NF-kappaB action regulated by reversible acetylation. *Science* **293:** 1653-1657.

Chiou, C.J., and Wu, M.C. 1990. Expression of human granulocyte-macrophage colony-stimulating factor gene in insect cells by a baculovirus vector. *FEBS Lett* **259:** 249-253.

Clemmons, D.R. 2001. Use of mutagenesis to probe IGF-binding protein structure/function relationships. *Endocr Rev* **22:** 800-817.

Cohen, H.Y., Lavu, S., Bitterman, K.J., Hekking, B., Imahiyerobo, T.A., Miller, C., Frye, R., Ploegh, H., Kessler, B.M., and Sinclair, D.A. 2004. Acetylation of the C terminus of Ku70 by CBP and PCAF controls Bax-mediated apoptosis. *Mol Cell* **13:** 627-638.

Creemers, A.F., Klaassen, C.H., Bovee-Geurts, P.H., Kelle, R., Kragl, U., Raap, J., de Grip, W.J., Lugtenburg, J., and de Groot, H.J. 1999. Solid state 15N NMR evidence for a complex Schiff base counterion in the visual G-protein-coupled receptor rhodopsin. *Biochemistry* **38:** 7195-7199.

Cress, W.D., and Seto, E. 2000. Histone deacetylases, transcriptional control, and cancer. *J Cell Physiol* **184:** 1-16.

Croall, D.E., and DeMartino, G.N. 1991. Calcium-activated neutral protease (calpain) system: structure, function, and regulation. *Physiol Rev* **71:** 813-847.

de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S., and van Kuilenburg, A.B. 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* **370:** 737-749.

Doverskog, M., Bertram, E., Ljunggren, J., Ohman, L., Sennerstam, R., and Haggstrom, L. 2000. Cell cycle progression in serum-free cultures of Sf9 insect cells: modulation by conditioned medium factors and implications for proliferation and productivity. *Biotechnol Prog* **16:** 837-846.

Duffy, S., Tsao, K.L., and Waugh, D.S. 1998. Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. *Anal Biochem* **262:** 122-128.

Dulic, V., Kaufmann, W.K., Wilson, S.J., Tlsty, T.D., Lees, E., Harper, J.W., Elledge, S.J., and Reed, S.I. 1994. p53-dependent inhibition of cyclin-dependent kinase activities in human fibroblasts during radiation-induced G1 arrest. *Cell* **76:** 1013-1023.

Dyson, N. 1994. pRB, p107 and the regulation of the E2F transcription factor. *J Cell Sci Suppl* **18:** 81-87.

Elledge, S.J., and Harper, J.W. 1994. Cdk inhibitors: on the threshold of checkpoints and development. *Curr Opin Cell Biol* **6:** 847-852.

El-Osta, A., Kantharidis, P., Zalcberg, J.R., and Wolffe, A.P. 2002. Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. *Mol Cell Biol* **22:** 1844-1857.

Esteller, M. 2003. Profiling aberrant DNA methylation in hematologic neoplasms: a view from the tip of the iceberg. *Clin Immunol* **109:** 80-88.

Finnin, M.S., Donigian, J.R., Cohen, A., Richon, V.M., Rifkind, R.A., Marks, P.A., Breslow, R., and Pavletich, N.P. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401:** 188-193.

Fischle, W., Dequiedt, F., Hendzel, M.J., Guenther, M.G., Lazar, M.A., Voelter, W., and Verdin, E. 2002. Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Mol Cell* **9:** 45-57.

Forsberg, E.C., and Bresnick, E.H. 2001. Histone acetylation beyond promoters: long-range acetylation patterns in the chromatin world. *Bioessays* **23:** 820-830.

Gabrielli, B.G., Johnstone, R.W., and Saunders, N.A. 2002. Identifying molecular targets mediating the anticancer activity of histone deacetylase inhibitors: a work in progress. *Curr Cancer Drug Targets* **2:** 337-353.

Gao, L., Cueto, M.A., Asselbergs, F., and Atadja, P. 2002. Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J Biol Chem* **277:** 25748-25755.

Gibbs, B.S., Wojchowski, D., and Benkovic, S.J. 1993. Expression of rat liver phenylalanine hydroxylase in insect cells and site-directed mutagenesis of putative non-heme iron-binding sites. *J Biol Chem* **268:** 8046-8052.

Glading, A., Lauffenburger, D.A., and Wells, A. 2002. Cutting to the chase: calpain proteases in cell motility. *Trends Cell Biol* **12:** 46-54.

Glaser, K.B., Staver, M.J., Waring, J.F., Stender, J., Ulrich, R.G., and Davidsen, S.K. 2003. Gene expression profiling of multiple histone deacetylase (HDAC) inhibitors: defining a common gene set produced by HDAC inhibition in T24 and MDA carcinoma cell lines. *Mol Cancer Ther* **2:** 151-163.

Goll, D.E., Thompson, V.F., Li, H., Wei, W., and Cong, J. 2003. The calpain system. *Physiol Rev* **83:** 731-801.

Gore, S.D., Weng, L.J., Figg, W.D., Zhai, S., Donehower, R.C., Dover, G., Grever, M.R., Griffin, C., Grochow, L.B., Hawkins, A., et al. 2002. Impact of prolonged infusions of the putative differentiating agent sodium phenylbutyrate on myelodysplastic syndromes and acute myeloid leukemia. *Clin Cancer Res* **8:** 963-970.

Grana, X., Garriga, J., and Mayol, X. 1998. Role of the retinoblastoma protein family, pRB, p107 and p130 in the negative control of cell growth. *Oncogene* **17:** 3365-3383.

Grignani, F., De Matteis, S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F.F., Zamir, I., et al. 1998. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. *Nature* **391:** 815-818.

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

Huang, L., Sowa, Y., Sakai, T., and Pardee, A.B. 2000. Activation of the p21WAF1/CIP1 promoter independent of p53 by the histone deacetylase inhibitor suberoylanilide hydroxamic acid (SAHA) through the Sp1 sites. *Oncogene* **19:** 5712-5719.

Ito, K., Barnes, P.J., and Adcock, I.M. 2000. Glucocorticoid receptor recruitment of histone deacetylase 2 inhibits interleukin-1beta-induced histone H4 acetylation on lysines 8 and 12. *Mol Cell Biol* **20:** 6891-6903.

Ito, M. 2000. Factors controlling cyclin B expression. *Plant Mol Biol* **43:** 677-690.

Jaques, G., Noll, K., Wegmann, B., Witten, S., Kogan, E., Radulescu, R.T., and Havemann, K. 1997. Nuclear localization of insulin-like growth factor binding protein 3 in a lung cancer cell line. *Endocrinology* **138:** 1767-1770.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N.P. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376:** 313-320.

Johnstone, R.W. 2002. Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. *Nat Rev Drug Discov* **1:** 287-299.

Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Grol, M., Demuth, D., Schumacher, R., Dony, C., Lang, K., et al. 1998. Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): Implications for IGF and IGF-I receptor interactions. *Embo Journal* **17:** 6558-6572.

Khandwala, H.M., McCutcheon, I.E., Flyvbjerg, A., and Friend, K.E. 2000. The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. *Endocr Rev* **21:** 215-244.

Kim, Y.M., and Ziegler, D.M. 2000. Size limits of thiocarbamides accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab Dispos* **28:** 1003-1006.

Kitazono, M., Rao, V.K., Robey, R., Aikou, T., Bates, S., Fojo, T., and Goldsmith, M.E. 2002. Histone deacetylase inhibitor FR901228 enhances adenovirus infection of hematopoietic cells. *Blood* **99:** 2248-2251.

Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J.W., Elledge, S., Nishimoto, T., Morgan, D.O., Franza, B.R., and Roberts, J.M. 1992. Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle. *Science* **257:** 1689-1694.

Kouzarides, T. 1999. Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev* **9:** 40-48.

Krauer, K.G., Burgess, A., Buck, M., Flanagan, J., Sculley, T.B., and Gabrielli, B. 2004. The EBNA-3 gene family proteins disrupt the G2/M checkpoint. *Oncogene* **23:** 1342-1353.

Kumar, M., Shanker, S., Rawat, V.P., and Das, R.H. 2001. Characterization of the antiapoptotic (p35) gene homologue of Spodoptera litura nucleopolyhedrosis virus (SlNPV). *Mol Biol Rep* **28:** 167-173.

Li, H., Thompson, V.F., and Goll, D.E. 2002. Effects of autolysis an properties of mu- and m-calpain. *Molecular Biology of the Cell* **13:** 304A-304A.

Li, W., Chen, H.Y., and Davie, J.R. 1996. Properties of chicken erythrocyte histone deacetylase associated with the nuclear matrix. *Biochem J* **314 (Pt 2):** 631-637.

Lindemann, R.K., Gabrielli, B., and Johnstone, R.W. 2004. Histone-deacetylase inhibitors for the treatment of cancer. *Cell Cycle* **3:** 779-788.

Luo, R.X., Postigo, A.A., and Dean, D.C. 1998. Rb interacts with histone deacetylase to repress transcription. *Cell* **92:** 463-473.

Marks, P.A., Richon, V.M., Miller, T., and Kelly, W.K. 2004. Histone deacetylase inhibitors. *Adv Cancer Res* **91:** 137-168.

Marks, P.A., Richon, V.M., and Rifkind, R.A. 2000. Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *J Natl Cancer Inst* **92:** 1210-1216.

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

Mellgren, R.L. 1997. Evidence for participation of a calpain-like cysteine protease in cell cycle progression through late G(1) phase. *Biochemical and Biophysical Research Communications* **236:** 555-558.

Miller, T.A., Witter, D.J., and Belvedere, S. 2003. Histone deacetylase inhibitors. *J Med Chem* **46:** 5097-5116.

Mitsiades, C.S., Mitsiades, N.S., McMullan, C.J., Poulaki, V., Shringarpure, R., Hideshima, T., Akiyama, M., Chauhan, D., Munshi, N., Gu, X., et al. 2004. Transcriptional signature of histone deacetylase inhibition in multiple myeloma: biological and clinical implications. *Proc Natl Acad Sci U S A* **101:** 540-545.

Moldoveanu, T., Hosfield, C.M., Lim, D., Jia, Z., and Davies, P.L. 2003. Calpain silencing by a reversible intrinsic mechanism. *Nat Struct Biol* **10:** 371-378.

Morgan, D.O. 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* **13:** 261-291.

Nobiron, I., O'Reilly, D.R., and Olszewski, J.A. 2003. Autographa californica nucleopolyhedrovirus infection of Spodoptera frugiperda cells: a global analysis of

host gene regulation during infection, using a differential display approach. *J Gen Virol* **84:** 3029-3039.

Ohtani, N., Yamakoshi, K., Takahashi, A., and Hara, E. 2004. The p16INK4a-RB pathway: molecular link between cellular senescence and tumor suppression. *J Med Invest* **51:** 146-153.

Ohtsubo, M., Gamou, S., and Shimizu, N. 1998. Antisense oligonucleotide of WAF1 gene prevents EGF-induced cell-cycle arrest in A431 cells. *Oncogene* **16:** 797-802.

Pagano, M. 1997. Cell cycle regulation by the ubiquitin pathway. *Faseb J* **11:** 1067-1075.

Patel, Y.M., and Lane, M.D. 2000. Mitotic clonal expansion during preadipocyte differentiation: calpain-mediated turnover of p27. *J Biol Chem* **275:** 17653-17660.

Pavletich, N.P. 1999. Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J Mol Biol* **287:** 821-828.

Payet, L.D., Firth, S.M., and Baxter, R.C. 2004. The role of the acid-labile subunit in regulating insulin-like growth factor transport across human umbilical vein endothelial cell monolayers. *J Clin Endocrinol Metab* **89:** 2382-2389.

Payet, L.D., Wang, X.H., Baxter, R.C., and Firth, S.M. 2003. Amino- and carboxyl-terminal fragments of insulin-like growth factor (IGF) binding protein-3 cooperate to bind IGFs with high affinity and inhibit IGF receptor interactions. *Endocrinology* **144:** 2797-2806.

Pijnappel, W.W., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R., and Stewart, A.F. 2001. The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev* **15:** 2991-3004.

Qiu, L., Burgess, A., Fairlie, D.P., Leonard, H., Parsons, P.G., and Gabrielli, B.G. 2000. Histone deacetylase inhibitors trigger a G2 checkpoint in normal cells that is defective in tumor cells. *Mol Biol Cell* **11:** 2069-2083.

Richon, V.M., Sandhoff, T.W., Rifkind, R.A., and Marks, P.A. 2000. Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. *Proc Natl Acad Sci U S A* **97:** 10014-10019.

Rountree, M.R., Bachman, K.E., Herman, J.G., and Baylin, S.B. 2001. DNA methylation, chromatin inheritance, and cancer. *Oncogene* **20:** 3156-3165.

Russo, A.A., Jeffrey, P.D., and Pavletich, N.P. 1996. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* **3:** 696-700.

Russo, A.A., Tong, L., Lee, J.O., Jeffrey, P.D., and Pavletich, N.P. 1998. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395:** 237-243.

Sandor, V., Senderowicz, A., Mertins, S., Sackett, D., Sausville, E., Blagosklonny, M.V., and Bates, S.E. 2000. P21-dependent g(1)arrest with downregulation of cyclin D1 and upregulation of cyclin E by the histone deacetylase inhibitor FR901228. *Br J Cancer* **83:** 817-825.

Shand, J.H., Beattie, J., Song, H., Phillips, K., Kelly, S.M., Flint, D.J., and Allan, G.J. 2003. Specific amino acid substitutions determine the differential contribution of the N- and C-terminal domains of insulin-like growth factor (IGF)-binding protein-5 in binding IGF-I. *Journal of Biological Chemistry* **278:** 17859-17866.

Shapiro, G.I., Edwards, C.D., Ewen, M.E., and Rollins, B.J. 1998. p16INK4A participates in a G1 arrest checkpoint in response to DNA damage. *Mol Cell Biol* **18:** 378-387.

Sherr, C.J. 1994a. G1 phase progression: cycling on cue. *Cell* **79:** 551-555.

Sherr, C.J. 1994b. Growth factor-regulated G1 cyclins. *Stem Cells* **12 Suppl 1:** 47-55; discussion 55-47.

Sherr, C.J. 1996. Cancer cell cycles. *Science* **274:** 1672-1677.

Sherr, C.J., and Roberts, J.M. 1999. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13:** 1501-1512.

Somoza, J.R., Skene, R.J., Katz, B.A., Mol, C., Ho, J.D., Jennings, A.J., Luong, C., Arvai, A., Buggy, J.J., Chi, E., et al. 2004. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure (Camb)* **12:** 1325-1334.

Strahl-Bolsinger, S., Hecht, A., Luo, K., and Grunstein, M. 1997. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev* **11:** 83-93.

Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395:** 917-921.

Vannini, A., Volpari, C., Filocamo, G., Casavola, E.C., Brunetti, M., Renzoni, D., Chakravarty, P., Paolini, C., De Francesco, R., Gallinari, P., et al. 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:** 15064-15069.

Villar-Garea, A., and Esteller, M. 2004. Histone deacetylase inhibitors: understanding a new wave of anticancer agents. *Int J Cancer* **112:** 171-178.

Wade, P.A. 2001. Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Hum Mol Genet* **10:** 693-698.

Wolffe, A.P., Urnov, F.D., and Guschin, D. 2000. Co-repressor complexes and remodelling chromatin for repression. *Biochem Soc Trans* **28:** 379-386.

Yang, W.M., Tsai, S.C., Wen, Y.D., Fejer, G., and Seto, E. 2002. Functional domains of histone deacetylase-3. *J Biol Chem* **277:** 9447-9454.

You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. 2001. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc Natl Acad Sci U S A* **98:** 1454-1458.

Zerfass-Thome, K., Zwerschke, W., Mannhardt, B., Tindle, R., Botz, J.W., and Jansen-Durr, P. 1996. Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene* **13:** 2323-2330.

Zeslawski, W., Beisel, H.G., Kamionka, M., Kalus, W., Engh, R.A., Huber, R., Lang, K., and Holak, T.A. 2001. The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5. *Embo Journal* **20:** 3638-3644.

Zhang, M., and Raveche, E.S. 1998. Apoptosis induction in fludarabine resistant malignant B-1 cells by G2-M cell cycle arrest. *Oncol Rep* **5:** 23-30.

Zhu, Z., Bulgakov, O.V., Scott, S.S., and Dalton, J.T. 2001. Recombinant expression and purification of human androgen receptor in a baculovirus system. *Biochem Biophys Res Commun* **284:** 828-835.

Zindy, F., Soares, H., Herzog, K.H., Morgan, J., Sherr, C.J., and Roussel, M.F. 1997. Expression of INK4 inhibitors of cyclin D-dependent kinases during mouse brain development. *Cell Growth Differ* **8:** 1139-1150.

Appendix: Abbreviations and Symbols

- Å — Ångstrom (10^{-10} m)
- aa — amino acid
- AcNPV — *Autographa californica* nuclear polyhedrosis virus
- ATP — adenosine triphosphate
- 1D — one-dimensional
- APS — ammonium peroxodisulfate
- BEVS — baculovirus expression vectors system
- bp — base pair
- B_0 — magnetic field
- B_{eff} — effektive Magnetic field
- BSA — bovine serum albumin
- BV — budded virus
- CAK — CDK activating kinase
- CDC — cyclin-dependent kinase
- cDNA — complimentary DNA
- CDK — cyclin-dependent kinase
- CIP — inhibitor of kinase
- COSY — correlation spectroscopy
- CSF — colony-stimulating factor
- δ — chemical shift
- Da — Dalton ($g\ mol^{-1}$)
- DHFR — dihydrofolate reductase
- DMSO — dimethylsulfoxide
- DNA — deoxyribonucleic acid
- DNaseI — deoxyribonuclease I
- DTT — Dithiothreitol
- E2F — factor interacting with adenovirus E2-promoter
- EDTA — ethylenediamine tetraacetic acid
- FID — free induction decay
- G — gravity ($9.81\ m\ s^{-2}$)
- GSH — reduced glutathione
- GSSG — oxidized glutathione

- •ΔG freie Reaktionsenergie
- • GST glutathione S-transferase
- • HAT histone acetyltransferase activity
- • HDAC histone deacetylase
- • HDACi histone deacetylase inhibitor
- • HSQC heteronuclear single quantum coherence
- • Hz Hertz
- • IGF insulin-like growth factor
- • IGFBP IGF binding protein
- • IGFBPrP IGFBP related protein
- • IGF-IR IGF receptor type I
- • INK4 inhibitor of CDK4 and CDK6
- • IPTG isopropyl-β-thiogalactopyranoside
- • IRS insulin receptor substrates
- • KIP inhibitor of kinase
- • LB Luria-Broth medium
- • M mol l^{-1}
- • MAD multiwavelength anomalous diffraction
- • MIR multiple isomorphous replacement
- • MM minimal medium
- • MW molecular weight
- •MOI multiplicity of infection
- • NiNTA nickel-nitrilotriacetic acid
- • NLS nuclear localization signal
- • NMR nuclear magnetic resonance
- • NOE nuclear Overhauser effect
- • NOESY nuclear Overhauser enhancement spectroscopy
- • OD optical density
- • PAGE polyacrylamide gel electrophoresis
- • PBS phosphate-buffered saline
- • POL DNA polymerase α
- • ppm parts per million
- • RB retinoblastoma susceptibility protein
- • RF RING finger motif, or radio frequency

- RMSD root mean square deviation
- RNaseA ribonuclease A
- SDS sodium dodecyl sulfate
- Sf Spodoptera frugiperda
- TEMED N,N,N',N'-tetramethylethylendiamine
- TK thymidine kinase
- TOCSY total correlation spectroscopy
- TS thymidylate synthase
- ZF zinc finger motif

Amino acids and nucleotides are abbreviated according to either one or three letter IUPAC code.