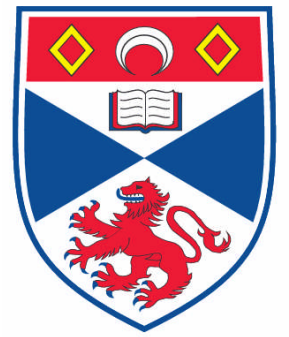
CHARACTERISATION OF SULFOLOBUS SOLFATARICUS ARD1, A PROMISCUOUS N-ACETYLTRANSFERASE

Dale Tara Mackay

A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



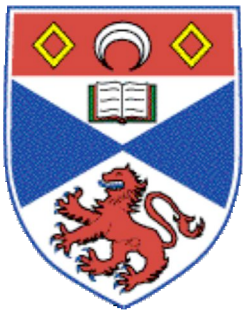
2008

Characterisation of *Sulfolobus solfataricus* Ard1, a promiscuous N-acetyltransferase

Dale Tara Mackay

A thesis submitted for the degree of Doctor of Philosophy

September 2007



University of St. Andrews

Figures & Tables

DECLARATION

I, Dale Tara Mackay, hereby certify that this thesis, which is approximately 35, 000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date …… Signature of candidate ……

I was admitted as a research student in October 2003 and as a candidate for the degree of PhD in October, 2004; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2006.

Date …… Signature of candidate ……..

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ……. Signature of supervisor…….

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and

supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date ……. Signature of candidate ……..

Abbreviations

3’	3 prime end (DNA)
5’	5 prime end (DNA)
3D PSSM/PHYRE	3D position-specific score matrix/protein homology recognition engine
$A_{276/280}$	Absorbance at 276/280nm etc.
AAC	Aminoglycoside antibiotic N-acetyltransferase
AANAT	Arylalkylamine N-acetyltransferase
AcCoA	Acetyl coenzyme A
Af	*Archaeoglobus fulgidus*
Alba	Acetylation lowers binding affinity
AML	Acute myeloid leukaemia
Ard1	Arrest defective protein 1
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
CBP	CREB binding protein
CCP4	Collaborative computational project number 4
CD	Circular dichroism
CDSSTR	Secondary structure determination algorithm for CD data analysis
CoASH	Coenzyme A
CYCI UAS2	Transcriptional regulator (oxidative phosphorylation)
DAPI	4,6-diamidino-2-phenylindole
DMSO	Dimethyl sulfoxide
dsDNA	Double stranded DNA

DSC	Differential scanning calorimetry
DTT	Dithiothreitol
E1A	Adenovirus oncoprotein
EDTA	Ethylenediaminetetraacetic acid
Elp3	Elongator protein 3
ESA1	Essential sas family acetyltransferase
ESRF	European synchrotron radiation facility
FMR1	Fragile-X mental retardation gene
GNAT	Gcn-5 related acetyltransferase
GST	Glutathione-S-transferase
HAT	Histone acetyltransferase
HBO1	MYST family acetyltransferase
HD	Huntington's disease
HDAC	Histone deacetyltransferase
HIV Tat	Human immunodeficiency virus transactivator
Hjc / Hje	Holliday junction resolvase (cutting / endonuclease)
HMG	Chromatin associated proteins
Hpa2/3	Histone [and other] protein acetyltransferase
H. sap	*Homo sapiens*
Hsp	Heat shock protein
IF_3	Translation initiation factor 3
IPTG	Isopropyl β-D-thiogalactopyranoside
k_{cat}	Overall catalytic rate (turnover number)
kDa	Kilo Dalton
k_m	Michaelis constant = substrate concentration at Vmax/2

MAD	Multiple wavelength anomalous dispersion
MALDI-ToF	Matrix-assisted laser desorption/ionization tome of flight
MBP	Maltose binding protein
MSDfold (SSM)	Molecular structure database (secondary structure matching)
MORF	Moz related factor
MOZ	Monocytic leukaemia zinc-finger protein
MPD	M-phenylenediamine
MRSA	Methicillin-resistant *Staphylococcus aureus*
Mth	*Methanothermobacter thermautotrophicus*
Mja	*Methanococcus jannaschii*
MW	Molecular weight
MyoD	Myogenic regulator D
MYST	Moz, Yfb2/Sas3, Sas2 and Tip60 superfamily
NAD^+	Nicotinamide adenine dinucleotide
NAT	N-acetyltransferase
NHL	Non-Hodgkin's lymphoma
NRMSD	Normalised root mean square deviation
OD	Optical density
ORC	Origin recognition complex
p21	Cyclin-dependent kinase inhibitor
p300	E1A-associated protein
p53	Tumour suppressor protein & general transcriptional activator
PCAF	p300/CBP associated factor
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction

PDB	Protein data bank
PEG	Polyethylene glycol
Pfu	DNA proof reading polymerase from *Pyrococcus furiosus*
Pho	*Pyrococcus horikoshii*
pI	Isoelectric point
RA	Rheumatoid arthritis
Rb	Retinoblastoma tumour suppressor protein
Rec	Recombinant
RFZ	Rotation function score
Rim	*E.coli* ribosomal protein acetyltransferase
rpm	Revolutions per minute
S.aci	*Sulfolobus acidocaldareus*
Sas 2/3	Something about silencing protein 2/3
S.cer	*Saccharomyces cerevisiae*
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Se-Met	Selenomethionine
Sirtuin (Sir)	Silent mating type information regulation two family
Spt10	Yeast transcriptional regulator
SRC-1	Steroid nuclear receptor co-activator
SSB	Single stranded binding protein
ssDNA	Single stranded deoxyribonucleic acid
Ssh	*Sulfolobus shibatae*
Sso	*Sulfolobus solfataricus*
*Sso*10b	*Sulfolobus solfataricus* Alba1
$TAF_{II}250$	TATA-binding protein-associated factor

TCA	Trichloroacetic acid
TFIIE/F	General transcription factor II E/F
TFZ	Translation function score
Tip 60	HIV Tat-interactive protein
TNF-α	Tumour necrosis factor alpha
UV	Ultra violet
V_{max}	Maximum velocity of reaction
Wt	Wild type
XPF	*Xeroderma pigmentosum* complementation group F

ABSTRACT

Compaction of DNA into chromatin is an important feature of every living cell. This compaction phenomenon is brought about and maintained by a variety of DNA binding proteins, which have evolved to suit the specific needs of the different cell types spanning the three kingdoms of life; the eukaryotes, prokaryotes and archaea. *Sulfolobus solfataricus*, a member of the crenarchaeal subdivision of the archaea, has two prominent DNA binding proteins known as Alba (1&2) and *Sso*7d. Alba1 is acetylated *in vivo* at two positions and this modification lowers its' affinity for binding DNA. Acetylation levels impact many cellular processes and in higher organisms play a critical role in the development of many cancers and other diseases.

This thesis documents the finding and characterisation of the N-terminal acetyltransferase (*ss*Ard1) of *Sso*Alba1, based on its' sequence homology to the catalytic subunits Ard1, Nat3 and Mak3 belonging to the larger eukaryal Nat complexes NatA, NatB and NatC, respectively. Mutagenesis studies revealed that *ss*Ard1 preferentially acetylates N-termini bearing a serine or alanine residue at position 1 (after methionine cleavage). It is also capable of acetylating other proteins with very different physical structures. These findings allow classification of *ss*Ard1 as a promiscuous acetyltransferase belonging to the Gcn5-N-acetyltransferase (GNAT) superfamily. The active site of the enzyme was examined through mutagenesis studies, revealing that the mechanism of acetylation is likely to proceed through a direct acetyl transfer involving a tetrahedral intermediate. Structural studies provided some insight into the molecular structure of *ss*Ard1.

Acknowledgements

A special thank you goes to my supervisors Professor Malcolm White and Professor Garry Taylor for all their teaching, understanding and support. I thank them especially for putting up with me (!) and always being available to encourage and give guidance when needed. I would also like to thank all my St. Andrews colleagues, too numerous to name, for their friendship, constructive criticism and for all their help particularly regarding the foreign field of protein crystallography. I really miss playing football with all you guys!

My gratitude goes to the mass spectrometry gurus Dr. Catherine Botting, Alex Houston and Dr. Robin Antrobus for the many samples they analysed for me. Thanks to Dr. Steve Bell (Cambridge) and previous lab member Ben Wardleworth for Alba constructs, Dr. Huanting Liu and Dr. Louise Major for HisTev and pLOU vectors. My thanks also goes to Clare Jelinska for being my mentor when I first arrived at St. Andrews.

My studies were very well supported and encouraged by all my family and I would like to thank them immensely, especially my mother who did so much more than I could ever have asked of her. I thank my husband Pete for his incredible patience, financial support and constant motivation tactics. I'd also like to acknowledge my flatmate (for the main part), William, and his parents for their friendship and for making life in St. Andrews a lot more pleasant. Thanks also to all my friends from home and University. Many thanks to Ian Glover for friendship, COFFEE and enthusiasm over my studies!

And finally but most importantly, thanks be to God, the greatest scientist man will ever know: "When I consider your heavens, your fingers work I trace, the moon and stars, which you have set in place, what is man that you are mindful of him, the

son of man that you care for him? You have set him a little lower than the heavenly beings and crowned him with glory and honour. You made him rule over the works of your hands; you put everything under his feet….O LORD our Lord, how majestic is your name in all the earth" Psalm 8.

This PhD was made possible through the generous funding of BBSRC.

Chapter 1: Introduction

Introduction

DNA compaction is vital for the cell. Eukaryotic cells can contain several metres of DNA and this DNA must fit into the nucleus in an orderly way so that it can be easily accessed for repair, transcription and DNA replication when required. In addition, the winding and coating of DNA by protein complexes affords protection against many forms of damage and sustains the integrity of the DNA molecule.

1.1 Histones and Chromatin

In eukaryotes, DNA compaction is achieved through a variety of interactions with histone proteins. There are five histones. Four of these histones, H2A, H2B, H3 and H4 combine to form a core octamer (two H2A/H2B dimers and one H3/H4 tetramer) consisting of two of each protein (Luger, 1997). Each core will coil 146 base pairs (bp) of DNA around itself 1.65 times in a left-handed super-helix. Short N- and C-terminal tails of varying lengths protrude from the histones (figure 1.1b) and the positive charges associated with residues such as lysine and arginine enable these tails to bind and hold the DNA. This configuration is known as the nucleosome, which is further stabilized by the binding of H1 (Zlatanova, 1998) to the outer surface (figure 1.1a). There are eight species of H1 (or H5) and exactly which one is found where depends on factors such as stage in the cell cycle, type of cell etc. A short piece of linker DNA, of approximately 50 bp, joins one nucleosome to the next giving the appearance of 'beads on a string' (figure 1.2a). Each piece of linker DNA is bound by one molecule of H1. This structure is known as euchromatin and the loose binding of the DNA means that transcription of genes is possible at this level.

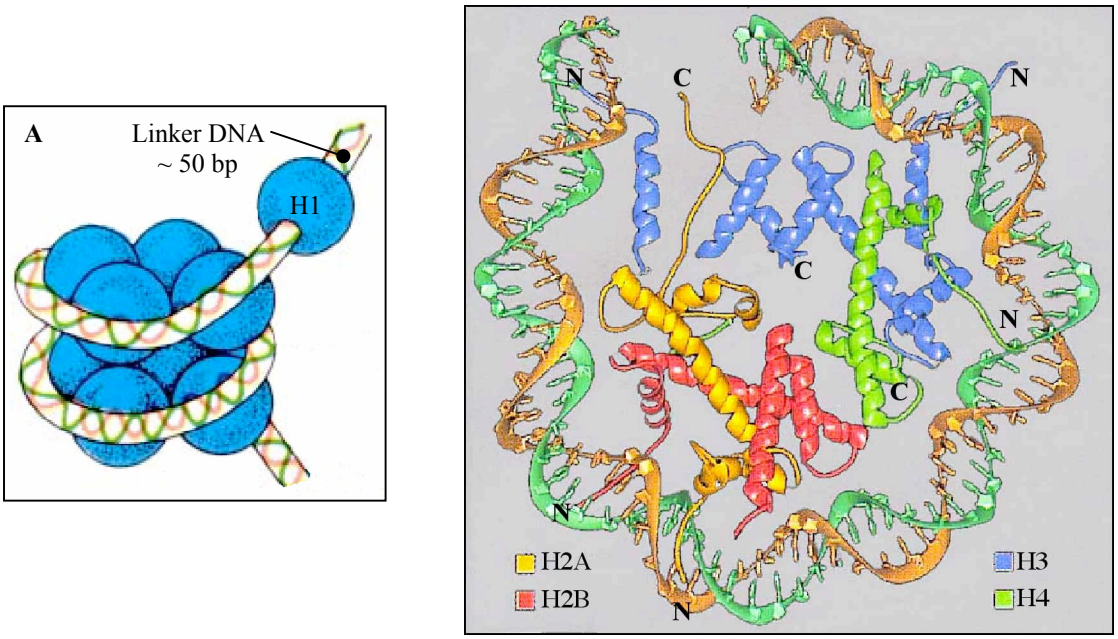


Figure 1.1: The Nucleosome.

A) Cartoon version of the nucleosome showing DNA wound around the core octamer of histone proteins. H1 is associated with the linker DNA between nucleosomes, serving to stabilise them. **B)** Protein Data Bank (PDB) structure of the nucleosome looking down through the superhelical axis (Luger, 1997) shows the positions of histone N and C terminal tails relative to the DNA. H2A (orange) and H2B (red) form two heterodimers, whilst H3 (blue) and H4 (green) form a tetramer.

After this initial stage of packaging, the chromatin is further packaged in a second stage (Figure 1.2b). This involves condensing of the nucleosomes into 30 nm thick fibres (called the solenoid) by the action of charged histone N-termini from one nucleosome binding to the DNA of a neighbouring nucleosome. Further recruitment of H1 aids stability of the solenoid structure. At this point, gene transcription is completely unfeasible and this silent form of chromatin is known as heterochromatin, which is easily distinguished from the euchromatin in a cell because it stains much darker with Giemsa (also known as G-banding), Feulgen and DAPI/Hoechst dyes. Heterochromatin only exists in eukaryotes, having probably evolved with the development of the nucleus in order to cope with the smaller volume to which it is confined. Finally, this solenoid structure is joined by a number of other proteins and compacted into the metaphase chromosome (figure 1.2c).

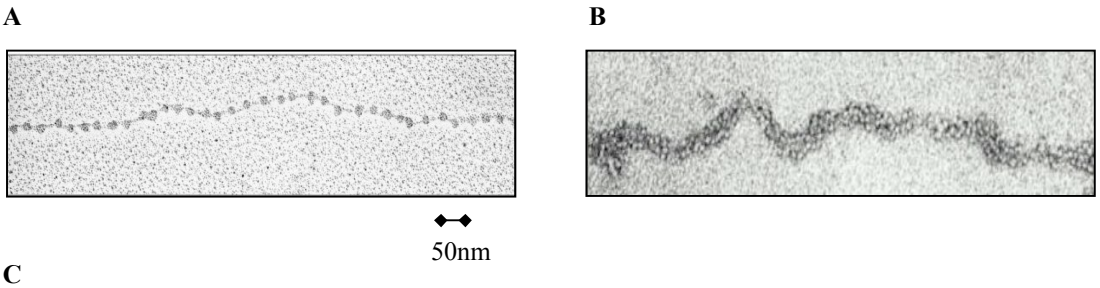


Figure 1.2: Electron micrographs of chromatin at different stages of compaction. (A) Chromatin at stage 1 of packaging to give 'beads on a string conformation'. (Alberts, B., 2002).
(B) Chromatin at stage 2 of packaging: 30 nm thick chromatin fibres. Picture by Barbara Hamkalo (Alberts, B., 2002).
(C) Metaphase chromosome (Bloom, W., 1994).

Chromatin has three main functions: (i) DNA compaction, allowing the vast amount of DNA to be systematically arranged within the nucleus, (ii) DNA protection against damaging agents and nucleases, and (iii) regulating the accessibility of the genome to transcription, replication and repair machinery. These processes are highly dependent on the precise organization of the chromatin fibres. Histone proteins therefore play an important role as control elements which are effectively switched on and off by modifications targeted at the internal globular structure or, more commonly, the protein N-termini. Lysine rich tails protruding from the nucleosome core are accessible to reversible covalent modifications (e.g. methylation, phosphorylation, ubiquitination and acetylation) resulting in purposeful remodelling of chromatin for the manipulation of processes such as transcription, gene silencing, cell proliferation, senescence and apoptosis (Cheung, 2000; Han, 2006). These modifications constitute what is hypothesised as the 'histone code' and in addition to other factors, including DNA methylation, they contribute to the epigenetic code that determines the differentiation pattern of a particular cell. Figure 1.3 shows the known post-translational modifications that have been commonly found at the N-terminal regions of the different histone proteins. H1/H5 is not shown

here because until recently, very little was known about its modification status (except for phosphorylation). Although H1 is an important determinant in DNA condensation, it was thought not to be regulated by acetylation. However, new research has revealed that H1 variants also incur a range of modifications similar to histones H2A, H2B, H3 and H4 (Wisniewski, 2006).

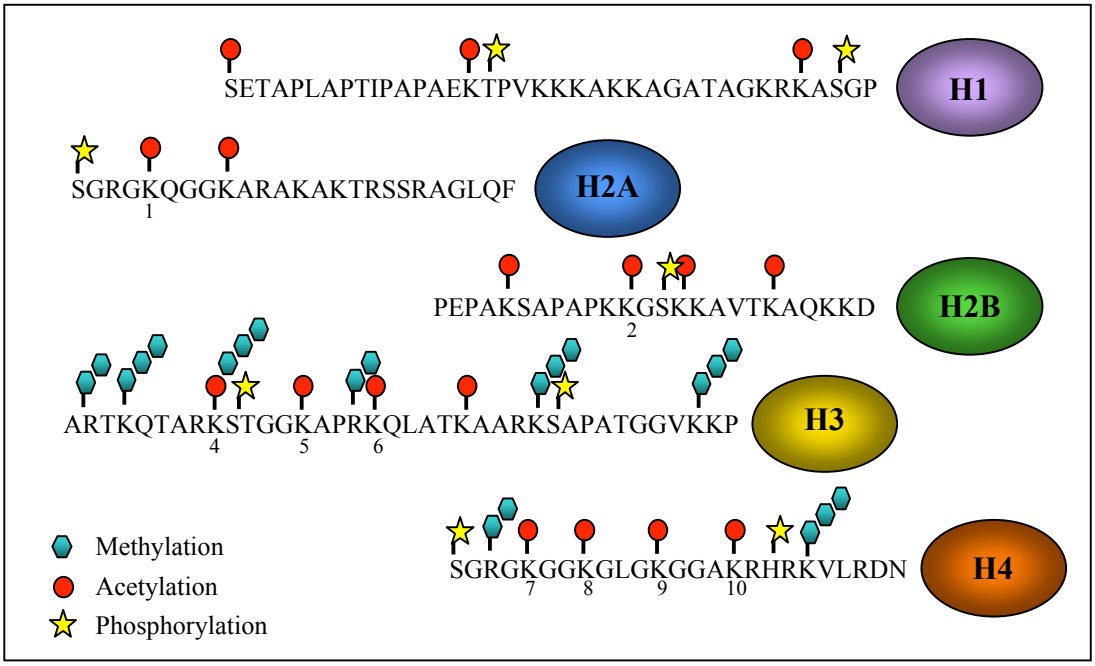


Figure 1.3: Common N-terminal Histone modifications.

Common modifications on the N-terminal tails of human histones H1 (Wisniewski, 2007), H2A, H2B, H3 and H4 (Croce 2005). *In Vitro* experiments have demonstrated that several known HAT proteins will preferentially target selected N-terminal lysines on histones H2A, H2B, H3 and H4 for acetylation.
GCN-5, PCAF, Hpa2, $TAF_{II}230$ ($TAF_{II}230$ is the *Drosophila* homolog of human $TAF_{II}250$ (Sterner & Berger, 2000), SRC-1 and p300/CBP have all been demonstrated to acetylate H3 at position denoted '5'. SRC-1 has been shown to target position denoted '4', p300/CBP targets multiple sites denoted '1', '2', '3', '5', '6', '7' and '8' , Esa1 targets position denoted '7', Hat1 acetylates at position denoted '9' and MOF acetylates at position denoted '10'. Gene identifications are as follows: H1 >gi|4885377|NP_005311.1| H1 histone family, member 3 [Homo sapiens], H2A >gi|121986|sp|P04908|H2AM_HUMAN Histone H2A.m, H2B >gi|20336754|NP_066402.2| H2B histone family, member R [Homo sapiens], H3 >gi|45768281|gb| AAH 67493.1| H3 histone family, member F [Homo sapiens] and H4 >gi|55930957|gb|AAH50615.1|H4 histone family, member J [Homo sapiens].

Histone methylation is typically associated with transcriptional repression (Curradi, 2002) but in certain cases i.e. methylation of lysine 4 on histone H3 and/or specific arginine residues of H3 and H4, it has been shown to activate transcription (Bernstein, 2002; Wang, 2001).

Phosphorylation of histone tails neutralises the basic charge of the histone and therefore reduces its affinity for DNA binding, again leading to transcriptional activation. Phosphorylation of serine 10 on histone H3 and other substrates has been shown to increase acetylation of these substrates by Gcn-5- acetyltransferase enzymes (Lo, 2000).

Ubiquitination of histones represents the most bulky of modifications considered so far. The function of this kind of histone modification is not well understood but mono-ubiquitinated substrates can initiate cell signalling by allowing other proteins with ubiquitin binding domains to interact with them. Many substrates targeted for ubiquitination have a chain of ubiquitin molecules added onto a lysine residue and this usually leads to degradation of the poly-ubiquitinated protein by the 26S proteasome (Fu, 1998).

Acetylation is among the most studied of the histone modifications. It appears that acetylation may be responsible for a vast amount of subtle signalling processes within the cell, however, no major immediate phenotypic defect can be seen in yeast by altering the levels of the enzymes responsible for N-terminal acetylation. Defects noted thus far include failure to enter G0 phase of growth and sensitivity to temperature fluctuation (Polevoda & Norbeck, 1999). In the 1960's, Allfrey and colleagues first demonstrated that acetylation of core histone tails affected the level of eukaryotic gene expression (Allfrey, 1964). Thus, the link to transcriptional activation was discovered and since then, many examples of increased or decreased

transcription of certain genes (or groups of genes) have been recorded in relation to this modification (Eberharter, 2002).

1.2 Acetylation

Acetylation is a process that involves the transfer of an acetyl radical derived from a molecule such as acetic anhydride (in the case of chemical acetylation) or an acetyl group from a cofactor such as acetyl-CoA (in the case of biological acetylation) to another molecule. This type of chemical reaction is known as a nucleophilic acyl substitution and usually involves two steps: the nucleophilic addition of a new group to form a tetrahedral intermediate and then the elimination of a leaving group. Acetylation is assumed to be an S_N1 reaction but there may be some exceptions depending on the exact enzymes/substrates in question. The process is catalyzed by a variety of enzymes, which have the ability to acetylate N-terminal α-amino groups (non-reversible) on proteins or molecules and the ε-amino groups (reversible) of specific lysine residues found on the N-terminal regions of the histone components of chromatin (figure 1.4). Acetylation of these histone tails, or indeed, other DNA binding proteins, neutralises the positive charge that they carry and is therefore thought to weaken interactions with the DNA. This gives rise to a relaxed chromatin structural conformation (Hansen, 1998; Walia, 1998) and thus allows RNA polymerase and other transcriptional machinery to gain access to the promoters belonging to genes located at that site (Nightingale, 1998; Tse and Sera *et al*, 1998). In contrast, genes that are located in a region of DNA that is wound tightly by histone molecules, due to little or no acetylation, show transcriptional repression (Marmorstein, 2001).

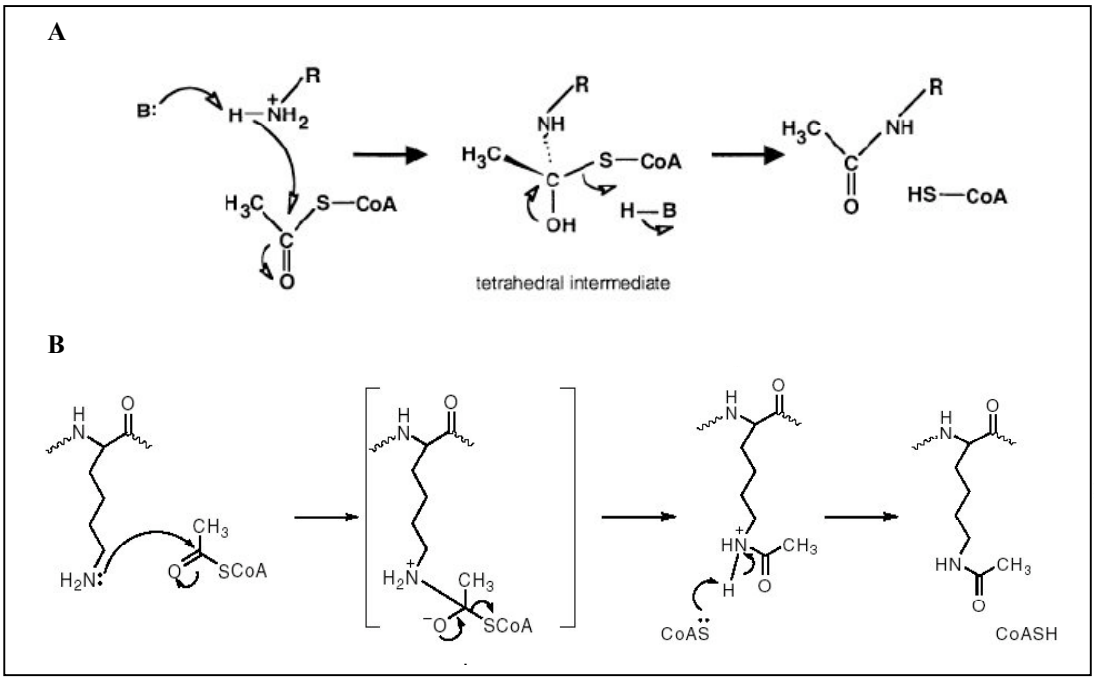


Figure 1.4: Mechanism of Biological Acetylation. (A) A possible mechanism for acetylation of protein Nα-terminal amino groups via direct nucleophilic substitution/attack on acetyl-CoA (Dyda, 2000). **(B)** Direct acetyl transfer from cofactor to ε-NH_2 of lysine side chain (adapted from Walsh, 2007).

Histone acetyltransferases (HAT's) are involved not only in transcriptional activation but also DNA repair, gene silencing and cell cycle progression (Carrozza, 2003). Therefore, a good understanding of the structure and function of N-acetyltransferase enzymes will improve our understanding of the regulation of gene expression and provide insight into the diseases in which anomalous HAT activity has been implicated as a possible cause (Marmorstein, 2001).

1.2.1 The enzymes responsible

There are thousands of enzymes that have acetyltransferase activity, very few of which have been structurally characterised to date. These enzymes fall into a

number of different families, many of which belong to the large superfamily of GNAT's (Gcn5-related N-Acetyltransferases) of which it has been noted that approximately 10,000 identified members exist (Vetting, 2004) across all the kingdoms of life. Other families include the MYST superfamily, the CBP/p300 family and transcriptional co activators. Each of these families includes enzymes with diverse functions and affinity for certain substrate types (Angus-Hill, 1999), e.g. histones, aminoglycoside antibiotics, arylalkylamines (Dyda, 2000) as well as puromycin and diamines. Histone acetyltransferase (HAT) enzymes, in particular, are abundant throughout the GNAT/MYST superfamilies and others including global co-activators such as p300/CBP and nuclear receptor co-activators (Table 1.1).

Acetylated proteins	**Residues**	**Acetyltransferases**
Majority of eukaryotic proteins.	α-Ser, α-Ala, etc.	NatA, NatB, and NatC
E. coli ribosomal proteins (S18, S5, and L12).	α -Ser, α-Ala	RimI, RimJ, and RimL
Regulatory peptides (β-endorphin, α-MSH, enkephalin, GHRF).	α-Tyr, α-Ser, α-Ala	Unknown
Histones (H2A, H2B, H3, H4).	ε-Lys	GNAT 's: Gcn5, PCAF,Hat1, Elp3 & Hpa2 MYST's: Esa1, MOF, Sas2, Sas3, Tip60 & MORF p300/CBP group Transcription factor's: $TAF_{II}250$ & $TF_{III}C$ Nuclear receptors cofactors: ACTR & SRC-1
Transcription factors: (p53, E2F1-3, EKLF, TFIIEβ, TFIIF, c-Jun, TCF, GATA1, MyoD, HMGI(Y), pRb, NF-E2(MafG) and ACTR).	ε -Lys	PCAF/GCN5, p300/CBP, $TAF_{II}250$, SRC-1?, MOZ, Tip60? & BRCA2?

HMG proteins (HMG1 and HMG2).	ε -Lys2 & ε-Lys11	p300/CBP & PCAF
Nuclear receptor HNF-4.	ε-Lys	p300/CBP
Nuclear import factors (importin-α7 and Rch1).	ε-Lys22	p300/CBP
α-tubulin.	ε-Lys40	62-67 kDa protein

Table 1.1: Acetylated proteins and the corresponding acetyltransferase enzymes that act on them. Abbreviations not mentioned in the text: BRCA2, breast cancer protein; Elp3, elongator protein, a subunit of the RNA polymerase II holoenzyme complex; Esa1, essential SAS2-related acetyltransferase; Gcn5, general control nonrepressible protein, a nucleosomal histone acetyltransferase; GHRF, growth-hormone-releasing factor; GNAT, Gcn5p-related amino-acetyltransferase superfamily; Hpa2, histone and other protein acetyltransferase; MOF, males absent on the first, an X-linked dosage-compensation protein in Drosophila; MORF, monocytic leukaemia zinc-finger protein related factor; MOZ, monocytic leukaemia zinc-finger protein; MYST group, named for the founding members MOZ, YBF2/SAS3 and Tip60; p53, a tumour-suppressor protein; pRb, retinoblastoma protein; Rch1, Rag1 cohort, human importin-α; Sas2, something about silencing protein, involved in silencing at telomeres and mating-type loci; SRC1, steroid nuclear receptor co-activator; Tip60, HIV Tat-interactive protein. A question mark indicates uncertainty. (Table adapted from Polevoda & Sherman, 2002, Diversity of acetylated proteins).

1.2.1.1 The GNAT's

The GNAT superfamily consists of a few distinct subfamilies (figure 1.5), designated GCN5 (general control non-repressed), PCAF, AANAT, AAC, ELP3, HAT1 and HPA2 (Pandey, 2002) among others (figure 1.5a). GNAT enzymes are characterised by their N-acetyltransferase domain, which consists of four motifs A-D. These motifs are ordered C, D, A and B from N- to C-terminus of the proteins (figure 1.5b). These motifs were first discovered by Tercero *et al* (1992) and further characterised by Neuwald & Landsman (1997). Motifs A and B are common to the general histone acetyltransferase population, motif A being the most highly conserved. It contains a consensus sequence (R/Q)XXGX(G/A) (where X is any amino acid), which is partly responsible for the recognition and binding of acetyl-CoA (Dutnall, 1998b; Wolf, 1998). Motif C is found only in the GNATs. It is not highly conserved at the sequence level but structural homology is well preserved

throughout subsets of the superfamily. The hypothesised function of motif C is substrate recognition/binding (Trievel, 1999). The typical secondary structure of a GNAT member will contain four α-helices and six to eight β-sheets (figure 1.5c).

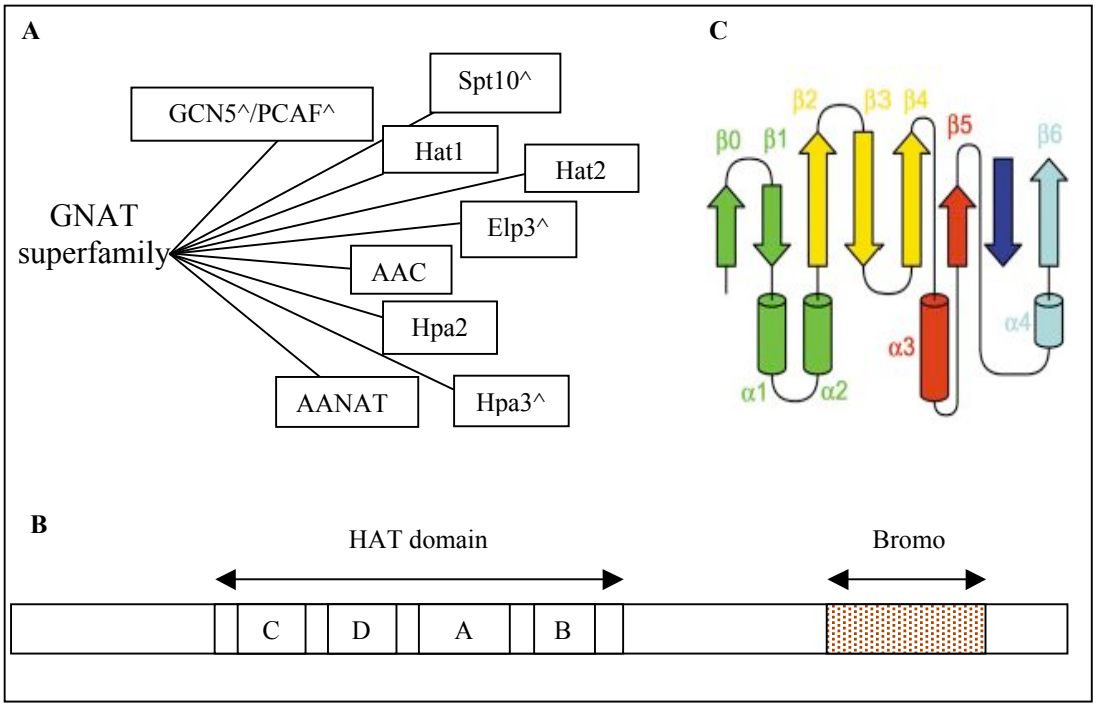


Figure 1.5: The GNAT family. (A) GNAT superfamily of enzymes showing a few well-characterised members. ^ denotes enzymes with additional function/s related to transcription. **(B)** Domains and motifs within the general GNAT protein. **(C)** Topology of the GNAT fold. Yellow and red regions highlight the conserved CoA binding domain, whilst green and blue show the less conserved N- and C-termini (Vetting, 2004).

GCN5 (also known as ADA4) is one of the founding members of the GNAT superfamily and is a type A HAT, i.e. localised to the nucleus (Mizzen, 1998). GCN5 contains a high number of charged residues; the N-terminal region is quite acidic and its mid section is more basic (Georgakopoulos, 1992). It includes the signature HAT domain of the GNAT family (figure 1.5b) and has many closely related homologues including *T. thermophila* P55 (the catalytic subunit of a previously identified HAT). Brownell and colleagues were the first to prove that GCN5 displayed HAT activity toward free histones *in vitro* (Brownell & Zhou *et al*, 1996), showing a marked

preference for Lys 14 of histone H3 and residues Lys 8 and Lys 16 in histone H4 (Kuo, 1996). Recombinant GCN5 cannot acetylate chromatin associated histones but it exists *in vivo* as a subunit belonging to the multimeric complexes ADA and SAGA, and as part of these complexes, GCN5 is able to acetylate nucleosomal histones (Grant, 1997). GCN5 activity causes remodelling of promoter DNA (Gregory, 1998) and assists in transcriptional initiation/activation through interaction with a variety of transcriptional regulators e.g. the GCN4 homodimer and the CYCI (encoding iso-1-cytochrome c) UAS2 element, as well as transcriptional activation through the heterotrimer complex HAP2-HAP3-HAP4. GCN4 is needed to transcribe amino acid biosynthetic genes in yeast cells when the supply of amino acids is lacking in the environment (Georgakopoulos, 1992). The CYCI UAS2 element is involved in the transcription of genes responsible for regulating the oxidative phosphorylation process (Trueblood, 1998). Transcription of several genes involved in *S. cerevisiae* respiratory function is brought about by the activity of the HAP2-HAP3-HAP4 complex.

PCAF (p300/CBP associated factor) is a close homologue to GCN5 which has been found to associate with the transcriptional co activators p300/CBP *in vivo* (Yang, 1996). As a HAT and a co activator, it has been implicated in growth factor-signalled activation (Xu, 1998) and receptor-receptor mediated activation (Blanco, 1998). PCAF not only acetylates histones but also non-histone proteins such as the activators p53, HIV Tat, MyoD and general transcription factors (TFIIE and TFIIF) and chromatin associated proteins (HMG17 and HMGI) (Sterner & Berger, 2000). During myogenesis, MyoD recruits PCAF and p300/CBP to the promoters of myogenic genes, where they all combine to form a complex that initiates the transcription of these genes. The HAT activity of PCAF is then essential for differentiation of cells into specific muscle tissue types (Puri, 1997). PCAF and

p300/CBP also form complexes with other proteins e.g. the multimeric activation complex with a nuclear co-activator ACTR, which has also been discovered to be a HAT (Chen, 1997).

The AANAT subfamily comprises arylalkylamine N-acetyltransferase and its homologues. AANAT is a ~23 kDa globular cytosolic protein that is expressed highly in both the pineal gland and the retinal photoreceptor of the eye (Klein & Coon, 1997). It forms a reversible complex with 14-3-3 proteins and its main function is to regulate the levels of melatonin that are produced during the circadian rhythmic cycle in response to levels of daylight. Its role in the retina is less understood but it may be involved in local signalling through N-acetyserotonin or melatonin (Pozdeyev, 2006). AANAT preferentially acetylates arylethylamines such as serotonin, dopamine and tryptamine. Diamines and monoamines are also acetylated but to a far lesser degree (Klein 2007).

The AAC subfamily (aminoglycoside antibiotic N-acetyltransferases) is another important group belonging to the GNAT superfamily. Aminoglycosides are bactericidal antibiotics that target the A-site of the 16S ribosomal RNA (Moazed & Noller, 1987). An increase in the amount of antibiotic resistance among pathogenic bacteria such as Gram positive enterococci and staphylococci has resulted from N-acetyltransferase activity on certain amino groups of aminoglycosides, which results in reduced affinity of the antibiotic for its target (Dyda, 2000).

To briefly mention the other GNAT enzymes outlined in figure 1.5a and their substrates:-

Hat1 is a B type HAT (responsible for predominant cytoplasmic HAT activity in yeast) and is associated with a second subunit Hat2. It was one of the first yeast HATs to be discovered (Kleff, 1995 and Parthun, 1996) but its function is thought perhaps to be redundant since null mutation confers no phenotypic defect. Previous

studies have shown its involvement in telomeric silencing and DNA-double strand break repair (Duntall, 1998b, Ruiz-Garcia 1998, Parthun 1996, Kelly 2000, Kleff 1995 and Qin 2002).

Elp3 (elongator protein 3, also known as Hpa1 or Tot3) targets H3 and H4 (Wittschieben, 1999; Winkler, 2002). Elp3 is part of the RNA polymerase II holoenzyme and plays a role in transcriptional elongation. *S. cerevisiae elp3* null mutants are temperature sensitive and show poor growth adaptation but are still viable (Sterner & Berger, 2000).

Hpa2 is capable of forming a dimer or a tetramer *in vitro* and is only found in fungi. Histones H3 and H4 are both substrates of Hpa2 as well as Elp3. Hpa2 also exhibits autoacetylation activity (Angus-Hill, 1999).

Hpa3 is a close homologue of Hpa2 but although it too can autoacetylate, it displays very poor HAT activity *in vitro* (Sterner and Berger, 2000).

Spt10 is a yeast transcriptional regulator that shares homology to GNAT proteins but as yet, no acetyltransferase activity has been demonstrated for it (Natsoulis, 1994).

1.2.1.2 The MYST's

The founding members of the MYST family are Moz, Yfb2/Sas3, Sas2 and Tip60. The MYST superfamily (figure1.6a) contains a HAT region identified by its homology to motif A of the GNATs and an unusual C2HC zinc finger domain (figure 1.6b). The other identifying properties of this group are the C-terminal bromodomain and the occasional inclusion of one or more zinc finger motifs or chromo-like domains near the N-terminus (Sterner and Berger, 2000). Moz was first identified as a putative acetyltransferase, found fused to the CREB binding protein (CBP) (Borrow, 1996), and is involved in oncogenic transformation (see section

1.2.4). Sas2 and 3 (Something about silencing) are connected to transcriptional silencing in yeast. Sas2 is involved in telomeric silencing, capable of either suppressing it at the HMR loci or promoting it at the HML loci (Ehrenhofer-Murray, 1997). Sas3 has weaker HAT activity than Sas2 but is still active in the silencing of mating type loci (Reifsnyder, 1996). Tip60, the first MYST protein to be discovered, is a 60 kDa protein that interacts with the activation domain of Tat, the HIV-1 transactivator protein (Kamine, 1996). Tip60 cannot acetylate chromatin-associated proteins very well but it is able to acetylate free histone proteins H2A, H3 and H4 (Kimura, 1998; Yamamoto, 1997). Tip60 allows positive regulation of the Tat-repressed gene for Mn-dependent superoxide dismutase (Westendorp, 1995) but once bound, Tat can cause inhibition of Tip60 acetyltransferase activity (Creaven, 1999).

Esa1, an essential sas family acetyltransferase required for cell cycle progression (Smith, 1998), is also unable to act on nucleosomal histones but like Tip60, it is able to modify free histones *in vitro*. It has a high specificity for Lys 5 of histone H4. Yeast cells that are deficient in Esa1 are halted in growth between the DNA replication and mitosis stage of the cell cycle (Clarke, 1999). MORF, a close homologue of Moz (Moz related factor; Champagne, 1999), exists as MORFα and MORFβ, each having slightly different insertions in the N-terminal repressed region of the protein. This region contains two zinc fingers and deletion of it causes the HAT activity of the enzyme to be enhanced, leading to increased *in vivo* transcription. It acetylates nucleosomal histone H4 strongly and will also acetylate free histones H3 and H4, but unlike its relative Moz, it has not yet been linked to any kind of cancer. HBO1 is a HAT that interacts with a specific subunit (ORC1) of the origin recognition complex, ORC (Iizuka, 1999), which is conserved throughout the eukaryotic kingdom. The main function of ORC is the binding of DNA replication

origins in order to initiate replication (Dutta, 1997). The subunits of the origin recognition complex enhance the otherwise weak HAT activity of HBO1 and its preferred substrates are free histones H3 and H4.

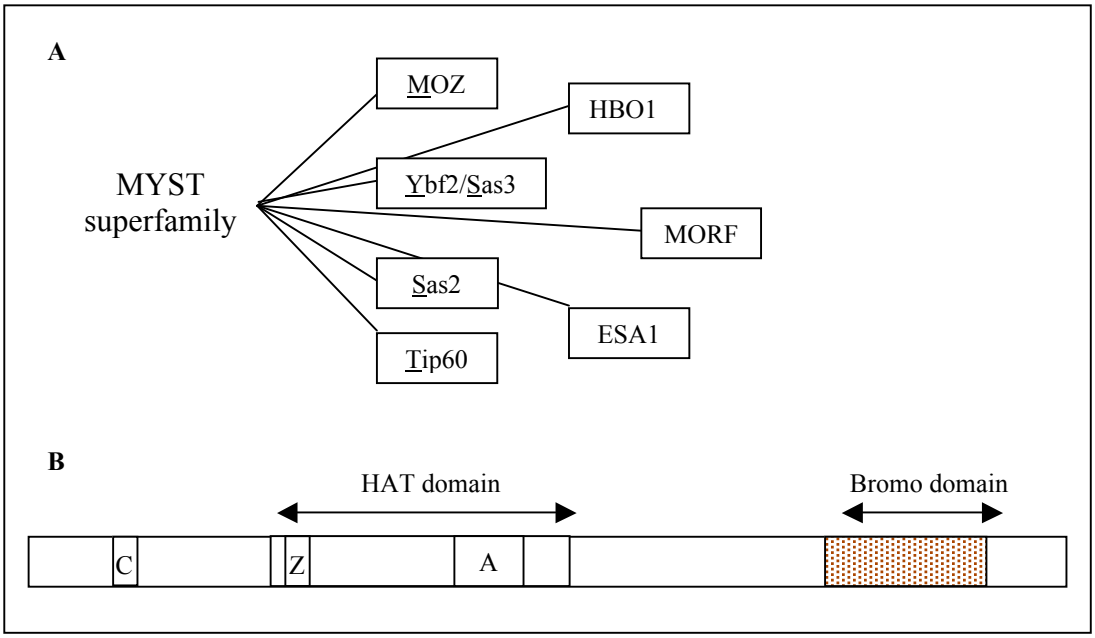


Figure 1.6: The MYST family. (A) The MYST family founding members and some others. **(B)** The general MYST domains. Note that one or more zinc fingers (Z) may be present depending on the actual enzyme. A chromo-like domain 'C' may exist near the N-terminus, again depending on the actual enzyme (Sterner & Berger, 2000).

1.2.1.3 p300/CBP

p300 and CBP are two very closely related proteins with a high level of sequence homology, both containing a transactivation domain at each terminus and an acetyltransferase domain in the middle (figure 1.7). They were first identified as belonging to a conserved family of transcriptional co-activators (Arany, 1994) and found to interact with many cellular proteins, either as modifying agents in cellular cascades or as subunits in a number of multi-subunit complexes. The adenovirus oncoprotein E1A targets p300 and CBP and the interaction between these proteins suggests involvement in DNA synthesis control and progression of yeast cells into

the S-phase of the cell cycle (Stein, 1990; Moran, 1993). These co-activators play a range of distinct roles in processes such as retinoic-acid-induced F9-cell differentiation (Kawasaki, 1998) and the cellular response to hypoxic conditions (Arany, 1996). Bannister and colleagues were among the first to notice that CBP and p300 are histone acetyltransferases (Bannister, 1996; Ogryzko, 1996) and they also demonstrated that p300/CBP acetylated a nuclear import factor called importin (Bannister, 2000).

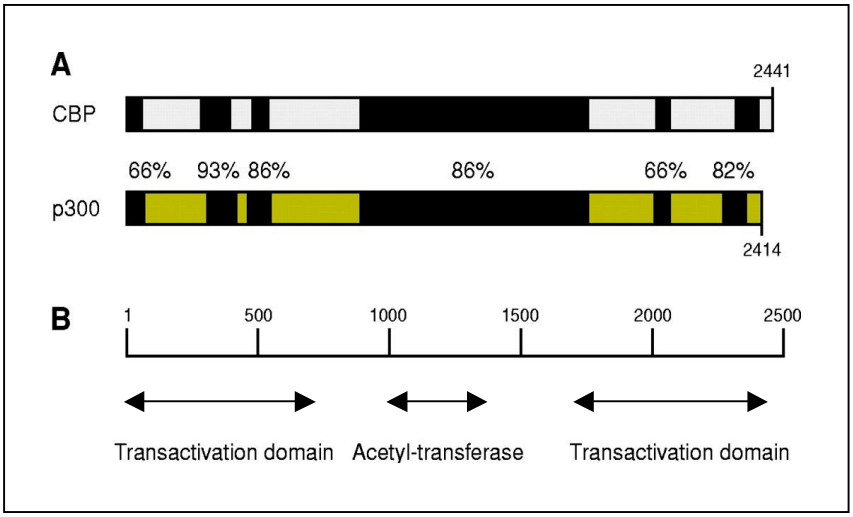


Figure 1.7: Organisation of p300/CBP proteins. (A) Comparison of p300 and CBP. The dark regions indicate the areas of highest homology, with the percentage amino acid identity between the two proteins indicated. The size of each protein, in number of amino acid residues, is indicated. **(B)** The functional domains in p300 are also indicated (adapted from Chan, 2001).

It was also observed that P300/CBP were recruited in p53 dependent pathways (Avantaggiati, 1997) and activation of p53 (Sakaguchi, 1998) has been demonstrated, *in vivo,* with several sites on this major protein becoming targets for acetylation by PCAF and p300 in response to DNA damage (Liu, 1999a). Among their many other roles, acetylation by p300/CBP is also responsible for the control of the retinoblastoma tumour suppressor protein (Chan, 2001). Therefore, p300/CBP are essential as transcriptional co-activators and HATs within the cell. They have been

described as 'partners in life and death' (Giordano, 1999) because they are major players implicated in so many pathways and no doubt, new functions will continue to come to light regarding this pair of ubiquitous enzymes.

1.2.2 Histone deacetyltransferase enzymes

Histone deacetyltransferases enzymes (HDACs) can generally be divided into three different groups (de Ruijter, 2003). Each group possess highly differing sequences, sharing no homology with each other. Class I members can be found localised to the nucleus and they share a considerable degree of homology to yeast RPD3 (reduced potassium dependency). Members of class I include HDACs 1, 2, 3 and 8. Class II affiliates are homologous to yeast Hda1 (histone deacetylase) and are approximately twice the molecular weight of their class I relatives. HDACs 4, 5, 6, 7, 9 and 10 fall into the class II and they can be found in both the nucleus and the cytoplasm of the cell. HDAC 11 shares in the properties of both class I and II enzymes and is sometimes referred to as class IV. A third class of HDACs consists of yeast Sir2 related proteins (sirtuins), Sirt1-Sirt7, and is based around their distinct NAD-dependent mode of action. Class III enzymes are also found in the nucleus and the cytoplasm. In the same way that some HATs have an additional role as co-activators, many HDACs also double up as transcriptional co-repressors and can be found as subunits belonging to large regulatory complexes. Histone deacetylation by HDACs is generally considered one of the main mechanisms controlling transcriptional repression in mammalian cells and one would guess that if HDACs were inhibited by the action of a specific compound that gene induction would rapidly increase. In fact, only about 2% of genes have been shown to be switched on by treatment with HDAC inhibitors (Van Lint, 1996).

1.2.2.1 Class I

HDAC1 interacts physically with the retinoblastoma tumor-suppressor protein (Rb) via the LXCXE sequence motif to cooperatively enhance its activity. Rb inhibits cell proliferation and progression from G1 to S phase of the cell cycle by repressing a subset of genes that are controlled by the E2F1 transcription factor. This complex plays a key role in the control of cell proliferation and differentiation (Magnaghi-Jaulin, 1998). In addition to this role, HDAC1 also interacts with metastasis-associated protein-2 in order to deacetylate the promiscuous p53 protein. HDAC1 and p53 are "antagonistic regulators of the cyclin-dependent kinase inhibitor gene, p21" (Lagger, 2003) and thus adapt it to affect cell growth and apoptosis. The p21 gene has been found to be regularly induced in differing cell types by the action of HDAC inhibitors (Marks, 2001). HDAC2 interacts with PELP1 (proline-, glutamic acid-, and leucine-rich protein 1), a co-activator of estrogen receptor (ER), which recruits p300/CREB-binding protein HAT to the target chromosome. HDAC2 acts to repress this activity (Choi, 2004) and also functions to a similar level in conjunction with numerous different factors. HDAC3, like HDAC1 can facilitate recruitment of the histone binding protein RbAp48 to Rb and leads to the gathering of a repressive complex to the E2F-regulating promoters (Nicolas, 2001). HDAC8 is induced by interleukin-2 and is important for the contractile ability of smooth muscle tissue by association with α-actin. It is co-expressed with α-actin and heavy chain myosin (Waltregny, 2005). HDAC8 regulates hEST1B protein stability and thus alters the enzymatic activity of telomerase in humans, resulting in a contribution to tumourigenesis (Lee, 2006).

1.2.2.2 Class II

HDAC4 is able to shuttle between the cytoplasm and nucleus, its subcellular localization being affected by phosphorylation at specific sites. It has been suggested that HDAC4 has a positive role in the differentiation of muscle tissue (Zhao, 2001). HDAC4 is also involved in cardiac cell signaling (leading to cardiac hypertrophy and/or arrest) as a downstream substrate of nuclear calcium/calmodulin-dependent protein kinase IIδ (Little, 2007). HDAC4 is also linked to caspase enzymes and therefore plays a part in the control of cell differentiation and death (Liu, 2004). Gβγ, a subunit of heterotrimeric G proteins, binds and inhibits the transcriptional co-repression activity of HDAC5 (Spiegelberg, 2005). HDAC5 can bind Ca^{2+}/calmodulin and this inhibits MEF2a (monocyte enhancer factor) binding, thus impacting the differentiation of cardiac and skeletal cells (Berger, 2003). Over-expression of HDAC5 has been found to inhibit tumour cell growth and lead to apoptosis of such cells (Huang, 2002). HDAC6 deacetylation of tubulin plays a key role in microtubule-dependent cell motility (Hubbert, 2002).

The activity of molecular chaperone heat shock protein 90 (Hsp90) "is regulated by reversible acetylation and controlled by the deacetylase HDAC6" (Kovacs, 2005). Hyperacetylation of Hsp90 by inhibition or deactivation of HDAC6 results in its dissociation from p23, an essential co-chaperone, and leads to a loss of chaperone activity. Glucocorticoid receptor (GR) maturation is dependent on Hsp90 and is compromised in HDAC6-deficient cells, giving rise to GR defective ligand binding, nuclear translocation, and transcriptional activation (Kovacs, 2005). HDAC7 is found mainly in the nucleus, as outlined for class II enzymes, but can also be found to a degree in the cytoplasm, indicating the likelihood of nucleocytoplasmic shuffling. It often associates with class I HDAC3 *in vivo* but cytoplasmic HDAC7 is inactive when not bound to HDAC3 (Fischle, 2001). Along with other class II

HDACs, HDAC7 contributes to the regulation of MEF-2 function in muscle and heart tissue (Dressel, 2001). HDAC9, also known as HDAC7B, is also a repressor of MEF-2 mediated transcription.

1.2.2.3 Class III

The Sirtuins (silent information regulator 2) family are composed of a core domain of 250 amino acids and share 25-60 % sequence identity (reviewed in Moazed, 2001). Sir2 is a unique deacetylase that removes acetyl groups from acetylated lysine residues on histones and other proteins in an NAD dependent mode. This makes it the first deacetylase of its kind since other members of the family catalyze an NAD-nicotinamide exchange reaction that requires the presence of acetylated lysine residues but they do not absolutely require NAD when it comes to the catalysis of histone deacetylation (Landry, 2000). Conservation of the Sir2 family throughout all domains of life, even in organisms that do not possess histones, indicates that it is an enzyme with a broad range of substrates and little specificity. In yeast and other eukaryotes, Sir2 has potential links to chromosome segregation, transcriptional silencing, cell ageing (Guarente, 2000), cell cycle progression and DNA repair. It interacts with the hypoacetylated H3 and H4 histone tails of the nucleosome and therefore hinders access of the DNA to replication, transcription and recombination machinery, hence it is known as a chromatin silencer.

1.2.3 Acetylation and disease

Abnormal levels of acetylation have been linked to several major diseases including acute myeloid leukaemia, Non-Hodgkins lymphoma, Epithelial cancers, Rheumatoid arthritis, Huntington's disease, Fragile-X syndrome and Rubenstein-

Taybi syndrome (see table 1.2 for specific acetylation defects in some of these diseases). However, the underlying causes of such aberrant hypo/hyperacetylation levels noticed in these diseases is not at all well understood. In most cases, it has been observed that HDAC levels are more prevalent and given that there are several known HDAC inhibitors already commercially available, this lends an ideal opportunity for potential therapeutic treatments tailored towards specific diseases. A quick overview of some of the above mentioned diseases follows in table 1.2.

Disease	Acetylation Defect	
Leukaemia	HAT fusions MOZ/CBP MOZ/P300 MOZ/TIF-2 MLL/CBP MLL/p300	HDAC mediated RAR/PML RAR/PLZF AML1 fusions
Epithelial cancers*	P300 mutations in colorectal, gastric, breast, pancreatic carcinomas/cancer cell lines	
RTS (Rubenstein-Taybi syndrome)	CBP haploinsufficiency due to mono-allelic CBP mutations	
Polysomy X	Incomplete acetylation of all X chromosomes	
Fragile X syndrome	FRM-1 mutations linked to deacetylation and hypermethylation of X chromosome	

Table 1.2: Acetylation defects in various human diseases. (Adapted from Timmermann, 2001). *Epithelial cancers are too numerous and diverse to cover in the above body of text. MOZ – monocytic leukaemia zinc-finger protein, CBP – CREB binding protein, p300 - ??, TIF2 – transcription factor, MLL – mixed lineage leukaemia, HDAC – histone deacetyltransferase, RAR – retinoic acid receptor, promonocytic leukaemia, PLZF – promyelocytic leukaemia zinc-finger protein, AML – Acute myeloid leukaemia, FRM-1 – fragile X gene.

Acute myeloid leukaemia (AML) is a disease affecting the myeloid/blast white blood cells, where they do not mature but continually divide and cause congestion in the bone marrow so that the production of red blood cells and platelets is also hindered. This can cause anaemia, bruising, repeated infections and aching joints. There are several classifications of AML, for which different chromosomal

translocations have been pinpointed. This results in an array of mutant fusion proteins that over-recruit co-repressor HDAC complexes. Inhibition of these HDAC complexes by the HDAC inhibitor trichostatin A can relieve transcriptional repression, arrest disease progression and induce remission (Collins, 1998; Zwiebel, 2000). **Non-Hodgkin's Lymphoma** (NHL) is a cancer of the immune system that primarily causes swelling of the lymph nodes. Again, there are many types of NHL but the largest subgroup is diffuse large B-cell lymphoma (DLBCL). In 35 % of DLBCL cases, a 95 kDa nuclear phosphoprotein protein and transcription factor, Bcl-6, is expressed constitutively. This protein recruits HDAC complexes and thus causes repression of genes important for the control of lymphocyte activation and differentiation. Acetylation down regulates the activity of Bcl-6 by hindering its interaction with HDACs (Pasqualucci, 2003).

Huntington's disease (HD) is an autosomal (dominant) inherited neurodegenerative disorder arising from an expansion of CAG trinucleotide (polyglutamine) repeats in the huntingtin protein (Taylor, 2002). It is thought that a transcriptional dysregulation is responsible for the disease. Mutant proteins with expanded polyglutamine are able to bind to and restrain CBP and cause cellular toxicity (Jiang, 2006). This obviously leads to problems with CBP co-factor dependent transcription and also results in diminished HAT activity within the cell (McCampbell, 2000). **Fragile X-syndrome** is a mental retardation disease occurring on a sliding scale from mild to severe depending on the number of CGG trinucleotide repeats found at the 5' end of *FMR1* (fragile X mental retardation). The 5' end of normal *FMR1* associates with acetylated histones H3 and H4 (Coffee, 1999), allowing transcription of the FMR1 RNA-binding protein, but in fragile X patients (>200 CGG repeats) the chromatin structure at the 5' end of *FMR1* is altered (Eberhart, 1996). In addition to

methylation, this repetitive area of the gene attracts HDAC complexes which reduce the acetylation levels of the histones and result in transcriptional silencing.

Rheumatoid arthritis shares many similarities with cancer e.g. hyper-proliferation of synovial tissue surrounding the joints and a dependency on angiogenesis. Nuclear acetylation levels are enhanced in rheumatoid synovial cells causing over activation of several transcription factors (Nakawaza, 2002). N-acetyltransferase polymorphisms (of NAT2) have been implicated in the disease. People with slow NAT2 acetylation genotypes are up to 4.39 times more likely to develop Rheumatoid arthritis (Pawlik, 2004; Astbury, 1995). High concentrations of tumour necrosis factor alpha (TNF-α) in synovial cells cause them to undergo abnormal rates of mitosis and become enlarged (hyperplasia). TNFα recruits (and therefore depletes) CBP to its promoter to activate further transcription of the gene (Cvoro, 2006). P53 acts to suppress local inflammation (Yamashini, 2002) but depleted CBP levels mean the p53 promoter cannot be activated, resulting in decreased transcriptional activation and little/no p53 induced apoptosis of hyperplastic cells. It has been reported that TNFα also causes increased acetylation of p53 (Nakazawa, 2002), inducing accumulation of an inactive form of the protein (Drané, 2002). It should be noted that p53 acetylation can also activate the protein e.g. by the HAT activity of CBP , depending on the site of the modification. Needless to say, this disease involves some extremely complex pathways, many of which are still poorly understood but it is evident that acetylation plays a key role.

Rubinstein-Taybi syndrome is an inherited mental retardation syndrome characterised by congenital malformations such as beaked nose, broad thumbs/big toes and cardiac irregularities (Petrij, 1995). It is an autosomal dominant condition arising from mutation of CBP on chromosome band 16p13.3. Murata *et al*

demonstrated that several mutations within the CBP gene (found in patients with the condition) had the ability to abolish the HAT activity of the enzyme (Murata, 2001). **Polysomy X**, as the name suggests, is the inheritance of an abnormal number of X chromosomes, which can be from either parent. The condition displays similar physical and mental symptoms as Down's syndrome and the greater the number of extra X chromosomes an individual has, the greater the severity of the disease. The extra X chromosomes are not activated and show incomplete H4 acetylation (Leal, 1998).

1.3 Archaeal Histones

Prokaryotes have less DNA to compact and do not need the same tight controls on gene regulation. Bacterial DNA is compacted by small basic DNA binding proteins e.g. Fis and Hu; histone-like nucleoid associated proteins. The activity of these proteins does not seem to be regulated by posttranslational modification (reviewed in White & Bell, 2002). Many species of archaea from the subdomain Euryarchaeota contain histone homologues. These archaeal histones have the same signature histone fold as eukaryotic histones but are shorter at both N and C-terminal ends, lacking the tail domains where targeted modification takes place in the eukaryal histones (White & Bell, 2002). Nevertheless, they do have the ability to form a protein core for binding and wrapping DNA in a similar way. It is possible that gene regulation may be aided by archaeal histone dimer formations, if for example the histone core contained various histone compositions with different DNA binding properties (reviewed in Reeve, 2003). Transcriptional regulation in the archaea may provide clues as to the nature of the last common ancestors with eukaryotes and bacteria (Whitman, 1999). None of the Crenarchaeota sequenced so

far have histones. Since *Sulfolobus solfataricus* belongs to the Crenarchaeota, it does not contain histone proteins but its genome is known to encode two extremely abundant DNA-binding proteins: Sul7 and Alba1 (White & Bell, 2002).

The evolutionary tree of life shows that archaea diverged from Eukaryotes at a later stage than they diverged from the bacteria. This suggests that archaea are more closely related to eukaryotes in their information processing pathways and in particular, their core machinery involved in transcription, DNA replication and repair (Bond *et al*, 2001; White & Bell, 2002). *Sulfolobus solfataricus* is an extreme thermophile that lives and metabolises in temperatures between 60-89°C, pH range 2-5 (Zillig, 1980). Its proteins are easy to study because they are often straightforward to express and are usually very stable due to the high temperature at which they normally function. In addition to this, they can be more easily isolated from other proteins in an *E.coli* expression system when the temperature factor is exploited since all proteins from the mesophile *E.coli* are non-heat stable and will form aggregates upon heating.

1.4 Sul7 enzymes

These proteins were first identified in the 1980's by Reinhardt and colleagues (Kimura, 1984; Grote, 1986; Choli, 1998) and found only in *Sulfolobus* species, resulting in the generic family name Sul7 (White, 2002). So far, Sso7d, Sac7d and Ssh7d have been identified and characterized (from *S. solfataricus*, *S. acidocaldarius* and *S. shibatae* respectively). These sequence independent DNA binding proteins are small, around 66 residues, and basic with a molecular weight of approximately 7 kDa and have almost identical sequences. They constitute about 5 % of total soluble protein in *Sulfolobus* cells (Baumann, 1994). The function of these proteins has not

yet been confirmed but they are believed to be important in forming higher ordered structures for DNA compaction and stabilization of the duplex at the high temperatures at which *Sulfolobus* function, since Tm increases by as much as 33 °C for solutions with an excess of Sac7d in thermal denaturation studies of both DNA and DNA-protein complexes (Edmondson, 2001). The Sul7 family have been reported to induce negative supercoiling of DNA of any topology, whether relaxed, positively or negatively supercoiled (López-Garcia, 1998). These proteins have been found to have cation dependent ATP-ase activity, allowing them to function as ATP-dependent protein chaperones that disassemble and renature protein aggregates (Guagliardi, 2000). Surprisingly, Sso7d also displays ribonuclease activity, whereas Sac7d has none according to experimental observation (Fusi, 1993). However, the likely scenario is that small molecular weight thermostable RNase proteins were co-purified with Sso7d and not detected, thus giving rise to a false conclusion.

Sequence independent DNA binding proteins and their interactions with duplex DNA were previously poorly understood because capture of the protein–DNA complex in crystalline form was not easy and few examples existed. In 1998, Gao and colleagues managed to crystallise the Sso7d bound to DNA. From this, it was devised that Sso7d binds the minor groove of duplex DNA without sequence specificity and intercalates two hydrophobic side chains at positions 29 (Met) and 26 (Val). This results in an approximate 61° single step sharp kink being introduced into the DNA, the largest of its kind known to date (Gao, 1998). Studies carried out on Sac7d by small angle X-ray scattering showed that although binding induced a large kink in the DNA, sequential binding of the protein to long lengths of DNA at contiguous 4 bp sites leads to little change in the topology of the DNA because the sequential bends are compensatory (Krueger, 1999). From NMR analysis, proteins belonging to the Sul7 family have an OB fold topology

consisting of an incomplete five stranded β-band capped by an amphipathic α-helix bordering the β3-β4-β5 strands (Murzin 1993; Gao 1998). High-resolution crystal structures of two different Sso7d-DNA complexes were solved and it was found that each Sso7d monomer bound indiscriminately to a site of DNA spanning just 4 bp in length (McAfee 1996, Gao, 1998; Reeve 2003). This binding was associated with significant unwinding of the helix and bending of the DNA resulting in an A-DNA type structure at the binding site (Edmondson, 2001; Gao, 1998). Although binding is non-cooperative, the distortions induced in the DNA require cooperative interactions between adjacent protein monomers (Edmondson, 2001). Sso7d contains a small hydrophobic core of 11 residues. Four glycine residues at positions 10, 27, 38 and 39 are found in the loop regions and there is a greater than 45% solvent accessibility to hydrophobic regions. Four hydrophobic amino acids located on the β3-β4-β5 sheet of the molecule appear to be sites of contact between protein and DNA (Trp 24, which is conserved in all Sul7 proteins (Baumann, 1995), Val 26, Met 29 and Ala 45). The fact that Sul7 members are able to bind DNA in a sequence independent matter may be due to the bridging water molecules that occupy the cavity area between protein and DNA, thus inhibiting any steric clash between G-C base pairs and N2 amino groups.

1.5 Alba

This small DNA binding protein was previously known as Sso10b when it was first identified (Grote *et al*, 1986) but has been renamed Alba for Acetylation lowers binding affinity (Bell *et al*, 2002). Alba is a dimer, each subunit M_r of 10 kDa, that makes up approximately 4-5 % of total soluble protein in the cell (Xue, 2000). *S. solfataricus* has two copies of the Alba gene (Alba1 and Alba2), the second of which

is expressed at a 10-20 fold lower concentration in stationary cells and has only 30-40 % identity with Alba1 (Jelinska, 2005). Hexagonal and tetragonal crystal structures of Alba1 have been isolated and solved by collaborating groups at St. Andrews University and Cambridge using MAD phasing techniques to a resolution of 2.6 Å (Wardleworth, 2002). The crystal structure of the Alba1 monomer showed that it contained two α-helices (A and B) and four β strands (a-d). Alba proteins are unrelated to the Sul7 family and homologues are present in many other thermophilic and hyperthermophilic Archaea (Xue, 2000). In addition, Alba orthologues have been found in some eukarya (Aravind, 2003). Structural studies on Alba1 have demonstrated that it shares a common fold with DNaseI and the C-terminal end of translation initiation factor 3 (IF_3), giving rise to the conclusion that its origins arose from an ancient nucleic acid-binding domain (Aravind, 2003). The Alba1 homodimer contains a central body with lysine residues at positions 16 and 17, flanked by two extended β-hairpins, which suggest three main points of contact with the DNA (Wardleworth, 2002).

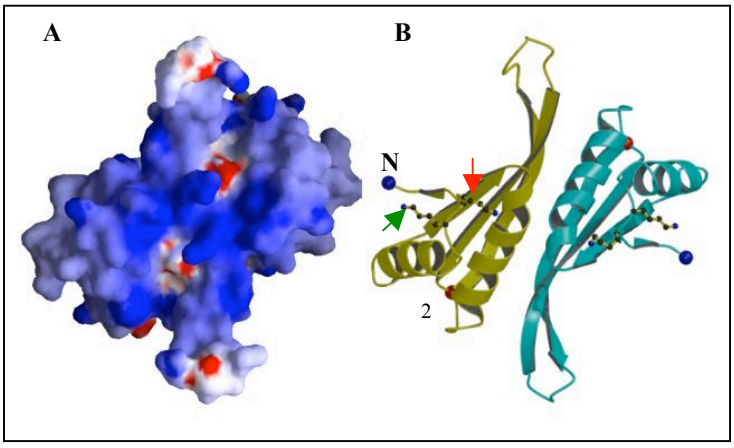


Figure 1.8: Structure of *Sso*Alba1. (A) Alba1 molecule coloured by electrostatic potential. The groove formed between the two loops containing Lys16 and Lys17 is apparent. **(B)** The Alba dimer with N and C terminal ends coloured blue and red. Lys16 (red arrow) and 17 (green arrow) are depicted by ball and stick notation (Wardleworth, 2002).

Native Alba1 subunits have a mass of 10, 538 Da, 84 mass units larger (after removal of N-terminal methionine in both cases) than recombinant Alba1 expressed in *E.coli*. This 84 mass unit difference is due to the addition of two acetyl groups; one at the N-terminal serine and the other site is the ε-amino group of Lysine 16 or 17. Mass spectrometry studies of Alba1 tryptic digests imply that the second acetylation site is Lys 16 (Bell, 2002).

The hairpin loop of monomer A is ordered and responsible for making contact with other symmetry related monomers in the crystal form. In contrast, the hairpin loop of monomer B is disordered and does not interact with other monomers. The dimer has a buried surface area of 690 $Å^2$ (the interface formed between strands c/d of monomer A and helix B of monomer B). There is an additional monomer association between helix A and the helix B C-terminal end, with a buried surface area of 460 $Å^2$. The significance of this association has yet to be understood (Wardleworth, 2002). Alba1 is highly conserved throughout both archaeal domains (figure 1.9), suggesting a conserved function and a fundamental role.

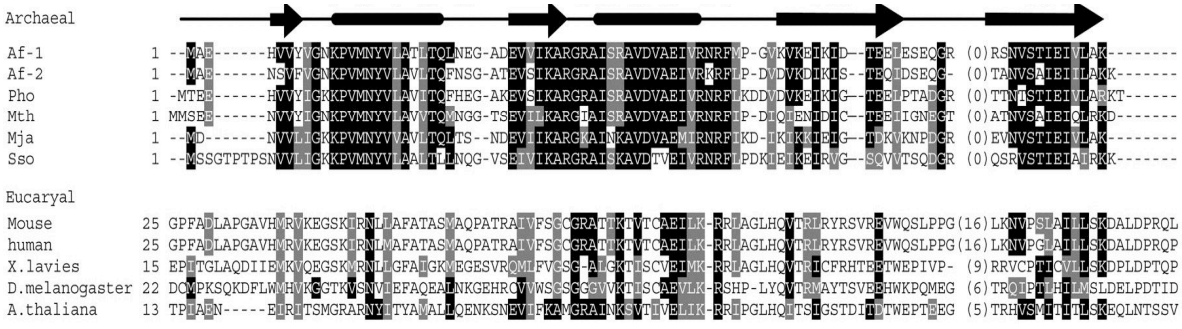


Figure 1.9: Sequence alignment of Alba1 proteins in Archaea and Eukaryotes. Strictly conserved residues in black and close substitutions in grey. Af-1 and Af-2, *A. fulgidus* (AF1956 and AF1067 respectively); Pho, *Pyrococcus horikoshii* (NP_143671); Mth, *Methanothermobacter thermautotrophicus* str. Delta H (AAB85958); Mja, *Methanococcus jannaschii* (Q57665); and Sso, *Sulfolobus solfataricus* (CAA67066) are included in the archaeal protein alignment (Zhao, 2003).

Most of the non-polar residues belonging to the hydrophobic core or dimer interface are completely conserved (Zhao, 2003). The loop containing Lys16 and 17 houses the most conserved residues. Gly15, Pro18, and Tyr 22 are responsible for stabilizing this loop. There are a number of salt bridges and ion pairs to be found stretching across the back of the dimer, again thought to be for stability reasons (Wardleworth, 2002).

1.5.1 Alba-DNA binding and function

Alba1 binds DNA with greater affinity than Sul7d members but in common with them, it seems to lack sequence specificity. The protein from *S. shibatae* has a DNA binding site of approximately 12 bp (Xue, 2000). From chromatin precipitation studies, it has been demonstrated that Alba1 is distributed along the chromosome in a ubiquitous and uniform fashion, suggesting a functional role in archaeal chromatin architecture (Wardleworth, 2002; Marsh, 2005). DNA binding affinities of both the native (acetylated) and recombinant forms of the protein dimer measured by electrophoretic mobility shift assays were shown to exhibit high affinity co-operative binding by the recombinant form (dissociation constant ~30 nM) whereas the native (acetylated) dimer bound with approximately 3 to 4 fold lower affinity (Jelinska, 2005; Guo, 2003). Mutation analysis and replacement of the lysine 16 and 17 residues with alanine and glutamate implicates them in having an important role in DNA binding. The ability of Alba1 to mediate transcriptional repression was also tested. It was found that acetylated Alba1 is unable to repress transcription as are the mutant forms of Alba1 K16A/E and K17E. It was then shown by Bell *et al* that binding of archaeal Sir2 to native Alba1 results in deacetylation, thus altering and

controlling Alba's DNA binding affinity and ultimately giving rise to transcriptional repression (Bell, 2002).

The Alba2 homodimer binds DNA with far weaker affinity than Alba1 but it is able to form a tight heterodimer with Alba1 (capable of binding DNA to a similar level as the Alba1 homodimer), which suggests that its function within the cell is enhanced by heterodimerization with Alba1 (Jelinska, 2005). Alba can bind DNA in two ways: (1) saturating concentrations of the Alba1 homodimer bind a single DNA duplex, giving rise to little DNA compaction (Lurz, 1986). This confirms a previous conclusion by Reinhardt *et al* that it did not supercoil DNA, nor was it involved in compaction (Dick and Reinhardt, 1986) and (2) low-intermediate concentrations of the Alba heterodimer can bind two DNA duplexes, resulting in a highly compacted DNA-protein complex (figure 1.10), indicating a possible role in the control of *in vivo* DNA packaging (Jelinska, 2005).

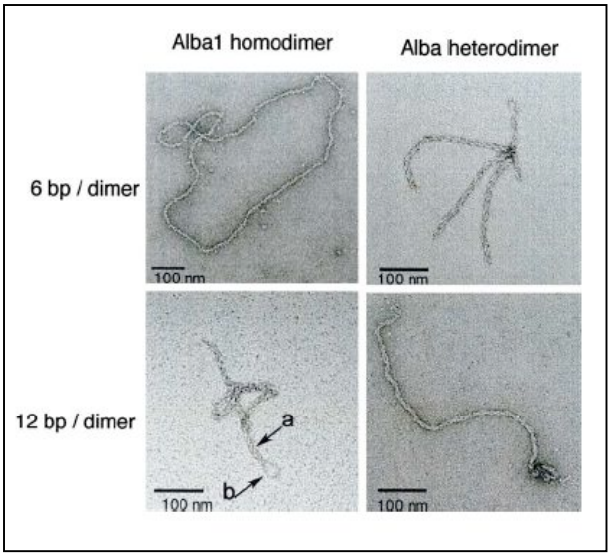


Figure 1.10: Electron Micrographs of nucleoprotein filaments formed by Alba(1) homodimer and Alba heterodimer binding to duplex DNA. DNA:protein ratios of 6bp/dimer and 12 bp/dimer are shown. The arrows in the bottom left EM where two duplexes are interwound, indicate stem (a) and loop (b) structures (Jelinska, 2005). Note: ds DNA is a nicked circular phiX174 plasmid.

1.5.2 A model for DNA-binding

Experiments involving the use of DAPI dye were carried out by Wardleworth and colleagues. DAPI binds DNA in the minor groove and has fluorescent properties. Addition of Alba1 to the DNA / DAPI complex results in loss of fluorescence from the

DNA suggesting that Alba binds DNA in the minor groove or at least distorts this area upon binding (figure 1.11). No evidence for increased DNA electrophoretic mobility was found and it was therefore assumed that little or no wrapping of the DNA by the protein occurred, this observation was also supported by EM studies (Wardleworth, 2002). However, interaction of DNA with protein will undoubtedly alter the net charge of the DNA resulting in an alternate rate of propagation through the agarose gel. In addition, compacted DNA will migrate faster. Thus, it is possible that the charge difference created by the addition of Alba to DNA (if it renders the molecule less negative) and any potential packing / compaction, could cancel each other out and consequently result in no change in movement through the gel, giving rise to the wrong conclusions.

The final stoichiometry for *S. solfataricus* Alba binding to DNA is 5-10 bp of double stranded DNA per Alba dimer. The first model for Alba set by Wardleworth *et al* suggests that Alba binds DNA in the minor groove, like DNaseI, but it has a longer β-hairpin that is angled in a different direction meaning that the DNA would have to be oriented in an alternative way to that found in the DNA-DNaseI complex. The two β-hairpin arms of the dimer are separated in length by ~40 Å, which corresponds with the length of one complete turn of B-form dsDNA. This means each hairpin could potentially interact with an equivalent part of the duplex in the minor groove and thus allow interaction of the central dimer body with the major

groove, giving three main points of contact with the DNA (figure 1.11). Small steric clashes in the model will possibly be overcome by localized distortion of the DNA or changes in the β - hairpin loop regions. However, recent findings by Jelinska have shown that the hairpin arms remain flexible upon binding and chemical shift data for these locations remained unchanged regardless of binding status. In addition, the residues concerned (positions 78-84) are not conserved and thus the loops are unlikely to be involved in contacting DNA (Jelinska, unpublished results). The model suggests that each dimer spans a region of ~15 bp duplex DNA. If additional Alba dimers bind along the duplex in a similar fashion with a rotation of 120° around the helix, this coincides with a stoichiometry of 5 bp in agreement with experimental observation. Such sheathing of the DNA by Alba affords protection against nucleases e.g. DNaseI. Therefore Alba may play a role in DNA protection (Wardleworth, 2002).

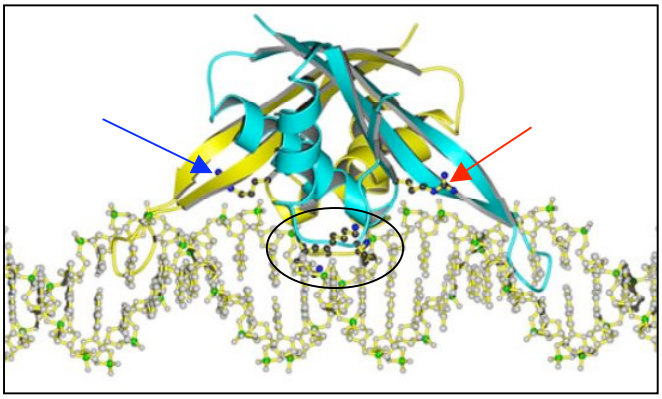


Figure 1.11: Model for Alba-DNA binding proposed by Wardleworth *et al.* Two hairpin arms interact with the minor grooves whilst the central body of the dimer makes contact with the major groove (Wardleworth, 2002). Lys16 and 17 are shown circled whilst conserved residues Arg42 and Phe60 are labelled with red and blue arrows, respectively

The molecular mechanism of Alba : DNA binding is not well understood but dimer - dimer contacts have revealed a positive regulatory role in DNA binding and it is thought that Alba actually binds DNA as a 'dimer of dimers' (Zhao, 2003), in

which binding of the Alba complex to DNA recruits consecutive dimers to the site and results in extended helical protein fibres (Jelinska, unpublished results, Jelinska, 2005). In contrast to the model of Wardleworth, Zhao and colleagues have proposed that the central body of the Alba dimer binds along the minor groove rather than the major groove. In addition to the putative roles of Alba in chromatin compaction, protection and transcriptional repression, it is predicted that Alba may also (or in some lineages of Archaea, exclusively) play a part in RNA metabolism since it shares a degree of structural homology with the RNA binding versions of the IF_3 – C fold (Aravind, 2003).

1.6 The deacetyltransferase of Alba, Sir2 (discussed in section 1.2.3.3).

In *Sulfolobus*, a protein doublet migrating at ~27 kDa was discovered when antisera raised against *Sso*Sir2 (M_r ~27.6 kDa) were used to probe western blots of whole cell extracts. After size exclusion chromatography, the doublet was found to elute at 50-60 kDa and 27 kDa. It was reasoned that the increase in mass was due to a stable protein-protein interaction forming between *Sso*Sir2 and an unknown protein of around 10 kDa. Addition of NAD to the extract resulted in disruption of this interaction. After isolation of the complex, the other protein concerned was found to be Alba. Verification of the findings were shown by an Alba-GST fusion complex, but not GST itself, interacting with recombinant *Sso*Sir2. *Sso*Sir2 interacts with Lys16 of the Alba dimer, causing a deacetylation reaction to take place and thus rendering Alba active to bind DNA with greater affinity (Bell, 2002).

1.7 Acetyltransferases of Alba

When the work described in this thesis first commenced, the deacetylase enzyme, *Sso*Sir2, had already been discovered and the acetyltransferase, *Sso*Pat (Marsh, 2005), responsible for modifying lysine 16 was in the process of being characterised but the N-acetyltransferase enzyme had not been identified. There are two genes in the *Sulfolobus solfataricus* P2 genome, *Sso0082* and *Sso0209* that may encode acetyltransferases of Alba1. The respective predicted proteins are small, *Sso*0082 being 151a.a long and *Sso*0209 167a.a long. From multiple sequence alignments, it appears that *Sso*0209 has eukaryotic N-acetyltransferase homologues as well as archaeal N-acetyltransferase homologues. *Sso*0082 shows sequence similarity to acetyltransferase enzymes from other prokaryotes (section 3.1 and appendix I). Both *Sso*0082 and *Sso*0209 share 30% identity with each other. Sequence alignments group *Sso*0209 as a putative N-acetyltransferase (NAT) enzyme. NAT's acetylate the α-amino group of other proteins but it has also been documented that they are capable of internal acetylation i.e. N^{ε}- acetylation of specific lysine residues (Polevoda and Sherman, 2003). This brings us to question whether *Sso*Alba is a target substrate for *Sso*0209 and if so, does the enzyme predominantly acetylate amino acids at the N^{α} or N^{ε} positions, or both? The closest homologue of *Sso*0209 is the Ard1 protein, sharing 37 % identity and 57 % similarity to the *H. sapiens* version (*h*Ard1). This sequence conservation allows *Sso*0209 to be renamed *ss*Ard1.

*h*Ard1 is a 26 kDa subunit belonging to a tri-subunit complex termed NatA, in which the other two proteins are NAT1 and NAT5. The NatA complex from *S. cerevisiae* has been well characterised and interesting findings from deletion mutant strains have shown that Ard1, the proposed catalytic subunit, cannot fulfil its

function as an acetyltransferase in the absence of NAT1. NAT1 has no capacity to acetylate polypeptides independently, as is the case in *ard-1Δ* mutants (Arnold, 1999). It was first demonstrated by Sigiura *et al* that mammalian NAT1 and mammalian Ard1 combine to form a functional acetyltransferase (Sugiura, 2003). Little is known about NAT5 apart from its homology to Ard1 and that it may be required for complex stability with the ribosome but it is not essential for activity (Gautschi, 2003). It has been speculated that NAT5p could be responsible for acetylation of a small subset of proteins that are too scarce to detect (Gautschi, 2003). It appears that NAT1 in the eukaryotic cell binds the ribosome and the emerging polypeptide chain to bring it into close proximity with Ard1 for possible modification (figure 1.12). Not all polypeptides are modified by Ard1 because acetylation depends largely on the N-terminal sequence as well as other factors that are poorly understood. It is possible that other residues within the N-terminal region may decrease or prevent acetylation (Moerschell, 1990) e.g. proline and the positively charged residues lysine or arginine occurring at the antepenultimate position i.e. the 3rd amino acid, after counting cleaved methionine (Polevoda & Sherman, 2003). Equally, the absence of certain residues may be responsible for non-modification of proteins displaying a viable target sequence (Polevoda & Norbeck, 1999). It has been noted that acidic residues in the N-terminal vicinity enhance acetylation in eukaryotes, but the reverse is true for archaea (Falb, 2006). NatA is responsible for acetylating sequences beginning with serine, alanine, glycine and threonine after removal of the initiator methionine. Acetylation of these residues by NatA accounts for the majority of N-terminal acetylated proteins within the cell.

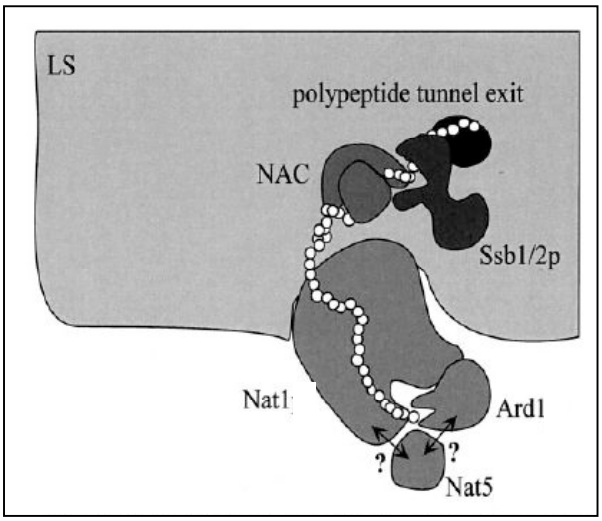


Figure 1.12: A model for co-translational Nα-acetylation by NatA (taken from Gautschi 2003). Nat1 binds to the large ribosomal subunit (LS) to stabilise the NatA complex and in doing so it contacts the emerging polypeptide chain, bringing it into close proximity with Ard1 for acetyl group transfer. It is unknown how Nat5 connects with the complex. NAC and Ssb1/2p constitute other factors that bind the ribosome and contact the polypeptide chain in advance of NatA. Their presence or absence seems to be of no consequence to the binding or function of NatA.

Complex	Residue Specificity	Subunits	Known functions
scNatA	ser, ala, gly, thr	Nat1 Ard1* Nat5	Influences multiple processes such as the cell cycle, heat-shock resistance, mating, sporulation, and telomeric silencing
scNatB	met-asp, met-glu (and a subclass of met-asn)	Mdm20 Nat3*	Involved in mitochondrial inheritance and actin assembly
scNatC	met-ile, met-leu, met-phe	Mak10 Mak31 Mak3*	Required for replication of dsRNA virus; expression is glucose-repressible
scNatD	histones H2A & H4	Nat4*	?

Table 1.3: Subunits, specificities and functions of the different *Saccharomyces cerevisiae* NAT complexes. *catalytic subunit of complex. All other subunits are associates essential for complex formation.

In addition to NatA, a further two major acetyltransferase complexes, NatB and NatC, exist in yeast (table 1.3). NatB substrates have Met-Asp- or Met-Glu- termini, whereas NatC targets substrates containing Met-Ile-, Met-Leu- or Met-Phe- termini (Arnold, 1999). The catalytic subunits belonging to NatB and NatC are known as Nat3 and Mak3, respectively. Both subunits bear noticeable overall % homology to Ard1. Another homologue, Nat4, also known as NatD, has been recently discovered to acetylate N-terminal sequences belonging to histones H2A and H4 (Song, 2003).

This protein does not seem to require additional subunits for acetyltransferase activity.

The existence of a complex in higher organisms allows speculation as to whether *ss*Ard1 can act alone or if it absolutely requires a binding partner in order to function. Another option is that the protein may have a binding partner to enhance its acetyltransferase activity. Sequence analysis shows that *ss*Ard1 is smaller in comparison to *h*Ard1 and *sc*Ard1 since it does not contain the extended C-terminal region (which is often less structured) of the eukaryotic homologues. The homologues posses a predicted coiled coil region from amino acids 194 to 224 (figure 1.13), which is thought to interact with NAT1, and supporting this theory, early research by Park and Szostak showed very little co-immunoprecipitation of NAT1 with a truncated version of Ard1 comprising only the first 194 amino acids of the protein (Park & Szostak, 1992). The fact that *ss*Ard1 lacks this potential binding region suggests that it probably does not bind a NAT1 homologue in the crenarchaea. In addition to this, similarity searches in *Sulfolobus* show absolutely no indication of any NAT1 or NAT5 homologues. However, recent research by Arnesen and colleagues (Arnesen, 2005) has suggested that the N-terminal amino acids 1-58 are important for binding NAT1.

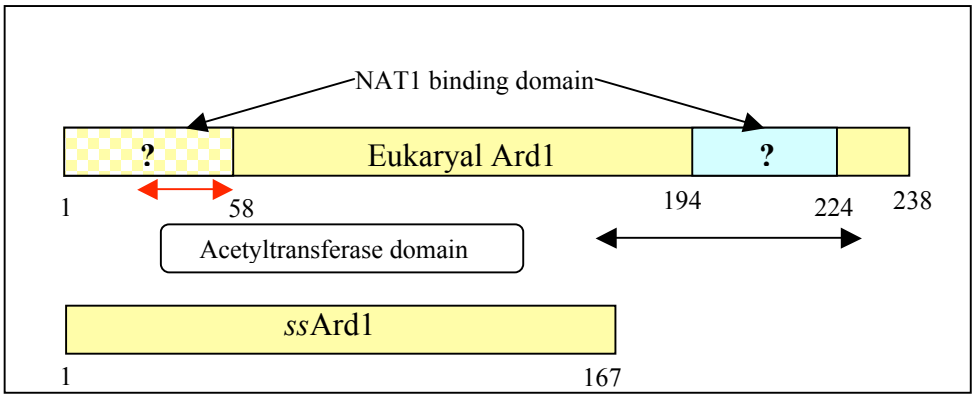


Figure 1.13: Schematic representation of *S. cerevisiae* and *S. solfataricus* Ard1. *sc*Ard1 contains a NAT1 binding domain at the N-terminal (1-58) region (Arnesen, 2005) *OR* the C terminus between residues 194 and 224 according to early research by Park & Szostak but *ss*Ard1 lacks this later region. The acetyltransferase domain spans residues 45-130 (Arnesen 2005). Red arrow indicates a string of negatively charged amino acids that may allow Ard1 to form a homodimer. Black arrow indicates a region of Ard1 that is able to form homodimer with other Ard1 molecules (Park & Szostak, 1992).

Similarity searches using BLAST for *ss*Ard1 homologues throughout the archaea gives rise to a large number of hypothetical and known protein sequences which share a high degree of sequence identity, especially in the Sulfolobales. *S. tokodaii* gene ST0258 and *S. acidocaldarius* gene Saci_0459 are clear *Ard1* homologues since they posses 77 % and 68 % identity, respectively, with *ssArd1*. A distribution of Ard1 homologues in the archaea is shown in the following dendogram (figure 1.14).

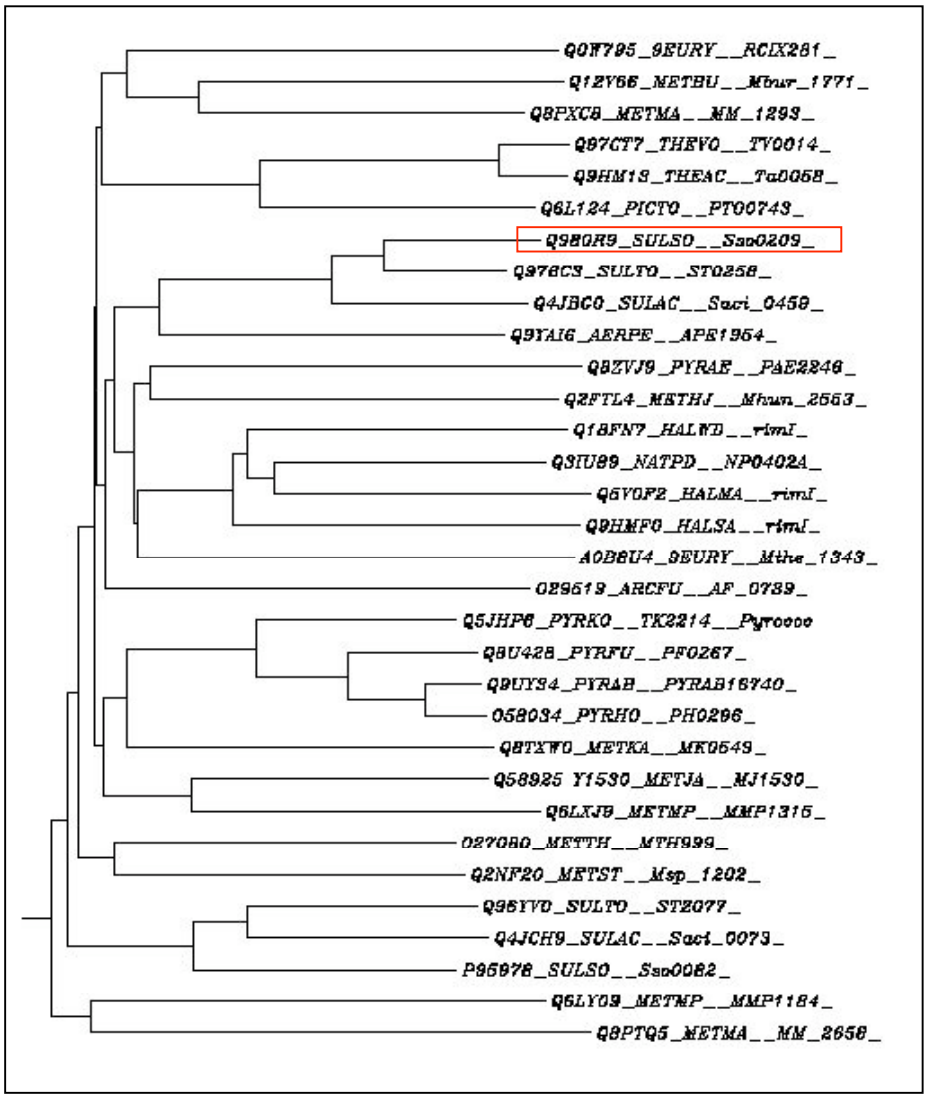


Figure 1.14: Phylogenetic tree displaying the evolutionary divergence of *ss*Ard1 homologues throughout the archaea (created at www.ebi.ac.uk/Tools/clustalw2/index.html**).**

*ss*Ard1 (*Sso*0209) is boxed in red. The combined distance of horizontal lines between two subjects indicates relative genetic similarity i.e. shorter distance equals greater similarity. Q0W795_9EURY_RCIX281 Predicted acetyltransferase -Uncultured methanogenic archaeon RC-I, Q12V66_METBU_Mbur_1771 SSU ribosomal protein S18P alanine acetyltransferase – *M. burtonii* (strain DSM 6242), Q8PXC8_METMA_MM_1293 Acetyltransferase – *M. mazei*, Q97CT7_THEVO_TV0014 ARD1 – *T. volcanium*, Q9HM13_THEAC_Ta0058 ard1 subunit – *T. acidophilum*, Q6L124_PICTO_PT00743 Ribosomal protein s18 alanine acetyltransferase – *P torridus,* Q980R9_SULSO_Sso0209 Ard –*S. solfataricus*, Q976C3_SULTO_ST0258 predicted NAT – *S. tokodaii*, Q4JBG0_SULAC_Saci_0459 predicted NAT – *S. acidocaldarius*, Q9YAI6_AERPE_APE1954 N-terminal acetyltransferase – *A. pernix*, Q8ZVJ9_PYRAE_PAE2246 N-acetyltransferase – *P. aerophilum*, Q2FTL4_METHJ_Mhun2553 SSU ribosomal protein S18P alanine acetyltransferase – *M. hungatei* (strain JF-1 / DSM 864), Q18FN7_HALWD_rimI – *H. walsbyi* (strain DSM 16790), Q3IU89_NATPD_NP0402A putative uncharacterized protein – *N. pharaonis* (strain DSM 2160 / ATCC 35678), Q5V0F2_HALMA_rimI N-terminal acetyltransferase – *H. marismortui*, Q9HMF0_HALSA_rimI N-terminal acetyltransferase – *H. salinarium*, A0B8U4_9EURY_Mthe_1343 GNAT family acetyltransferase – *M. thermophila* (strain DSM 6194 / PT), O29519_ARCFU_AF_0739 Ribosomal protein S18 alanine acetyltransferase – *A. fulgidus*, Q5JHP6_PYRKO_TK2214 RimI homolog – *P. kodakaraensis*, Q8U428_PYRFU_PF0267 Ribosomal protein s18 alanine acetyltransferase – *P. furiosus*, Q9UY34_PYRAB_16740 N-terminal

acetyltransferase – *P. abyssi*, O58034_PYRHO_PH0296 hypothetical acetyltransferase – *P. horikoshii*, Q8TXW0_METKA_MK0549 Acetyltransferase – *M. kandleri*, Q58925 Y1530_METJA_MJ1530 uncharacterised acetyltransferase – *M. jannaschii*, Q6LXJ9_METMP_MMP1315 rimI NAT– *M. maripaludis* (genome segment 4/5), O27080_METTH_MTH999 ARD1 – *M. thermoautotrophicum*, Q2NF20_METST_Msp_1202 Predicted acetyltransferase – *M. stadtmanae* (strain DSM 3091), Q96YV0_SULTO putative uncharacterized protein ST2077 – *S. tokodaii*, Q4JCH9_SULAC_Saci_0073 Acetyltransferase – *S. acidocaldarius*, P95978_SULSO_Sso0082 Orf c04040 rimI related NAT – *S. solfataricus*, Q6LY09_METMP_MMP1184 N-acetyltransferase related protein – *M. maripaludis*, Q8PTQ5_METMA MM_2658 Acetyltransferase – *M. mazei*.

1.8 Research Aims

Acetylation is a fundamental biological modification that has evolved to meet the individual requirements of many diverse cellular processes. In this case, it regulates the ability of Alba to bind DNA, whether directly or as a low level molecular signal to other proteins. This thesis will primarily describe the identification and biological characterisation of an acetyltransferase enzyme responsible for modifying the N-terminus of the Alba1 protein. In addition, kinetic parameters and reaction rates between Alba1 and the acetyltransferase will be examined. Acetyltransferase activity towards several other *Sulfolobus* proteins will also be tested to assess whether the enzyme is promiscuous or specific to Alba. The active site and the substrate binding site of the putative N-acetyltransferase will be explored using mutational analysis and crystallographic studies.

Chapter 2: Materials and Methods

2.1 Isolation of *Sulfolobus Solfataricus* genomic DNA

Genomic DNA was isolated from *Sulfolobus* cells (-80 °C frozen pellet) using the animal tissue protocol from the QIAGEN DNeasy® tissue kit (50). A 1:10 dilution of this DNA was used for Polymerase Chain Reaction (PCR).

2.2 Cloning the genes and expressing protein

2.2.1 Cloning and Vectors

The *Sso0082* gene region was amplified from *Sulfolobus solfataricus* (P2 strain) genomic DNA by PCR using Pfu (proof reading) polymerase (Promega) and the following specifically designed primers (Operon):

5' primer: 5'- CGTCGGATCCCCATGGTAATTATAACTGATGCAACAG-3'

3' primer: 3'- CCGGGGATCCGTCGACTTAATCCAGATATTTATAATATAAATTC-3'

In order to subclone the gene into the pET28c vector (Novagen) for native protein expression, 5 µl of PCR product or plasmid was used per 20 µl reaction along with restriction enzymes *Nco*I (1µl of 10U/µl) and *Bam*HI (0.5µl of 10U/µl) and 2µl 10X *Bam*HI buffer (Fermentas). Samples were incubated at 37 °C for 2 hours with *Nco*I and then overnight at 28 °C on addition of *Bam*HI. Ligations were then carried out overnight at 5 °C using 4 µl plasmid DNA, 10 µl digested product, 1 µl T4 DNA ligase and 2 µl T4 DNA ligase buffer. The gene was also subcloned into the pLOU3 (construct prepared by Dr. Louise Major, St. Andrews University. See figure 15) *Nco*I / *Sal*I site and pGEX-5X-3 (Amersham Pharmacia Biotech) *Bam*HI / *Sal*I site to

generate protein fused with maltose binding protein (MBP) and glutathione-S-transferase (GST), respectively.

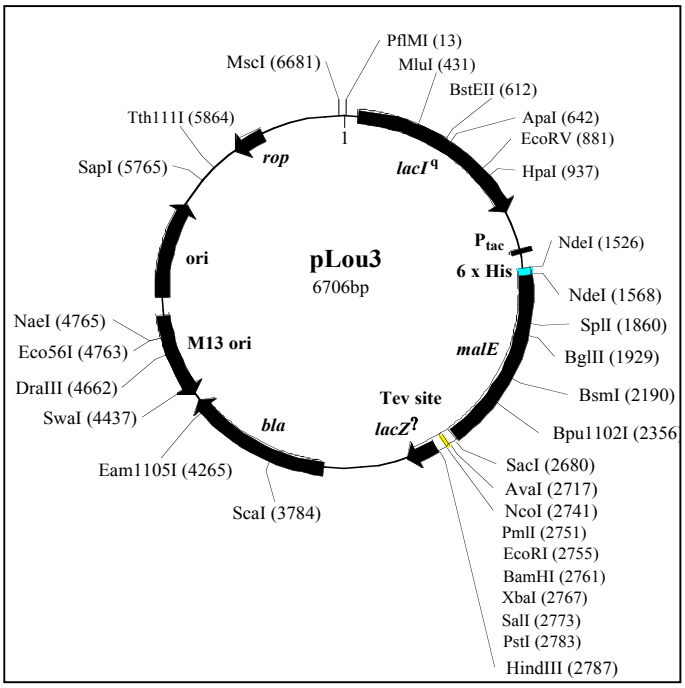


Figure 2.1: pLOU3 (or pHIS_MalC2-Tev):

pLou3 was derived from the NEB vector pMAL-c2X, and thus has cytoplasmic expression of the fusion protein. New features are N-terminal 6xHis tag before MBP and a Tev protease cleavage site between MBP and the (modified) polylinker.
Unique restriction sites are shown on the map (+ NdeI which cuts twice).

The *ssArd1* (*Sso0209)* gene was amplified from *Sulfolobus solfataricus* genomic DNA using Pfu polymerase and the following primers:

5' primer: 5'- CGTCGGATCCCCATGGAGCTCGCTGAAAAAGATAAAGG-3'

3' primer: 5'- CCGGGGATCCGTCGACTTAGAGTGGTCTTGCCATAAGATATGC-3'

The PCR product was cloned into the *Nco*I/*Bam*HI site of pET28c for native protein expression and the *Nco*I/*Bam*HI site of the pHisTev30a vector (designed by Dr. Huanting Liu, St. Andrews University) to create a poly-histidine tag on the N-terminal region of *ss*Ard1.

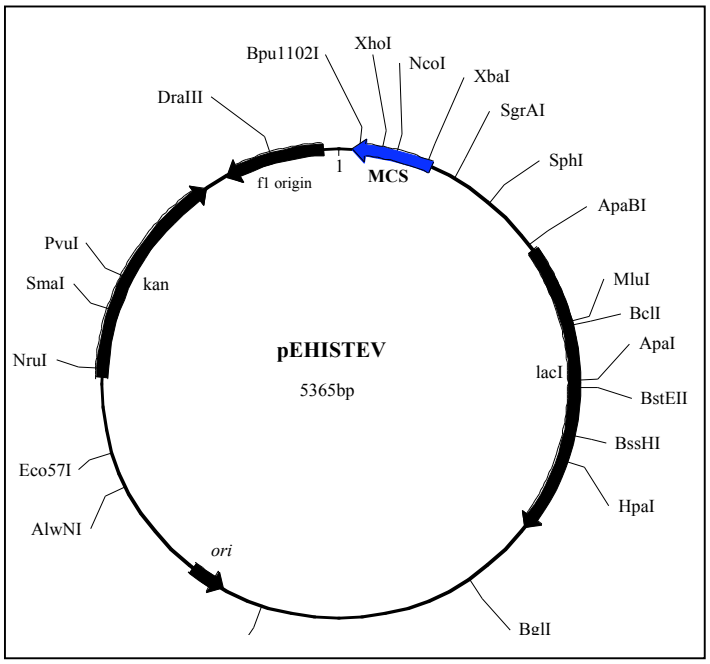


Figure 2.2: Vector pHISTEV30a.

pHisTev30a designed by Dr. Huanting Liu.

The *Sir2* gene was amplified from *Sulfolobus solfataricus* using KOD HiFi polymerase (Novagen) and the following primers:

5' primer: 5'- CGTCGGATCCCCATGGTATACGAGAAAGTAGCTGAAG-3'

3' primer: 5'- CCGGGGATCCGTCGACTTACGAAAGTATTTTCTGTCTCACG-3'

The PCR product was cloned into the *Nco*I/*Bam*HI site of pET28c.

A pET30 construct containing Alba1 (Sso10b) for native expression was obtained from Dr. Steve Bell (MRC Cancer Cell Unit, Cambridge). Three mutant versions, K16A (purified protein only), K16EpET29 and K17EpET29 were obtained from Dr. Ben Wardleworth (former St. Andrews lab member).

All products were cleaned up after PCR and digestion using a QIAquick purification kit. Plasmids were then transformed into an *E.coli* Top10 strain (Invitrogen) for amplification. The amplified plasmids were extracted via miniprep kit (QIAprep® Miniprep, QIAGEN) and the constructs sent for DNA sequencing (Dundee Sequencing service).

2.2.2 Site Directed Mutagenesis

Mutants of Alba at the N-terminus (S2A, S2E, S2G, S2L, S2P, S2T and S2V) were created by site directed mutagenesis using Stratagene® quickchange site directed mutagenesis kit. Stratagene SDM involves the use of individually synthesised primers, (each containing the mutation of interest) a high fidelity polymerase i.e. *Pfu Turbo*®, DNTPs and a PCR machine. The primers anneal to the plasmid-gene construct and *Pfu Turbo*® fully replicates both strands of plasmid DNA (figure 2.3). The parent DNA is methylated and, as such, can be digested at the end of thermal cycling using *Dpn*I. The construct containing the mutated gene is then transformed into supercompetent cells.

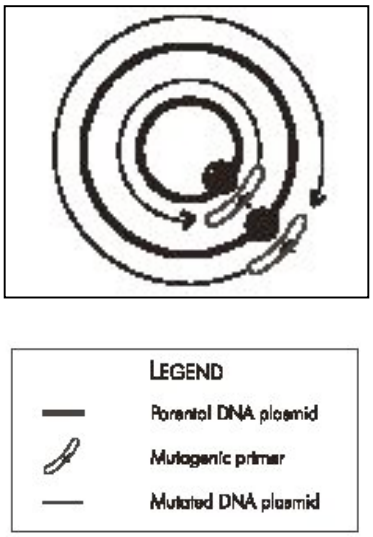


Figure 2.3: Basis of Stratagene Site Directed Mutagenesis.
Plasmid-gene construct. Primers anneal at complimentary DNA site on gene and replication occurs, generating high copy construct containing mutant gene.

The following primers (Operon) were used to generate mutations in Alba1:

5’ primer: 5’-GGAGATATACATATGGCGAGCGGAACCCCAACTCC-3’ Alanine

3’ primer: 5’-GGAGTTGGGGTTCCGCTCGCCATATGTATATCTCC-3’ (MASG…)

5’ primer: 5’-GGAGATATACATATGGAAAGCGGAACCCAACTCC-3’ Glutamic acid

3’ primer: 5’-GGAGTTGGGGTTCCGCTTTCCATATGTATATCTCC-3’ (MESG…)

5’ primer: 5’-GGAGATATACATATGGGTAGCGGAACCCCAACTCC-3’ Glycine

3’ primer: 5’-GGAGTTGGGGTTCCGCTACCCATATGTATATCTCC-3’ (MGSG…)

5’ primer: 5’-GGAGATATACATATGCTGAGCGGAACCCCAACTCC-3’ Leucine

3’ primer: 5’-GGAGTTGGGGTTCCGCTCAGCATATGTATATCTCC-3’ (MLSG…)

5’ primer: 5’-GGAGATATACATATGCCCAGCGGAACCCCAACTCC-3’ Proline

3’ primer: 5’-GGAGTTGGGGTTCCGCTGGGCATATGTATATCTCC-3’ (MPSG…)

5’ primer: 5’-GGAGATATACATATGACCAGCGGAACCCCAACTCC-3’ Threonine

3’ primer: 5’-GGAGTTGGGGTTCCGCTGGTCATATGTATATCTCC-3’ (MTSG…)

5’ primer: 5’-GGAGATATACATATGGTGAGCGGAACCCCAACTCC-3’ Valine

3’ primer: 5’-GGAGTTGGGGTTCCGCTCACCATATGTATATCTCC-3’ (MVSG…)

Point mutations were made at specific sites in *ss*Ard1 that had been identified by sequence alignments to be absolutely or highly conserved. The chosen residues were D21, H88, S131_N132, D156 and E158. The following primer pairs were used:

5’ primer: 5’-CGCAAGAATGGATGCAATAGACCAAATAATAAAAATA-3’ D21A

3’ primer: 5’-TATTTTTATTATTTGGTCTATAGCATCCATTCTTGCG-3’

5’ primer: 5’-CGTTAGTCAGAAAAGGAGCTGTGGTTTCAATAGCAG-3’ H88N

3’ primer: 5’-CTGCTATTGAAACCACAGCTCCTTTTCTGACTAACG-3’

5’ primer: 5’-TTAGAAGTAAGAGTCGCAGCATATCCCGCAATAGCCCTA-3’ S131A_

3’ primer: 5’-TAGGGCTATTGCGGGATATGCTGCGACTCTTACTTCTAA-3’ _N132A

5’ primer: 5’-GGTTATTATGCTGCTGGGGAAGACGCATATC-3’ D156A

3’ primer: 5’-GATATGCGTCTTCCCCAGCAGCATAATAACC-3’

5’ primer: 5’-GGTTATTATGCTGATGGGGCAGACGCATATCTTATGGC-3’ E158A

3’ primer: 5’-GCCATAAGATATGCGTCTGCCCCATCAGCATAATAACC-3’

Transformation was into XL1-Gold supercompetent cells and the amplified constructs were sent for sequencing.

2.2.3 Protein Expression

Each plasmid in turn was transformed into Rosetta BL21 cells (Novagen). Mini cultures were set up in 5 ml L-broth containing the appropriate antibiotic (35 μg/ml kanamycin or 100 μg/ml ampicillin) to test for best protein production. These were induced with 0.1-0.2 mM IPTG for 3 hours. 1ml of each culture was centrifuged and the pellet lysed in 100 μl lysis buffer (20 mM Tris-HCl, pH 8.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) with a few granules of benzamidine. The samples were centrifuged again to separate insoluble matter from the supernatant. The best expressing cultures were then scaled up and the bacterial pellets were frozen until use.

2.3 Protein Purification

2.3.1 Purification of ssArd1

Frozen pellets were defrosted on ice by adding 30-40 ml lysis buffer (20 mM Tris-HCl, pH 8.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT and 1 mM benzamidine). The pellet was lysed by sonicating for 3 X 2 min on ice during which a few granules of DNaseI were added. The soluble fraction was separated from cell debris by centrifugation for 30 minutes at 75, 600 g and 4 °C. Supernatant was diluted by $^1/_6$ in buffer A (20 mM Tris-HCl, pH 8, 1 mM EDTA, 1 mM DTT) to lower salt concentration and syringe filtered through a 2 μm Acrodisk® syringe filter.

The protein sample was heat treated at 65 °C for 25 minutes and centrifuged again at 4 °C, 75, 600 g for 20 minutes. This step eliminates non-heat stable *E.coli* proteins. The sample was then loaded onto an anion exchange Hi Trap 5 ml HP Q-sepharose column (Amersham Biosciences), which had been washed with buffer B (20 mM Tris-HCl, pH 8, 1 M NaCl, 1 mM EDTA and 1 mM DTT) and equilibrated with buffer A. Protein was eluted at 5 ml/min over a 400 ml linear gradient of 0-1000 mM NaCl. Fractions and flow through were analysed by SDS-PAGE. The 530 ml of flow through sample contained little salt and was loaded onto a cation exchange HiTrap 5 ml HP heparin column (Amersham Biosciences), which had been equilibrated with buffer B. Samples were collected over a 400 ml linear gradient of 0-1000 mM NaCl and fractions were again analysed on SDS-PAGE. The resulting630 ml of flow through that was collected was concentrated using ammonium sulphate precipitation at 80 % saturation: proteins were pelleted by centrifuging at 15, 900 g at 4°C for 30 minutes and resuspended in 40 ml of buffer C

(20 mM BisTris, pH 6, 0.3 M NaCl and 1 mM EDTA). The solution was dialysed overnight in 3L buffer C to remove Ammonium Sulphate and persistent contaminant proteins of less than 14 kDa and remaining proteins were separated by size on a HiLoad® 26/60 Superdex® gel filtration column (Amersham Biosciences) that had been equilibrated with buffer C. Fractions were checked on SDS-PAGE gel and the band suspected as *ss*Ard1 was cut from the gel and sent for matrix assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry for identification.

Any protein solution still containing contaminants was then dialysed in buffer D (20 mM MES, pH 6) and passed through a Bioscale CHT2-1 hydroxyapatite column (Bio-Rad) or another cation exchange 30 Sp-Resource column (Amersham Biosciences) equilibrated with buffer D. Fractions eluted over 60 minutes with a salt gradient of 0-1000 mM NaCl.

2.3.2 Purification of Alba1

Recombinant, K16A, K16E and K17E mutants:

Pellets were defrosted on ice and lysed in 20 mm Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM $MgCl_2$ and 1 mM benzamidine. Samples were sonicated 3 X 2 min on ice and DNaseI was added and allowed to incubate for 10 minutes on ice. Lysate was then centrifuged at 75, 600 g, 4 °C, for 20 minutes. Supernatant was collected and heated for 20 minutes at 70 °C to precipitate unwanted proteins and then cooled down to 4 °C again during a repeat centrifugation step. Supernatant was diluted 1/5 with 20 mM Tris- HCl, pH 8.0, 1 mM EDTA and 1 mM DTT, and loaded onto a cation exchange heparin column that had been equilibrated with the same buffer. Elution conditions were as follows: 5 ml/min over 40 minutes with a linear incline of

0-100% buffer B (dilution buffer containing 1M NaCl). Fractions containing Alba were concentrated in a 5, 000 MW vivaspin column and then loaded onto a HiLoad® 26/60 Superdex® gel filtration column for the final purification step.

N-terminal mutants:

The procedure was the same as that outlined above but the heat step was not included or was instead carried out after cation exchange.

2.3.3 Purification of Sir2

Pellets were defrosted and lysed in ~10 ml 20 mM Tris-HCl, pH 8.1, 200 mM NaCl, 1 mM $MgCl_2$ and 1 mM benzamidine. After sonicating for 3 X 2 minutes, samples were incubated on ice for 10 minutes with DNaseI. Samples were centrifuged as usual and checked on SDS PAGE gel. A band of ~27 kDa was excised and sent to mass spec for identification. In the mean time, the protein solutions were tested for heat stability at temperatures ranging from 50 °C to 70 °C for varying time periods. Non heat-treated sample was diluted 1/6 in buffer without salt (buffer A) and then loaded onto a Q-sepharose column. Elution conditions were as follows; fractions were collected at 2 ml/min intervals over a 56 minute gradient consisting of 100 % buffer A with increasing 1M salt buffer B (to 40%) over the first 20 minutes, then increasing to 100% buffer B over 20 minutes.

2.3.4 MALDI-TOF and tryptic digest analysis

All protein samples were sent for analysis to the mass spectrometry unit at St. Andrews University for identification, whole protein mass and confirmation of

mutagenesis. For impure samples, main bands were excised from SDS-PAGE gel for ID. 20 μl of 10 μmolar pure protein samples were sent for whole protein mass analysis and cleavage by the protease enzyme trypsin. This was done by adjusting the pH of each sample to pH 8.0 with 1M Tris, pH 8.0. 0.5 ml (0.1mg) trypsin was then added and the samples were incubated overnight at 37 ºC. "The resultant peptides were diluted 1:5 into 5 % formic acid and then separated using an UltiMate nanoLC (LC Packings, Amsterdam) equipped with a PepMap C18 trap & column using a 3 and a half hour gradient of increasing acetonitrile concentration. The eluent was sprayed into a Q-Star Pulsar XL tandem mass spectrometer (Applied Biosystems, Foster City, CA) and analysed in Information Dependent Acquisition (IDA) mode. MS/MS data for doubly and triply charged precursor ions were analysed using ProID software (Applied Biosytems), searching against a *Sulfolobus solfataricus* protein database containing entries with and without the N-terminal methionine. The data was searched with tolerances of 0.2 Da for the precursor and fragment ions, trypsin as the cleavage enzyme, one missed cleavage, and acetylation and methionine oxidation selected as possible modifications. Unambiguous acetylation was assigned if the peptide match gave a confidence score of 99" (Mackay, 2007).

2.3.5 Concentrating the samples

Concentrating of recombinant Alba and mutants was straightforward. This was carried out in 3,000 or 5,000 MWCO vivaspin concentrators (Vivascience) in an eppendorf centrifuge 5810 at 4000 rpm and 4 °C. For *ss*Ard1, however, dialysis tubing was used in the concentrating of the samples. 12-14 kDa of pre-treated membrane was washed in buffer and filled with ~30 ml sample. This was laid flat in

a small container, covered with large amounts of polyethylene glycol (PEG) 8K and left at room temperature for 2-3 hours to allow the excess buffer to cross the membrane whilst keeping the protein inside the tubing. It was necessary to achieve high concentrations of *ss*Ard1 to generate protein crystals for structural analysis.

Optical density A_{276} (*ss*Ard1 extinction coefficient 24250 M^{-1} cm^{-1}, absorbance 1.247 (0.1% = 1g/L)) or A_{280} (*Sso*Alba1 extinction coefficient 1490 M^{-1} cm^{-1}, abs 0.141 (0.1% = 1g/L)) were taken for all concentrated samples. These initial results were double checked by Bradford assay. 0-10 μg/μl of BSA (bovine serum albumin) was bound to Bradford dye reagent and used to generate a standard curve showing absorbance levels at A_{550}. A small amount e.g. 1 μl of sample protein is then tested in the same way and the A_{550} reading is compared with the standard curve, to give an accurate reading. Alba does not bind Bradford dye in the same manner as BSA because it contains 2X the number of arginine residues (primary reactant sites) causing an increased number of interactions with Bradford dye molecules. Therefore, a pure concentrated sample of Alba of known concentration (obtained from Claire Jelinska) was used to generate a standard graph for the comparison.

2.3.6 Glutaraldehyde crosslinking

50 μg *ss*Ard1 was incubated for 10 minutes with 50 mM $NaPO_4$, pH 7.6 at room temperature. Glutaraldehyde was added to a final concentration of 1% and the solution incubated for a further 1-2 minutes. 2 M $NaBH_4$ was used to quench the reaction over 20 minutes. Deoxycholate and TCA were used to precipitate the reaction before centrifuging for 20 minutes at 4 °C. The pellet was washed with cold acetone and centrifuged again, after which, it was resuspended in 20 μl loading buffer (NuPage) and loaded onto SDS-PAGE (4-12 %) gel.

2.4 Identifying a possible binding partner for *ss*Ard1

2.4.1 Affi-gel 10 column

Two columns were prepared as stated in the manufacturers protocol (BioRad). 1 ml of 6.5 mg/ml *ss*Ard1 was diluted to 3 ml and dialysed into 20 mM MES, pH 6.5, 100 mM NaCl, 1 mM EDTA and 1 mM DTT. *ss*Ard1 was then bound as a ligand to one column and dH_2O was added to the control column. 8 ml of *Sso* lysate (prepared from approximately 2 g of cells lysed in 20 mM MES, 100 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 6.5) was filtered through the columns in low salt (50 mM) buffer. Columns were washed with low salt buffer, then twice with 1 ml volumes of 300 mM buffer, 600 mM buffer and 1 M salt buffer. 1 ml fractions were collected from each wash and 20 μl applied to SDS-PAGE (4-12 %) gel. The gel was sypro-ruby stained for increased clarity.

2.4.2 Western Blot analysis

A semi-dry blot was performed using a dry blot machine at 25V (400 mA) for 25 minutes. The nitrocellulose membrane was then placed in blocking buffer for 10 minutes and primary antibodies for Alba1 and Alba2 were added in 1:1000 dilution. After incubation on a shaker machine for 1 hour (and a further wash step) secondary antibody (anti rabbit-host-Goat IgG-antigen) was added in a 1:5000 dilution. This was allowed to incubate for 30 minutes. After washing, the blot was drained and pico-detection solution (Pierce) was applied for 1 minute.

2.5 Acetylation assays

2.5.1 Acetylating Alba with ssArd1

[Assay adapted from Takei *et al*, EMBO reports, 2001, Vol21, pg119-123.]

Six 25 µl reactions were set up in acetylation buffer (50 mM Tris-HCl, pH 8, 0.1 mM EDTA, 10 mM Na butyrate, 1 mM DTT and 10 % glycerol), see table 2.1.

Alba (10 µM)	*ss*Ard1 (500 nM)	Acetylation buffer	AcCoA (20µM)
Native	-	+	-
Recombinant	-	+	-
K16A	-	+	-
Recombinant	+	+	+
Recombinant	+	+	-
K16A	+	+	+

Table 2.1: Acetylation assay sample constituents. Acetylation of native (wt) *Sso*Alba1, recombinant *Sso*Alba1 and K16A mutant Alba1 protein by *ss*Ard1.

These samples were incubated at 70°C for 1 hour before being flash frozen to terminate the reaction. All samples were sent to St. Andrews University mass spectrometry service for analysis. Assays were repeated over 2 hours and mass spectrometry analysis was carried out:

Alba (10 µM)	*ss*Ard1 (500 nM)	Acetylation buffer	AcCoA (20µM)
Old Recombinant	-	+	-
New Recombinant	-	+	-
New Recombinant	+	+	-
New Recombinant	+	+	+

Table 2.2: Acetylation assay sample constituents. Old recombinant Alba1 sample compared to newly purified recombinant Alba1.

2.5.2 Radiolabelled acetylation assay on SsoAlba1 & SsoAlba2

An initial acetyltransferase assay was set up using 0.5 μM *ss*Ard1, 10 μM Alba and 35.7 μM ^{14}C AcCo-A. The same 10 μl reactions were set up for Alba 2, single stranded DNA binding protein SSB (Cubeddu, 2005), obtained from Lisa Cubeddu, and xeroderma pigmentosum group F endonuclease XPF (Roberts *et al*, 2003), obtained from Jana Rudolf. Controls (minus *ss*Ard1) were also prepared. Each reaction was carried out at 60°C for 30 minutes.

Focusing on Alba 1 acetylation, time points were reduced to 1, 5, 10, 15, 20 and 30 minutes. The 5 minute time point was chosen in order to determine the effects of varying *ss*Ard1 and Alba concentrations to establish optimal conditions. This was carried out using 18 μM ^{14}C AcCo-A. 250 nM *ss*Ard1 and 10 μM recombinant Alba, the optimal conditions selected for further experiments, were incubated with 18 μM ^{14}C AcCoA at 55 °C over varying time points up to 5 minutes. Each time point was repeated at least 3 times.

To investigate kinetics of the reaction, the same assay was set up over 4 minutes using 250 nM *ss*Ard1 and up to 80 μM Alba. An excess of cold AcCoA (9/10) was added to the radiolabelled AcCoA before adding to the samples.

2.5.3 Curcumin assay (HAT inhibitor)

9.2 mg curcumin powder was dissolved in 2.5 ml DMSO to a stock concentration of 10 mM. Samples of 10 μM Alba, 250 nM *ss*Ard1 and varying curcumin concentrations (0, 50, 100, 200, 500 μM) were incubated at 55 °C for 3 minutes using 18 μM of ^{14}C AcCoA.

2.5.4 Radiolabelled assay on Alba1 mutants

Alba mutants were diluted to the same concentration as the recombinant version (checked by A_{276} and bradford assay using pure Alba of known concentration as a standard). All mutants at 10 μM underwent the same assay as in 2.5.2.

2.5.5 Radiolabelled acetylation assay on other substrates

Acetylation assays were carried out as described in 2.5.2 with a variety of proteins replacing Alba (holliday junction cleavage and holliday junction endonuclease enzymes Hjc & Hje (Kvaratskhelia and White, 2000) obtained from Jo Parker, proliferating cell nuclear antigen subunits PCNA1 & 2 (Dionne *et al*, 2003) obtained from Jen Roberts, Sso7d obtained from Biljana Petrovic-Stojanovska and DNA repair helicase XPB1 (Richards *et al*, 2007) obtained from Jodi Richards), all of which have different N-terminal sequences. Each sample was repeated 2 times for an average result and one control was carried out for each protein.

2.5.6 Radiolabelled assay on <u>Sulfolobus</u> cell lysate

A scaled up version of the previous acetylation assay was carried out using 54 μM ^{14}C acetylCoA, 2.5 μM *ss*Ard1, 1X acetylation buffer and 10 μl of non-diluted cell lysate. *E.coli* cell lysate was also used as a positive control to show how much acetylation occurs in prokaryotes. The positive control was incubated at 37 °C, the optimum temperature for *E.coli* growth as was *Sso* lysate for a direct comparison. *Sso* lysate was also incubated at 70 °C to show acetylation extent under its ecological growth temperature. Total soluble protein was obtained from three more cell types (HeLa (immortal cell line derived from *H. sapien*), MRSA (methicillin resistant *S.*

aureus) and *S. cerevisiae*). Bradford assay was used to determine the protein concentration and all samples were diluted to 0.47 mg/ml. Each sample was subject to the same acetylation assay as previously described but 36 μM ^{14}C AcCoA and 2 μM *ss*Ard1 were used because the excess radiolabelled AcCoA interfered with scintillation counting. For the same reason, the frequency of acetylation was reduced by decreasing *ss*Ard1 concentration.

2.5.7 Radiolabelled acetylation assay using ssArd1 mutants

*ss*Ard1 mutants were diluted to stock concentrations of 2 μM for accurate dilution to 250 nM. Acetylation assays were carried out, as described in 2.5.2, for each mutant and again with the recombinant wt version.

2.6 Stability of acetylated Alba1

2.6.1 Acetylation of Alba N-terminus for Differential Scanning Calorimetry

We used 2 mg/ml Alba (i.e. 96 μM) in a total of 6 ml. 177.6 μM acetyl CoA, and 2.4 μM *ss*Ard1 were used (a direct scale up from 20 μl sample) and the sample was incubated at 55 °C for 1$_{1/2}$-2 hours. Gel filtration was carried out to separate *ss*Ard1 and acetyl CoA from acetylated Alba. The acetylated Alba was concentrated in a viva spin column to ~1 ml, giving a protein concentration of ~2 mg/ml. The sample was checked on SDS-PAGE gel to confirm purity and concentration before being sent to Glasgow University for differential scanning calorimetry using a Microcal-MCS system (carried out by Margaret Nutley). 1 mg/ml protein sample was degassed and heated (20-100 ºC at a scan rate of 60 ºC/hour) in a calorimeter cell at a constant rate along side an identical reference cell containing the same solution

buffer as the sample. Differences in heat capacity between the contents of the two cells yields thermodynamic data on the sample of interest (Cooper, 2000). Scans were repeated to ascertain refolding ability of protein and experimental reproducibility.

2.6.2 Circular Dichroism (CD) spectroscopy of Alba

Recombinant and acetylated Alba were each dialysed in buffer appropriate for CD spectroscopy (10 mM Bis-Tris, 200 mM NaF, 0.5 mM EDTA) and further diluted to a 1 mg/ml concentration, the required concentration for near UV analysis. The aforementioned buffer was used as a blank. Protein solution was inserted into a cuvette with a 0.5 cm pathlength and spectra were recorded between 450 nm and 260 nm at a scan rate of 20 nm/min in a JASCO J-810 spectalpolarimeter using Spectra Manager Software. Far UV spectra were recorded between 260 nm and 180 nm at a scan rate of 50 nm/min using a protein concentration of 0.5 mg/ml in a circular cuvette with a 0.02 cm pathlength. Each sample was scanned three times for average values and the blank was subtracted. The data were submitted to the Dichroweb (Lobley, 2002; Whitmore, 2004) server for analysis (http://www.cryst.bbk.ac.uk/cdweb/html/home.html).

2.7 Crystallisation

2.7.1 Crystallisation of ssArd1

*ss*Ard1 was concentrated to a maximum of ~9 mg/ml (in 20 mM Tris-HCl, pH 8.5, 200 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT buffer) and snap frozen before

storing at - 80 °C to prevent precipitation. A pre-crystallisation test was performed to test the optimal concentration of *ss*Ard1 to use in subsequent trials. 96 well sitting drop plates were used for crystal trials, where drops were set up in a total volume of 4 µl (2 µl protein : 2 µl reservoir solution) so that the drops did not dry out quickly and allowed time for vapour diffusion. The reservoir volume was usually set to ~180 µl. Trials were set up with Hampton research: crystal screens I and II, Pegion, PEG/LiCl grid screen, cryo, Natrix, Lite, Ammonium sulphate grid screen, Index and Salt-RX. Additional screens used were Emerald biosciences: screens I and II, Wizard screens I and II (decode genetics), and jMAC (developed in-house kit). Screens were set up at 4 °C and 19 °C under conditions that allowed minimum variation in temperature and movement. All screens were repeated with a mix of Acetyl-CoA to *ss*Ard1 in a ratio of 1:1 and also 2:1 as there may be as many as two molecules required per protein dimer to stabilise it due to predicted structural comparisons from the 3D PSSM program.

Initial crystals were obtained with Hampton Pegion solution 39 (0.2 M NaH_2PO_4, 20 % (w/v) PEG3350, final pH 4.5), and a fine grid screen was set up to optimise crystal growth (see table 6.1 in chapter 6 for details of specific solution constituents). Each optimised solution was set at pH of 4.25, as well as 4.7 and 5.2. All conditions were set up (1:1) using 1.5 µl of *ss*Ard1 protein solution at 5.4 mg/ml, 6.9 mg/ml and 7.17 mg/ml *ss*Ard1 with (2X and 1X) and without acetyl coenzyme A. Further plates were set up using these solutions with Hampton Research additive screens I, II and III. Oil (a mix of 50:50 paraffin oil with al's oil) was also dispersed as a layer over the top of the reservoir's to allow slower vapour diffusion, to produce more robust crystals. Trials that gave rise to crystals were repeated with addition of 2.5-5 % ethanol, 2.5 % (v/v) PEG 400/1500 or 1 mM pentaglycine, additives that increase crystal size and also aid the growth of 2D crystals in the 3rd dimension. All

other solutions that yielded crystals with potential (i.e. size and morphology) for X-ray diffraction were optimised in the same way.

2.7.2 Diffraction and data analysis

Potential protein crystals that were plentiful in number were stained with izit dye to distinguish them from common salt crystals. Protein crystals were cryo-protected in their reservoir solution with addition of 20 % glycerol, MPD, or PEG400 for a period of 2 minutes. These crystals were then mounted on the X-ray machine (MSC/Rigaku MicroMax-007 rotating anode X-ray generator with Osmic optics, set up to serve two image plate detectors, a HTC detector with X-Stream cryohead on the right port, and an R-Axis IV++ with X-Stream cryohead on the left) and placed in the beam path. Capillary mounting of the crystal was also performed to exclude the cryo-protecting step. Initial data was collected at 0° and 90° to ensure the crystal diffracted at all angles. Data was then collected with a crystal to detector distance of 200 mm and a 1° rotation of the crystal for each image. The diffraction pattern was analysed by crystallographic software MOSFLM (Leslie, 1992). Crystals were also taken to the ESRF synchrotron (beamline 14.1) to achieve better resolution.

2.7.3 CD spectrometry on ssArd1

*ss*Ard1 was diluted 1/10 to a final concentration of ~1.3 mg/ml in 10 mM Na-phosphate buffer, pH 7, 200 mM Na-fluoride, 0.5 mM EDTA. The sample was then dialysed into the same buffer to remove excess Tris-HCl and NaCl. The aforementioned buffer was used as a blank. Protein solution was inserted into a cuvette with a 0.5 cm pathlength and spectra were recorded between 450 nm and 260

nm at a scan rate of 20 nm/min in a JASCO J-810 spectalpolarimeter using Spectra Manager Software. Far UV spectra were recorded between 260 nm and 180 nm at a scan rate of 50 nm/min using a protein concentration of 0.5 mg/ml in a circular cuvette with a 0.02 cm pathlength. Each sample was scanned three times for average values and the blank was subtracted. The data were submitted to the Dichroweb (Lobley, 2002, Whitmore, 2004) server for analysis (http://www.cryst.bbk.ac.uk/cdweb/html/home.html).

2.7.4 Seleno-methionine incorporation

3 ml of *ss*Ard1 overnight culture was gently centrifuged and resuspended in 3 ml of warmed M9 media.

M9 media preparation: 1L of autoclaved dH_2O containing 15 mM NH_4Cl, 20 mM KH_2PO_4, 20 mM $Na_2HPO_4.7H_2O$. pH was adjusted to 7.4 with 5 M NaOH. 100 ml of filtered (non-autoclaved) solution containing 20 % glucose, 0.3 %$MgSO_4$, 10 mg $Fe_2(SO_4)_3$ and 10 mg thiamine was added to this. Finally, 35 μg/ml kanamycin and 50 mg L-Se-Met was added.

This was aliquoted into 4X 250 ml volumes of M9 in 1L flasks to allow for plentiful aeration and cultures were allowed to grow at 37 °C until A_{595} reached 0.3. Amino acids were then added to induce SeMet-*ss*Ard1 protein production. Cells were lysed by usual method (section 2.3). Lysate was analysed by SDS-PAGE and the *ss*Ard1 band was excised for mass spectrometry analysis for confirmation of seleno-methionine incorporation.

2.7.5 ssArd1 and Alba in complex

Using a ratio of 4.3 μM: 2.14 μM: 1μM, a mix of *ss*Ard1:AcCoA:Alba was produced and sent to St. Andrews University SPoRT laboratory, where 100 nL drops of protein mix and crystal screen were dispensed by robotic arm and incubated at 19 °C. These samples were digitally photographed periodically for signs of nucleation and/or crystal growth.

2.7.6 Molecular replacement

The 3DPSSM/Phyre programme (Kelley, 2000) uses threading to predict homologous structures based on the protein sequence input. For *ss*Ard1, this programme recommended several acetyltransferase proteins that were overlaid using SSM (Krissinel, 2004) as implemented in the MSD fold website at EBI. The overlaid structures suggested a common structural core to the acetylase that was used as a search model in molecular replacement with the program PHASER (McCoy, 2007).

Chapter 3: Protein expression and Purification

3.1 Introduction

In the first instance, the *Sulfolobus Solfataricus* genome was scanned to identify possible acetyltransferase encoding genes (section 1.7). Sequence analysis placed the focus of attention onto two specific genes, *Sso0082* and *Sso0209*. From multiple sequence alignments, it appears that *Sso*0082 has sequence similarity to bacterial acetyltransferase enzymes e.g. RimI. *Sso*0209 has eukaryotic N-acetyltransferase homologues as well as archaeal acetyltransferase and N-acetyltransferase homologues and most notably its closest match with a protein of known function in the human proteome is (*h*)Ard1. Therefore, we have renamed *Sso*0209 as *ss*Ard1 (as discussed in section 1.7). This chapter describes the cloning of these two genes, and also *Sso*Sir2, the deacetylase, and the subsequent expression and purification of proteins from the successful clones. The gene products were then investigated through a series of assays to establish their capacity (if at all) to acetylate Alba and to characterise them further. In addition to the putative acetyltransferase enzymes, assays with the known deacetylase (*Sso*Sir2) were also desirable to affirm whether it reversed the acetylation process carried out by these particular enzymes. *Sso*Sir2 deacetylates the K16 residue of the Alba protein (Bell 2002), allowing it to bind DNA with an approximate three to four fold higher affinity than the acetylated version *in vitro* (Jelinska, 2005).

3.2 Sso0082

The complete *Sso0082* gene region was amplified successfully by gradient PCR (figure 3.1) using the following primers:

5’ primer: 5’- CGTCGGATCCCCATGGTAATTATAACTGATGCAACAG-3’

3’ primer: 3’-CCGGGGATCCGTCGACTTAATCCAGATATTTATAATATAAATTC-3’

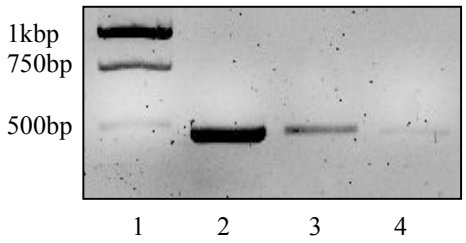


Figure 3.1: Gradient PCR on *Sso0082* at different annealing temperatures.

Lane 1 = size markers, 2 = 50 °C, 3 = 57 °C and 4 = 65 °C.

An annealing temperature of 50 °C produced the highest yield of specific *Sso0082* product. Restriction digests allowed the fragment to be inserted into the *Bam*HI/*Nco*I site of pET28c, the *Nco*I/*Sal*I site of pLOU3, the *Bam*HI/*Sal*I site of pGEX-5X-3 and the *Nco*I/*Bam*HI site of pHisTev30a. Competent *E.coli* Top 10 cells were transformed with each of the above constructs and colonies were recovered for all constructs except pHisTev30a. Repetition of digests, ligation and transformation yielded a small number of colonies for the pHisTev30a construct from which 3 were chosen for colony PCR. PCR on all samples including controls failed each time and hence, the plasmid was purified and checked by restriction digests. A plasmid of the expected size (plasmid + insert) was indeed observed on the agarose gel for some transformants, whilst PCR on the Sso0082 constructs showed a band of ~ 453 bp for *Sso0082* from 2 of the samples (results not shown). The *E.coli* strain Rosetta BL21 was transformed with the correct construct and colonies for all constructs were recovered for expression studies.

3.3 ssArd1

The *ssArd1* gene (501bp) was amplified successfully by gradient PCR (figure 3.2) using the following primers:

5' primer: 5'- CGTCGGATCCCCATGGAGCTCGCTGAAAAAGATAAAGG-3'

3' primer: 5'- CCGGGGATCCGTCGACTTAGAGTGGTCTTGCCATAAGATATGC-3'

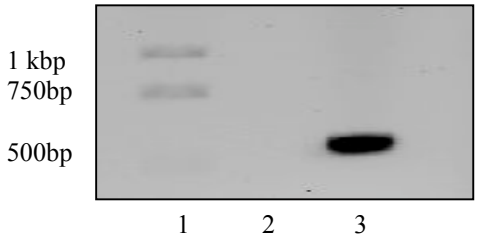


Figure 3.2: PCR on *ssArd1* at different annealing temperatures.

Lane 1 = size markers, 2= 62 °C and 3 = 65 °C

An annealing temperature of 65 °C produced the highest yield of specific *ssArd1*. The amplified product was cloned into the *Nco*I/*Bam*HI sites of pET28c and the *Nco*I/*Bam*HI sites of the pHisTev30a vector. In addition, *ssArd1* was inserted into the pGEX –5X-3 vector to give a Glutathione-S-transferase (GST) tagged version of the protein for purification via GST pull-down assay. Each construct was transformed into Top10 and Rosetta cells. After plasmid purification from Top10 colonies, PCR on pure pGEX-5X-3/*ssArd1* and pET28c/*ssArd1* constructs confirmed that gene insertion was successful (figure 3.3a). However, the very low signal intensity of the PCR results obtained from the pGEX construct may be due to cross contamination during the reaction and therefore, does not confirm the existence of the gene. Restriction enzyme digests to verify these findings only returned a positive result for the insertion in pET28c and not for pGEX-5X-3 (figure 3.3b). Mini inductions of Rosetta cultures were carried out before sequencing in order to identify positive colonies that produced a protein band of the expected size on SDS-PAGE gel.

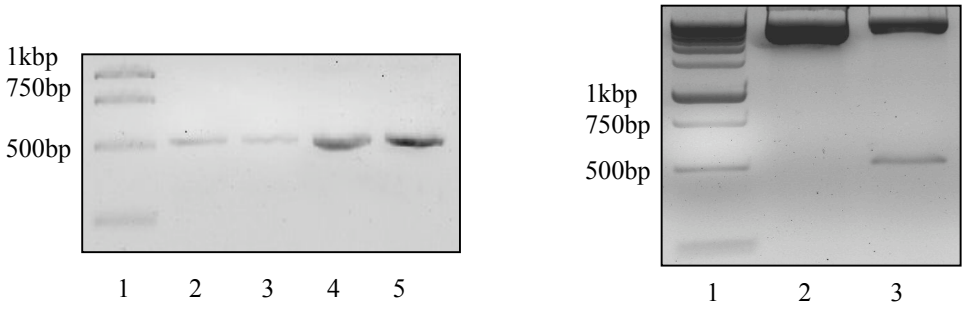


Figure 3.3a: PCR of pGEX-5X-3/*ssArd1* and pET28c/*ssArd1*. 1 = size markers, 2 and 3 = pGEX-5X-3 PCR (duplicate) and lanes 4 and 5 = pET28c PCR (duplicate).

Figure 3.3b: Digests on pGEX-5X-3/*ssArd1* and pET28c/*ssArd1*. 1 = size markers, 2 = restriction digest of pGEX-5X-3, 3 = restriction digest of pET28c construct.

3.4 SsoSir2

The *SsoSir2* gene was amplified by gradient PCR (figure 3.4) using the following primers:

5' primer: 5'- CGTCGGATCCCCATGGTATACGAGAAAGTAGCTGAAG-3'

3' primer: 5'- CCGGGGATCCGTCGACTTACGAAAGTATTTTCTGTCTCACG-3'

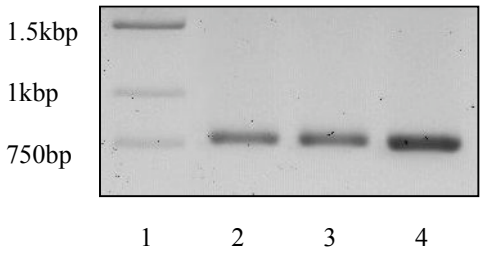


Figure 3.4: PCR on *SsoSir2* at different temperatures.

1 = size markers, 2 = 50 °C, 3 = 55 °C and 4 = 60 °C

An annealing temperature of 60 °C produced the highest yield of specific *SsoSir2* product. The PCR fragment was cloned into the *Nco*I/*Bam*HI sites of pET28c and the resulting construct was used to transform *E.coli* Top10 cells for plasmid production. However, sequencing results for vector/insert showed that the gene cloned was not

SsoSir2 but in fact *SsoPCNA1*. The procedure was repeated twice with many controls but the same results were obtained. The primer sequences were analysed against those required for *SsoPCNA1* amplification and no resemblance was found. *SsoPCNA1* was under study in our laboratory at the same time and, thus, the likely explanation is that the competent cell stock was contaminated with a construct containing this gene insertion. This highlights the importance of maintaining stringent aseptic techniques in the laboratory environment at all times. Further attempts to clone *SsoSir2* were put on hold.

*3.5 Purification of Sso*0082

The vector, pHisTev30a, contains a poly histidine tag of 6 consecutive residues at the insert site that enable expression of a given protein attached to a (6X) his-tag at the N-terminus. This tag allows one-step purification of the protein to take place on a nickel nitrilotriacetic acid (Ni-NTA) resin. The tag can then be cleaved off and separated from the desired protein. Mini inductions of 3 ml cultures containing pHisTev30a/*Sso*0082 gave insoluble protein, which was initially suspected due to larger pellet size and proven to be the case after SDS-PAGE analysis of supernatants (not shown) and pellets (figure 3.5a). After many efforts to solubilise the protein e.g. alteration of buffer constituents to include either glycerol, triton or 2 % (v/v) ethanol and/or greater salt concentration, changing IPTG concentration and/or the OD at which induction was instigated, growing the culture at 25 °C or using heat shock/cold shock on the cells, the protein was still insoluble.

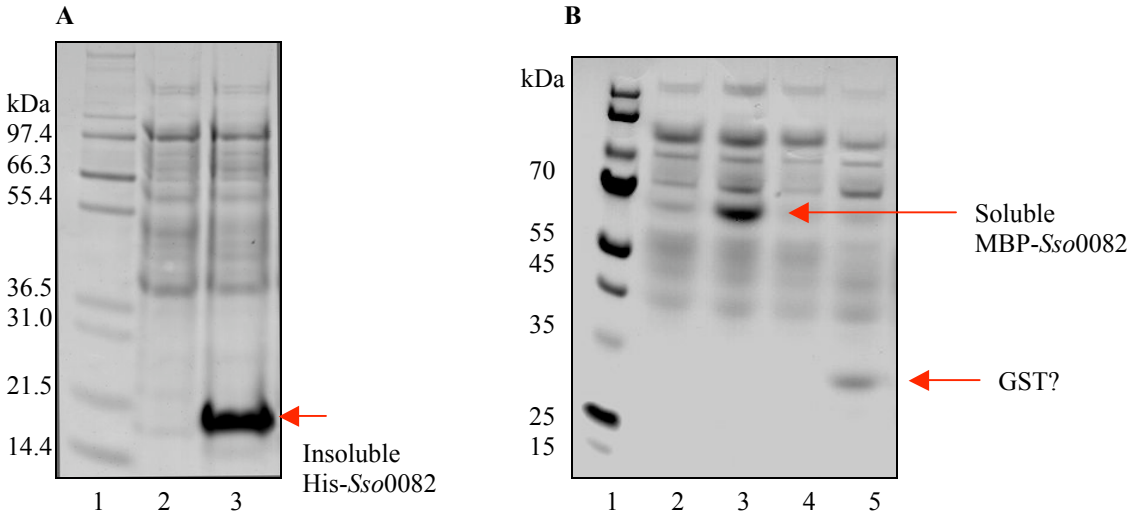


Figure 3.5a: Expression of *Sso*0082 from pHisTev30a. Lane 1 = size markers, 2 = pellet from uninduced control, 3 = pellet from induced control showing insoluble proteins.

Figure 3.5b: Expression of MBP-*Sso*0082 fusion and *Sso*0082-GST fusion proteins. Lane 1 = size markers, 2 = pLOU3 control, 4 = induced pLOU3, 3 = pGEX-5X-3 control, 4 = induced pGEX5X-3.

Mini expression of MBP-*Sso*0082-fusion protein from pLOU3 (figure 3.5b) yielded a small amount of soluble protein at approximately 60 kDa but scaling up of the culture to 1L resulted in insoluble protein. All efforts to maintain it in the soluble fraction failed. Despite DNA sequencing proving that the gene was present in the pGEX-5X-3 vector, no soluble fusion of ~ 43.6 kDa was found. A faint band can be observed at ~26 kDa, which may be GST expressing on its own (figure 3.5b). Since *Sso*0082 has solubility problems when expressed on its own *in vitro*, there is a high probability that it exists in complex with a binding partner (a possible homologue of the Nat type proteins found in the eukaryotes) *in vivo* or alternatively it may be membrane associated. At this point, it was decided to terminate the work with *Sso*0082.

*3.6 Purification of ss*Ard1

Mini induction of Rosetta cultures from colonies containing the pET28c/*ss*Ard1 construct (this vector contains a histidine tag but it is not utilised in this instance) indicated reasonable solubility for recombinant *ss*Ard1 (figure 3.6).

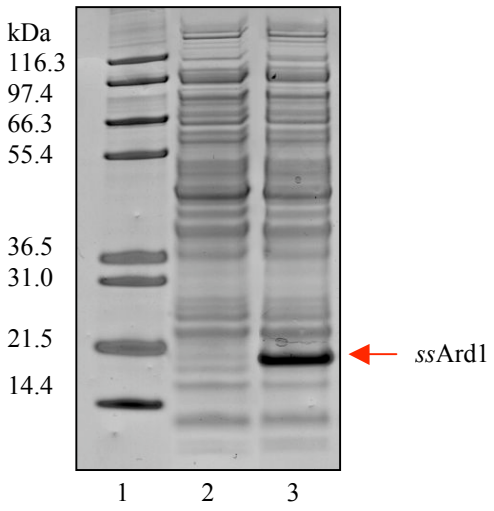


Figure 3.6: Expression of native *ss*Ard1 from pET28c.

Lane 1 = size markers, 2 = uninduced control, 3 = supernatant from induced culture.

A 6 L scale up culture was successful in producing a similar level of soluble protein. Heat treatment at 65 °C for 20 minutes eliminated the majority of *E.coli* proteins (figure 3.7a). Ion exchange chromatography was then utilized as a second purification step, on the basis that buffering the protein to an appropriate pH to give it an overall positive or negative charge will allow it to bind to a resin of opposite charge. *ss*Ard1 has a theoretical isoelectric point (pI) of ~7.6 calculated from the pK values of its amino acid content (http://www.expasy.org/tools/protparam.html). This calculation is based on the analysis of polypeptide migration in an immobilised gel environment between pH 4.5 and pH 7.3 (Bjellqvist *et al*, 1993; Bjellqvist *et al*, 1994). Based on the pI value, the protein was buffered to pH 8.5 to induce an overall negative charge on the molecule but it did not bind to the anion exchange resin (figure 3.7a). Buffering the protein to pH 9 made no difference to the lack of binding.

Likewise, cation exchange on the heparin column at pH 6.0 did not cause *ss*Ard1 to bind either and the majority eluted again in the flow through (figure 3.7b). Further buffer alterations did not make any difference to the binding affinity for either resin. The discrepancy between predicted behaviour and actual behaviour of the protein on the columns could be due to several factors including weak buffer capacity since *ss*Ard1 is a fairly small molecule, post-translational modification or blockage of the N-terminus, and the protein morphology i.e. existence as monomer and/or dimer, all of which would heavily influence the isolelectric point.

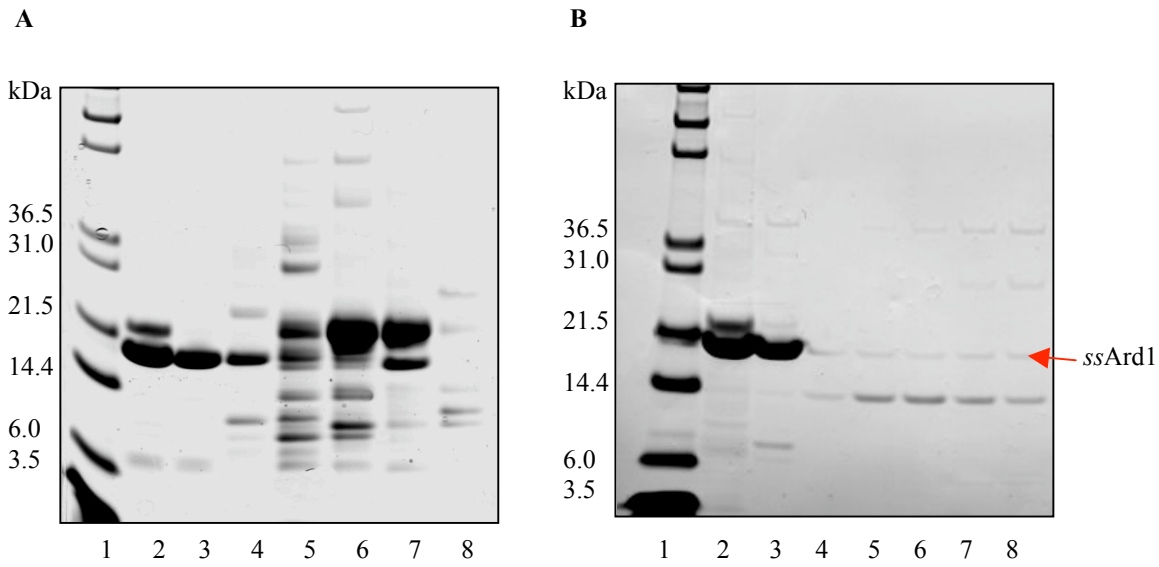


Figure 3.7a: Anion exchange on *ss*Ard1 using a 5 ml Q-sepharose Hi-Trap column at pH 8.5. Lane 1 = marker, 2 = sample after heat, 3 = column flow through, 4-8 = column fractions representing each peak.

Figure 3.7b: Cation exchange on *ss*Ard1 using a 5 ml heparin Hi-Trap column at pH 6. Lane 1 = marker, 2 = sample after Q-sepharose, 3 = flow through, 4-8 = column fractions representing each peak.

Ammonium sulphate (80 % w/v) precipitation was used to reduce the volume of the flow through (unbound protein solution) from the heparin column. The volume was decreased from 660 ml to approximately 45 ml, therefore concentrating the

sample as well as purifying it further (figure 3.8a). Overnight dialysis into 20 mM BisTris, pH 6, 300 mM NaCl and 1 mM EDTA was followed by gel filtration chromatography (HiLoad® 26/60 Superdex® column), separating proteins on the basis of their size. The column was first calibrated with protein standards of known molecular weight. The *ss*Ard1 absorbance trace showed two main peaks at 280 nm (figure 3.8b). The first peak eluted at a volume of 208 ml, which corresponds to a molecular weight of ~40 kDa. *ss*Ard1 is present here as a dimer but the majority of the protein elutes at 236 ml in monomeric form (~19.5 kDa) along with some persistent minor contaminants (figure 3.8c). To separate *ss*Ard1 from the contaminants, SP-resource and hydroxyapatite (ion exchange) resins were used but yet again *ss*Ard1 did not bind to either column although most of the contaminants did. The protein was now pure enough to be used for biochemical and structural analysis.

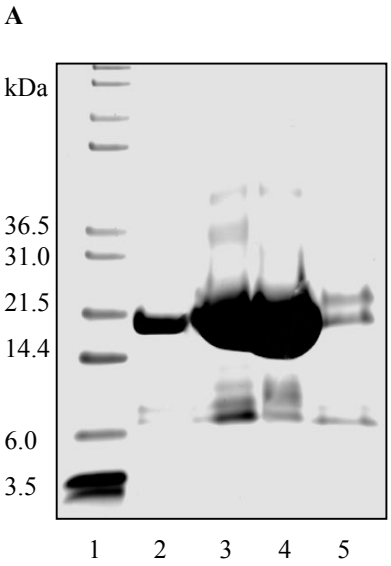


Figure 3.8a: Ammonium sulphate precipitation (80 % w/v) of *ss*Ard1 supernatant.

Lane 1 = size markers, 2 = dilute flow through from 5 ml heparin column (i.e. before procedure), 3 & 4 = resuspended precipitate from two different protein samples, 5 = supernatant containing left over proteins that did not precipitate.

B

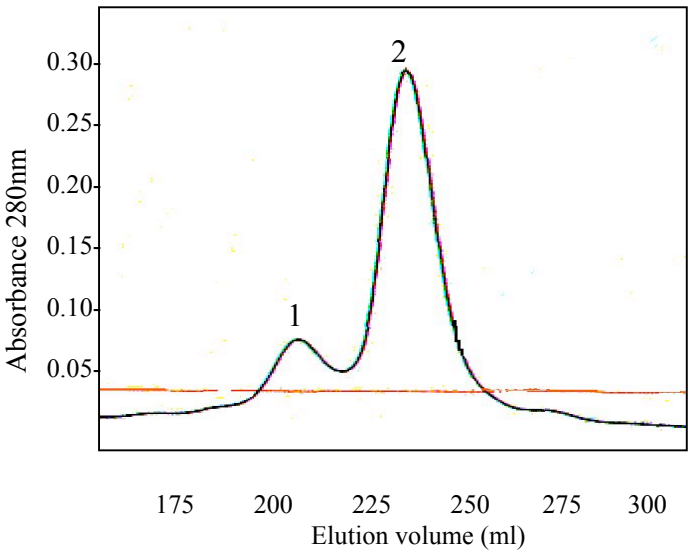


Figure 3.8b: Typical gel filtration (HiLoad® 26/60 Superdex® column) trace for *ss*Ard1.

Peaks correspond to dimeric and monomeric forms of the protein. Peak 1 is ~ 40 kDa and peak 2 is ~ 20 kDa.

C

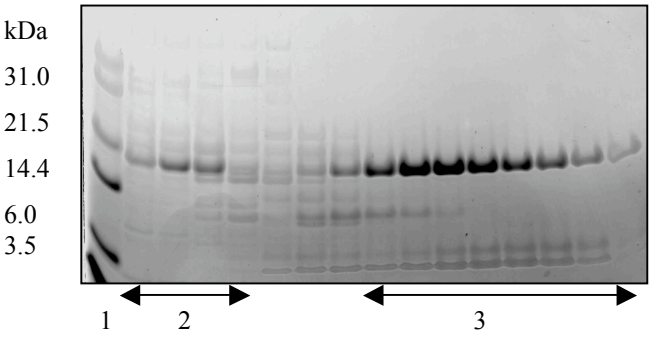


Figure 3.8c: SDS-PAGE gel showing *ss*Ard1 gel filtration fractions.

Fractions are from peaks 1 & 2 of figure 3.8b; 1 = size markers, 2 = fractions covering peak 1 and 3 = fractions covering peak 2.

In order to ease the purification, attempts were made to create a tagged version of the protein as discussed in section 3.3. The three new constructs provided potential fusion proteins: (6X) His-*ss*Ard1, *ss*Ard1-GST and MBP-*ss*Ard1. Mini expression trials yielded a band of ~ 22 kDa for (6X) His-*ss*Ard1. This was identified by mass spectrometry as *ss*Ard1 and scaled up for greater protein production. The tagged protein was subsequently purified on a 5 ml IMAC column (figure 3.9a) and incubated with TEV protease to cleave off the His-tag. Sample was compared on a gel before (it seems that the tag has degraded slightly as a small amount of untagged

*ss*Ard1 appears here) and after cleavage but although a fainter band appeared at about 22 kDa showing that the protease had worked, no increase in the amount of untagged protein at around 19.5 kDa was observed (figure 3.9b). After centrifugation, it was clear that protein had precipitated. Sequencing results for the His-*ss*Ard1 construct suggested that His-*ss*Ard1 was not present in the construct.

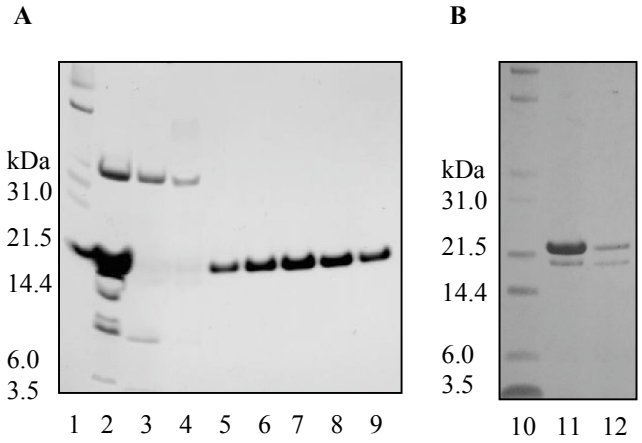


Figure 3.9a: Expression of His-*ss*Ard1 and purification on a 5ml IMAC column. Lanes: 1 = size markers. 2 = crude sample after heating, 3 = column flow through, 4 = 1st small peak, 5 to 9 = *ss*Ard1 peak.

Figure 3.9b: Before and after treatment with TEV protease. Lanes: 1 = size markers, 2 = pre cut his-*ss*Ard1, 3 = post cut his-*ss*Ard1

His-tagged *ss*Ard1 isolated from the final column (gel filtration) was prepared for analysis by mass spectrometry, however, the protein precipitated prior to analysis. Further attempts to create a (6X) His tagged version failed. Despite sequence results confirming the presence of the gene in the pGEX-5X-3 vector, GST was found to express on its own with no trace of *ss*Ard1. The pLOU3 vector did yield a high amount of fusion protein (results not shown) but the enzyme responsible for cleaving the tag off was prohibitively expensive and since native expression in pET28c did give a reasonable amount of *ss*Ard1 expression, it was resolved to continue working with this construct and optimise the purification conditions. This involved increasing the heat treatment to 25 minutes, then slowly pumping small volumes of the filtered supernatant through Q-sepharose and heparin columns using buffers of varying pH. Buffer containing 20 mM Tris, pH 8.5, 50 mM NaCl, 1 mM EDTA and 1 mM DTT

worked well for both columns. The flow through was collected and the columns were washed with 1 M salt buffer. Almost all *ss*Ard1 eluted in the flow through and most of the contaminating proteins were excluded. The protein was concentrated by ammonium sulphate precipitation as before and the final purification step on the gel filtration column yielded pure protein (figure 3.10).

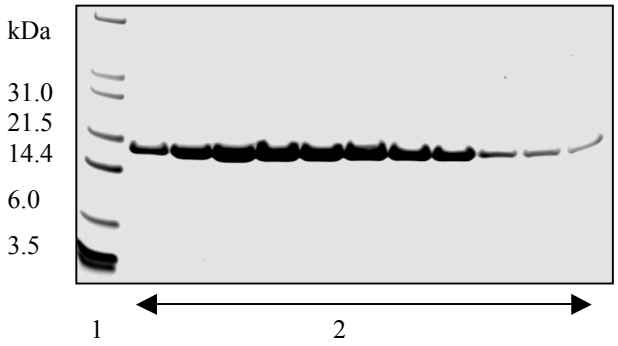


Figure 3.10: SDS-PAGE gel of pure *ss*Ard1 from gel filtration chromatography (HiLoad® 26/60 Superdex® column). Pure *ss*Ard1 eluted from gel filtration over one peak. 1 = size marker, 2 = fractions collected for monomeric peak of pure *ss*Ard1 eluted at 225-250 ml volume from size exclusion chromatography.

*3.7 Concentrating pure ss*Ard1

Almost all *ss*Ard1 in any given sample was lost due to sticking to the membrane of the low protein binding vivaspin concentrator. A method was therefore devised where the sample was placed in dialysis tubing and saturation of the tubing with high MW PEG enabled slow diffusion of buffer out of the tubing. Concentrations of between 5-13 mg/ml were achieved by this method but the protein tended to precipitate. The concentrated sample was immediately frozen in liquid nitrogen and storage at –80 °C until required.

Research into *h*Ard1 shows that the pure protein forms protofilaments (which can resolubilise) with a large β-sheet content under physiological pH and temperature and that this process is accelerated by thermal denaturation as well as high protein concentration (Sanchez-Puig, 2006). This could be largely due to the flexible unstructured C-terminal tail of *h*Ard1. *ss*Ard1 does not contain a similar

flexible C-terminal region. Obviously, *ss*Ard1 is not thermally denatured and the physiological conditions are very different to those of *h*Ard1 but it seems that high protein concentration causes *ss*Ard1 to form aggregates as well, which do not go back into the soluble state even on gentle heating.

3.8 Circular Dichroism

Circular Dichroism (CD) spectroscopy is a measure of the differential absorption (usually in UV range) of right- and left- circularly polarised light. The far UV spectra (between 180 and 220 nm) is dominated by electronic transmissions of the protein backbone and typical signals are listed in table 3.1.

α-Helix	Negative signal at 222 nm Negative and positive signals at 208 and 190 nm respectively
β-Sheet	Negative signal at ~215 nm Positive signal at ~198 nm

Table 3.1: Secondary structure signals relating to α–helix and β–sheet conformation of *ss*Ard1.

The CDSSTR program from CDPro (http://lamar.colostate.edu/~sreeram/CDPro/main.html) was used to generate the spectrum (figure 3.11) showing the absorbance of polarised light by *ss*Ard1 and to then calculate the secondary structure (table 3.2) of the protein based on this signal information. CDSSTR has estimated that *ss*Ard1 contains 44 % (56 % including distorted helix) α-helical structure and 9 % (15 % including distorted sheet) β-sheet structure.

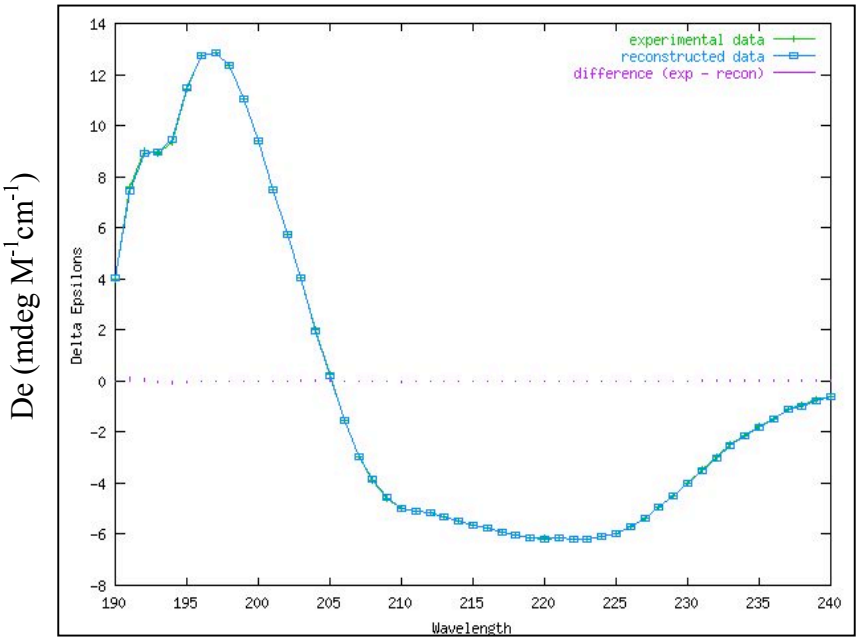


Figure 3.11: Far UV CD spectra of *ss*Ard1. NRMSD: 0.006. Helix segments per 100 residues: 2.981, Strand segments per 100 residues: 3.347. Average helix length per segment: 18.694, Ave strand length per segment: 4.816.n4.98 helix = (4.98 * 18.694) 93.1 residues (55.75 %), 5.59 sheet= (5.59 * 4.816) 26.9 residues (16 %).

(α) Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.44	0.12	0.09	0.07	0.13	0.15	1

Table 3.2: Proportion of protein (multiply by 100 for %) estimated to have a certain structure. Helix1 is standard α-helix, Helix2 is distorted α-helix, Strand1 is standard β-sheet and strand2 is distorted β-sheet (as determined in Sreerama, 1999).

According to the CD data reported for *h*Ard, it contains a random coil content of 44 % (Sanchez-Puig, 2006), which is similar to the ~37 % predicted (by PHYRE) random coil content of *ss*Ard1. However, CD spectra only identifies ~15 % of the *ss*Ard1 structure as unordered but if the turns are included, in a manner similar to PHYRE predictions, the value rises to a more agreeable ~28 %. The *h*Ard1-ΔC (residues 1-178) version reportedly shows a random coil content of 40 % (CD signals for *h*Ard1 and *h*Ard1-ΔC are shown for comparison in figure 3.12), which seems rather high considering the removal of more than 80 residues, 82 % of which are estimated to exist as unordered by the JNET server (Cuff & Barton, 1999).

Regardless, this truncated version is a better comparison to *ss*Ard1 because it does not contain a flexible C-terminal region. The spectra of *h*Ard1-C and *ss*Ard1 show a high α-helical content and they superimpose with only slight variation between them (after unit conversion) thus giving evidence of a very similar globular structure.

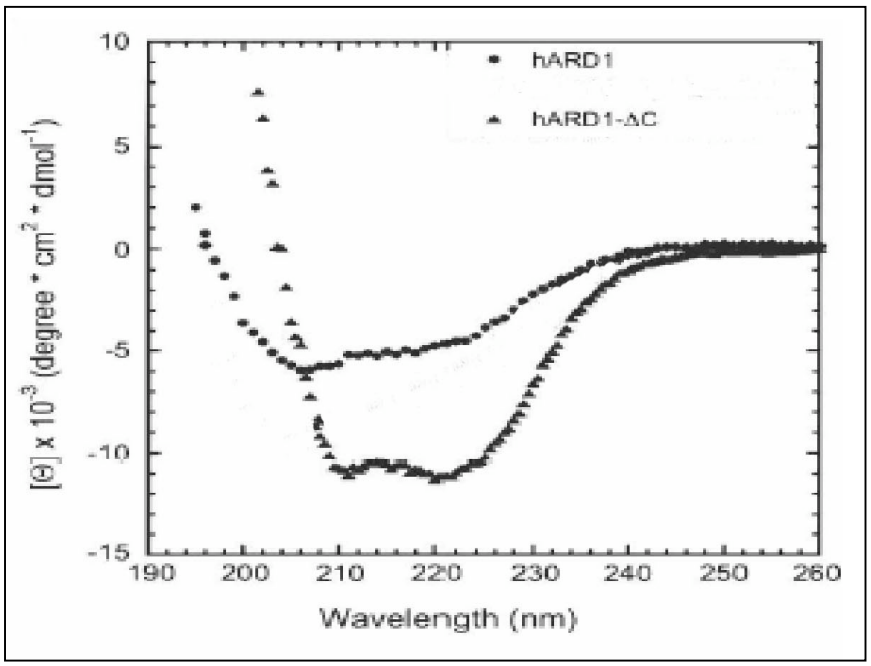


Figure 3.12: Far UV CD spectra for *h*Ard1 and *h*Ard1-ΔC. (adapted from Sanchez-Puig, 2006). Note: conversion of these units to Delta Epsilon (De), the units used in figure 3.11, is achieved by the following formula: De = θ / 3298.

3.9 Does ss*Ard1 have a binding partner in vivo?*

The existence of a complex in higher organisms allows speculation as to whether *ss*Ard1 can act alone or if it absolutely requires a binding partner in order to function. Another option is that the protein may have a binding partner to enhance its acetyltransferase activity. Sequence analysis shows that *ss*Ard1 is smaller in comparison to *h*Ard1 and *sc*Ard1 since it does not contain the extended C-terminal region (which is often less structured) of the eukaryotic homologues. It was not

possible to make a knock out of *ss*Ard1 in *S. solfataricus*. This would have been the best way to evaluate the enzyme's target substrate(s). A GST-pulldown assay was not feasible either since the enzyme could not be expressed with the N-terminal tag (see section 3.5), therefore *ss*Ard1 was bound to an affi-gel column matrix and *S. solfataricus* cell lysate was applied to the column. Unbound proteins were washed off and bound proteins were eluted via a number of high molarity salt washes. Fractions were tested on SDS-PAGE gel and sypro-ruby stained for increased visualisation. Any bound protein (either binding partner or substrate) may be identified. Although *ss*Ard1 exhibits acetyltransferase activity on it's own, it is indeed possible that *ss*Ard1 may have a specific binding partner e.g. a Nat1 homologue, which could enhance it's catalytic efficiency.

A large number of *Sulfolobus* proteins were eluted in the wash stage of the purification, however, many of these non-specifically bound proteins were still present in the higher salt washes (figure 3.13a). It is unlikely that *ss*Ard1 interacts with all of these and it could be the case that this method is just not efficient i.e. native *ss*Ard1 would have been present in the cell lysate and any binding partners would likely be already interacting with this free form of the enzyme, leaving fewer molecules available to bind the column, thus making detection difficult. In addition, any likely substrates of *ss*Ard1 have already been modified in the *Sulfolobus* cell because we could not knock the gene out. There were no suspect bands to indicate that *ss*Ard1 is part of a complex in the archaea as it is in eukaryotes and similarity searches for a Nat1 homologue in the *S. solfataricus* P2 genome failed to identify a protein with significant homology over the 854 residue sequence, although a hypothetical *S. solfataricus* protein (gi|15899426|ref|NP_344031.1 *Sso*2710) with 22 % identity was found corresponding to a 245 amino acid segment of *sc*Nat1. Western blot analysis using antibodies against Alba1 and Alba2 showed possible

binding of both proteins (figure 3.13b). A control experiment using an affigel column in the absence of *ss*Ard1 did not produce a similar response. In the presence of an active *ss*Ard1, the majority of Alba1 and/or Alba2 is probably already acetylated.

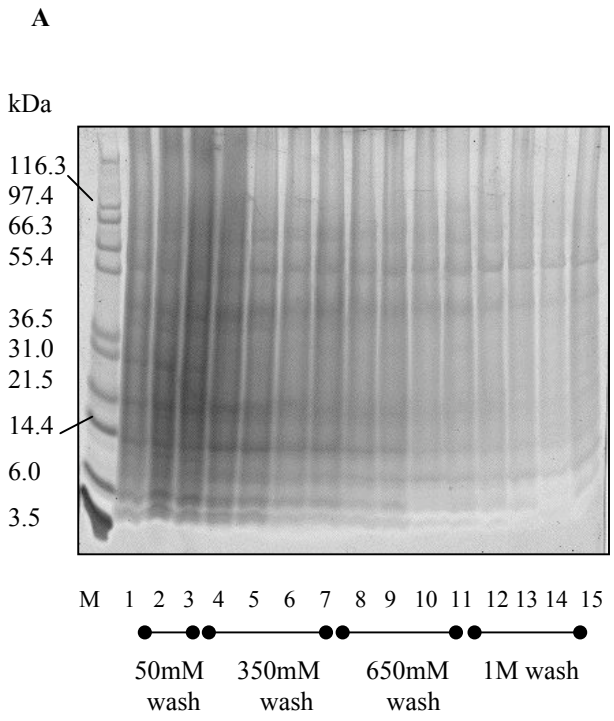


Figure 3.13a: Sypro-Ruby Stained SDS-PAGE gel of affigel 10 cell lysate fractions.

M = size marker, 1 = *Sso* lysate before addition to column.

All even numbers show fractions obtained from specific salt washes (50 mM, 350 mM, 650 mM and 1M) of the control column. All remaining odd numbers show fractions obtained from the specific salt washes of the *ss*Ard1 column.

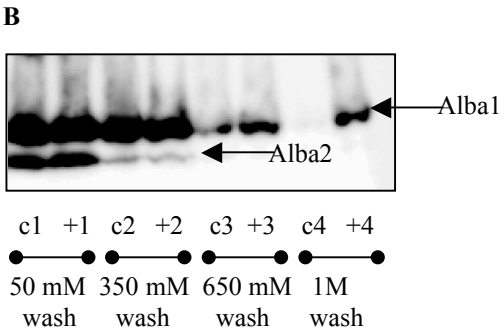


Figure 3.13b: Western blot (of SDS-PAGE gel from figure 3.13a) with antisera raised against Alba1 and Alba 2.

C = control column lysate, + = *ss*Ard1 column lysate, 1 = 50 mM salt wash, 2 = 300 mM salt wash, 3 = 600 mM salt wash, 4 = 1M salt wash. Alba1 and Alba2 are indicated by arrows.

3.10 Summary and conclusions

Despite trying a variety of conditions, *Sso*0082 could not be rendered soluble and work on this protein was terminated. This may be because it exists in complex

with another protein *in vivo*. Cloning of a (6X) histidine tagged version of *ss*Ard1 for simplified purification failed and the only viable route was to purify the recombinant protein. Purification was not simple because the protein failed to bind ion exchange resins under favourable conditions and eluted with the flow through. Therefore a five-step purification procedure was necessary. In brief, this involved heating the protein sample at 65 °C for 25 minutes, pumping lysate through a cation exchange column and then an anion exchange column, precipitating the protein by ammonium sulphate addition and finally subjecting it to size exclusion chromatography. There were difficulties with concentrating purified *ss*Ard1 as this aggregated readily, similar to the properties of *h*Ard1. Circular dichroism studies suggested that *ss*Ard1 contains a majority secondary structure of α-helices, accounting for 44 – 56 %, in agreement with *h*Ard1-ΔC. β-sheet content seems to be much lower at only 9-15 %. This value is rather low in comparison to the predicted ~ 30 % β-sheet content by the PHYRE programme (section 1.5). And finally, the investigation into the existence of a putative binding partner for *ss*Ard1 offered little revelation, probably because there isn't one but it would be naïve to completely rule it out. It may be worth investigating the small *Sso*2710 hypothetical protein in future.

Chapter 4: Acetylation of Alba1 by *ss*Ard1

4.1 Introduction

Native *Sso*Alba1 is 84 daltons larger than recombinant Alba1. This is due to the acetylation of two residues, serine 2 and lysine 16 (Bell, 2002). It has already been demonstrated that *ss*Pat is the acetylase responsible for modifying K16 but the N-terminal acetylase enzyme is still unknown. It has also been documented that NAT's are in some cases capable of internal acetylation as well i.e. N^{ε}- acetylation of specific lysine residues (Polevoda and Sherman, 2003). This brings us to question whether *Sso*Alba1 is a target substrate of *ss*Ard1 and if so, does the enzyme predominantly acetylate amino acids at the N^{α} or N^{ε} positions, or both? As a homologue of the yeast NatA subunit *sc*Ard1, we predict that *ss*Ard1 will also be an N-terminal acetylase which will acetylate sequences beginning with serine, alanine, glycine and threonine after removal of the initiator methionine. With an N-terminal sequence of SSGT, *Sso*Alba1 may be a target for *ss*Ard1. Acetylation largely depends on the N-terminal sequence (Polevoda and Sherman, 2003), several factors may affect the addition of the acetyl group. It is possible that other residues within the N-terminal region may decrease or prevent acetylation (Moerschell, 1990). Equally, the absence of certain residues may be responsible for non-modification of proteins displaying a viable target sequence (Polevoda & Norbeck, 1999).

4.2 Acetylation of *Sso*Alba1 by *ss*Ard1 and non-labelled AcCoA

Matrix Assisted Laser Desorption Ionisation – Time-of-Flight (MALDI-ToF) mass spectrometry analysis of recombinant wt *Sso*Alba expressed in *E.coli* showed that the protein was not modified, with exception to the N-terminal methionine, which was cleaved off during the expression phase. Recombinant Alba1 has an intact

mass of 10454 Da. Incubation of recombinant Alba with *ss*Ard1 in the absence of acetyl-CoA served as a negative control, also showing the same intact mass of 10454 Da. In the presence of a small amount of substrate, acetyl CoA, an increase in mass of approximately 42 Da larger was observed. This could be explained as the addition of aceto-nitrile adducts from the mass spectrometry process unless a small amount of acetyl-CoA had been co-purified with one of the proteins but this seems unlikely due to the rigorous purification steps. In hindsight, it would have been appropriate at this point to do a negative control with alba and acetyl-CoA to test for non-enzymatic acetylation. Non-enzymatic acetylation is a common occurrence in proteins with cysteine residues, the only amino acid that can act as an acetyl acceptor under conditions that lack the addition of an acetylase enzyme (Dormeyer, 2005), however, *Sso*Alba1 does not contain any cysteine residues. After addition of the *ss*Ard1, a 42 Da increase representing one acetyl group was evident for approximately 80 % of the substrate (see figure 4.1).

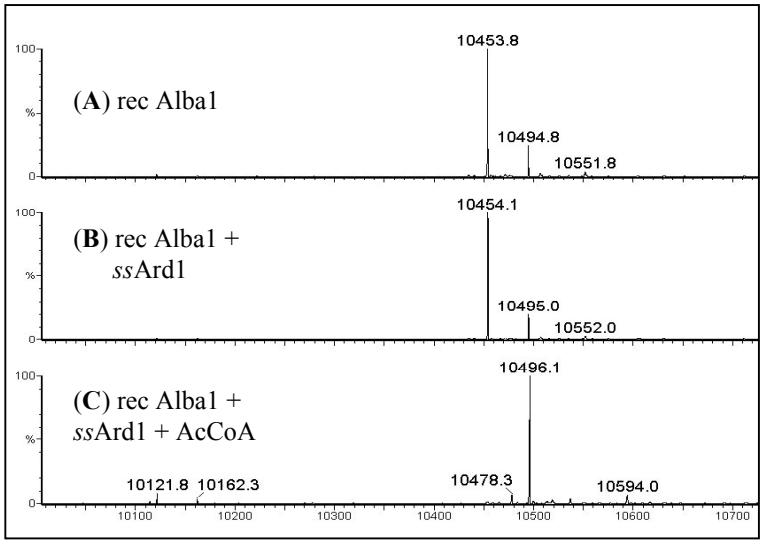


Figure 4.1a: MALDI-ToF Intact mass on *Sso*Alba1.

a] 20 µM recombinant wt Alba + acetylation buffer, b] 20 µM rec wt Alba + acetylation buffer + 500 nM *ss*Ard1, c] 20 µM recombinant wt Alba + acetylation buffer +500 nM *ss*Ard1 + acetyl CoA. Each Sample was incubated at 70 °C for 2 hours.

Although this procedure confirms the likely acetylation of Alba by *ss*Ard1 and gives an estimate of the amount of Alba modified, the peak height is not necessarily quantitative because the ionisation process is competitive and protein solutions from different purification batches will likely contain a small number of contaminants. The size and chemistry of these components will result in different extents of ionization. Incubations at higher temperatures or over longer time periods showed evidence of up to four acetyl group additions to the substrate. Tryptic digests (figure 4.2) showed a mixture of acetylated sites including the N-terminal serine, N-terminal serine and lysine 16, lysine 48 and lysine 64 or 68. Peptides where K64 and K68 appeared separately and unacetylated were evident but unfortunately the mass for peptides containing either acetylated residue were too small or too large to be seen. However, it appears that these residues may have been modified due to saturation of the initial site and in any case K48, K64 and K67. Native Alba is only acetylated on S2 and K16, therefore, acetylation of K48, K64 and K67 may only occur under assay conditions in which non-specific acetylation is promoted.

Peptides seen	Theoretical mass	Actual Mass
2-16(-)SSGTPTPSNVVLIGK(K)	1456.8	1498.8
2-17(-)SSGTPTPSNVVLIGKK(P)	1584.9	1668.9
45- 57 (R)AISKAVDTVEIVR(N)	1400.81	1442.8
60-71 (R)FLPDKIEIKEIR(V)	1500.88	1542.9

Table 4.1: Peptides produced by tryptic digestion of *Sso*Alba1. The theoretical mass (i.e. from unmodified protein) and actual mass (i.e. after acetylation) are shown for each peptide. Potentially modified residues in each peptide are shown in red.

Fragment 2-17 seems an unlikely occurrence because proline in position P1´ (i.e. residue following Lys), being a bulky residue, usually exerts a strong negative effect on peptide cleavage. Under certain circumstances it would be possible for trypsin to cleave at K17 e.g. when a lysine residue is present in the P2 position, as demonstrated by Keil, 1992 (figure 4.1b). This explains the occurrence of peptide 2-17 from samples where K16 is acetylated.

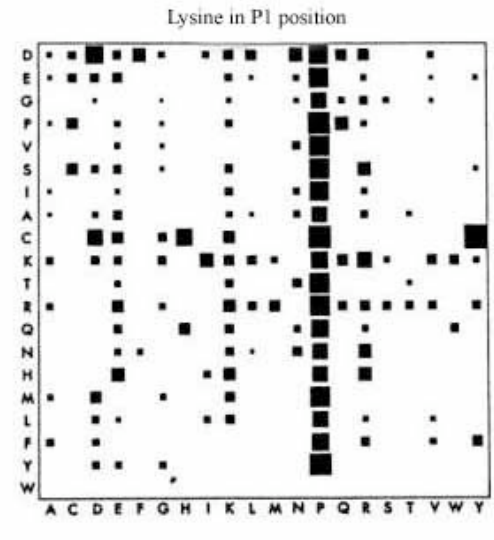


Figure 4.1b: Negative influence of residues in position P2 and P1´ on the cleavage of lysine at position P1 by trypsin. P2 shown on vertical scale, P1´ on horizontal scale. The area of the black square represent the percentage of inhibition in the corresponding cross-section for each tripeptide sequence P2-Lys-P1´ (according to Keil, 1992).

Analysis of fragments from a K16E mutated versions of Alba1 treated with *ss*Ard1 showed that the double acetylated fragment 2-17 did not occur. Only a single acetylated fragment was detected, presumably modified at S2.

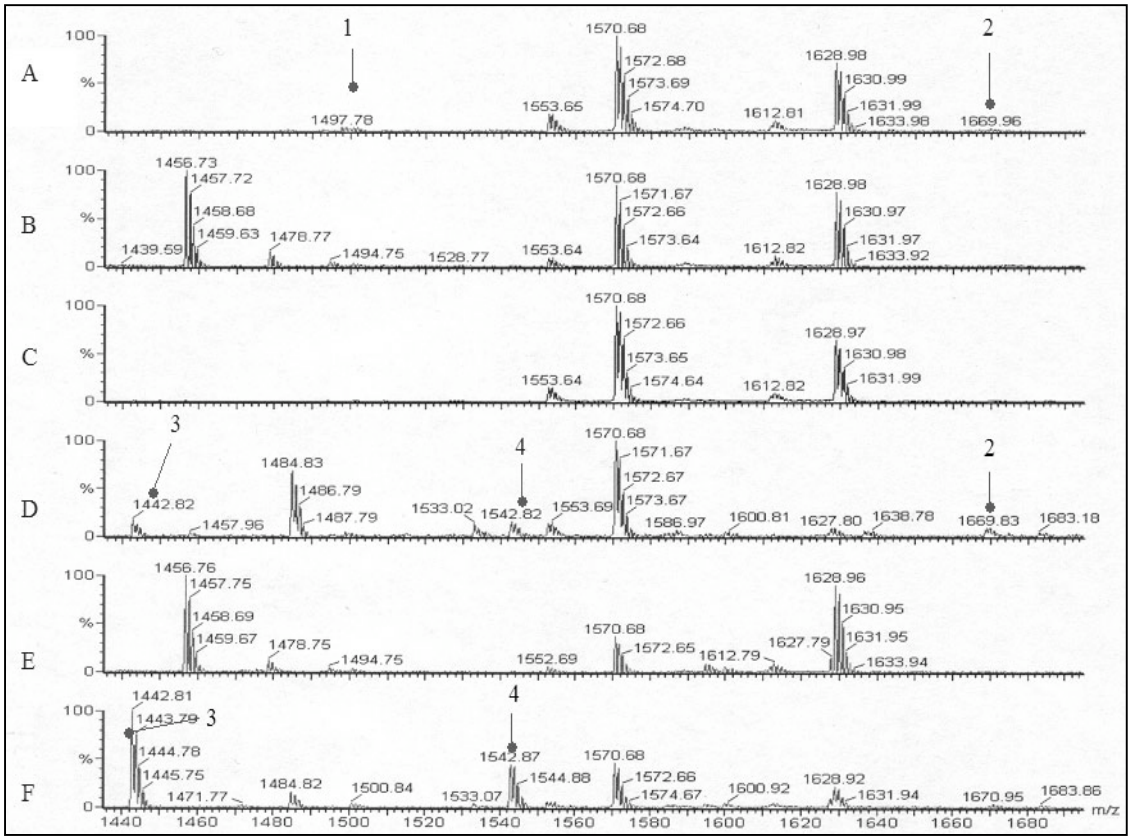


Figure 4.2: MALDI-TOF spectra of tryptic digests. Samples A-F (see section 2.5.1) were incubated for 1 hour at 70 °C. **A** = 10 μM native Alba, **B** = 10 μM rec Alba, **C** = 10 μM K16A mutant, **D** = 10 μM rec Alba + 0.5 μM *ss*Ard1 + 20 μM AcCoA, **E** = 10 μM rec Alba + 0.5 μM *ss*Ard1, **F** = 10 μM K16A + 0.5 μM *ss*Ard1 + 20 μM AcCoA. 1 = N-term acetyl group, 2 = N-term + K16 acetyl groups, 3 = K48 acetyl group, 4 = K64/68 acetyl group.

Further analysis of sample D (10 μM recombinant Alba1, 0.5 μM *ss*Ard1 and 20 μM AcCoA) (figure 4.3) showed that a mixed population of acetylated sample was present. This demonstrates the acetyltransferase ability of *ss*Ard1 and shows that recombinant Alba1 contains several possible sites for modification. The substrate also appears to be oxidised. The enzyme's order of preference for the above sites were probed by using a range of incubation times from 15 minutes to several hours at a lower temperature of 55 °C. The results (not shown) suggested that S2, rather than K16 was the first site to be targeted. Radiolabelled assays validated these findings and allowed better quantification (section 4.4).

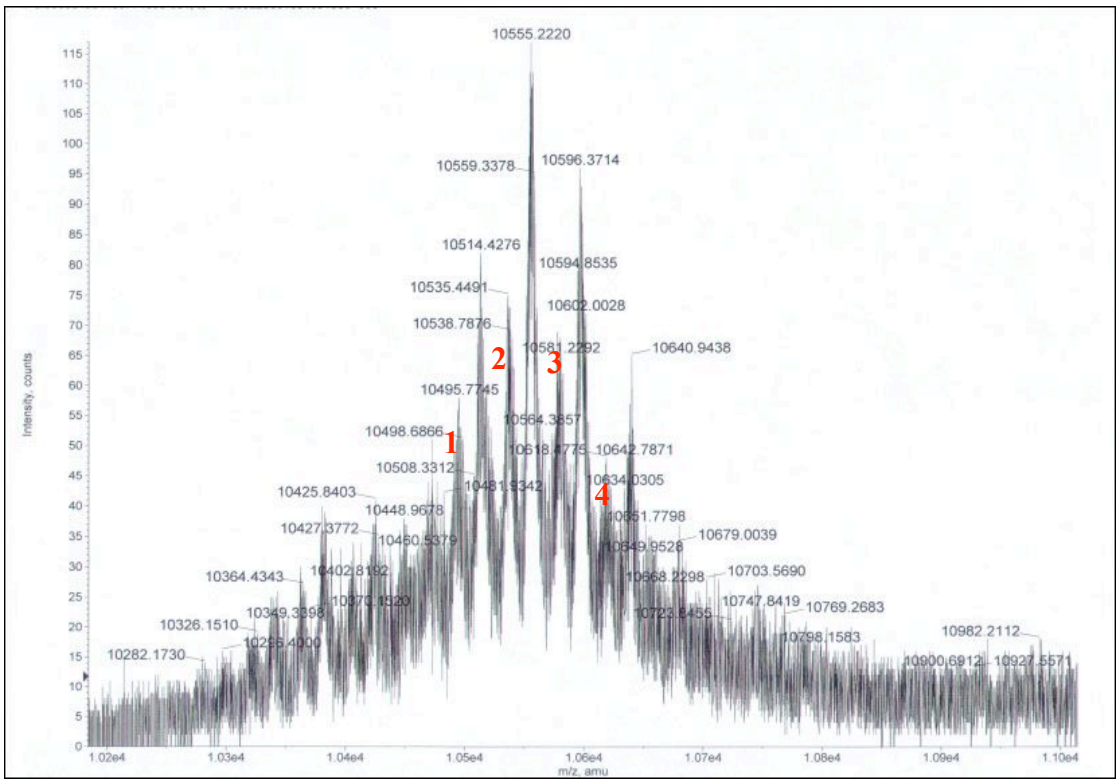


Figure 4.3: MALDI-TOF analysis of sample D. (1) 1 acetyl group (10495.77), **(2)** 2 acetyl groups (10538.78), **(3)** 3 acetyl groups (10581.23), **(4)** 4 acetyl groups (10618.48).

4.3 Optimum temperature

Since more control over the reaction rate can be gained by lowering the temperature of incubation it was necessary to explore the effect of temperature on enzyme activity. An SDS-PAGE gel on samples labelled with ^{14}C AcCoA over a temperature gradient from 25-75 °C showed a general increase in acetylated substrate with increasing temperature (figure 4.4a). A plot of acetylation vs temperature (figure 4.4b) reveals a slightly higher amount of substrate was acetylated at 55 °C than at 75 °C (the physiological temperature at which the organism exists). Whether this difference is real or down to error, the graph shows that 55 °C is the optimum temperature at which further experiments should proceed. It is likely that some evaporation of the 10 μl samples may have occurred at the highest temperatures

because it is not possible to overlay such small volumes with oil and this could in fact account for the lower amount of acetylated product.

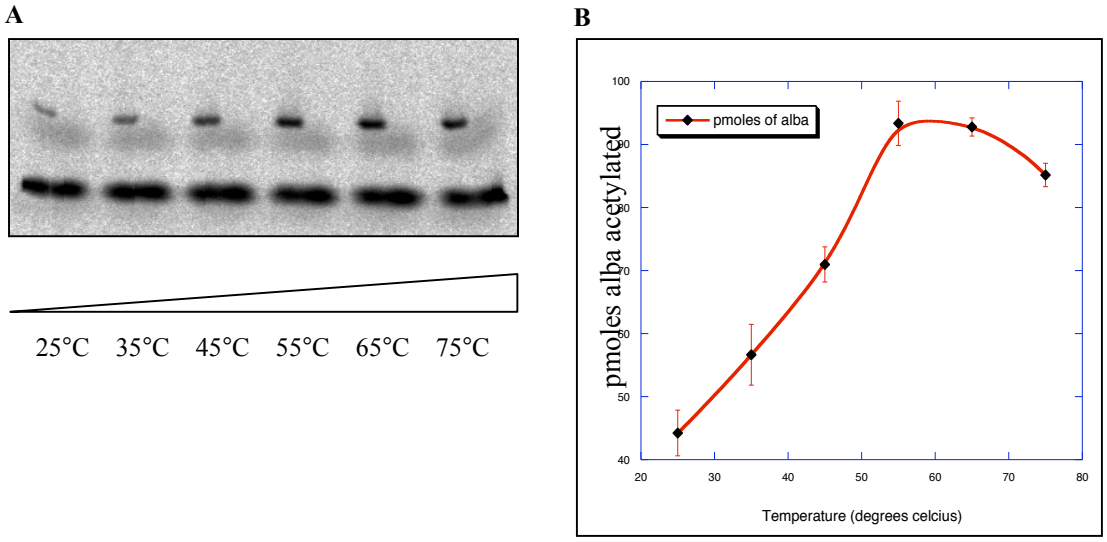


Figure 4.4a Phosphorimage of SDS-PAGE analysis of radiolabelled acetylated Alba1. Each sample was incubated at different temperatures for 10 minutes and contained 20 μM Alba1, 250 nM *ss*Ard1 and 18 μM ^{14}C acetyl-CoA. Each positive sample is followed by a control.

Figure 4.4b: Quantification of acetylated Alba1. Data was calculated to show pmoles of Alba1 acetylated.

4.4 Determination and quantification of preferred acetylation site.

As previously stated, it appears that *ss*Ard1 preferentially targets S2 for acetylation. To confirm this finding, a radiolabelled assay was carried out in triplicate for a variety of Alba1 samples. These included wt recombinant Alba1 as well as K16E and S2P mutated versions. The K16E mutated protein was acetylated to almost the same level as the wt version, which suggests that K16 is not a major target for modification by *ss*Ard1. In contrast the S2P mutated version showed very little acetylation, thus verifying the hypothesis that *ss*Ard1 does function as an N^{α}-acetyltransferase of *Sso*Alba.

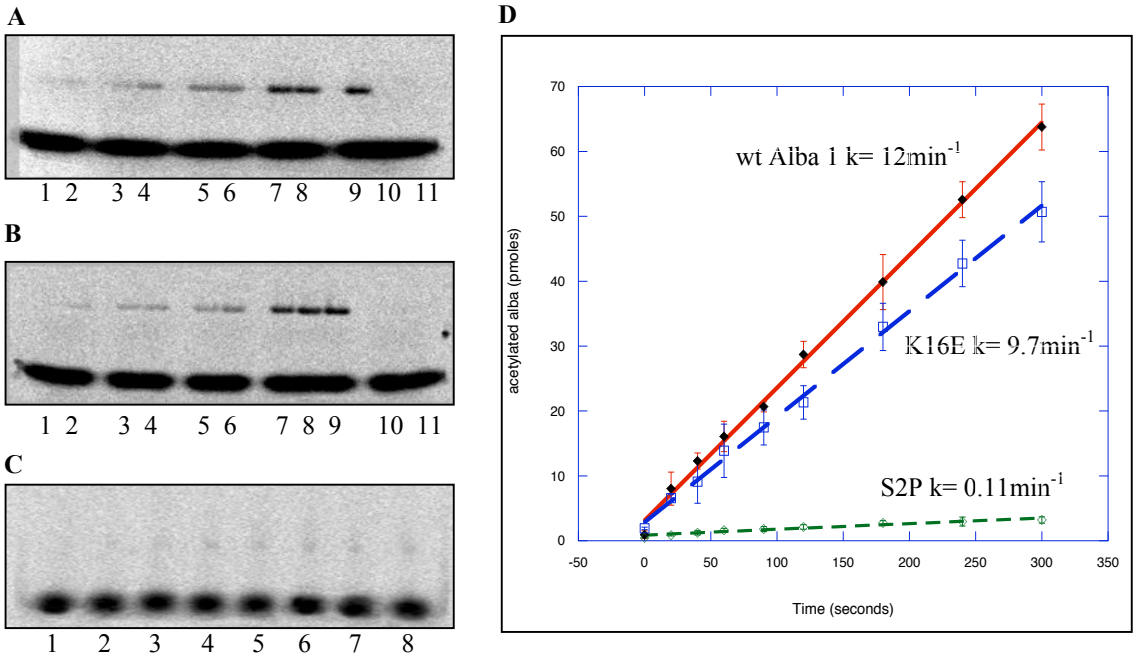


Figure 4.5: Acetylation of recombinant Alba1, K16E mutant Alba1 and S2P mutant Alba1 by *ss*Ard1 at 55 °C over time. (A) 20 µM recombinant Alba, 250 nM *ss*Ard1 and 18 µM ^{14}C Acetyl-CoA. 1 = 0, 2 = 20, 3 = 40, 4 = 60, 5 = 90, 6 =120, 7 = 180, 8 = 240, 9 = 300 10 = 90 control, 11 = 300 control (Time in seconds). **(B)** 20 µM K16E, 250 nM *ss*Ard1 and 18 µM ^{14}C Acetyl-CoA. Numbered same as in **(A)**. **(C)** 20 µM S2P, 250 nM *ss*Ard1 and 18 µM ^{14}C Acetyl-CoA. 1 = 0, 2 = 20, 3 = 40, 4 = 60, 5 = 90, 6 = 120, 7 = 180, 8 = 300 (Time in seconds). Control samples are not shown. **(D)** Graph comparison of wt recombinant Alba acetylation with K16E and S2P mutated versions. Band intensities were calculated by imagegauge software and converted to pmoles of Alba acetylated.

4.5 Kinetics of reaction

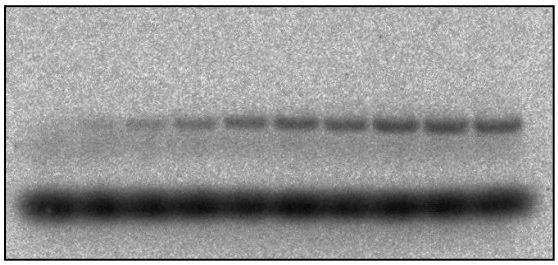
Reactions were carried out in triplicate using a constant concentration of *ss*Ard1 (250 nM) and increasing concentrations of recombinant Alba up to 80 µM. All samples were incubated for 4 minutes at 55 °C, a timescale which allows accurate detection and quantification of acetylated Alba. 9/10 cold acetyl-CoA and 1/10 ^{14}C acetyl-CoA was used to ensure that the isotope was not a limiting factor in the reaction. Radioactivity was counted by scintillation from dried SDS page gels on which the samples had been run (figure 4.6A) and results were quantified as nmoles of alba acetylated per minute per mg of *ss*Ard1, the standard units for easy comparison of K_m and V_{max} values with other acetylases, and represented on a graph (figure 4.6B). The data were fitted to the Michaelis-Menten equation (below) by

using a non-linear least squares approach (Kaleidagraph™-Macintosh computer program) to 'minimize the square of the error between the original data and the values predicted by the curve equation' (Kaleidagraph manual version 3.6).

Michaelis-Menten Equation

$$v = Vm \,.\, S \,/\, (Km + S)$$

A



1 2 3 4 5 6 7 8 9 10

Figure 4.6a: Phosphorimage of SDS-PAGE gel showing increasing acetylation with increasing Alba1 concentration. Incubation was carried out at 55 °C over 4 minutes using 250 nM *ss*Ard1 and 180 μM acetyl-CoA, 1/10 of which was ^{14}C labelled. Lanes: 1= 0, 2= 4, 3= 8, 4= 16, 5= 24, 6= 36, 7= 48, 8=60, 9= 70 and 10= 80 (μM Alba1).

B

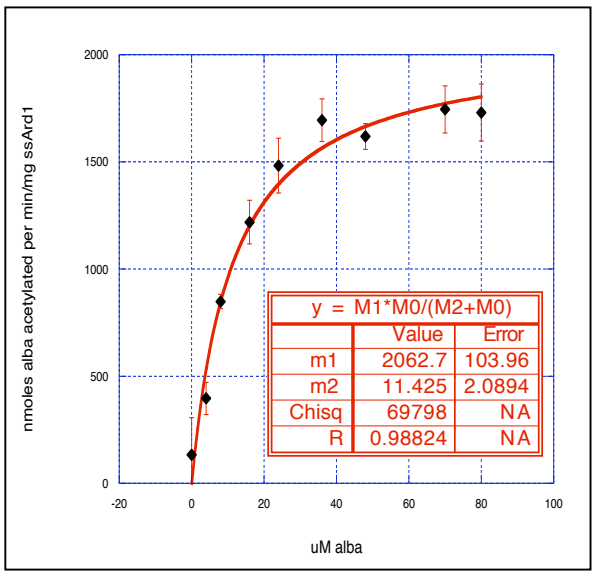


Figure 4.6b: Michaelis-Menten curve. Band intensities were calculated by image gauge software to show μmol/min Alba1 acetylated during the reaction. V*max* = 2062 ± 100 nmoles/min/mg enzyme.

Arrhenius and Van't Hoff have shown that:

$$\text{Temperature quotient } (Q_{10}) = \frac{\text{Velocity of reaction at } T^{n+10}}{\text{Velocity of reaction at } T^{n}} = \sim 2$$

i.e. an increase in 10 °C will approximately double (depending on the activation

energy) the velocity (v_0) of reaction (van't Hoff, 1884; Arrhenius, 1889), thus V_{max} must be multiplied by 4 to achieve a result corresponding to the real physiological conditions (~75 °C) of the organism. Hence V_{max} is estimated to be 8250 nmoles/ min/mg.

$k_{cat} = V_{max} / [E]$
$= 8250 / 51$
$= 160\ min^{-1}\ (3\ sec^{-1})$

NOTE: [E] is expressed here as nmoles enzyme (*ss*Ard1) in 1mg.

k_{cat} represents the number of substrate molecules that each molecule of enzyme converts to product per minute and from the above calculation it is evident that multiple turnover is achieved by *ss*Ard1. This calculation is based on the monomeric unit but if *ss*Ard1 is a dimer, which is likely, the value of [E] would be halved, giving rise to a higher k_{cat} value of 320 min^{-1} (3 sec^{-1}). Each sample was incubated for 4 minutes and in hindsight it would have been better to do the assay over lesser time such as 1 minute. But even if we calculate the amount of acetylated Alba after 1 minute based on the above data, this is unlikely to be a true value since substrate in the initial samples at 4 minutes were almost saturated. We find that almost half of the substrate is used in the early samples e.g. the y-axis value for sample 2, containing 4 µM (0.004 nmoles total) Alba1 and 0.0025 nmoles *ss*Ard1, reads 398 nmoles/min/mg acetylated Alba, thus 398/51 x 0.0025 = 0.0195 nmoles Alba1 were acetylated in a sample containing only 0.04 nmoles. And for sample 3, containing 8 µM Alba1, the y-axis reads 801 nmoles/min/mg acetylated Alba, thus 801/51 x 0.0025 = 0.0393 nmoles Alba1 were acetylated in a sample containing 0.08 nmoles. It is clear from this data that substrate was a limiting factor early on in the assay, making it difficult to determine the rate of the reaction. The accuracy of the software for the measurement of band intensities also comes into question in this case due to

the need for excessive isotope in order to achieve a reasonable acetylation signal in the early samples. The problem is that the unincorporated label (lower band) appears to be so intense that it is reasonable to suspect that the software has a limit to the intensity that it can measure and therefore the % incorporation of acetyl by Alba1 may actually be much lower than that calculated. On the other hand, a signal showing ~133 nmoles/min/mg acetylation could be detected in the zero sample where no Alba1 was present. So, it is quite clear that the measurements for the reaction are not ideal and it would be better to develop another means of assaying acetylation kinetics. In addition, it would be very difficult to achieve a realistic *Km* using the radiolabelled assay above no matter how many parameters we change because there are intrinsic limitations of the assay. We can be confident however, that the V_{max} is a realistic measure of the reaction because the later samples include excess Alba1. k*cat*/*Km* ($M^{-1}.min^{-1}$) is the apparent second rate constant and a measure of catalytic efficiency. We cannot calculate this value since we do not have an accurate *Km* but we can be certain that the *Km* is below 11 μM and based on this we can speculate that the catalytic efficiency of *ss*Ard1 for the acetylation of Alba1 will be very high and therefore Alba1 is a good substrate for the enzyme. [Note: Substrates having a k*cat/Km* ratio in the range 10.5-10.6 M-1.sec-1 are considered to be very good substrates i.e good affinity and rapid turn over.]

A paper published in 2000 describes the development of a non-radioactive assay as a means of testing acetylation. Kim *et al* monitored the amount of free CoASH generated during their histone acetylation reaction by adding α-ketoglutarate dehydrogenase or pyruvate dehydrogenase and α-ketoglutarate or pyruvate, respectively, as a target substrate for oxidation by CoASH (Kim, 2000). This reaction is coupled by the reduction of NAD to NADH, which was measured by spectrophotometric analysis at 340 nm (Kim, 2000). The kinetic parameters

calculated from each of the enzyme coupled reactions were largely in agreement with each other and strongly contradicted the values obtained from a radiolabelled filter binding assay (using [^{3}H]acetyl-CoA and histone H3 as a substrate) that was carried out simultaneously. The results showed that the radiolabelled assay gave 1.5 to 4 fold lower k_{cat} and 1.5 to 5 fold higher K_m values. The discrepancy was shown to be due to technical difficulty in washing product off the membrane *and* variable scintillation counting efficiency. Although filter binding was not employed in the ^{14}C-AcCoA assays, scintillation counting was not entirely accurate and more reliable results may have been produced by using an enzyme coupled assay of the kind described in Kim *et al*, 2000.

Several other NAT's have been evaluated for kinetic data (Table 4.1) but few of them bear any sequence or substrate similarity to *ss*Ard1 and, as such, cannot be used as a good comparison. NAT1 V_{max}/K_m= 418 and NAT2 V_{max}/K_m= 0.101 values can be obtained from a publication by Dupret *et al*, 1994. NAT1 is far more efficient than Nat2 but it is not uncommon for acetyltransferase V_{max}/K_m values to be in the range of NAT2 (Vatsis, 1999). Human serotonin AANAT has a value of V_{max}/K_m = 0.0269 μl/min/mg (Ferry, 2000).

N-acetyltransferase	Substrate	K*m* μM	k*cat* (min^{-1})	V*max* nmol/min/mg
GST*- Human Serotonin N-acetyltransferase (Ferry *et al*,2000)	Serotonin Tryptamine	1350 ± 115 880 ± 75	N/D N/D	132 ± 12 302 ± 25
Human Arylamine N-acetyltransferase 1 Human Arylamine N-acetyltransferase 2 (Dupret *et al*, 1994)	SMZ[1] PAS[2] SMZ[1] PAS[2]	4850 ± 550 8 ± 1 114 ± 27 7100 ± 470	"V*max* provides rough estimates of k*cat*"	610 ± 50 3340 ± 190 11.5 ± 1.0 3.6 ± 0.1
Klebsiella pneumoniae AANAT (Chung *et al*, 1997)	2-aminofluorene	1080±50	N/D	9.17 ± 0.11
M.tuberculosis AAC(2′)-Ic (Hegde *et al*, 2001)	KanamycinA KanamycinB Sisomycin	320 ± 29 1.4 ± 0.05 3.4 ± 0.3	15.2 ± 0.4 32 ± 0.3 258 ± 6.3	? ? ?
*ss*Ard1 (S.sol)	Sso10b	<11.4	160 ± 1.8 (monomer)	8250 ± 412
yGCN5 (kim, 2000)	H3 frag (aketo) H3 frag (pyruv)	130.2 ± 52.9 113.7 ± 24.4	27 ± 4.2 25.8 ± 1.8	? ?
S.typhimurium RimL (Vetting, 2005)	Ribosomal protein L12	1.0 ± 0.2	21 ± 1 min^{-1}	?
E.coli RimL (Miao, 2007)	L12	0.55	25.71 min^{-1}	?

Table 4.2: Comparison of V*max* and k*cat* with other acetylase enzymes. [1]PAZ, p-aminosalicylic acid, [2] SMZ, sulfamethazine.

4.6 Specificity of *ss*Ard1 for N-terminal sequence

In addition to Alba mutants K16E and S2P, another seven mutations were created by site directed mutagenesis where the N-terminal serine (after the cleavable methionine) was substituted with alanine, glutamine, glutamic acid, leucine, proline, threonine or valine. Each mutated plasmid was used to transform *E.coli* Top10 cells. The mutated sequences were confirmed by sequence analysis. Positive samples were selected for transformation into Rosetta BL21 cells and several colonies were picked for mini-expression trials. Alba mutants were purified in the same way as the wt recombinant version as described in (Wardleworth, 2002). Purified mutant proteins were analysed by mass spectrometry for an intact mass to

allow us to determine which, if any, did not have the N-terminal methionine cleaved off (table 4.3). Recombinant wt alba, S2A and S2G do not have this residue. S2E and S2L still maintain the methionine, whilst the others appear to exist in both forms.

	Molecular Weight Da	N-term Met
Rec Alba	10454	X
S2A	10438	X
S2E	10627	√
S2G	10424	X
S2L	10611	√
S2T	10468/10599	X/√
S2V	10466/10597	X/√

Table 4.3: wt Alba Vs mutants. Table displays intact mass for each protein and presence or lack of N-terminal methionine.

Purified mutant proteins were concentrated and diluted to approximately 50 μM. Each sample dilution was subject to Bradford assay using a known concentration of pure wt Alba as a standard for the dye reagent to bind. 10 μl of each 50 μM sample is shown on SDS-PAGE gel (figure 4.7A). These samples were assayed in the same way as the wild-type recombinant Alba and acetylation levels were quantified from phosphorimage of SDS-PAGE gels by imagegauge software. The results (figure 4.7B) show that alanine is the next preferred residue for N-terminal acetylation after serine. Surprisingly the glutamic acid and leucine substitutions, both of which maintain the initial methionine, exhibit a similar level of modification. Although the residue acetylated in both cases is methionine, it is important to note that the subsequent residue will usually impact sequence recognition by the enzyme. Leucine and glutamic acid have very different properties so it would perhaps be expected that the acetylation levels would differ somewhat. The valine substitution also displays a moderate amount of acetylation, whereas threonine and glycine are the least acetylated residues.

A

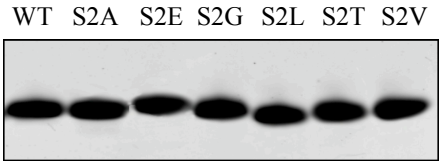


Figure 4.7a: SDS-PAGE gel on Alba1 variants. 50 µM wt recombinant Alba in comparison with pure N-terminal mutant Alba proteins to show purity of samples and similar concentration. % activity of each mutated version of Alba is shown.

B

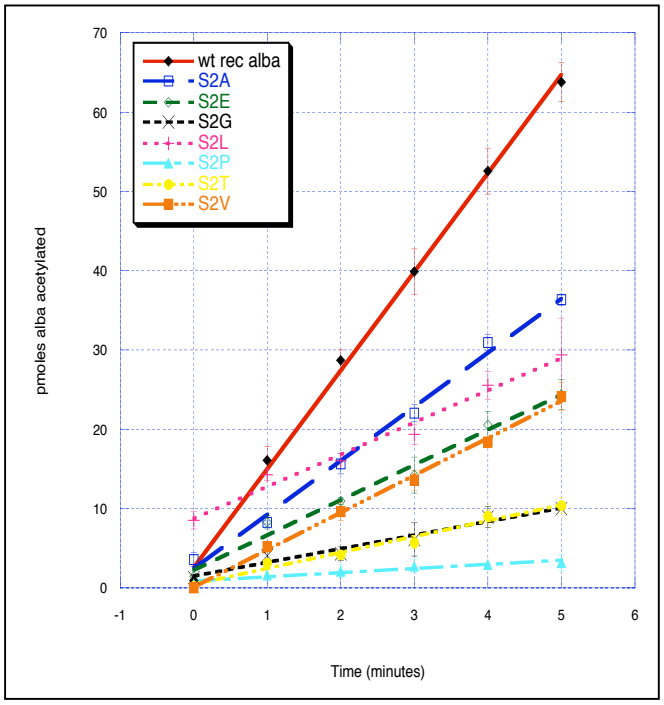


Figure 4.7b: Acetylation levels of wt Alba1 and mutated Alba1 compared. Graph shows pmole acetylation levels of 20 µM Alba and N-terminal mutant proteins over 5 minutes at 55 °C, using 250 nM *ss*Ard1 and 18 µM ^{14}C acetyl-CoA. Each result is an average of triplicate samples. Imagegauge software was used to calculate % acetylation from band intensities on SDS-PAGE gels.

In addition to *ss*Ard1 displaying a high degree of homology with yeast and human Ard1 (the catalytic subunit of NatA), it also exhibits a preference to acetylate N-terminal serine and alanine residues in that order, in a manner similar to NatA. Threonine, glycine and valine are also substrates for NatA, whereas met-glu is a substrate of NatB and met-leu is a substrate of NatC (table 4.4). This evidence, taken together implies that *ss*Ard1 may be an ancestral Nat protein with relaxed sequence specificity from which three distinct eukaryal Nat's have evolved.

Frequencies of acetylated and non-acetylated yeast proteins with various N-terminal sequences

Acetylation	+	0	%			+	0	%
NatA substrates					NatB substrates			
Ser-	124	12	91	10	Met-Glu- and Met-Asp-	13[a]	0	100
Ala-	19	44	30	0	Met-Lys- and Met-Arg-[b]	0	13	0
Gly-	3	13	23	99	NatC substrates			
Thr-	15	44	25	67	Met-Leu-	8	7	53
Cys-	1	6	14	80	Met-Phe-	3	2	60
Val-	1	32	3	10	Met-Ile-	4	7	36
Pro-[b]	0	24	0	0	Met-Trp-	1	1	50

+, Complete or partial acetylation; 0, no acetylation; %, percent acetylated proteins. Proteins whose acetylation status is questionable are not included in this Table.

[a] One of the proteins having N-terminal sequence MDPLA is partially acetylated.

[b] Not a NAT substrate; placed in NAT substrate section only for comparison purpose.

Table 4.4: Nat substrate specificity in *S. cerevisiae*. Values outlined in blue show % acetylation in mammals. Adapted from Polevoda and Sherman, 2003.

4.7 Differential Scanning Calorimetry

Solution Differential Scanning Calorimetry (DSC) is a procedure designed to measure the uptake of energy generated by gradually increasing the temperature of a protein sample, causing it to unfold. This is compared with a reference sample (the buffer minus protein), which is used to create a baseline. Accurate sample concentrations are essential for fully quantitative analysis and several scans per sample are typical. Heat energy uptake is labelled as C_p endotherm and measured in calories per degree Celsius.

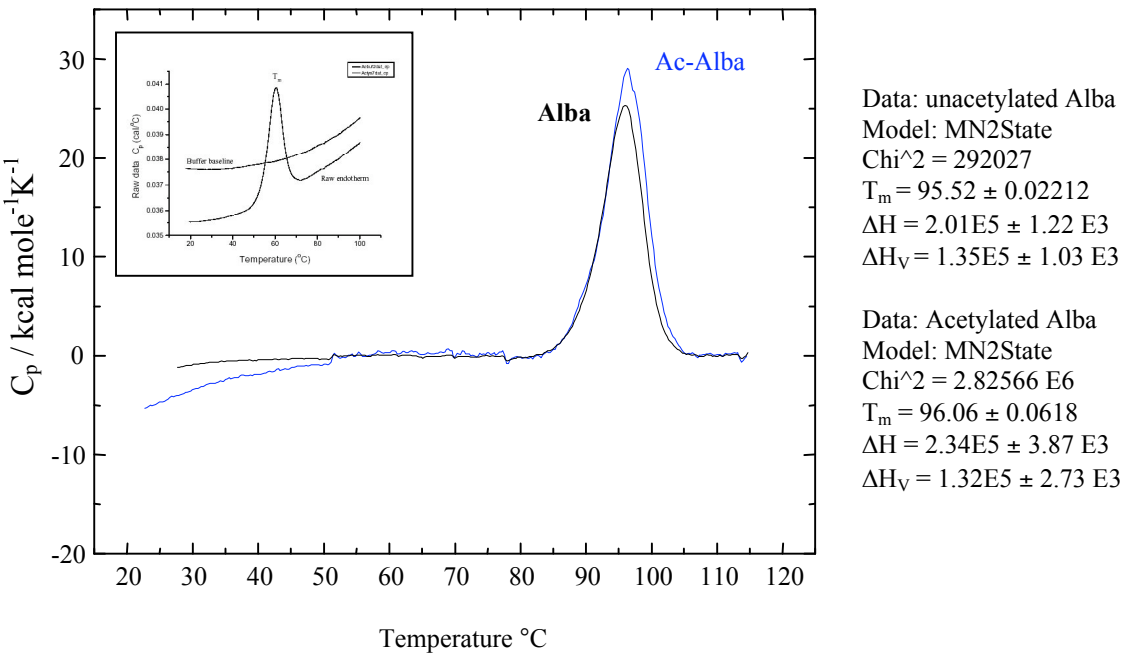


Figure 4.8: Thermogram showing DSC data for non-acetylated and acetylated Alba1.

Temperature stability (excess heat capacity, Cp, as a function of temperature) of native recombinant Alba is represented by a black line, whilst acetylated Alba is represented by a blue line. Inset is an example of how the raw data appears before deduction of the baseline. Scan rate 60 ºC per hour. ΔHcal (kJ/mol) is the calorimetric enthalphy representative of the total heat energy uptake by the sample under transition and is equal to the total area below the thermogram peak. ΔHVH is the van't Hoff enthalpy, an independent estimate of the enthalpy of transition based on an assumed 2-state model for the process. It is the area under the peak at any given temperature, divided by the total area under the peak and measures the extent of unfolding that has occurred at that specific temperature.

T_m is the transition temperature at which approximately 50 % of protein molecules are either unfolded or in an 'ideal dynamic reversible two-state equilibrium' where any molecule spends half its time folded and the other half unfolded. Raw data is normalised by deducting the baseline and streamlining the sample concentrations. N-terminally acetylated Alba did not appear to be any more or less sensitive to temperature denaturation as there is only a small difference of 0.54 ºC between the two samples (figure 4.8). The negative portion at the start of the curves is due to the high heat capacity of water in the buffer, which is reflected only at temperatures well below the onset of thermal unfolding. Theoretically, $ΔH_{cal}$ and $ΔH_{VH}$ should be equal. Variation between the two values may be caused by error in

sample concentration and the level of purity. However, they can also be compared so as to give information on the apparent mechanism of unfolding and the reversibility of the process. If the unfolding transition is not two-state (i.e. it consists of one or more intermediate steps) ΔH_{cal} will be larger than ΔH_{VH}. This may be the case with both samples here but it could also be down to error in the concentration (underestimation) since Bradford reagent was used to determine the protein concentration. Both samples were carefully purified and exhibit similar calorimetric enthalpies, thus suggesting that the unfolding mechanism is perhaps not two-state. But it would be naïve to read too much into these values. Le Chatelier's principle shows that binding of any small molecule to the native form of the protein will stabilise the folded state, which is what seems to happen in the case of acetylated Alba, albeit in minute quantity. The ability of the acetylated sample to absorb more heat energy implies several possibilities: this sample is more concentrate, it is more stable and/or the presence of the ligand cofactor itself is responsible for the increase in ΔH_{cal}. Rescanning the samples (not shown) proved that the thermal process was reversible given that approximately 75% of the protein in both samples was able to refold correctly. The melting temperatures of other thermostable proteins within the *Sulfolobus* genera vary, as do their refolding ability. DSC studies on Sso7d have given a Tm of 98 ºC (Knapp, 1996), whilst the Tm of Sac7d is reported to be 90.7 ºC (McCrary, 1996). Data on the thermostability of other acetylated proteins is not available but in terms of general stability, *in vivo* acetylation of the site-specific DNA-binding transcription factor E2F1 prolongs the half-life of the enzyme (Martínez-Balbás, 2000). Another example is the increased stability of the microtubule against depolymerizing reagents upon acetylation of α-tubulin (Takemura, 1992).

4.8 Summary and Conclusions

*ss*Ard1 displays acetyltransferase activity optimally at 55°C *in vitro* and is responsible for acetylating the N-terminal serine of *Sso*Alba. It also seems to be able to acetylate several lysine sites upon long incubation periods but its primary target is Ser2. Kinetic parameters were not easily obtained in the radioactive assay due to the reaction speed and the need for excess acetyl-CoA. Some areas of the procedure may be prone to additional error (out with the expected error of sample handling etc) such as the image reading software's capacity to determine accurate amounts at highly saturated scintillation levels (e.g. for unused acetyl-CoA). We can confidently assume that V_{max} parameters are correct since the substrate is plentiful in later samples but K_m is undetermined, suffice to say that it is lower than 11 μM. The k_{cat} value shows multiple turnover and catalytic efficiency (k_{cat}/K_m) is estimated to be >1600 $M^{-1}.sec^{-1}$. It may also have been useful to employ Electrospray ionisation (EI) mass spectrometry as a means quantitative analysis for comparison with this assay. Mutational analysis determines that the preferred N-terminal residues of *ss*Ard1 are serine > alanine > met-glu > met-leu > valine > threonine > glycine > proline. The only discrepancy here between *ss*Ard1 and NatA preference is valine. This could be due to error or in fact be a real property belonging to *ss*Ard1. The fact that methionine is also acetylated by this enzyme shows that it displays functional similarity to NatB and NatC in addition to NatA, and this supports the theory that *ss*Ard1 is an ancestral Nat from which the three distinct eukaryal forms have evolved.

DSC analysis of the both acetylated and unacetylated Alba shows that both forms of the protein have a melting temperature (T_m) in the region of 95-96 ºC but gives little other information on the apparent benefit of existing in one form or the

other. It was certainly expected that native Alba from a hyperthermophilic organism would not denature at temperatures below 80 ºC and in this case it does not appear that acetylation affords any extra protection from thermal denaturation or vice versa. It is likely then that the acetylation of Alba serves as an intracellular signal in several ways; (1) to initiate the binding of other proteins/complexes to itself, (2) to increase/decrease the half life of the protein, and (3) to integrally lower its affinity for binding DNA, the later of which has already been proven. Binding of other proteins to Alba may be investigated by raising antisera against *Sso*Alba1 from *Sulfolobus* cell lysate or by binding the purified protein to gel beads in the same way as the affi-gel assay was carried out for *ss*Ard1. Protein half life could possibly be assessed using an epitope (e.g. TAP) tagged system and western blot analysis.

Chapter 5: *ss*Ard1 active site and promiscuity for other substrates

5.1 Introduction

Ard1 is a member of the gcn5-like N-acetyltransferase superfamily (GNAT's). It shares a reasonable amount of sequence homology (up to 37% identity) with a large number of hypothetical proteins and several known GNAT's, which is evident from sequence alignments (Appendix 1). Although many bases may have mutated over time within different species, significant domain homology between the proteins is a common feature. On comparison of *ss*Ard1 with yeast Gcn5 HAT, we can predict that the core HAT fold has been maintained in *ss*Ard1 by aligning the two proteins with each other. Domains A, B and D are clearly evident but domain C is far less conserved and is not found in these two cases. Domains A and B contain residues that are known to form contacts through side chains, the main chain peptide backbone and via hydrogen bonding through water molecules to acetyl CoA. Only a few HAT structures have been solved with acetyl CoA and substrate. The acetyl CoA binding seems to be very similar throughout the GNAT family but residues responsible for substrate binding are not conserved because, as previously discussed, GNAT's have a wide range of substrates, from histones and proteins to short peptides and antibiotic compounds. Little is known about how these enzymes distinguish between substrates because identification of amino acids responsible for substrate recognition and binding is difficult since these residues are not conserved between different GNAT subsets belonging to the superfamily. Neither is it simple to pinpoint residues involved in the catalysis of acetyl group cleavage and addition to the substrate. In many HAT's, there is a cysteine residue close enough to the active site that can thus act as a strong general base for the catalysis of the reaction. *ss*Ard1 does not contain any cysteine residues

that could be implicated as part of the catalytic site, therefore a sequence alignment must be used to detect highly conserved residues.

5.2 Identification of highly conserved residues

*ss*Ard1 was aligned with known Ard1 homologues from several species including *S. cerevisiae*, *H. sapiens* and *C. elegans*. The alignment also included two other homologous sequences, *sc*NAT3 and *sc*MAK3, the catalytic subunits of NatB and NatC respectively (Figure 5.1). Identical residues are shown highlighted in black and several were chosen for mutational analysis (based on their physical properties).

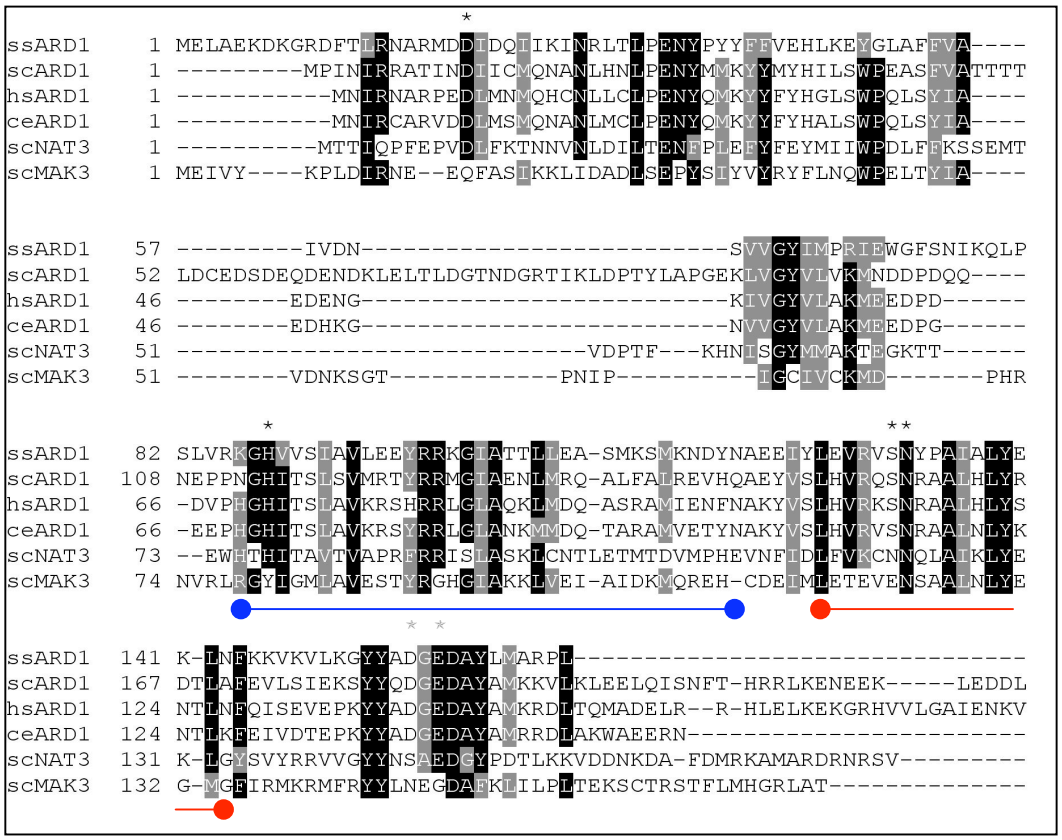


Figure 5.1: Sequence alignment of known Ard1 proteins and homologues. Motifs responsible for Acetyl-CoA binding are underlined. Motif A (in blue) spans *S.cerevisiae* residues 112-146 and motif B (in red) spans residues 152 to ~170 (Song, 2003).

Each of these residues: D21, S131-N132, D156 and E158 were mutated to alanine, whilst H88 was mutated to asparagine (sustaining the polar characteristic of histidine but without the charged side chain). D156 was chosen because although it is not conserved completely in the alignment, it is found in all four Ard1 homologues and therefore may be involved in substrate recognition. Primer design for each mutation can be found in chapter 2 and purification of the mutated proteins proceeded in the same manner as that used for recombinant *ss*Ard1 in chapter 3. Each mutant protein was diluted to a 2 μM concentration and analysed by SDS-PAGE (figure 5.2a). The wild type and mutated proteins were incubated as described in section 4.4, with 20 μM *ss*Alba1 and 18 μM acetyl-CoA at 55 °C. Samples were monitored in triplicate over ten time points (0, 1, 2, 3, 4, 5, 10, 15, 20, 25) and the results were plotted (figure 5.2b). The initial rates of reaction are difficult to determine since the data provides a curved rather than a straight line. Hence, the first three time points were plotted separately (not shown) to give the initial linear rate and the gradient of each line after deduction of the control value was used to calculate the % activity of each mutated *ss*Ard1 enzyme as compared to the wild type. Individual % activities are specified at the bottom of figure 5.2a.

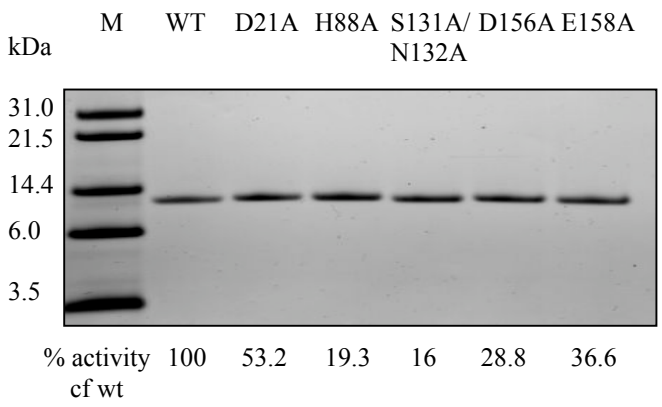


Figure 5.2a: SDS-PAGE gel on *ss*Ard1. Gel shows pure wt *ss*Ard1 versus pure mutants of *ss*Ard1; D21A, H88N, S131A-N132A, D156A and E158A. Each protein is at the same concentration of 2 μM. % acetyltransferase activity that each mutated protein possesses is shown underneath the gel.

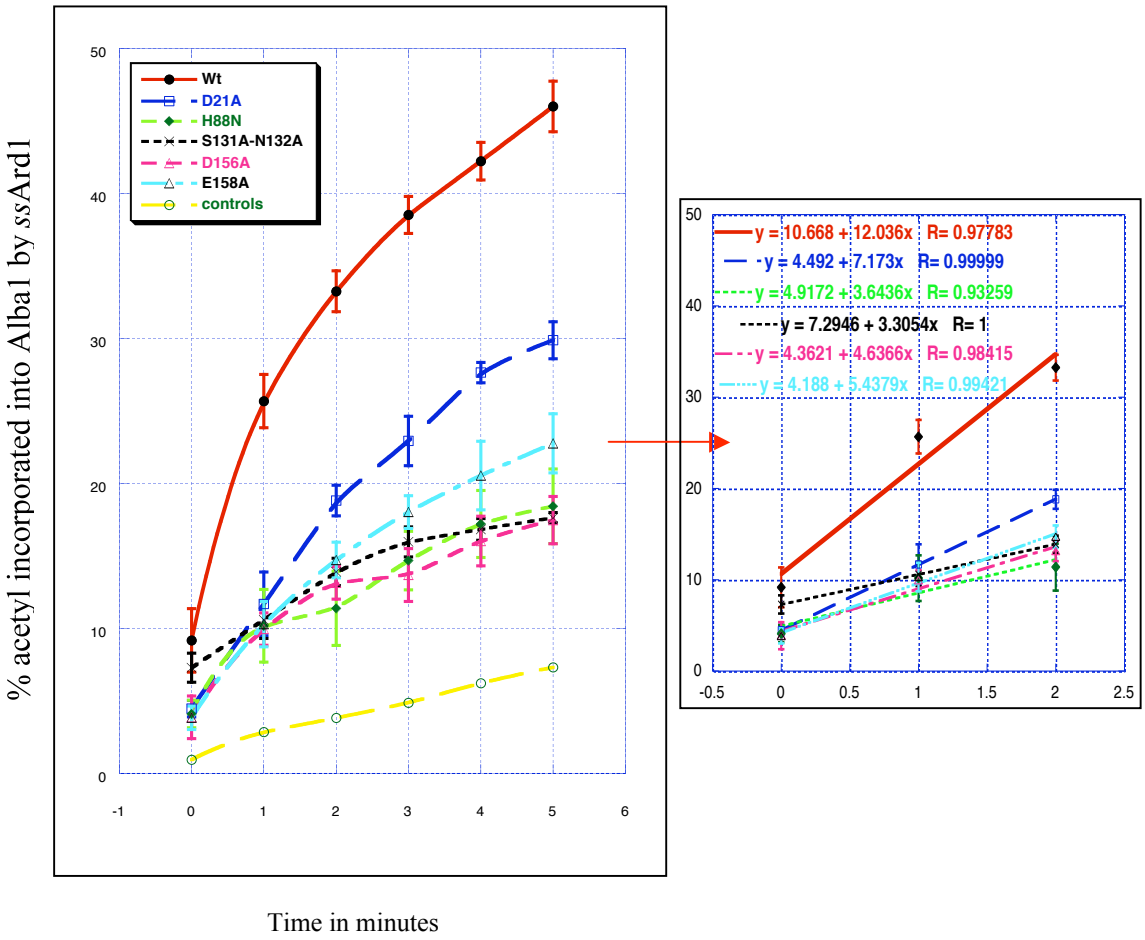


Figure 5.2b: Wild type *ss*Ard1 activity Vs activity of *ss*Ard1 mutants.

Graph showing 1-5 minute acetylation levels for wt *ss*Ard1 versus mutated *ss*Ard1 proteins. Later time points were excluded because they are not important for calculating the initial rate of reaction. Non-enzymatic (background rate) acetylation is shown in yellow and denoted as 'controls'. Linear plot from initial data points depicted on the right hand side allows calculation of % activity for each mutant Alba protein for comparison with wild type recombinant Alba protein.
Note: Y-axis is % acetyl incorporated into *ss*Alba1 rather than pmoles as only a percentage comparison is required for this particular assay.

Surprisingly, all *ss*Ard1 mutated proteins retained some level of acetylase activity but two of the mutations, H88N and S131A-N132A, exhibited very low levels of acetyltransferase activity (less than 1/5 of the wild type). It may be that these particular residues are involved in catalysis or substrate recognition. S131-N132 correspond to D214 in yGcn5, which occurs on the loop between β5 and α4. This loop is reoriented upon acetylCoA binding allowing D214 to be moved slightly out of the active sight so that substrate (H3) can bind. Hydrogen bonding between D214 and residues 215-217 grants additional stability to the complex (Langer,

2001). D156A also showed a substantial decrease in activity and could play a role in substrate recognition/binding. To understand more about the effect of the mutations, it is necessary to examine yeast (y)Gcn5 since it is one of the best studied members of the GNAT family and it aligns well with *ss*Ard1 (figure 5.3).

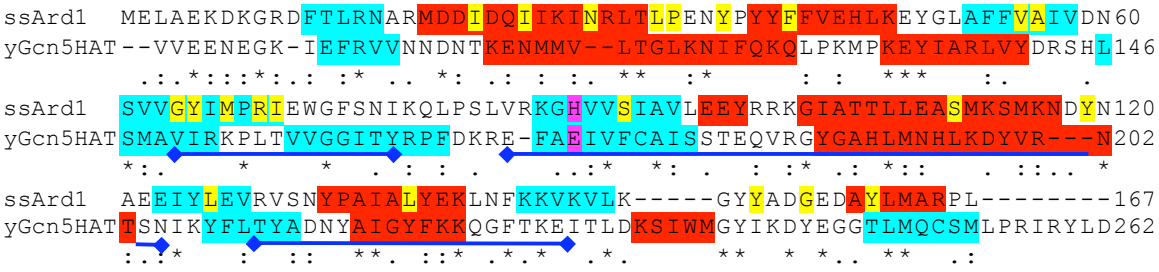


Figure 5.3: Alignment of *ss*Ard1 with *S. cerevisiae* Gcn5 HAT domain.

Key: Identical residues (*), highly similar residues (:), similar residues (.). Underlined are motifs D, A and B in that order. Italic residues are point mutations that reduce yGcn5 activity. Bold residues make contact with AcCoA.(Duntall, 1998b). Highlighted ssArd1 sequence shows the structure predicted by the 3D-PSSM program. Blue and red highlights show known β-sheet and α-helix secondary structure, respectively, as determined by crystallographic data for yGcn5 (Trievel, 1999). Further analysis of sixteen Ard1 homologues (appendix II) identified several residues (highlighted in yellow) that are absolutely conserved and that do not bind AcCoA. These residues are not found in yGcn5.

3D-PSSM/Phyre programme (see sections 2.7.6 and 6.7) was used to predict secondary structural features based on the primary sequence of *ss*Ard1 and compare these to previously solved structures of similar enzymes. Looking at figure 5.3, it is obvious that the predicted secondary structure of *ss*Ard1 compares favourably with the structure of yGcn5. β1, α1, α2, β2 and β3 correspond well to those of Gcn5, with only a little variation in the positions of the sheets and helices due to slightly longer or shorter loops connecting them. Of course, the prediction tool is not 100% accurate so there may be some error in the precise locations of the predicted conformations. β4 containing the putative catalytic base-H88 (discussed later) aligns perfectly. Prediction of a very short helix in *ss*Ard1 after this sheet may not be accurate and its presence would only have a small impact on the overall structure.

α4, β5 and α5 match well, showing the presence of a structurally conserved central region for acetyl-CoA binding. The end of the sequences differ in that *ss*Ard1 shows the presence of a sixth beta sheet (β6) before the final alpha helix (α5). Therefore, this last part of the structure may play a role in substrate selectivity.

Trievel *et al* (1999*)* expressed *y*Gcn5 with a large number of block mutations, spanning three or four residues (Trievel, 1999). Of these mutant versions, several were found to affect the activity of the enzyme and these are listed below:
KQL 126-128, PKM 129-131, YIA 135-137, FAE 171-173, RGY 186-188, GAHL 189-192, KDY 197-199, DNVA 214-217, IGY 218-220 and FKK 221-223.
From the information they gathered, they selected several residues within the blocks for point mutation and those that affected enzyme activity but did not contact acetyl-CoA are detailed in table 5.1.

Point mutations of yGcn5	Corresponding a.a's of *ss*Ard1	Similar properties	Putative role
H145	D59	-	Substrate binding and catalysis
E173	H88	-	catalysis
MN 193-194	LE 108-109	√√	?
Y220	L138	x	Substrate binding
KK 222-223	EK 140-141	√√	?
GF 225-226	NF 143-144	-√	?
G 239	G152	√	Substrate binding
I 241	Y 154	x	Substrate binding
YE 244-245	GE 157-158	x√	Substrate binding
TL 248-249	YL 161-162	x√	Substrate binding

Table 5.1: Point mutations that affect yGcn5 activity through mechanisms other than cofactor binding. Sequence alignment reveals the corresponding residues of *ss*Ard1 (column 2). These amino acids are marked in terms of their resemblance to one another. All residues are located within the binding pocket of the enzyme except MN193-194 which is located on α2 and depending on rotation, may be close enough for substrate contact, KK222-223 which is located on α3 and GF225-226 on the solvent exposed loop between α4-α5.

Key: (-) somewhat similar, (√) very similar or identical, (x) no similarity.

Omitting the similar residues (√), we are left with D59 ((Gcn5) H145 counterpart), a position which has retained its' polar properties because it is on a solvent exposed loop, yet it is different enough to implicate it in substrate binding. In addition to this, a string of other Ard1 conserved amino acids (i.e. not found in Gcn5) appear before and after D59 i.e. V55, A56, G64, Y65, M67, R69 and I70, further strengthening the importance of this region. L138 on β5 is possibly associated with substrate binding but N143 is situated right next to the AcCoA binding site and as such is probably not involved. Y154 and G157 occur in an outer (structural details contained in Ch6) loop and are probable candidates, as is the nearby Y161. A more detailed alignment (appendix II) of sixteen Ard1 proteins and predicted homologues (not used for figure 5.1 because they are only putative/not confirmed) implicates further conserved and highly similar residues that, according to the previous research on other GNATs, are assumed not to bind AcCoA and neither do they appear conserved in Gcn5. These are N29, L33, P34, Y37, A56 (already flagged), GY 64-65, S91, L126, L138 (already flagged), Y154, G157 and Y161. All of these residues could potentially be future mutagenesis targets for disrupting substrate recognition and/or binding.

Method of catalysis:

The catalytic mechanism of (y)Gcn5 could potentially proceed in two ways:

1) Via a ping-pong mechanism where the acetyl group is momentarily transferred to a cysteine residue in the vicinity of the active site, giving rise to an acetylated enzyme intermediate which is covalently bound, before catalysis of acetyl group transfer to the substrate (e.g. in Dupret, 1992, Watanabe, 1992) (figure 5.4).

2) Direct acetyl transfer via nucleophilic attack on the acylcarbon of the cofactor by the substrates primary amine, leading to transfer of the acetyl group from the co-factor to the substrate (figure 5.4) (Dyda, 2000).

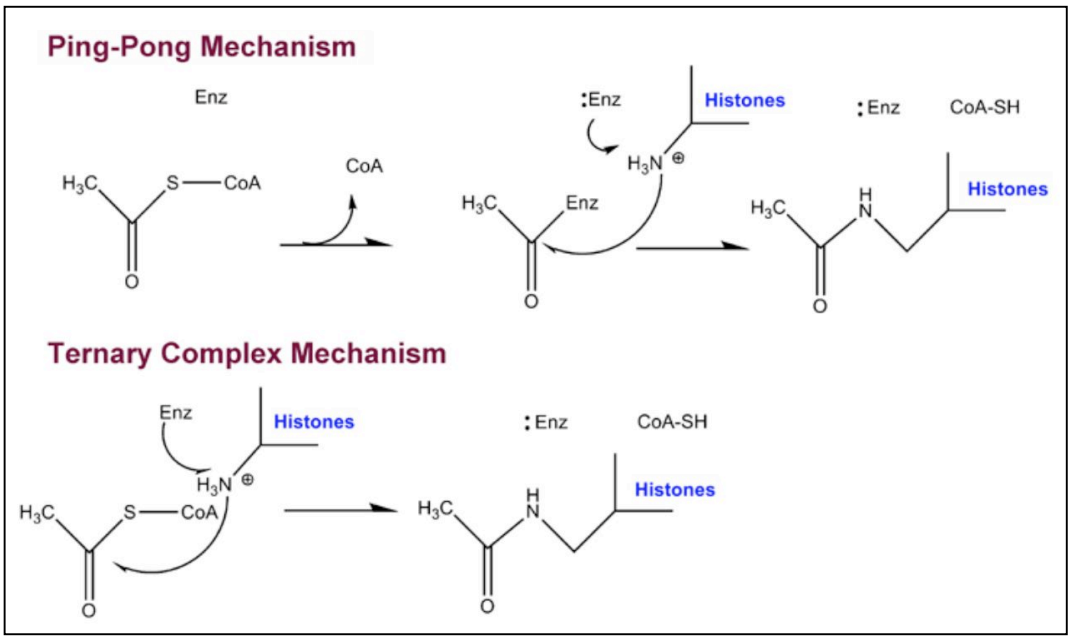


Figure 5.4: Catalytic mechanism of HATs. Ping-pong and ternary complex catalytic mechanisms as proposed for the MYST HATs and Gcn5/PCAF, respectively, are illustrated. (Hodawadekar and Marmorstein, 2007).

However, (y)Gcn5 does not contain a cysteine residue close enough to the active site to accept the acetyl group, which means that the reaction is likely to proceed via direct nucleophilic attack on the cofactor. E173 has been postulated to act as a general base for catalysis of the reaction (Langer, 2001), causing deprotonation of substrates at the N-terminal amine group. Many biochemical and kinetic experiments carried out by various groups (De Angelis, 1998; Duntall, 1998; Tanner, 1999) as well as the existence of ternary complexes found by crystallography (Hickmann, 1999; Rojas, 1999) support the theory that the catalytic mechanism for the gcn5 family involves the initial formation of a ternary complex, in which E173 (or similar) catalyses the deprotonation of the ε-amine (α-amine in

the case of Ard1) leaving it ready to engage in a nucleophilic attack on the bound cofactor carbonyl carbon. The acetyl group is then transferred to the appropriate site on the acceptor substrate, be it lysine or serine (Tanner and Trievel, 1999). *ss*Ard1 does not possess any cysteine residues but it does contains a histidine residue at position 88 corresponding to E173 in the alignment. Histidine can also act as a general base and it is reasonable to assume that this residue is important, although perhaps not essential, for catalysis within the *Sulfolobus* cell where the cytosolic pH of ~6.0 (She, 2001) means that the R-group belonging to this residue has a 50 % chance of existing in a protonated state. It remains to be proven whether H88 acts as a general base under the *in vitro* conditions (pH 8.5) utilised during this project. Mutagenesis studies show that acetylation levels fall when this residue is substituted by alanine. It must therefore play a substantial role.

5.3 Promiscuity of *ss*Ard1 for other substrates

In order to determine whether *ss*Ard1 acetylation is dependent only on N-terminal sequence, a variety of other recombinant proteins, Hje, Hjc, PCNA1&2, SSB, Sso7d, XPB1 and XPF were incubated as described in section 4.4 (Sso7d was purified from *S. solfataricus*). 20 μM of each sample was tested on SDS-PAGE gel (figure 5.5a) for purity and approximate concentration. 20 μM concentrations of each protein were used for assaying (Note: this is the monomer concentration because it is the acetylation of the accessible N-terminal tails that is being analysed). Some of these proteins had an alanine residue substituted at their N-terminus (table 5.1) but vary in terms of structure.

A

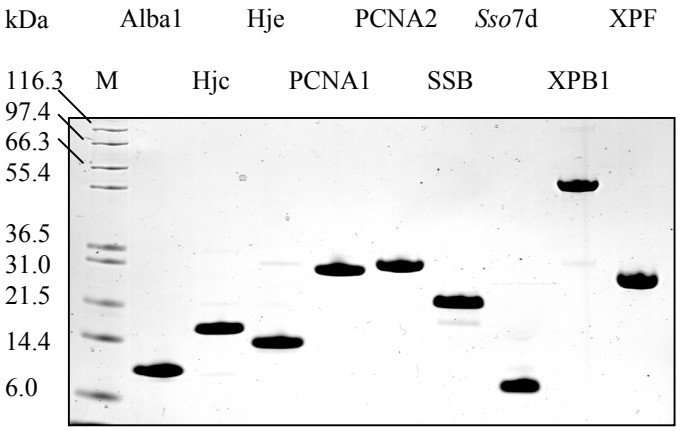
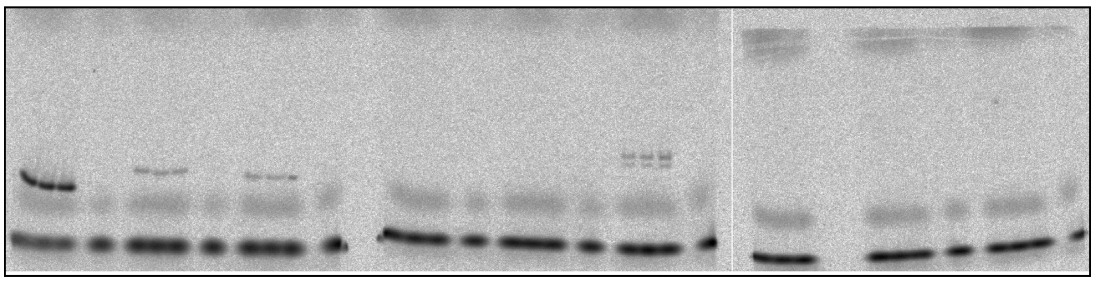


Figure 5.5a: SDS-PAGE gel of several *Sulfolobus* proteins.

20 µM of *S. solfataricus* recombinant* proteins (in monomeric form). Each sample was diluted to match *ss*Alba1 concentration. **Sso*7d was purified as the native form.

B



Wt Alba c Hjc c Hje c PCNA1 c PCNA2 c SSB c *Sso*7d XPB1 c XPF c

Figure 5.5b: Phosphorimage of SDS-page gels displaying acetylation levels of various *Sulfolobus* proteins. Gels show the intensity of radioactive signal from acetyl incorporated into different *Sulfolobus* proteins by *ss*Ard1 over a 5 minute time period. Protein samples include *Sso*Alba1, Hjc, Hje, PCNA1, PCNA2, SSB, *Sso*7d, XPB1 and XPF. C = control sample for the previous protein trio. Note: the control for *Sso*7d is not shown in this figure.

The average intensities were converted to % acetyl incorporated into each sample and are compared in the following bar chart (figure 5.6). For a complete comparison, S2A has also been included.

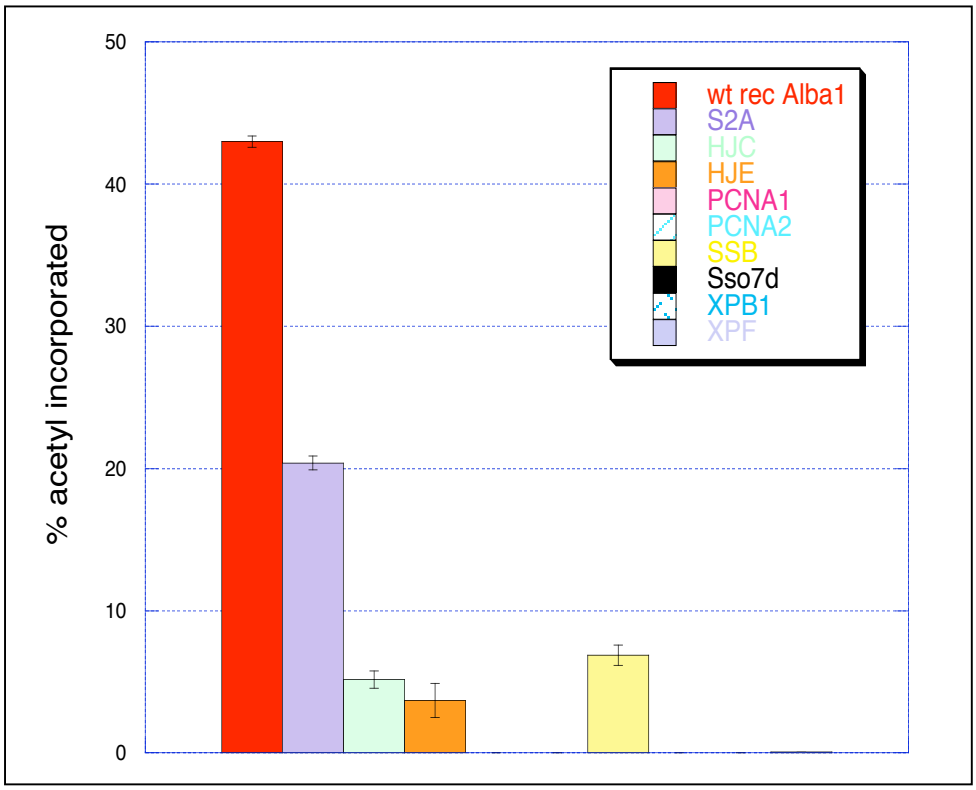


Figure 5.6: Bar Chart showing acetylation of various *S. solfataricus* proteins. Each sample (20 µM monomer concentration) was evaluated in triplicate to display the respective error for acetylation levels by 250 nM *ss*Ard1 at 55 °C over 5 minutes using 18 µM ^{14}C AcCoA.

Hjc and Hje (holliday junction resolving enzymes) were acetylated to a reasonable level by *ss*Ard1. This is not surprising since they have the preferred alanine residue available for modification. There is little difference between the amount of acetylation occurring in these samples, presumably because they have similar structures and functions. The data also suggests that the change in the subsequent amino acid from a small hydrophobic alanine to a larger, positive arginine (table 5.2) does not make much difference to the action of *ss*Ard1. The residues following this position also have quite different properties e.g. a hydrophobic residue in place of a hydrophilic residue when aligned but on the whole, the N-termini contain the same ratios of such residues. Only a little variation can be seen, which may be partly due to error in the determined sample concentrations.

SSB (single stranded binding protein) was also acetylated to a similar degree. The N-terminal sequence of SSB is MEEK… (table 5.2) which is not a eukaryal Ard1 substrate but is instead a Nat3 substrate. This data provides evidence to support the theory that *ss*Ard1 is a promiscuous ancestral NAT from which the three eukaryal NAT's have evolved. Native *Sso*7d (which is not acetylated *in vivo*) is not an *in vitro* substrate for *ss*Ard1 despite harbouring an alanine residue. The subsequent threonine is small like alanine (in Hjc) but also polar like arginine (in Hje) and, hence, would not have a significant affect on the acetylation level. This suggests that *ss*Ard1 recognises protein structure as well as N-terminal sequence.

Protein	% activity	N-term seq
Alba	100	SSGTPT
SSB	16	MEEKVG
Hjc	12	AAKKRG
Hje	8.6	ARDIGK
XPF	0.2	VIRIYA
PCNA1	0	MFKIVY
PCNA2	0	MKAKVID
Sso7d	0	ATVKFK
XPB1	0	VRLRYF

Table 5.2: *Sulfolobus* protein N-terminal sequences.

The first few amino acids partly determine the ability of the protein to become a substrate for acetylation. The % acetylation value (activity) is shown for those proteins that are indeed acetylated by ssArd1 in Vitro and compares them to Alba1.

XPF displays only limited modification which is at a level similar to the control. Overall, Alba1 is the preferred substrate of *ss*Ard1.

5.4 Acetylation of cell lysates

Since a knock out of *ss*Ard1 was not possible, an alternative approach was to inhibit the acetyltransferase action in the cell so that the enzyme could be added to cell lysate at a later stage to monitor the effects. This would give a better

indication of the number of potential targets of *ss*Ard1. A search for inhibitors of NAT's was carried out but there are very few commercially available and any that were available are quite specific for certain enzymes. The only inhibitor accessible was Curcumin. Increasing concentrations were added during the normal assay but no inhibition was shown, therefore it was only possible to assay lysates that had been subject to modification within the cell. However, proteins that are targets for acetylation will not be 100% modified within the cell due to the action of deacetyltransferases.

A variety of cell lysates from prokaryotic and eukaryotic origin were incubated with *ss*Ard1 for 1 hour at 37 °C. *Sso* lysate was also incubated at 70 °C. Each sample was analysed on SDS-PAGE gel (figure 5.7) and a number of bands can be seen representing proteins acetylated by *ss*Ard1.

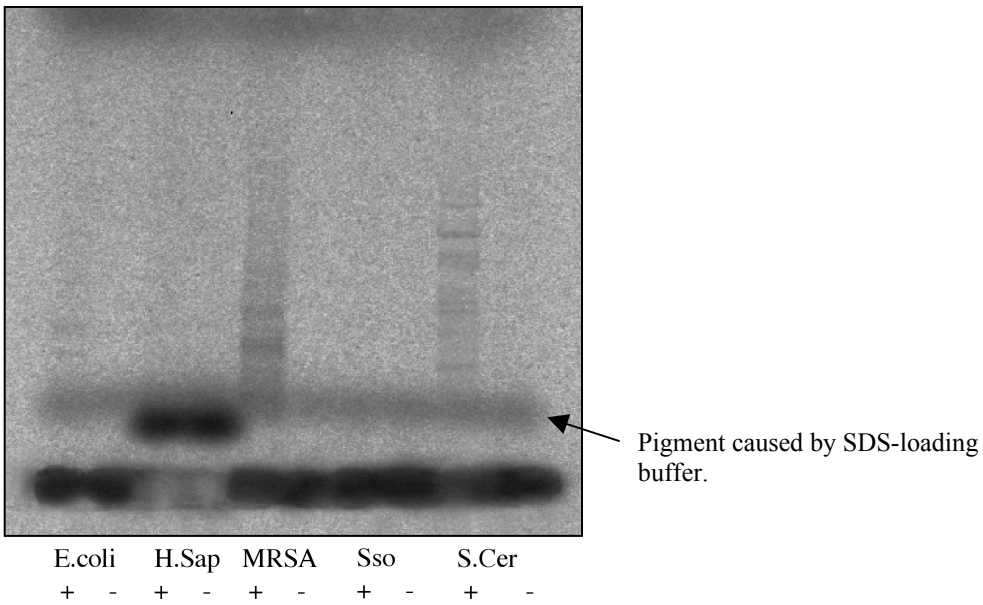


Figure 5.7: Acetylation of cell lysates from *E.coli*, *H. sapiens*, MRSA, *S. solfataricus* and *S. cerevisiae.* Each lysate concentration is approximately 0.47 mg/ml. Positive samples contain 36 μM ^{14}C AcCoA and 2 μM *ss*Ard1. Negative samples contain 36 μM ^{14}C AcCoA without *ss*Ard1. Each sample was incubated at 37 °C over 1 hour.

In figure 5.7, *Sso* lysate does not appear to show bands indicative of acetylated protein but close up analysis reveals a few acetylation occurrences (one being just above the darker pigment caused by the SDS-loading buffer, which makes it difficult to see), some of which will be natural substrate. The assay was repeated at 70°C to give an indication of the amount of acetylation that occurs at the organisms biological temperature and more bands were visible than at 37 °C (results not shown). The majority of bands appear in the yeast and MRSA extracts, showing that *ss*Ard1 is capable of acetylating a number of prokaryotic and eukaryotic proteins. The *E.coli* extract also shows a number of modified proteins. Unfortunately though, there was no easy way to identify the proteins that were modified but it does provide an idea of the extent of acetylation possible in the different organisms. Several other HATs are known to show a degree of promiscuity regarding the substrates they acetylate, especially those with dual function as transcriptional co-activatiors such as p300/CBP, which acetylates H2A, H2B, H3 and H4 as well as HMG proteins, p53, GATA-1, HIV Tat etc. PCAF targets H3, HMG proteins, p53, MyoD, E2F etc (Sterner & Berger, 2000).

5.5 Summary and Conclusions

A sequence alignment of Ard1 homologues highlighted several regions of conservation within the sequences. Five specific points were chosen for mutational analysis to pinpoint particular residues that could potentially be involved in catalysis and/or substrate binding. It is surprising that none of the five mutations (D21A, H88N, S131A-N132A, D156A and E158A) were sufficient to completely abolish the acetyltransferase activity of *ss*Ard1. However, H88N and S131A-N132A did reduce the enzyme activity by more than 80 %. On comparison with yGCN5, it is

highly likely that H88 acts as the catalytic base and that the reaction proceeds not via a ping pong mechanism but through nucleophilic attack on the acylcarbon leading to the formation of a ternary complex before the CoA is finally transferred to the N-terminal residue of the substrate. S131A-N132A are close to the acetylCoA binding domain so may be involved in substrate binding, but N132 is very similar to the AcCoA binding residue at the corresponding position in yGCN5, therefore it is likely that the decrease in activity could be attributed to an affect on cofactor binding. S131 as far as we know does not bind CoA but it is next door to N132 and is again highly similar to it's counterpart in yGCN5. Mutation of D156 also gave rise to a large decrease in activity and was thus postulated to engage in substrate recognition or catalysis. To determine between the roles of binding and catalysis, kinetic experiments would need to be carried out to define the catalytic efficiency, V_{max}/k_m values for the mutants.

*ss*Alba1 is not the only substrate of *ss*Ard1, though perhaps the main one. Evidence shows that the enzyme is capable of modifying Hjc, Hje and SSB. It is especially interesting to note that the N-terminal sequence of SSB would not be targeted by eukaryal Ard1 (in the NatA complex) because this sequence is preferred by Nat3 (of the NatC complex). This supports the theory that *ss*Ard1 may be an ancestral NAT from which the three modern day eukaryal NAT complexes have evolved. In addition to these findings, *ss*Ard1 not only targets specific N-terminal sequences but seems to recognise the structure of potential substrates. As for enzyme promiscuity, a number of proteins from prokaryotic and eukaryotic cell lysate were successfully modified by *ss*Ard1 *in vitro*. Further studies will be required to identify individual targets.

Chapter 6: Structural analysis of *ss*Ard1 by crystallography

6.1 Introduction

The Ard1 protein family has yet to be structurally characterized. One can predict from sequence alignments that they contain the common features of the GNAT superfamily i.e. domains A-D, the specific order of the domains and the common acetyl-CoA binding motif. The structure of these domains is fairly similar throughout the GNAT family but each enzyme differs in its mechanism of substrate binding, which is of course not only due to its structural conformation but also the amino acid sequence it possesses. So far, the structural information received for many members of the GNAT family centre around the ability of certain residues such as cysteine or glutamine to act as a catalytic base for the procession of the acetylation reaction but in the case or *ss*Ard1, no cysteine residues exist and a glutamine residue cannot be found in the proposed catalytic site, but sequence analysis highlights a histidine residue which could be used to achieve the same end means (see chapter 5). Structural information, obtained by crystallising the pure protein will test this hypothesis. It is also desirable to crystallise *ss*Ard1 in complex with a cofactor and substrate to give a full representation of the interaction occurring at the molecular level, thus identifying residues responsible for contact with the substrate and to give insight into how the enzyme performs catalysis of the acetylation reaction. This will provide essential information in regards to substrate binding, conformational change and the reason behind different enzymes having differing substrate preferences. It will also give an indication as to whether the catalytic mechanism is generally the same throughout the superfamily.

6.2 Introduction to Crystallography

Protein crystallography is an essential technique for the determination of molecular structures and is often used to analyse and understand the interaction between enzymes and substrates, which among other things, gives scope for therapeutic drug intervention. The procedure utilises electro-magnetic radiation (X-rays) to view atomic detail at a wavelength of approximately 1 Å (0.1 nm) but the nature of the technique means it is not possible to physically focus the X-ray diffraction pattern using a normal type lens (e.g. in the case of light microscopy) or a magnet (e.g. in the case of electron microscopy), so it must be done via mathematical equations using specialised software programmes. The diffraction pattern produced is recorded on a detector (an image plate or charge coupled device (CCD)) and then analysed by the software.

In order to achieve a strong enough diffraction pattern with good signal to noise ratio, an array of molecules must be used rather than just one. The molecules need to be in the same three-dimentional orientation (order) and hence the reason for converting solution protein into a more solid structure; a 3D array of molecules bonded non-covalently to form a protein crystal (Rhodes, 1993). The internal order of the crystal determines the angles of diffraction and the resolution achieved. As the X-rays hit the atom, the electron cloud surrounding it is disturbed and this results in diffraction of the X-rays at the same frequency of which they were emitted (Rayleigh scattering). Scattered waves in the same phase will combine to heighten the signal, resulting in a detailed 3D map showing the electron distribution throughout the protein. Both constructive (in phase) and destructive (out of phase) interference effects arise from the electron. In-phase rays give rise to the appearance of discreet reflections (spots) on the detector but out-of-phase rays will cancel each other out.

Since electrons are quite close to the nuclei of the atoms they belong to, it often results in a good interpretation of the location of specific atoms within the molecule.

Each reflection contains information on all the atoms in the structure and conversely each atom also contributes to the intensity of each reflection. The amplitude and phase of the X-ray waves relating to each individual reflection are important in reassembling the image of the molecule but phase information cannot be recorded experimentally and only the amplitude is known. This is known as 'the phase problem' and as such, must be solved before an interpretable electron density map can be calculated. Constructive interference (conditions that produce diffraction) occurs when the phase shift is proportional to 2π, and is expressed by Braggs law, which takes into account the relationship between the scattering angle and the interplanar space (explained in figure 6.1).

Bragg's law: where d = interplanar spacing between atomic lattice, hkl = set of parallel planes (index of hkl-Miller indices), λ = wavelength, θ = angle of incidence, n = integer.

$$2d_{hkl}\sin\theta = n\lambda$$

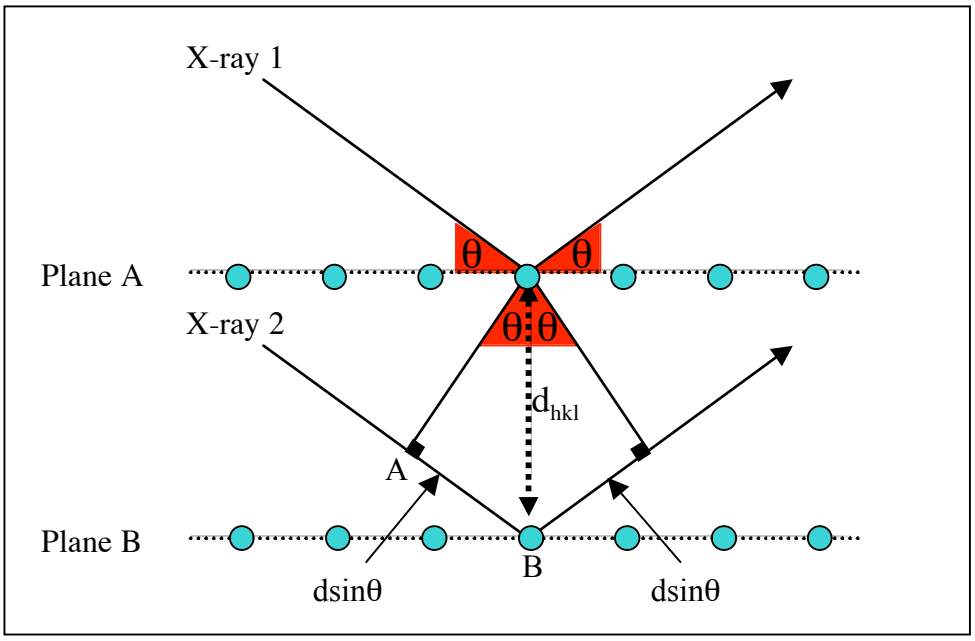


Figure 6.1: Example of the diffraction of two parallel X-rays after coming into contact with the atoms of a crystal lattice. dhkl represents the interplanar spacing between two rows of atoms, Θ is the angle of diffraction relative to the plane of atoms and dsinΘ is the value between A and B.

Previous to 1934, dried protein crystals were used for diffraction but the results were unreliable (Tulinsky, 1999). Bernal and Crowfoot-Hodgkin then discovered that crystals suspended in their mother liquor yielded better diffraction patterns (Bernal and Hodgkin, 1934). To enable protein crystals to form, the aqueous protein solution must be placed in an environment with the correct stimuli and a limited supply of oxygen to allow diffusion of the chosen buffer into the protein solution, causing slow precipitation of the protein whilst also preventing it from drying out. Protein purity is very important for the success of crystal growth, the solution must be at least 97 % homogeneous. The concentration of the protein solution is also important but can vary largely from protein to protein. Approximately 10 mg/ml is the recommended starting point but a quick test can be done to identify the approximate concentration at which reasonable precipitation occurs. The incubation temperature is also crucial as some proteins prefer to precipitate at room temperature (~19 °C) whereas others form better crystals at 4 °C (and this may also lead to better crystal packing if

precipitation happens too rapidly at room temperature). In addition to these factors, pH and precipitant type is crucial. Variation in pH can lead to different packing orientations of the molecules. Hence, it is obvious that the process of crystallography is not at all straightforward, especially when it concerns a protein that has never been crystallised before. Once crystals are produced, they are extremely fragile because the irregular surfaces of the molecules mean that large solvent containing channels are formed throughout the crystal and the non-covalent linkages are thus weakened (Rhodes, 1993). A typical protein crystal will contain approximately 50% solvent. Although the resulting fragility can be a problem, there is also a huge advantage to these solvent channels in that they are capable of soaking up certain metal ions that enable the solving of the phase problem. Likewise, they can be useful for the binding of substrates and inhibitors. Salt crystals are common and it is difficult to identify a salt crystal from a protein crystal by eye. Salt crystals are more tightly packed because they lack the solvent channels and will not take up Izit™ blue dye as opposed to protein crystals. This is fine for crystals that are not of suitable quality for x-ray diffraction but those that do look suitable should be tested by other means e.g. under a polarised lens because protein crystals commonly refract light. Alternatively they can be handled by a nylon loop to check for fragility (salt is robust due to the many strong ionic interactions) and finally they can be placed in the beam line of the x-ray generator for diffraction analysis. The resulting pattern will confirm the presence of salt or protein (salt crystals yield a only few dispersed, very high density reflections).

6.3 Methods for crystallising protein

There are several techniques that have been developed and applied for crystal growth but what works for one protein doesn't always work for another. One must weigh up the properties of the prepared solution first e.g. air and temperature sensitivity. Listed below are some of the more common methods (in bold) used:

Application of a **high pressure** environment is a method that was found useful for growing glucose isomerase crystals (Visuri, 1990). **Slow evaporation** can be useful for samples that are not air sensitive, where a large volume of sample is plated out and left exposed to an air supply by loosely covering with tinfoil for example**. Solvent diffusion (**also known as the layering technique) involves using a syringe to dribble a chosen (less dense) solvent into the protein solution. It should sit on top of the first solution to form a discreet layer. A common solvent combination to try here is CH_2Cl_2/Et_2O (if your compound is insoluble in ether). **Sublimation** requires sealing of the sample inside a glass tube under a vacuum and placing it at relatively high temperatures in an oven or likewise. **Vapour Diffusion** is probably the most common technique utilized today from which two methods have evolved: the sitting drop (figure 6.2a) and the hanging drop (figure 6.2b). The latter is the simplest setup but the sitting drop allows for a greater volume to be used if required. Both methods rely on the extraction of water from the drop into the reservoir solution, which increases the protein and precipitant concentration in the drop. At the same time, the elements of the reservoir diffuse into the drop until an equilibrium is reached. Reservoir solution will typically contain a buffer to maintain the pH e.g. Tris-HCl (Branden & Tooze, 1999), an ionic precipitant such as sodium chloride and an organic precipitant such as a high molecular weight polyethylene glycol (PEG). All of these compounds will be at a higher concentration in than in the original drop

(which usually contains similar constituents). The equilibrium hopefully creates optimum conditions that will remain until the crystallisation process is complete (McRee, 1993; Rhodes, 1993).

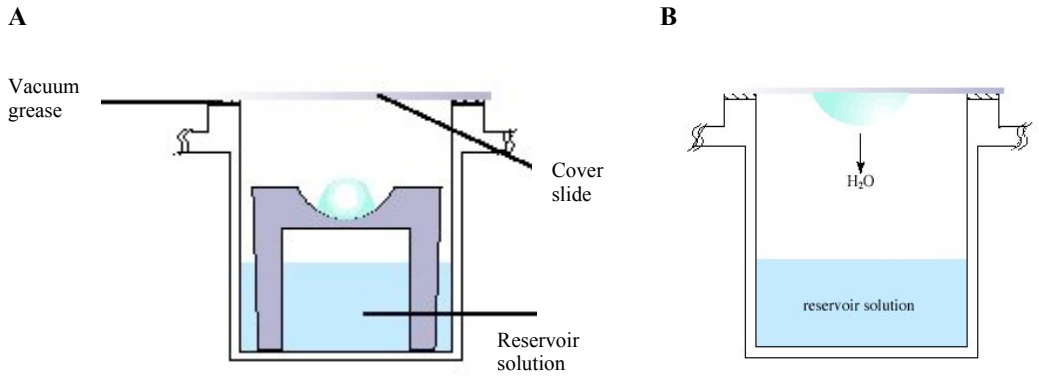


Figure 6.2: Vapour Diffusion. (A) Sitting drop method of vapour diffusion. **(B)** Hanging drop method of vapour diffusion. The very same labels apply (not shown) as in 6.1a. Both are closed systems to prevent the drop and the reservoir from drying out. Typically, solute is mixed in a 1:1 µl ratio with mother liquor (reservoir solution) and reservoir solution may be between 100-200 µl.

6.4 Protein crystals of *ss*Ard1

Concentrating *ss*Ard1 was difficult because it readily precipitated. Initially the highest possible concentration achieved was 3.15mg/ml. Crystal screens from Hampton and Wizard were utilised for preliminary testing. Hampton I, II and Peg/Ion solvents were used along with Wizard I and II solvents. An aesthetically pleasing crystal was obtained from Wizard I condition 44 but diffraction analysis showed it to be salt. Plenty of non-crystalline precipitate was observed and several conditions produced tiny needles but after several weeks of incubation no other crystals had formed. Many of the drops looked clear under the microscope and it seemed likely therefore, that the protein concentration was not high enough. Further

expression and purification of *ss*Ard1 at the higher pH of 8.5 (where the pH is not close to the pI of ~7.6) resulted in more concentrated protein yields.

Three separate batches of purified *ss*Ard1 protein solution were concentrated in dialysis tubing from approximately 100 ml to about 2 ml each and analysed on SDS-PAGE gel for purity (figure 6.3). Each of the samples showed good purity, although lane 2 shows up three very slight bands below the main one. These could be contaminants but it is likely that they are breakdown products of *ss*Ard1. Bradford assay determined the relative concentrations to be 8.1 mg/ml, 9.8 mg/ml and 6.4 mg/ml.

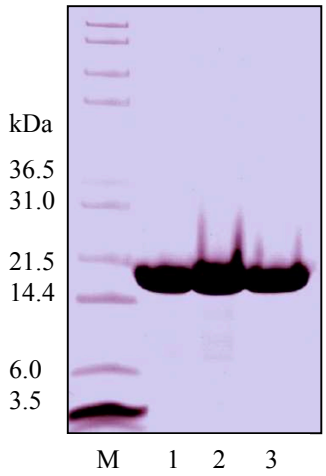


Figure 6.3: SDS-PAGE gel on pooled and concentration fractions from all 3 samples. M = mark 12 protein marker, lane 1 = 8.1 mg/ml (415 μM), lane 2 = 9.8 mg/ml (502 μM) and lane 3 = 6.4 mg/ml (329 μM).

Acetyl Co-enzyme A was added to half of the *ss*Ard1 samples (0.5 μl of 12.4 μM per 1 μl of enzyme) in the hope that it would stabilise the enzyme and make it more likely to crystallize. Many more crystal screens were set up manually using the different protein concentrations and the samples with added AcCoA in ratios of 1:1, 2:1 and 3:1 enzyme : reservoir solution. In addition to the previously used screens (Hampton Crystal Screen I and II, Hampton PegIon, Wizard Crystal screen I and II); Hampton Index, Peg/LiCl, Ammonium sulphate, Salt Rx, Natrix, Crystal Screen

Cryo, Crystal Screen Lite and Emerald Bioscience Crystal Screen I and II were used this time. After 2-3 weeks, the following crystals appeared (figure 6.4).

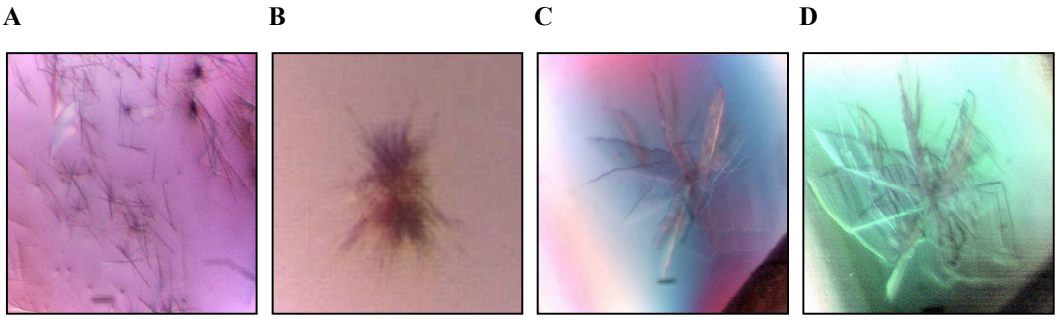


Figure 6.4: Crystals of *ss*Ard1 under polarizing light. (A) Needles obtained from 5.9 mg/ml *ss*Ard1 in Hampton Crystal screen I condition 12 (30 % (v/v) isopropanol, 0.1 M Na Hepes pH 7.5, 0.2 M $MgCl_2$). **(B)** Starbursts obtained from 8.1 mg/ml *ss*Ard1 + AcCo-A in Emerald Wizard I condition 35 (20 % (v/v) 1,4-butanediol, 0.1 M acetate pH 4.5). **(C)** & **(D)** Same plates shown under different light. Obtained from 8.1 mg/ml *ss*Ard1 + AcCo-A in Hampton Crystal screen I condition 6 (30 % (v/v) PEG 4000, 0.1 M Tris-HCl pH 8.5, 0.2 M $MgCl_2$).

The needle-type crystals stained dark blue with Izit dye and when micro seeded they did reproduce. Thus, the specific condition was chosen and a grid of solutions was made where each component of the crystal screen was varied very slightly one at a time. These were incubated at 4 °C and 19 °C but crystal morphology did not improve sufficiently. The starbursts easily broke apart when tapped gently and they were also floating near the top of the drop as opposed to being heavy and rigid like salt, verifying protein crystals but macro seeding in mother liquor was not a success. A few days later the crystals dissolved completely. Attempts to reproduce these crystals by optimising the condition (at 4 °C and 19 °C) failed. The plates had nice sharp edges but also contained irregular regions and refracted light well, which is a positive indication. But they did not grow suitably in the third dimension and were too thin to be used in the X-ray beam line. They did not appear again when the condition was reproduced and neither did they stain with Izit dye, both negative

signals signifying probable salt crystals. After several repetitions of the screens, a few crystals of decent size (figure 6.5) grew from 8.1 mg/ml *ss*Ard1 + AcCoA in Hampton Peg/Ion 39 (0.2 M sodium dihydrogen phosphate (monobasic, initial pH ~4.25), 20 % PEG 3350 (final pH ~4.5) after three weeks of incubation at 19 °C.

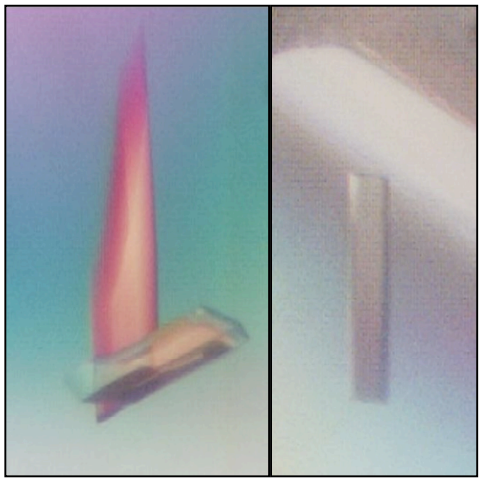


Figure 6.5: *ss*Ard1 Crystals formed in Hampton Pegion, solution 39. Approximate dimensions: Width ~ 0.15 mm, Height ~ 1.2 mm, Depth is difficult to measure but may be about 0.02 mm.

Producing a suitable cryoprotectant for the crystal was difficult. Crystals seemed to split in 10 %, 20 % or 25 % glycerol even when soaked slowly with increasing concentration. They dissolved quickly in MPD. Synchrotron diffraction pattern results were similar to that which was gathered in-house. Diffraction of the 25 % glycerol protected crystal occurred (figure 6.5) to a reasonable resolution of ~ 3 Å but this did not provide enough detail at the molecular level. The smeared double dots in figure 6.5 confirmed that the crystal had split (or was already twinned with another crystal although it looked like a single formation).

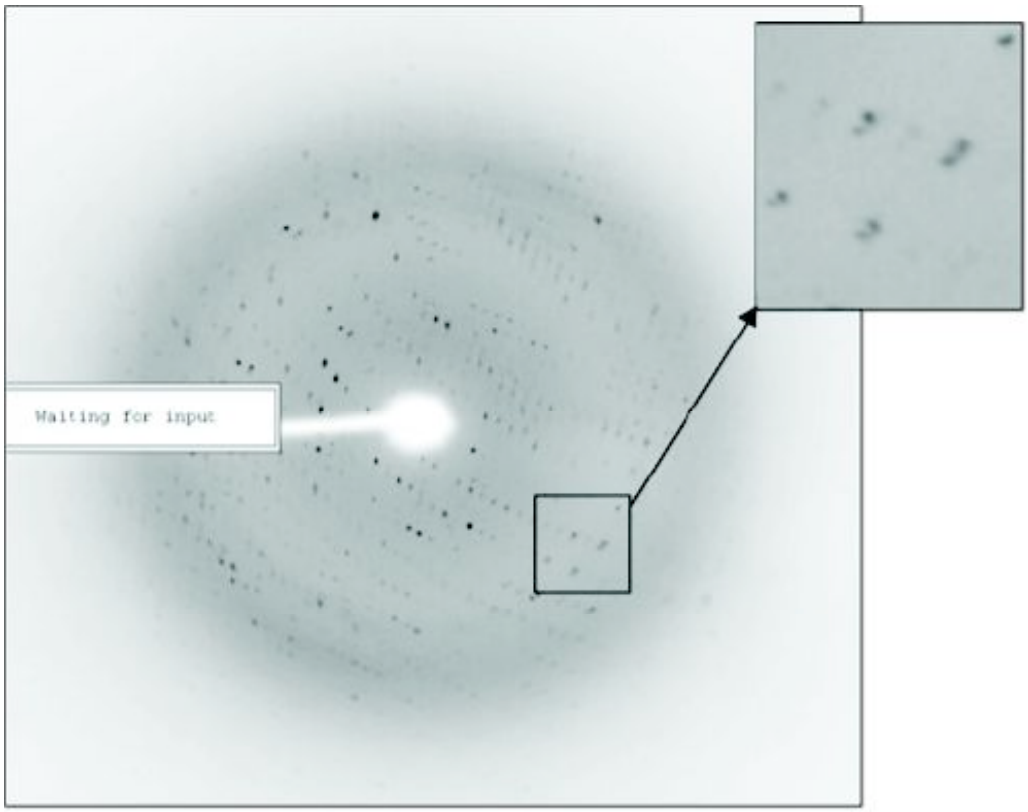


Figure 6.6: Diffraction pattern from an *ss*Ard1 crystal. Detector distance = 200mm, 45 second exposure per 1° angle rotation through 180°, Resolution to 2.9 Å. Other figures can be found in table 6.1.

Table 6.1 shows slight variations of Pegion 39 that were made in the attempt to optimize crystal growth. All solutions were at the initial pH values of 4.25, 4.7 and 5.2. All conditions were set up in a 1:1 ratio using 1.5µl of *ss*Ard1 protein solution at concentrations of 5.4 mg/ml, 6.9 mg/ml and 7.17 mg/ml each with the addition of 2X and 1X (molecules Acetyl-CoA per molecule *ss*Ard1) and without acetyl coenzyme A.

0.3M/30%	0.25M/30%	0.2M/30%	0.15M/30%	0.1M/30%
0.3M/25%	0.25M/25%	0.2M/25%	0.15M/25%	0.1M/25%
0.3M/20%	0.25M/20%	0.2M/20%	0.15M/20%	0.1M/20%
0.3M/15%	0.25M/15%	0.2M/15%	0.15M/15%	0.1M/15%
0.3M/10%	0.25M/10%	0.2M/10%	0.15M/10%	0.1M/10%

Table 6.1: Optimised Hampton Pegion 39 conditions. E.g. 0.3M sodium dihydrogen phosphate and 30 % PEG 3350. Each made at pH 4.25, 4.7 and 5.2.

Little was observed in the original condition but all 30 % PEG3350 concentrations (pH4.7) gave very small needles for the higher protein concentrations when used in a ratio of 1:2 enzyme to AcCoA. 0.1 M Na dihydrogen phosphate and 20 % PEG3350 at pH5.2 gave larger needles but unfortunately, all efforts to reproduce further worthy crystals with this solution failed.

The protein samples (*ss*Ard1 and *ss*Ard1 + AcCoA) were later sent to the St. Andrews SPoRT laboratory to be plated out in 100 nM droplets by robotic arm. Several screens, some similar to those above, were added in a 1:1, 2:1 and 1:2 ratio. These individual droplets were automatically photographed routinely to highlight any crystal growth or nucleation. A single crystal rod appeared from the JMAc screen (produced by colleagues at St. Andrews University) condition 52 (20 % PEG 8000, 0.2 M $MgCl_2$ and 0.1 M MOPS pH 6.5). Although very small, it gave diffraction at two angles (0° and 90°) confirming that it was protein. After many optimisations, the best crystals (fig 6.7) grew in (1:1 µl ratio) solution of 20 % PEG 8000, 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5 with addition of 0.5 µl Hampton Additive screen II condition 6 (30 % (w/v) D(+) glucose monohydrate). Several other additives worked well also.

Figure 6.7: *ss*Ard1 crystals. Formed in 20 % (v/v) PEG 8K, 0.2 M $MgCl_2$, 0.1 M Tris-HCl, pH 8.5 with addition of 30 % (w/v) D(+) glucose monohydrate.

*ss*Ard1 crystals were now being produced in their hundreds over a period of just three days but none of them diffracted to the same or better quality than those

obtained from peg 39. They were simply too thin in the 3rd dimension to obtain data for a full 180° rotation. Further optimisations failed, as did trying increase their size by slowing their growth via incubation at 4 °C. A layer of oil was also dispensed on top of the room temperature drops to slow the vapour diffusion rates. Many star burst crystals formed by this method but, as such, were useless. Therefore, the best data achieved was to 2.8 Å from the Hampton pegion 39 solution.

Data was collected at ESRF synchrotron (ID 14.1) for 120 images, each representing a 1° rotation of the crystal. The data were then interpreted using MOSFLM (Leslie, 1992) and scaled using SCALA in the CCP4 software package (Collaborative Computational Project, Number 4, 1994). The scaled statistics from SCALA are given below (figure 6.8). The processed data gives a unit cell with dimensions a = 42.3 Å, b = 69.4 Å and c = 58.0 Å, and angles $\alpha = \gamma = 90°$, $\beta = 97.8°$. The predicted space group is P21 with a mosaicity of 1.1°

Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA I/sigma	sd Mn(I)/sd	Nmeas	Nref	Ncent	FRCBIAS	Nbias
7.91	.065	.000	.065	.000	0	6833.	885.1	7.7	252. 34.7	724	275	13	-.059	217
5.59	.053	.000	.059	.000	0	3340.	301.4	11.1	151. 28.3	1447	495	10	-.031	484
4.56	.052	.000	.056	.000	0	4839.	436.2	11.1	222. 27.7	1826	623	7	-.021	630
3.95	.060	.000	.057	.000	0	4602.	455.6	10.1	240. 24.2	2160	742	7	-.001	720
3.54	.079	.000	.061	.000	0	2677.	358.8	7.5	210. 16.7	2457	829	6	-.020	890
3.23	.111	.000	.066	.000	0	1476.	248.4	5.9	200. 10.5	2733	929	6	.016	1027
2.99	.176	.000	.072	.000	0	712.	171.3	4.2	201. 5.5	2983	1017	4	-.001	1079
2.80	.295	.000	.078	.000	0	372.	146.7	2.5	208. 3.0	3129	1069	1	-.015	1216
2.64	.453	.000	.085	.000	0	242.	146.2	1.7	218. 1.9	3398	1173	1	-.011	1304
2.50	.676	.000	.093	.000	0	162.	145.7	1.1	231. 1.2	3570	1245	1	.027	1321
iline graphs use a Java browser</applet>														
	.093	.000	.093	.000	0	1795.	305.7	5.9	213. 11.0	24427	8397	56	-.016	8888
	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA I/sigma	sd Mn(I)/sd	Nmeas	Nref	Ncent	FRCBIAS	Nbias

Dmin	Nmeas	Nref	Ncent	%poss	C%poss	Mlplct	AnoCmpl	AnoFrc	AnoMlt	Rmeas	RmeasO	(Rsym)	PCV	PCVO
7.91	749	300	20	80.1	80.1	2.5	.0	.0	.0	.082	.082	.065	.095	.095
5.59	1515	563	22	84.4	82.8	2.7	.0	.0	.0	.066	.066	.053	.073	.073
4.56	1939	736	23	85.8	84.2	2.6	.0	.0	.0	.065	.065	.052	.070	.070
3.95	2309	891	27	87.0	85.2	2.6	.0	.0	.0	.074	.074	.060	.082	.082
3.54	2634	1006	25	88.0	86.0	2.6	.0	.0	.0	.100	.100	.079	.108	.108
3.23	2929	1125	28	89.1	86.7	2.6	.0	.0	.0	.138	.138	.111	.154	.154
2.99	3217	1251	29	89.3	87.2	2.6	.0	.0	.0	.219	.219	.176	.245	.245
2.80	3375	1315	24	90.3	87.8	2.6	.0	.0	.0	.367	.367	.295	.387	.387
2.64	3657	1432	28	90.3	88.2	2.6	.0	.0	.0	.564	.564	.453	.603	.603
2.50	3856	1531	26	90.9	88.6	2.5	.0	.0	.0	.840	.840	.676	.913	.913
ie graphs use a Java browser</applet>														
all	26180	10150	252	88.6	88.6	2.6	.0	.0	.0	.116	.116	.093	.128	.128
	Nmeas	Nref	Ncent	%poss	C%poss	Mlplct	AnoCmpl	AnoFrc	AnoMlt	Rmeas	RmeasO	(Rsym)	PCV	PCVO

Figure 6.8: SCALA generated data.

From the above (figure 6.8) the data are reasonable to 2.8 Å resolution where the Rmerge is 0.296 and Mn(I) / δ(I) is 3.0. The data are ~88 % complete to the resolution with a redundancy of ~2.5.

A monoclinic $P2_1$ space group has molecules related by a 180º (i.e. two-fold) screw axis with a translation of half of the lattice vector (figure 6.9a and b). The equation for a certain point in the lattice is x, y, z. The equivalent point where the reflection appears is -x, y + $^1/_2$, -z, thus positioning it in a neighbouring unit cell.

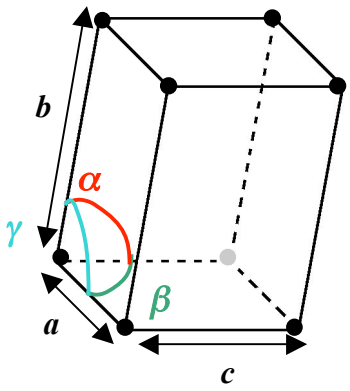


Figure 6.9a: A typical $P2_1$ monoclinic unit cell. Important features are a, b and c cell lengths. There are no constraints in this model i.e. sides can be different lengths. Angles α and γ = 90°, β varies.

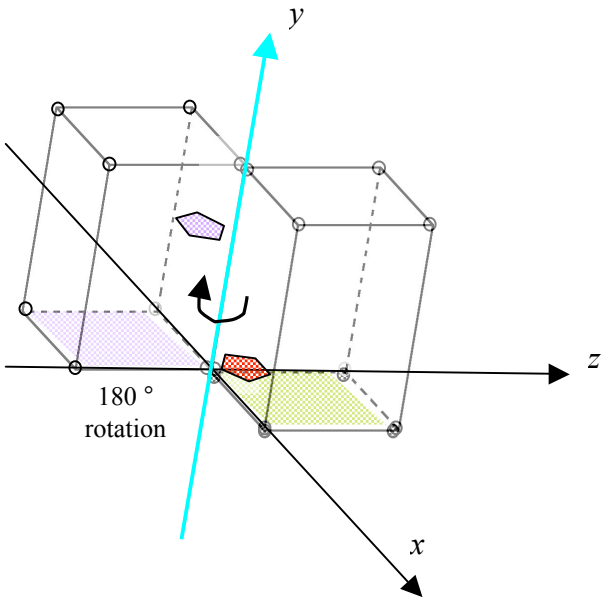


Figure 6.9b: Example of primitive monoclinic unit cell with a 2-fold screw axis. The arbitrarily chosen unit cell lengths are: a, b, c = 6, 10, 6. An object occuring at coordinate x, y, z (1, 2, 1) is shown in red and its reflection appears at –x, y+1/2, -z (-1, 7, -1), shown in lilac. The object has rotated 180° around the y-axis giving rise to two-fold symmetry, which is perpendicular to both the x and z axes. In addition, it has been translated by half a unit cell (b/2).

Mosaicity is a measure of the degree of long-range order of the unit cells within the crystal and of the order in the molecule itself. Lower mosaicity (<1°)

indicates better ordered crystals and hence better diffraction. This is estimated at 1.1° for the *ss*Ard1 crystal, showing that it is not so well ordered. The rate of the flash cooling step involved in cryogenic data collection can disrupt the dimensions of the unit cell and cause an increase in crystal mosaicity (Teng and Moffat, 1998). This may have contributed to the observed disorder factor, especially considering the nature of the cryoprotectant used (i.e. that it was not ideal for protection and likely caused the crystal to split).

R_{sym} is the error between reflection intensity measurements of symmetry related reflections, which should theoretically be identical (Rhodes, 2000). R_{merge} is a similar calculation but includes "all the measurements of symmetry related partners" over different data frames (Blow, 2002). Many authors use R_{sym} and R_{merge} interchangeably because there is little difference between the values. One hopes to attain a small R_{merge} value but it is highly dependent on how many times the reflections are measured i.e. the data redundancy (= number of reflections measured / number of unique reflections). The value will therefore increase with the number of measurements taken. Hence, a new calculation has been devised, R_{meas}, which is more accurate (Diederichs and Karplus, 1997) but many researchers still use R_{merge}. The *ss*Ard1 data display an R_{merge} of ~8 % at low resolution (>3.5 Å) and ~30 % at the high resolution shell (3.0-2.8Å), which is typical for protein crystals. But the data quality turns out to be quite reasonable given the overall value of ~8 %.

The Matthews coefficient, Vm is the volume of cell occupied per Dalton of protein and is on average ~2.4 $Å^3$ per Dalton, although it has been known for certain protein structures to vary from 1.9 to 4.2 $Å^3$/ Da (Blow, 2002). Given the MW of 19.5 kDa for *ss*Ard1, a single monomer in the asymmetric unit would give a Vm = 4.32 $Å^3$/ Da with a 72 % solvent content, whereas with two molecules in the

asymmetric unit the Vm = 2.16 $Å^3$ / Da with a 43 % solvent content. The asymmetric unit therefore most likely contains a dimer.

The atomic B-factor (also known as Debye-Waller factor) is a measure of the atomic disorder and is made up from the actual atomic motion of an atom and its statistical disorder throughout the crystal. A typical B-factor for a well ordered molecule may be about 20 $Å^2$ (Blow, 2002), allowing reasonably detailed resolution in the region of 2 Å. The Wilson B-factor is derived from a statistical analysis of the diffraction intensities and gives an idea of the overall atomic B-factor of the molecule. For the *ss*Ard1 dataset, the Wilson B-factor is ~60 $Å^2$. This means the crystal is very disordered (a combination of static disorder and that which is caused by thermal vibration), in agreement with the mosaicity value.

6.5 Co-crystallization of *ss*Ard1 with *ss*Alba1

Several hundred samples containing *ss*Ard1, *ss*Alba1 and acetyl-CoA, with the addition of many different precipitant solutions, were set up manually and by robotic assistance in order to grow crystals that would result in better diffraction and higher resolution data. Not only was it thought that this complex may provide a more structurally stable lattice, but it would also be interesting to identify the site of interaction between *ss*Ard1 and *ss*Alba1 at the molecular level to clarify and add to the observations made in Chapter 5.

Many nucleation sites occurred and about 50 % of the conditions produced needles. Attempts (as above in section 6.4) to change the morphology of the needles into more substantial rods failed. Some small blocks and rods did eventually appear but they remained very small. The crystals below (fig 6.10) were taken to the

Daresbury synchrotron for X-ray diffraction because they were far too small to be tested in-house.

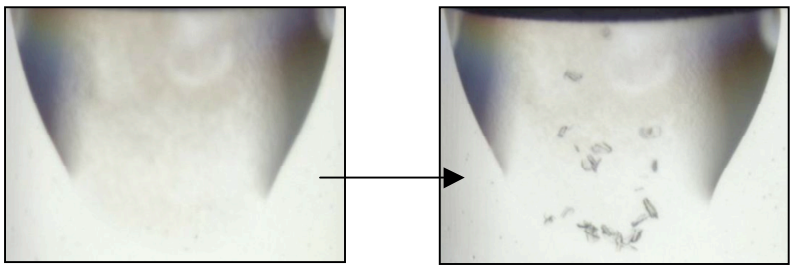


Figure 6.10: *ss*Ard1: *Sos*Alba1 co-crystals(?). 100 nL drop containing *ss*Ard1 (331.1 μM monomer = 7.3 mg/ml), *ss*Alba1 (152.2 μM monomer = 1.82 mg/ml) and acetyl-CoA (708 μM) in a precipitant solution of 0.05 M Cadmium sulphate, 0.1 M HEPES pH 7.5, 0.5 M Sodium acetate from (Qiagen) NeXtal Classics Lite.

However, no diffraction pattern was obtained to confirm the crystal identity to that of either salt or protein.

6.6 Solving the phase problem - are you MAD?

Seleno-methionine is often substituted in place of methionine when producing protein for crystallography. This is done because Selenium is chemically similar to Sulphur, but contains 34 electrons to sulphur's 16. This allows the use of anomalous scattering to identify the various locations of the Selenium atoms. Seleno-methionine has roughly the same properties as normal methionine and does not affect the protein structure in any way. This procedure is called MAD (multiple wavelength anomalous dispersion) phasing.

*ss*Ard1 was expressed in a methionine auxotroph in medium containing seleno-methionine and other amino acids. The protein was shown to express well (fig 6.11).

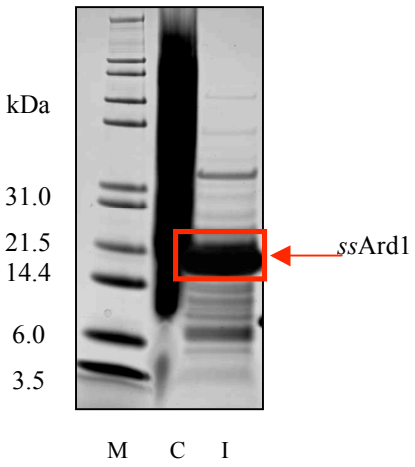


Figure 6.11: SDS-PAGE gel showing expression of *ss*Ard1 from a methionine auxotroph. *ss*Ard1 highlighted in red at molecular weight of ~19.5 kDa. M = Mark 12 molecular weight marker, C = uninduced control, I = Induced protein expression.

The 19.5 kDa band was excised and sent to the St. Andrews university mass spectrometry unit for analysis. No seleno-methionine residues were found to be incorporated into the enzyme. The whole procedure was repeated two more times but continued to fail. This experiment was not taken further.

6.7 Molecular Replacement

This technique relies upon the previously deposited data for proteins of similar function, which presumably share sequence homology and/or structural homology. One can use sequence alignments to come up with similar sequences that are already known. The PHYRE program attempts to fold the sequence of the target protein onto all of the unique protein structures in the protein database. It will then alert the user to a list of the most likely structural homologues within the protein data bank (PDB). These structural models can then be used in a program called PHASER (McCoy, 2007) to carry out molecular replacement against the diffraction data from *ss*Ard1. Molecular replacement uses a model structure to calculate "diffraction data" which is then compared to the actual diffraction data from *ss*Ard1. The process is divided into two stages: rotation and translation. PHASER produces a list of possible orientations

for the model in the *ss*Ard1 unit cell, and then attempts to find the position of each of these within the unit cell. PHASER orders its solutions by their Z-scores i.e. the number of standard deviations above the mean score. Each solution has a rotation function score (RFZ) and a translation function score (TFZ). PHASER then determines if the solutions pack in the unit cell i.e. whether there are any clashes between symmetry related molecules. In the case of *ss*Ard1, PHASER was asked to find two molecules in the asymmetric unit. The PHASER documentation suggests that if the Z-score is <5, the structure is not solved. If Z is 5-6 it is unlikely. If Z is 6-7 it is possible, whilst 7-8 is probable. A Z-score of >8 means the structure is definitely solved.

Six proteins were highlighted by PHYRE as having structural homology with *ss*Ard1. The PDB codes for these structures are: 1mk4 (probable acetyltransferase from *B. subtilis*), 1u6m (GNAT acetyltransferase from *E. faecalis*), 1ghe (Tabtoxin resistance protein, GNAT related from *P. syringae*), 1s3z, 1wk4 and 1z4e. Each of the above individual search models were used, as well as a combined model where PHASER is supplied with several models that have been superimposed. This has the effect of reinforcing the search model where there is similarity and reducing the weight of parts of the model that deviate. Running these ensembles in PHASER gives a translation function (TF) Z-score and a rotation function (RF) Z-score. TFZ is the first one to observe and if the value is above 5-6, the RF score is examined next. Table 6.2 lists the TF-Z scores for each model.

PDB file number	% identity	TFZ-score
1mk4	18	4.6
1u6m	19	4.0
1ghe	15	4.6
1s3z	15	4.0
1wk4	18	5.0
1z4e	19	5.0
With all cores		5.3

Table 6.2: PDB file numbers for structures with homology to *ss*Ard1. MSDfold SSM (Krissinel, 2004) was used to align the monomers for % sequence identity with *ss*Ard1. A TF Z-score <5 = structure not solved. 5-6 = unlikely.

None of the above models provided a suitable solution but it was possible to create a common core structure (figure 6.12) by trimming the superimposed models so that elements that deviated between the models were removed, resulting in a small increase in the TF-Z score. At only 5.3, it is unlikely (but not impossible) that we have a solved structure and it is likely that the basic core of *ss*Ard1 resembles that which is shown in figure 6.12.

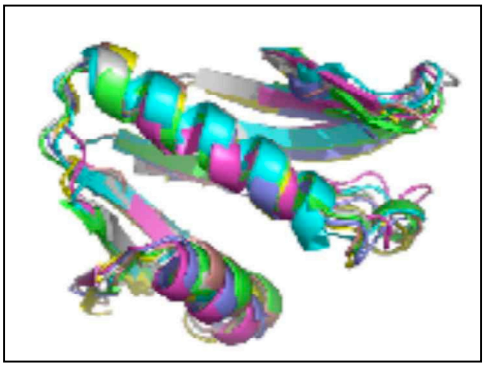


Figure 6.12: Superimposed core domain of GNAT related enzymes. Based on structural homology alignment of ssArd1 and seven other acetyltransferase enzymes (some with known function, some putative) by the Phyre programme, the core domain of the ssArd1 monomer will look very similar to what is shown on the left. 3 α-helical regions and 5 β-strands that are common in all the structures and very closely positioned when superimposed can be identified.

In late 2006 another acetylase was deposited in the protein databank, 2ob0, described as a human MAK3 homolog that shares 28 % sequence identity with *ss*Ard1. This structure was used as a search model in PHASER. When used as a complete structure there was no significant solution, with the highest TFZ score

being 5.0. However, a pruned search model was created using the program CHAINSAW which included complete side chains where there was identity with ssArd1, and truncated side chains (to Cβ or Cγ) where there was non-identity. Running PHASER with this model and including data between 15.0 and 4.0 Å gave a single solution for two monomers in the asymmetric unit with a TFZ score of 7.1 (figure 6.13). The two molecules pack well in the cell, but it was not possible to refine the solution, which may be due to the relatively poor quality of the diffraction data.

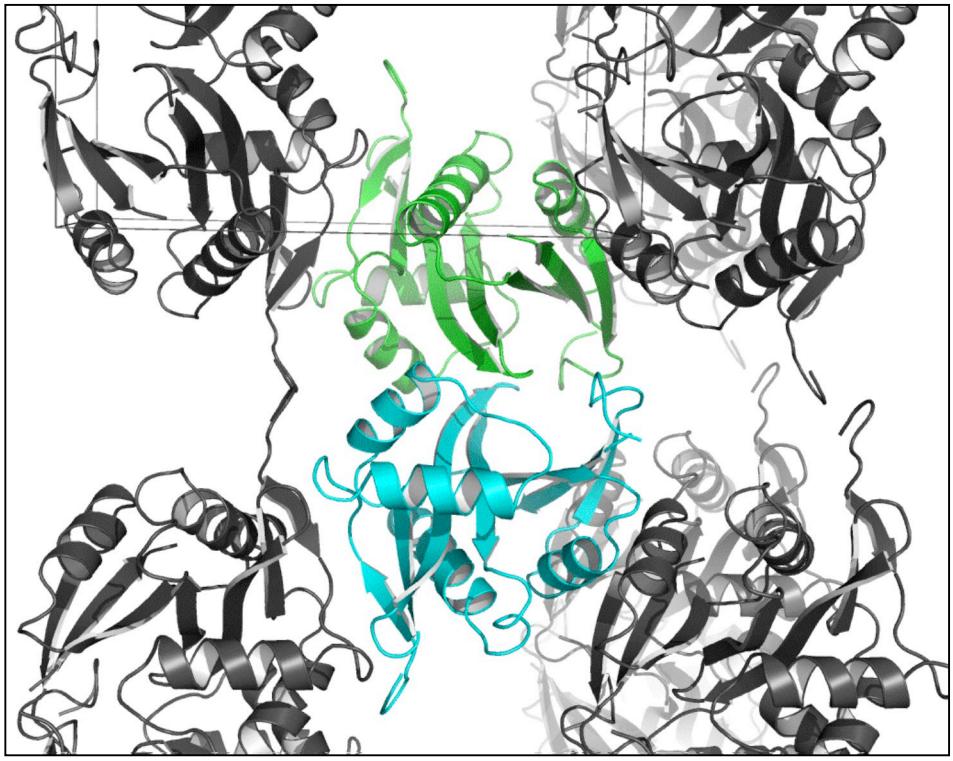


Figure 6.13: Crystal structure of *ss*Ard1 obtained from molecular replacement.

Search model PDB code: 2ob0. The packing of the two monomers in the unit cell are shown (green and cyan) with symmetry-related molecules shown in grey.

6.8 Summary and Conclusion

*ss*Ard1 seems to crystallise better in the presence of acetyl-CoA, indicating that this cofactor may help to stabilise the protein. Crystallisation proved to be very

difficult and the only viable data obtained was from a diffraction pattern with some slightly skewed reflections due to the combined nature of the disordered crystal, possible twinning and the fact that the cryo solution was not 100 % suitable for preventing damage to the lattice during the freezing process. The data suggests a $P2_1$ monoclinic unit cell with a volume suitable for containing two protein molecules per asymmetric unit. Data is complete to ~88 % and resolution was achieved to 2.8 Å. Attempts to gain phase information by incorporation of the heavy element selenium failed. Molecular replacement provided a plausible solution using a pruned model of a human acetylase, but unfortunately the model did not refine or progress any further, so the search is on for better quality crystals.

Chapter 7: Conclusions and Future work

7.1 Conclusions

Attempts to express soluble *Sso*0082 were terminated after several months. It is possible that this enzyme exists *in vivo* as a subunit of a protein complex. *ss*Ard1 expressed well but purification was initially problematic. A five step purification program was finally developed for this protein. Upon concentration of the protein solution, much precipitation occurred. For this reason, Circular Dichroism was carried out to ensure that aggregation was not a consequence of misfolded protein. These studies confirmed the presence of a folded globular protein wherein approximately half of its structural conformation was designated α-helical in nature. β-sheet content was estimated to be only 9-15 %, which is quite low compared to the predicted structure on the basis of sequence properties (PHYRE, section 1.5) and homology with a variety of proteins with solved structures e.g. yGcn5 (see Ch 4 & 5).

Since eukaryal Ard1 is the catalytic subunit belonging to a protein complex, NatA, it was fitting to acknowledge that the crenarchaeal enzyme may also have a binding partner(s). However, *ss*Ard1 exhibited acetyltransferase activity in the absence of other proteins. Further to this, no protein other than *Sso*Alba1, (and to a lesser degree *Sso*Alba2) interacted with column-bound *ss*Ard1. The majority of N-terminal acetylation in the eukarya is generally carried out by the three main NAT complexes NatA, B and C. Other known NAT complexes include AANAT/14-3-3, HAT1/HAT2 and PCAF, discussed in chapter 1. But there are also a variety of NATs and HATs that act alone e.g. Hpa2 and AAC.

*ss*Ard1 primarily modifies the N-terminal serine belonging to *ss*Alba1 and acetyltransferase activity is optimal at 55 °C *in vitro*. Longer incubation periods resulted in additional acetylation at several lysine sites. Secondary to a preference for

N-terminal serine; the enzyme is capable of modifying N-terminal alanine. *ss*Ard1 residues preferences are serine > alanine > methionine (-glu) and methionine (-leu) > valine > threonine > glycine > proline. Thus, not only does *ss*Ard1 possess an affinity towards NatA substrates, it also modifies NatB and NatC substrates, suggesting that it represents a common ancestor from which these complexes have evolved.

Kinetic experiments to attain values relative to enzyme activity made it possible to establish V*max* as ~8250 nmoles/min per mg of enzyme but it was not feasible to calculate an accurate value for K*m* because the velocity of reaction was so high that Alba1 was almost if not completely depleted in the earliest samples. *Km* is ≤ 11.425 μM. k*cat* shows multiple turnover and the calculated catalytic efficiency of *ss*Ard1 is estimated to be greater than 1600 $M^{-1}.sec^{-1}$.

Differential scanning calorimetry on acetylated Alba1 and non-acetylated Alba1 did not give any major insight into whether the protein was more stable in its acetylated form. Both samples had the same melting temperature, give or take 0.5 °C. The reason for *in vivo* acetylation of Alba1 probably lies within the realms of cell signalling and it is likely that acetylation affects the half life of the protein in addition to the more obvious consequences of lowering its binding affinity for DNA.

Point mutations of *ss*Ard1 resulted in the identification of the putative catalytic base histidine at position 88. The sequence homology with yGCN5 strongly suggests a similar catalytic mechanism. This would consist of deprotonation of the substrate primary amine by H88, which in turn causes the primary amine to carry out a direct nucleophilic attack on the acylcarbon of acetyl Co-enzyme A. This creates a tetrahedral intermediate between substrate, acetyl-CoA and enzyme before the acetyl group is ultimately transferred to the substrate.

S131-N132 lie close to the acetylCoA binding domain so may be involved in substrate binding, but it is likely that the decrease in activity in the S131A-N132A mutant could be attributed to poor cofactor binding. Mutation of D156 also results in a decrease in activity and has thus been postulated to engage in substrate recognition/binding. It is conserved in *y*GCN5 where it is situated on a loop with no co-factor contact. This loop (prior to β6) makes numerous contacts with substrate as does the loop between α1 and α2 (Wang, 1998). In accordance with the sequence alignment then, residues between 32 to 48 and 52 to 61 in *ss*Ard1 will be may be involved in substrate recognition.

*ss*Ard1 is a promiscuous N-terminal acetyltransferase that recognises protein structure as well as N-terminal sequence. It represents an ancestral form of NAT from which the three main acetyltransferase complexes NatA, NatB and NatC have derived their specific functions. Of the proteins it has been shown to modify thus far, *ss*Alba1 remains to be the preferred substrate. In addition, the enzyme is capable of acetylating many different proteins from eukaryotic and prokaryotic cell lysates but none of these could be identified during the course of this work.

Crystallisation of *ss*Ard1 was an arduous process, from which very few diffraction quality crystals developed. The data achieved was just under 3 Å resolution but disruption to the order of the crystal created problems during the data analysis steps. The calculated model implies a $P2_1$ space group with two monomers per asymmetric unit. Molecular replacement largely failed but did permit insight into the core structure of *ss*Ard1, presenting an arrangement of three α-helices and five β-strands.

7.2 Future work

Based on the comparison between *ss*Ard1 and *y*GCN5 sequences, further mutational studies would be useful in identifying regions specific for substrate recognition, expressly concentrating on residues located within the two loop sections that were mentioned previously in this chapter. As well as this, there are a string of other residues that need to be examined for their role in substrate binding i.e. N29, G64, Y65, M67, R69 and I70, S91, L126, L138 Y154, G157 and Y161.

It would be beneficial to carry out a non-radioactive kinetic experiment, perhaps using a coupled assay, to achieve better kinetic data and overcome the problems with substrate depletion. It was not possible to create a knockout of *ss*Ard1 in *Sulfolobus* but this can now be achieved due to recent scientific advances (Albers, 2007). This would enable changes in protein acetylation to be examined in a more convenient and reliable way, whilst also providing insight into the effect of ΔArd1 on cell viability.

And in terms of solving the crystal structure of the enzyme, it may be worth creating a truncated version of the protein because this has been tried and tested for many problematic samples in the past, often resulting in a solved structure. Among many examples of successfully crystallized truncated structures are the N-terminally truncated form of the DegS protein (a stress sensor of the bacterial periplasm) where crystals of the full length protein were not attainable (Grininger, 2004) and AcP (acylphosphatase) from *S. solfataricus* (Zuccotti, 2005).

References

Adler, E. M. (2007). "Making Life or Death Choices with CBP: CBP phosphorylation by IKK may be critical to the choice between proliferation and apoptosis." Sci. STKE **2007**(382): 132.

Albers, S. V. and Driessen, A. J. M. (2007). "Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome." Archaea **2**: 145-149.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2002). "Molecular Biology of the Cell."

Allfrey, V. G., Faulkner R., Mirsky, A. E. (1964). "Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis." *Proc Natl Acad Sci U S A*. **51**(5):786–794.

Alva, V., Ammelburg, M., Söding, J. and Lupas, A. N. (2007). "On the origin of the histone fold." BMC Structural Biology **7**: article number 17.

Angus-Hill, M. L., Dutnall, R. N., Tafrov, S. T., Sternglanz, R., Ramakrishnan, V. (1999). "Crystal structure of the histone acetyltransferase Hpa2: A tetrameric member of the Gcn5-related N-acetyltransferase superfamily." J Mol Biol **294**(5): 1311-25.

Arany, Z., Huang, L. E., Eckner, R., Bhattacharya, S., Jiang, C., Goldberg, M. A., Bunn, H. F., Livingston, D. M. (1996). "An essential role for p300/CBP in the cellular response to hypoxia." PNAS **93**: 12969-12973.

Aravind, L., Iyer, L. M., Anantharaman, V. (2003). "The two faces Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism." Genome biol **4**(10): R64.

Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M., Varhaug, J. E., Lillehaug, J. R. (2005). "Identification and characterization of the human ARD1–NATH protein acetyltransferase complex." Biochem J. **386**(Pt3): 433-443.

Arnold, R. J., Polevoda, B., Reilly, J. P., Sherman, F. (1999). "The action of N-terminal acetyltransferases on yeast ribosomal proteins." J Biol Chem. **274**(52): 37035-37040.

Arrhenius, S. (1889). Ueber die Dissociationswiirme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. Zeitschrift, F., Physik-Chemie, Bd. 4, S. 226.

Astbury, C., Bird, H. A., Beyeler, C. (1995). "Polymorphic acetylation: lack of influence of rheumatic disease activity and concomitant drug administration." Rheumatology Int. **14**(6): 257-260.

Avantaggiati, M. L., Ogryzko, V., Gardner, K., Giordano, A., Levine, A. S. and Kelly, K. (1997). "Recruitment of p300/CBP in p53-dependent signal pathways." Cell **89**: 1175-1184.

Bannister, A. J., and Kouzarides, T. (1996). "The CBP co-activator is a histone acetyltransferase." Nature **384**(6610): 641-643.

Bannister, A. J., Miska, E. A., Görlich, D., Kouzarides, T. (2000). "Acetylation of importin-alpha nuclear import factors by CBP/p300." Curr Biol. **10**(8): 467-470.

Barbaric, S., Walker, J., Schmid, A., Svejstrup, J. Q., Horz, W. (2001). "Increasing the rate of chromatin remodeling and gene activation--a novel role for the histone acetyltransferase Gcn5." EMBO J **20**(17): 4944-51.

Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R., Härd, T. (1995). "DNA-binding Surface of the *Sso*7d Protein from *Sulfolobus solfataricus*." J. Mol. Biol. **247**: 840-846.

Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R., Härd, T. (1994). "Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*." Nat Struct Biol. **1**(11): 808-819.

Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., White, M. F. (2002). "The Interaction of Alba, a Conserved Archaeal Chromatin Protein, with Sir2 and Its Regulation by Acetylation." Science **296**(5565): 148-151.

Bell, S. P., and B. Stillman. (1992). "ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex." Nature **357**: 128-134.

Berger, I., Bieniossek, C., Schaffitzel, C., Hassler, M., Santelli, E., Richmond, T. J. (2003). "Direct interaction of Ca2+/calmodulin inhibits histone deacetylase 5 repressor core binding to myocyte enhancer factor 2." J Biol Chem. **278**(20): 17625-17635.

Bernal, J. D., Crowfoot, D. (1934). "X-ray photographs of crystalline pepsin." Nature **133**: 794-795.

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T. and Schreiber, S. L. (2002). "Methylation of histone H3 Lys 4 in coding regions of active genes." PNAS **99**(13): 8695-8700.

Bjellqvist, B., Basse, B., Olsen, E., Celis, J. E. (1994). "Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. "Electrophoresis **15**(1):529-539.

Bjellqvist, B., Hughes, G., J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J-C. Frutiger, S., and Hochstrasser, D. (1993). "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences." Electrophoresis **14**(1):1023-1031

Blanco, J. C., Minucci, S., Lu, J., Yang, X. J., Walker, K. K., Chen, H., Evans, R. M., Nakatani, Y. and Ozato, K. (1998). "The histone acetylase PCAF is a nuclear receptor coactivator." Genes Dev **12**(11): 1638-1651.

Bloom, W., Fawcett, D. W. (1994). "A Textbook of Histology."

Blow, D. (2002). Outline of Crystallography for Biologists., Oxford University Press.

Bond, C. V., Kvaratskhelia, M., Richard, D., White, M. F., Hunter, W. M. (2001). "Structure of Hjc, a Holliday junction resolvase, from *Sulfolobus solfataricus*." PNAS USA. **98**(10): 5509-5514.

Borrow, J., Stanton, V. P. J., Andresen, J. M., Becher, R., Behm, F. G., Chaganti, R. S., Civin, C. I., Disteche, C., Dube, I., Frischauf, A. M., Horsman, D., Mitelman, F., Volinia, S., Watmore, A. E. and Housman. D. E. (1996). "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein." Nat. Genet. **14**(1): 33-41.

Branden, C. and Tooze, J. (1999). Introduction to Protein Structure. New York, Garland.

Brown, C. E., Lechner, T., Howe, L., Workman, J. L. (2000). "The many HATs of transcription coactivators." Trends Biochem Sci **25**(1): 15-9.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., Allis, C. D. (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-51.

Caillaud, A., Prakash, A., Smith, E., Masumi, A., Hovanessian, A. G., Levy, D. E., Marie, I. (2002). "Acetylation of interferon regulatory factor-7 by p300/CREB-binding protein (CBP)-associated factor (PCAF) impairs its DNA binding." J Biol Chem **277**(51): 49417-21.

Candau, R., Zhou, J. X., Allis, C. D., Berger, S. L. (1997). "Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo." EMBO J **16**(3): 555-65.

Carapeti, M., Aguiar, R. C., Goldman, J. M. and Cross, N. C. (1998). "A novel fusion between MOZ and the nuclear receptor coactivator TIF2 in acute myeloid leukemia." Blood **91**: 3127-3133.

Carrozza, M. J., Utley, R. T., Workman, J. L., Côté, J. (2003). "The diverse functions of histone acetyltransferase complexes." Trends Genet. **19**(6): 321-329.

Champagne, N., Bertos, N. R., Pelletier, N., Wang, A. H., Vezmar, M., Yang, Y., Heng, H. H. and Yang, X. J. (1999). "Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein. J. Biol. Chem. 274: 28528–28536." J Biol Chem **274**(40): 28528-28536.

Chan, H. M., Krstic-Demonacos, M., Smith, L., Demonacos, C., La Thangue, N. B. (2001). "Acetylation control of the retinoblastoma tumour-suppressor protein." Nat

Cell Biol. **3**(7): 667-674.

Chen, H., Lin, R. J., Schiltz, R. L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M. L., Nakatani, Y. and Evans, R. M. (1997). "Nuclear receptor co-activator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300." Cell **90**(3): 569-580.

Cheung, W. L., Briggs, S., Allis, C. D. (2000a). "Acetylation and chromosomal functions." Curr Opin Cell Biol **12**(3): 326-33.

Cheung, P., Allis, C. D. and Sansone-Corsi, P. (2000b). Signalling to chromatin through histone modifications. Cell, **103**: 263-271.

Choi, Y. B., Ko, J. K., Shin, J. (2004). "The transcriptional corepressor, PELP1, recruits HDAC2 and masks histones using two separate domains." J Biol Chem **279**(49): 50930-50941.

Choli, T., Wittmann-Liebold, B. and Reinhardt, R. (1998). "Microsequence analysis of DNA-binding proteins 7a, 7b, and 7e from the archaebacterium *Sulfolobus acidocaldarius."* J. Biol. Chem. **263**(15): 7087-7093.

Chung, J. G., Lo, H. H., Hsieh, S. E., Yen, Y. S. (1997). "Ibuprofen inhibits arylamine N-acetyltransferase activity in the bacteria *Klebsiella pneumoniae*." Curr Microbiol. **35**(4): 195-200.

Chung, Y. L., Lee, M. Y., Wang, A. J., and Yao, L. F. (2003). "A Therapeutic Strategy Uses Histone Deacetylase Inhibitors to Modulate the Expression of Genes Involved in the Pathogenesis of Rheumatoid Arthritis." Mol. Ther. **8**(5): 707-717.

Clarke, A. S., Lowell, J. E., Jacobson, S. J. and Pillus, L. (1999). "Esa1p is an essential histone acetyltransferase required for cell cycle progression." Mol Cell Biol. **19**(4): 2515-2526.

Clements, A. and Marmorstein, R. (2003). "Insights into structure and function of GCN5/PCAF and yEsa 1 histone acetyltransferase domains." Methods Enzymol **371**: 545-64.

Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., Marmorstein, R. (1999). "Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A." EMBO J **18**(13): 3521-32.

Coffee, B., Zhang, F., Warren, S. T. and Reines, D. (1999). "Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells." Nature Genetics **22**(2): 98-101.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Cryst. **D50**: 760-763.

Collins, S. J. (1998). "Acute promyelocytic leukemia: relieving repression induces remission." Blood **91**(8): 2631-2633.

Cooper, A., Nutley, M. A. and Wadood, A. (2000). "Differential scanning microcalorimetry in S.E. Harding and B. Z. Chowdhry (Eds.), Protein-Ligand Interactions ; hydrodynamics and calorimetry." Oxford University Press, Oxford New York, 287-318.

Creaven, M., Hans, F., Mutskov, V., Col, E., Caron, C., Dimitrov, S. and Khochbin, S. (1999). "Control of the histone-acetyltransferase activity of Tip60 by the HIV-1 transactivator protein. Tat. Biochemistry 38: 8826–8830." Biochemistry **38**(27): 8826-8830.

Cubeddu, L., White, M. F. (2005) "DNA damage detection by an archaeal single-stranded DNA-binding protein. " J. Mol. Biol. **353**(3):507–516.

Cuff, J. A., Barton, G. J. (1999). "Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction." Proteins **40**: 502-511.

Curradi, M., Izzo, A., Badaracco, G. and Landsberger, N. (2002). "Molecular Mechanisms of Gene Silencing Mediated by DNA Methylation." Mol. Cell Biol. **22**(9): 3157-3173.

Cvoro, A., Tzagarakis-Foster, C., Tatomer, D., Paruthiyil, S., Fox, M. S. and Leitman, D. C. (2006). "Distinct roles of unliganded and liganded estrogen receptors in transcriptional repression." Mol Cell **21**(4): 555-564.

De Angelis, J., Gastel, J., Klein, D. C., Cole, P. A.. (1998). "Kinetic analysis of the catalytic mechanism of serotonin N-acetyltransferase." J Biol Chem. **273**(5): 3045-3050.

de Ruijter, A. J., van Gennip, A. H., Caron, H. N., Kemp, S., van Kuilenburg, A. B. (2003). "Histone deacetylases (HDACs) HDACs): characterization of the classical HDAC family." Biochem J. **370**(Pt 3): 737-749.

Dick, J., and Reinhardt, R. (1986). "The structure of DNA-binding proteins from eu- and archaebacteria. In Bacterial chromatin." (eds. C.O. Gualerz and C.L. Pon): 185-218.

Diederichs, K., Karplus, P. A. (1997). "Improved R-factors for diffraction data analysis in macromolecular crystallography." Nat Struct Biol. **4**(4): 269-275.

Dionne, I., Nookala, R. K., Jackson, S., P., Doherty, A. J. and Bell, S. D. (2003). "A Heterotrimeric PCNA in the Hyperthermophilic Archaeon *Sulfolobus solfataricus*." Mol. Cell **11**(1): 275-282.

Drané, P., Leblanc, V., Miro-Mur, F., Saffroy, R., Debuire, B., May, E. (2002). "Accumulation of an inactive form of p53 protein in cells treated with TNF alpha." Cell Death Differ. **9**(5): 527-537.

Dressel, U., Bailey, P. J., Wang, S. C., Downes, M., Evans, R. M., Muscat, G. E. (2001). "A dynamic role for HDAC7 in MEF2-mediated muscle differentiation." J

Biol Chem. **276**(20): 17007-17013.

Dupret, J. M., Goodfellow, G. H., Janezic, S. A., Grant, D. M. (1994). "Structure-Function Studies of Human Arylamine N-Acetyltransferases NATl and NAT2. Functional analysis of recombinant NAT1/NAT2 chimeras expressed in Escherichia coli." J Biol Chem. **269**(43): 26830-26835.

Dupret, J. M., Grant, D. M. (1992). "Site-directed mutagenesis of recombinant human arylamine N-acetyltransferase expressed in Escherichia coli. Evidence for direct involvement of Cys68 in the catalytic mechanism of polymorphic human NAT2." J Biol Chem. **267**(11): 7381-7385.

Dutnall, R. N., Tafrov, S., Sternglanz, R., Ramakrishnan, V. (1998b). "Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related N-acetyltransferase superfamily." Cell **94**(4): 427-38.

Dutnall, R. N., Tafrov. S., Sternglanz, R. and Ramakrishnan, V. (1998a). "Structure of the yeast histone acetyltransferase Hat1: insights into substrate specificity and implications for the Gcn5-related N-acetyltransferase superfamily. Cold Spring Harb Symp Quant Biol 63:501-7 **63**: 501-7.

Dutta, A., and Bell, S. P. (1997). "Initiation of DNA replication in eukaryotic cells." Ann. rev. cell. dev. biol. **13**: 293-332.

Dyda, F., Klein, D. and Hickman, A. B. (2000). "GCN5-RELATED N-ACETYLTRANSFERASES: A Structural Overview." Ann. rev. biophys. & biomol. structure **29**: 81-103.

Eberhart, D. E. and Warren, S. T. (1996). "Nuclease sensitivity of permeabilized cells confirms altered chromatin conformation at the fragile X locus." Somat. Cell Mol. Genet. **22**(6): 435-441.

Eberharter A, S. D., Schieltz D, Hassan A, Yates JR 3rd, Berger SL, Workman JL. (1999). "The ADA complex is a distinct histone acetyltransferase complex in Saccharomyces cerevisiae." Mol Cell Biol. **19**(10): 6621-31.

Eberharter, A. and Becker, P. B. (2002). "Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics." EMBO Rep. 3(3): 224–229.

Edmondson, D. G., Davie, J. K., Zhou, J., Mirnikjoo, B., Tatchell, K., Dent, S. Y. (2002). "Site-specific loss of acetylation upon phosphorylation of histone H3." J Biol Chem **277**(33): 29496-502.

Edmondson, S. P., Shriver, J. W. (2001). "DNA binding proteins Sac7d and Sso7d from *Sulfolobus*." Methods Enzymol. **334**: 129-145.

Ehrenhofer-Murray, A. E., Rivier, D. H. and Rine, J. (1997). "The role of Sas2, an acetyltransferase homologue of Saccharomyces cerevisiae, in silencing and ORC function." Genetics **145**(4): 923-934.

Falb, M., Aivaliotis, M., Garcia-Rizo, C., Bisle, B., Tebbe, A., Klein, C., Konstantinidis, K., Siedler, F., Pfeiffer, F., Oesterhelt, D. (2006). "Archaeal N-terminal protein maturation commonly in involves N-terminal acetylation: a large-scale proteomics survey." J Mol Biol. **362**(5): 915-924.

Fan, Q., An, L., Cui, L. (2004). "Plasmodium falciparum histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling." Eukaryot Cell **3**(2): 264-76.

Ferry, G., Loynel, A., Kucharczyk, N., Bertin, S., Rodriguez, M., Delagrange, P., Galizzi, J. P., Jacoby, E., Volland, J. P., Lesieur, D., Renard, P., Canet, E., Fauchère, J. L., Boutin, J. A. (2000). "Substrate specificity and inhibition studies of human serotonin N-acetyltransferase." J Biol Chem **275**(12): 8794-8805.

Firestein, G. S., Echeverri, F., Yeo, M., Zvaifler, N. J. and Green, D. R. (1997). "Somatic mutations in the p53 tumor suppressor gene in rheumatoid arthritis synovium." PNAS **94**(20): 10895-10900.

Fischle, W., Dequiedt, F., Fillion, M., Hendzel, M. J., Voelter, W., Verdin, E. (2001). "Human HDAC7 histone deacetylase activity is associated with HDAC3 in vivo." J Biol Chem. **276**(38): 35826-35835.

Forsberg, E. C., Lam, L. T., Yang, X. J., Nakatani, Y., Bresnick, E. H. (1997). "Human histone acetyltransferase GCN5 exists in a stable macromolecular complex lacking the adapter ADA2." Biochemistry **36**(50): 15918-24.

Fu, H., Sadis, S., Rubin, D. M., Glickman, M., Nocker, S. V., Finley, D. and Vierstra, R. D. (1998). "Multiubiquitin Chain Binding and Protein Degradation Are Mediated by Distinct Domains within the 26 S Proteasome Subunit Mcb1." J. Biol. Chem. **273**(4): 1970-1981.

Fusi, P., Tedeschi, G., Alivert, A., Ronchi, S., Tortora, P., Guerritore, A. (1993). "Ribonuclease from the extreme thermophilic archaebacterium *Sulfolobus solfataricus*." Eur. J. Biochem. **211**: 305-310.

Gao, Y. G., Su, S. Y., Robinson, H., Padmanabhan, S., Lim, L., McCrary, B. S., Edmondson, S. P., Shriver, J. W., Wang, A. H. (1998). "The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA." Nat Struct Biol. **5**(9): 782-786.

Gautschi, M., Just, S., Mun, A., Ross, S., Rucknagel, P., Dubaquie, Y., Ehrenhofer-Murray, A., Rospert, S. (2003). "The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides." Mol Cell Biol. **23**(20): 7403-14.

Gaymes, T. J., Padua, R. A., Pla, M., Orr, S., Omidvar, N., Chomienne, C., Mufti, G. J., Rassool, F. V. (2006). "Histone deacetylase inhibitors (HDI) cause DNA damage in leukemia cells: a mechanism for leukemia-specific HDI-dependent apoptosis?" Mol. Cancer Res. **4**(8): 563-573.

Georgakopoulos, T. and Thireos, G. (1992). "Two distinct yeast transcriptional

activators require the function of the GCN5 protein to promote normal levels of transcription." EMBO J. **11**(11): 4145-4152.

Giles, R. H., Peters, D. J. and Breuning, M. H. (1998). "Conjunction dysfunc-dysfunction: CBP/p300 in human disease." Trends Genet. **14**: 178–183.

Giordano, A., and M. L. Avantaggiati. (1999). "p300 and CBP: partners for life and death." J. Cell. Physiol. **181**(2): 218-230.

Grant, P. A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F., Berger, S. L., Workman, J. L. (1997). "Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex." Genes Dev **11**(13): 1640-50.

Gregory, P. D., Schmid, A., Zavari, M., Liu, L., Berger, S. L. and Horz, W. (1998). "Absence of Gcn5 HAT activity defines a novel state in the opening ofchromatin at the PHO5 promoter in yeast." Mol Cell **1**: 495-505.

Grininger, M., Ravelli, R. B. G., Heider, U. and Zeth, K. (2004). "Expression, crystallization and crystallographic analysis of DegS, a stress sensor of the bacterial periplasm" *Acta Cryst* D60(8):1429-1431.

Grossman, S. R. (2001). "p300/CBP/p53 interaction and regulation of the p53 response." Eur. J. Biochem. **268**: 2773-2778.

Grote, M., Dijk, J. and Reinhardt., R. (1986). "Ribosomal and DNA binding proteins of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*." Biochim. Biophys. Acta. **873**(3): 405-413.

Grozinger, C. M., Hassig, C. A. and Schreiber, S. L. (1999). "Three proteins define a class of human histone deacetylases related to yeast Hda1p." Proc. Natl Acad. Sci. USA. **96**: 4868-4873.

Guagliardi, A., Cerchia, L., Moracci, M., Rossi, M. (2000). "The chromosomal protein *Sso*7d of the crenarchaeon *Sulfolobus solfataricus* rescues aggregated proteins in an ATP hydrolysis-dependent manner." J Biol Chem. **275**(41): 31813-31818.

Guarente, L. (2000). "Sir2 links chromatin silencing, metabolism, and aging." Genes Dev.(14): 9.

Guo, R., Xue, H., and Huang, L. (2003). "Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo." Mol. Microbiol. **50**(5): 1605–1615.

Han, X., Berardi, P. and Riabowol, K. (2006). "Chromatin Modification and Senescence: Linkage by Tumor Suppressors? " Rejuvenation Research, 9(1): 69-76.

Hansen, J. C., Tse, C. and Wolffe, A. P. (1998). "Structure and Function of the Core Histone N-termini: More than Meets the Eye." Biochemistry **37**: 17637-17641.

Harp, J. M., Timm, D. E. and Bunick, G. J. (1998). "Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography." Acta Crystallogr D Biol Crystallogr. **54**(4): 622-628.

Haynes, S. R., Dollard, C., Winston, F., Beck, S., Trowsdale, J. and Dawid. I. B. (1992). "The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. Nucleic Acids Res. 20: 2603." Nucleic Acids Res. **20**(get paper): 2603.

Hegde, S. S., Javid-Majd, F., Blanchard, J. S. (2001). "Overexpression and mechanistic analysis of chromosomally encoded aminoglycoside 2'-N-acetyltransferase (AAC(2')-Ic) from Mycobacterium tuberculosis." J Biol Chem. **276**(49): 45876-45881.

Herrera, J. E., Bergel. M., Yang, X. J., Nakatani, Y., Bustin, M. (1997). "The histone acetyltransferase activity of human GCN5 and PCAF is stabilized by coenzymes." J Biol Chem **272**(43): 27253-8.

Hickman, A. B., Klein, D. C., Dyda, F. (1999). "Melatonin Biosynthesis: the Structure of Serotonin N-Acetyltransferase at 2.5 A Resolution Suggests a Catalytic Mechanism." Mol. Cell **3**: 23-32.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. and Lucchesi, J. C. (1997). "mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila." EMBO J **16**: 2054-2060.

Hodawadekar, S. C. and Marmorstein, R. (2007). "Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design." Oncogene **26:** 5528–5540.

Hu, E., Chen, Z., Fredrickson, T., Zhu, Y., Kirkpatrick, R., Zhang, G. F., Johanson, K., Sung, C. M., Liu, R. and Winkler, J. (2000). "Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor." J. Biol. Chem. **275**: 15254-15264.

Huang, Y., Tan, M., Gosink, M., Wang, K. K., Sun, Y. (2002). "Histone deacetylase 5 is not a p53 target gene, but its overexpression inhibits tumor cell growth and induces apoptosis." Cancer Res. **62**(10): 2913-2922.

Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., Yoshida, M., Wang, X. F., Yao, T. P. (2003). "HDAC6 is a microtubule-associated deacetylase." Nature. **417**(6887): 455-458.

Iizuka, M., and Stillman, B. (1999). "Histone acetyltransferase HBO1 inter- interacts acts with the ORC1 subunit of the human initiator protein." J. Biol. Chem. **274**(33): 23027-23034.

Ito, A., Lai, C. H., Zhao, X., Saito, S., Hamilton, M. H., Appella, E., and Yao, T. P. (2001). "p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2." EMBO J **20**(6): 1331-1340.

Jacobson, S., and Pillus, L. (2004). "Molecular requirements for gene expression mediated by targeted histone acetyltransferases." Mol Cell Biol. **24**(13): 6029-39.

Jeanmougin, F., Wurtz, J.-M., Le Douarin, B., Chambon, P., and Losson., R. (1997). "The bromodomain revisited." Trends Biochem. Sci. **22**(get paper): 151-153.

Jelinska, C., Conroy, M. J., Craven, J. C., Hounslow, A., Bullough, P. A., Waltho, J. P., Taylor, G. L., White, M. F. (2005). "Obligate Heterodimerization of the Archaeal Alba2 Protein with Alba1 Provides a Mechanism for Control of DNA Packaging." Cell **13**(7): 963-971.

Jelinska, C. Unpublished results (p.35).

Jiang, H., Poirier, M. A., Liang, Y., Pei, Z., Weiskittel, C. E., Smith, W. W., DeFranco, D. B. and Ross, C. A. (2006). "Depletion of CBP is directly linked with cellular toxicity caused by mutant huntingtin." Neurobiol Dis. **23**(3): 543-551.

Kadosh, D. and Struhl, K. (1998). "Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo." Mol. Cell. Biol. **18**: 5121-5127.

Kamine, J., Elangovan, B., Subramanian, T., Coleman, D. and Chinnadurai, G. (1996). "Identification of a cellular protein that specifically interacts with the essential cysteine region of the HIV-1 Tat transactivator." Virology **216**(2): 357-366.

Kao, H. Y., Downes, M., Ordentlich, P. and Evans, R. M. (2000). "Isolation of a novel histone deacetylase reveals that class I and class II deacetylases promote SMRT-mediated repression." Genes Dev. **14**: 55-66.

Kawasaki, H., Song, J., Eckner, R., Ugai, H., Chiu, R., Taira, K., Shi, Y., Jones, N., and Yakoyama, K. K. (1998). "p300 and ATF-2 are components of the DRF complex, which regulates retinoic acid- and E1a-mediated transcription of the c-jun gene in F9 cells." Genes Dev. **12**: 233-245.

Keil, B. (1992). "Specificity of proteolysis", Springer-Verlag, Berlin, pp. ix, 336.

Kelley, L. A., MacCallum, R. M., Sternberg, M. J. E. (2000). "Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM." J Mol Biol **299**: 499 –522.

Kelley, R. L., and Kuroda, M. I. (1995). "Equality for X chromosomes. Science." Science **270:** 1607–1610.

Kelly, T. J., Qin, S., Gottschling, D. E., Parthun, M. R. (2000). "Type B histone acetyltransferase Hat1p participates in telomeric silencing." Mol Cell Biol. **20**(19): 7051-8.

Kim, I. A., Shin, J. H., Kim, I. H., Kim, J. H., Kim, J. S., Wu, H. G., Chie, E. K., Ha, S. W., Park, C. I., Kao, G. D. (2006). "Histone Deacetylase Inhibitor-Mediated Radiosensitization of Human Cancer Cells: Class Differences and the Potential

Influence of p53." Clin. Cancer Res. **12**: 940-949.

Kim, S., Ohkuni, K., Couplan, E., Jazwinski, S. M. (2004). "The histone acetyltransferase GCN5 modulates the retrograde response and genome stability determining yeast longevity." Biogerontology **5**(5): 305-16.

Kim, Y, Tanner. K. G., Denu, J. M. (2000). "A continuous, nonradioactive assay for histone acetyltransferases." Anal Biochem **280**(2): 308-14.

Kimura, A., and Horikoshi, M. (1998). "Tip60 acetylates six lysines of a specific class in core histones *in vitro*." Genes Cells **3**: 789-800.

Kimura, M., Kimura, J., Davie, P., Reinhardt, R. and Dijk, J. (1984). "The amino acid sequence of a small DNA binding protein from the archaebacterium Sulfolobus solfataricus." FEBS Lett. **176**(1): 176-178.

Kleff, S., Andrulis, E., Anderson, C. W. and Sternglanz, R. (1995). "Identification of a gene encoding a yeast histone H4 acetyltransferase." J Biol Chem **270**(42): 24674-7.

Klein, D. C. (2007). "Arylalkylamine N-Acetyltransferase: "the Timezyme"." J Biol. Chem. **282**(7): 4233-4237.

Klein, D. C., Coon, S. L., Roseboom, P. H., Weller, J. L., Bernard, M., Gastel, J. A., Zatz, M., Iuvone, P. M., Rodriguez, I. R., Bégay, V., Falcón, J., Cahill, G. M., Cassone, V. M., Baler, R. (1997). "The melatonin rhythm-generating enzyme: molecular regulation of serotonin N-acetyltransferase in the pineal gland." Recent Prog Horm Res. **52**: 307-357.

Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B. and Ladenstein, R. (1996). "Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile *Sulfolobus solfataricus*." Journal Mol. Biol. **264**(5):1132-1144.

Korzus, E., Torchia, J., Rose, D. W., Xu, L., Kurokawa, R., McInerney, E. M., Mullen, T. M., Glass, C. K., and Rosenfeld. M. G. (1998). "Transcription factor-specific requirements for coactivators and their acetyltransferase functions." Science **279**: 703-707.

Kovacs, J. J., Murphy, P. J. M., Zhao, X., Wu, J-T., Nicchitta, C. V., Yoshida, M., Toft, D. O., Pratt, W. B. and Yao, T-P. (2005). "HDAC6 Regulates Hsp90 Acetylation and Chaperone-Dependent Activation of Glucocorticoid Receptor." Mol. Cell. **18**: 601-607.

Krissinel, E. and Henrick, K. (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr. D Biol. Crystallogr. **60**: 2256-2268.

Kristjuhan, A., Walker, J., Suka, N., Grunstein, M., Roberts, D., Cairns, B. R., Svejstrup, J. Q. (2002). "Transcriptional inhibition of genes with severe histone h3 hypoacetylation in the coding region." Mol Cell **10**(4): 925-33.

Krueger, J. K., McCrary, B. S., Wang, A. H.-J., Shriver, J. W., Trewhella, J., and Edmondson, S. P. (1999). "The Solution Structure of the Sac7d/DNA Complex: A Small-Angle X-ray Scattering Study." Biochem **38**(32): 10247-10255.

Kuo M. H., Brownell, J. E., Sobel, R. E., Ranalli, T. A., Cook, R. G. and Edmondson. D. G. *et al.* (1996). "Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines." Nature **383**: 269-272.

Kuo, M. H., Allis, C. D. (1998). "Roles of histone acetyltransferases and deacetylases in gene regulation." Bioessays **20**(8): 615-26.

Kuo, M. H., vom Baur, E., Struhl, K., Allis, C. D. (2000). "Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription." Mol Cell **6**(6): 1309-20.

Kuo, M. H., Zhou, J., Jambeck, P., Churchill, M. E. A., and Allis, C. D. (1998). "Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo." Genes Dev **12**(5): 627-39.

Kvaratskhelia, M. and White, M. F. (2000) "Two holliday junction resolving enΩymes in *Sulfolobus solfataricus*." J. Mol. Biol. **297**(4): 923-932.

Lagger, G. Doetzlhofer, A., Schuettengruber, B., Haidweger, E., Simboeck, E., Tischler, J., Chiocca, S., Suske, G., Rotheneder, H., Wintersberger, E., Seiser, C. (2003). "The tumor suppressor p53 and histone deacetylase 1 are antagonistic regulators of the cyclin-dependent kinase inhibitor p21/WAF1/CIP1 gene." Mol Cell Biol. **23**(8): 2669-2679.

Landry, J., Sutton, A., Tafrov, S. T., Heller, R. C., Stebbins, J., Pillus, L., Sternglanz, R. (2000). "The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases." Proc Natl Acad Sci U S A. **97**(11): 5807-5811.

Langer, M. R, Fry. C. J., Peterson, C. L., Denu, J. M. (2002). "Modulating acetyl-CoA binding in the GCN5 family of histone acetyltransferases." J Biol Chem **277**(30): 27337-44.

Langer, M. R., Tanner, K. G., Denu, J. M. (2001). "Mutational analysis of conserved residues in the GCN5 family of histone acetyltransferases." J Biol Chem **276**(33): 31321-31.

Leal, C. A., Ayala-Madrigal, M. L., Figuera, L. E. and Medina, C. (1998). "Histone H4 acetylation analyses in patients with polysomy X: implications for the mechanism of X inactivation." Hum. Genet. **103**(1): 29-33.

Lee, F. J., Lin, L. W., Smith, J. A. (1989). "N-alpha acetylation is required for normal growth and mating of *Saccharomyces cerevisiae*." J Bacteriol **171**(11): 5795-802.

Lee, H., Sengupta, N., Villagra, A., Rezai-Zadeh, N., Seto, E. (2006). "Histone deacetylase 8 safeguards the human ever-shorter telomeres 1B (hEST1B) protein from ubiquitin-mediated degradation." Mol Cell Biol. **26**(14): 5259-5269.

Lemercier, C., Verdel, A., Galloo, B., Curtet, S., Brocard, M. P. and Khochbin, S. (2000). "mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity." J. Biol. Chem. **275**: 15594-15599.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB News-letter on Protein Crystallography, No. 26. **26**.

Leuba, S. H., Bustamante, C., Zlatanova, J. and van Holde, K. (1998). "Contributions of Linker Histones and Histone H3 to Chromatin Structure: Scanning Force Microscopy Studies on Trypsinized Fibers." Biophysics **74**(6): 2823-2829.

Liang, J., Prouty, L., Williams, B. J., Dayton, M. A. and Blanchard, K. L. (1998). "Acute mixed lineage leukemia with an inv(8)(p11q13) resulting in fusion of the genes for MOZ and TIF2." Blood **92**(get paper): 2118-2122.

Lin, Y., Fletcher. C. M., Zhou, J., Allis, C. D. and Wagner, G. (1999). "Solution structure of the catalytic domain of GCN5 histone acetyltransferase bound to coenzyme A." Nature **400**(6739): 86-9.

Little, G. H, Bai, Y., Williams, T., Poizat, C. (2007). "Nuclear calcium/calmodulin-dependent protein kinase IIdelta preferentially transmits signals to histone deacetylase 4 in cardiac cells." J Biol Chem. **282**(10): 7219-7231.

Liu, F., Dowling, M., Yang, X. J., Kao, G. D. (2004). "Caspase-mediated specific cleavage of human histone deacetylase 4." J Biol Chem. **279**(33): 34537-34546.

Liu, L., Scolnick, D. M., Trievel, R. C., Zhang, H. B., Marmorstein, R., Halazonetis, T. D., Berger, S. L. (1999). "p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage." Mol Cell Biol. **19**(2): 1202-1209.

Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R., Berger, S. L. (2000). "Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14." Mol Cell **5**(6): 917-26.

Lobley, A., Whitmore, L. and Wallace, B. A. (2002). "DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra." Bioinformatics **18**(1): 211-212.

López-Garcia, P., Knapp, S., Ladenstein, R. and Forterre, P. (1998). "In vitro DNA binding of the archaeal protein Sso7d induces negative supercoiling at temperatures typical for thermophilic growth." Nucleic Acids. Res. **26**(10): 2322-2328.

Lu, J., McKinsey, T. A., Nicol, R. L. and Olson, E. N. (2000). "Signal-dependent activation of the MEF2 transcription factor by dissociation from histone deacetylases." Proc. Natl Acad. Sci. USA. **97**: 4070-4075.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Lurz, R., Grote, M., Dijk, J., Reinhardt, R. & Dobrinski, B. (1986). "Electron microscopic study of DNA complexes with proteins from the Archaebacterium Sulfolobus acidocaldarius." EMBO J **5**: 3715-3721.

Magnaghi-Jaulin, L., Groisman, R., Naguibneva, I., Robin, P., Lorain, S., Le Villain, J. P., Troalen, F., Trouche, D., Harel-Bellan, A. (1998). "Retinoblastoma protein represses transcription by recruiting a histone deacetylase." Nature **391**(6667): 601-605.

Mai, A., Rotili, D., Tarantino, D., Ornaghi, P., Tosi, F., Vicidomini, C., Sbardella, G., Nebbioso, A., Miceli, M., Altucci, L., Filetici, P. (2006). "Small-molecule inhibitors of histone acetyltransferase activity: identification and biological properties." J Med Chem **49**(23): 6897-907.

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.

Marks, P. A., Rifkind, R. A., Richon, V. M. and Breslow, R. (2001). "Inhibitors of histone deacetylase are potentially effective anticancer agents." Clin. Cancer Res. **7**: 759-760.

Marmorstein (2001). "Structure of histone acetyltransferases." J Mol Biol **311**(3): 433-44.

Marsh, V. L., Peak-Chew, S. Y., Bell, S. D. (2005). "Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba." J Biol Chem. **280**(22): 21122-21128.

Martınez-Balbás, M. A., Bannister, A. J., Martin, K., Haus-Seuffert, P., Meisterernst, M. and Kouzarides, T. (1998). "The acetyltransferase activity of CBP stimulates transcription." EMBO J. 17:2886–2893

Martínez-Balbás M. A., Bauer U-M., Nielsen S. J., Brehm, A. and Kouzarides, T. (2000). "Regulation of E2F1 activity by acetylation". EMBO J, 19:662–671.

McAfee, J. G., Edmondson, S. P., Zegar, I. & Shriver, J. W. (1996). "Equilibrium DNA Binding of Sac7d Protein from the Hyperthermophile *Sulfolobus acidocaldarius*: fluorescence and circular dichroism studies." Biochemistry **35**(13): 4034-4045.

McCampbell, A., Taylor, J. P., Taye, A. A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G., Fischbeck, K. H. (2000). "CREB-binding protein sequestration by expanded polyglutamine." Hum. Mol. Genet. **9**(14): 2197-2201.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). "Phaser crystallographic software." J. Appl. Cryst. **40**: 658-674.

McCrary, B. S., Edmondson, S. P. and Shriver, J. W. (1996). "Hyperthermophile protein folding thermodynamics: Differential scanning calorimetry and chemical denaturation of Sac7d". Journal Mol. Biol. **264**(4): 784-805.

McRee, D. E. (1993). Practical Protein Crystallography. San Diego, Academic Press.

Miao, L., Fang, H., Li, Y., Chen, H. (2007). "Studies of the in vitro N?-acetyltransferase activities of *E. coli* RimL protein." Biochem Biophys Res Commun. **357**(3): 641-647.

Mizzen, C. A. and Allis, C.D. (1998). "Linking histone acetylation to transcriptional regulation." Cell Mol. Life Sci. (CMLS) **54**(1): 6-20.

Moazed, D. (2001). "Enzymatic activities of Sir2 and chromatin silencing." Curr. Op. Cell Biol. **13**: 232-238.

Moazed, D. and Noller, H. F. (1987). "Interaction of antibiotics with functional sites in 16S ribosomal RNA." Nature **327**: 389-394.

Moerschell, R. P., Hosokawa, Y., Tsunasawa, S. and Sherman, F. (1990). "The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo: processing of altered iso-1-cytochromes c created by oligonucleotide transformation." J Biol Chem. **265**: 19638-19643.

Moran, E. (1993). "DNA tumour virus transforming proteins and the cell cycle." Curr. Opin. Genet. Dev. **3**: 63-70.

Mullen, J. R., Kayne, P., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F., Sternglanz, R. (1989). "Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast." EMBO J **8**(7): 2067-75.

Muller-Ladner, U. and Nishioka, K. (2000). "p53 in rheumatoid arthritis: friend or foe?" Arthritis Res. **2**: 175-178.

Murata, T., Kurokawa, R., Krones, A., Tatsumi, K., Ishii, M., Taki, T., Masuno, M., Ohashi, H., Yanagisawa, M., Rosenfeld, M. G., Glass, C. K., Hayashi, Y. (2001). "Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome." Hum Mol Genet. **10**(10): 1071-1076.

Murphy, M., Ahn, J., Walker, K. K., Hoffman, W. H., Evans, R. M., Levine, A. J., George, D. L. (1999). "Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a." Genes Dev. **13**: 2490-2501.

Nakajima T, Aratani, S., Nakazawa, M., Hirose, T., Fujita, H., Nishioka, K. (2004). "Implications of transcriptional coactivator CREB binding protein complexes in rheumatoid arthritis." Mod Rheumatol. **14**(1): 6-11.

Nakazawa, M., Aratani, S., Hatta, M., Araya. N., Daitoku, H., Kawahara, K., Watanabe, S., Nakamura, H., Yoshino, S., Fujii, R., Fujita, H., Fukamizu, A., Nishioka, K. and Nakajima, T. (2002). "TNFalpha induces acetylation of p53 but

attenuates its transcriptional activation in rheumatoid synoviocytes." Int. J Mol. Med. **10**(3): 269-275.

Natsoulis, G., Winston, F., Boeke, J. D. (1994). "The SPT10 and SPT21 genes of Saccharomyces cerevisiae." Genetics **136**(1): 93-105.

Neuwald, A. F., Landsman, D. (1997). "GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein." Trends Biochem Sci **22**(5): 154-155.

Nicolas, E., Ait-Si-Ali, S., Trouche, D. (2001). "The histone deacetylase HDAC3 targets RbAp48 to the retinoblastoma protein." Nucleic Acids Res. **29**(15): 3131-3136.

Nightingale, K. P., Wellinger, R. E., Songo, J. M. and Becker, P. B. (1998). "Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin." EMBO J **17**(10): 2865-2876.

Ogryzko, V. V., Schlitz, R. L., Russanova, V., Howard, B. H. and Nakatani, Y. (1996). "The transcriptional coactivators p300 and CBP are histone acetyltransferases." Cell **87**(5): 953–959.

Osada, H., Tatematsu, Y., Saito, H., Yatabe, Y., Mitsudomi, T., Takahashi, T. (2004). "Reduced expression of class II histone deacetylase genes is associated with poor prognosis in lung cancer patients." Int. J. Cancer **112**(1): 26-32.

Pandey, R., Müller, A., Napoli, C. A., Selinger, D. A., Pikaard, C. S., Richards, E. J., Bender, J., Mount, D. W. and Jorgensen, R. A. (2002). "Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes." Nucleic Acids Res. **30**(23): 5036-5055.

Park, E. C., Szostak, J. W. (1992). "Ard1 and NAT1 proteins form a complex that has N-terminal acetyltransferase activity." EMBO J **11**(6): 2087-93.

Parthun, M. R., Widom, J., Gottschling, D. E. (1996). "The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism." Cell **87**(1): 85-94.

Pasqualucci, L., Bereschenko, O., Niu, H., Klein, U., Basso, K., Guglielmino, R., Cattoretti, G., Dalla-Favera, R. (2003). "Molecular pathogenesis of non-Hodgkin's lymphoma: the role of Bcl-6." Leuk. Lymphoma **44**(3): S5-12.

Pawlik, A., Ostanek, L., Brzosko, I., Gawroska-Szklarz, B., Brzosko, M. and Dbrowska-amojcin, E. (2002). "Increased genotype frequency of N-acetyltransferase 2 slow acetylation in patients with rheumatoid arthritis." Clinical Pharmacology & Therapeutics **72**(3): 319-325.

Petrij, F., Giles, R. H., Dauwerse, H .G., Saris, J. J., Hennekam, R. C., Masuno, M., Tommerup, N., van Ommen, G. J., Goodman, R. H., Peters, D. J. *et al.* (1995). "Rubenstein-Taybi syndrome caused by mutations in the transcriptional co-activator

CBP." Nature **376**(6538): 348-351.

Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., Sherman, F. (1999). "Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae." EMBO J **18**(21): 6155-68.

Polevoda, B., Sherman, F. (2001). "NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p." J Biol. Chem. **276**(23): 20154-9.

Polevoda, B., Sherman, F. (2002) "Diversity of Acetylated Proteins." Genome Biol. **3**(5): reviews 0006.1-0006.6.

Polevoda, B., Sherman, F. (2003). "Composition and function of the eukaryotic N-terminal acetyltransferase subunits." Biochem. Biophys. Res. Commun. **308**(1): 1-11.

Poux, A. N., Cebrat., M., Kim, C. M., Cole, P. A., Marmorstein, R. (2002). "Structure of the GCN5 histone acetyltransferase bound to a bisubstrate inhibitor." Proc Natl Acad Sci USA **99**(22): 14065-70.

Pozdeyev, N., Taylor, C., Haque, R., Chaurasia, S. S., Visser, A., Thazyeen, A., Du, Y., Fu, H., Weller, J., Klein, D. C., and Iuvone P. M. (2006). "Photic Regulation of Arylalkylamine N-Acetyltransferase Binding to 14-3-3 Proteins in Retinal Photoreceptor Cells." J. Neurosci.

Puri, P. L., Sartorelli, V., Yang, X. J., Hamamori, Y., Ogryzko, V. V., Howard, B. H., Kedes, L., Wang, J. Y., Graessmann, A., Nakatani, Y. and Levrero, M. (1997). "Differential roles of p300 and PCAF acetyltransferases in muscle differentiation." Mol Cell **1**: 35-45.

Qin, S. and Parthun., M. R. (2002). "Histone H3 and the histone acetyltransferase Hat1p contribute to DNA double-strand break repair." Mol Cell Biol. **22**(23): 8353-65.

Reeve, J. N. (2003). "Archaeal chromatin and transcription." Mol. Micro. **48**(3): 587-598.

Reifsnyder, C., Lowell, J., Clarke, A., and Pillus, L. (1996). "Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases." Nat. Genet. **14**: 42-49.

Rhodes, G. (1993). Crystallography made crystal clear. San Diego, Academic Press.

Rhodes, G. (2000). Crystallography Made Crystal Clear: A guide for users of macromolecular models. San Diego, Academic Press.

Richards, J. D., Cubeddu, L., Roberts, J., Liu, H. and White, M. F. (2007). "The archaeal XPB protein is a ssDNA dependent ATPase with a novel partner" J. Mol. Biol. **376**(3): 634-644.

Roberts, J. A., Bell, S. D. and White, M. F. (2003). "An archaeal XPF repair

endonuclease dependent on a heterotrimeric PCNA." Mol. Micro. **48** (2), 361-371.

Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D., Marmorstein, R. (1999). "Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide." Nature **401**(6748): 93-98.

Roth, S. Y., and Allis, C. D. (1996). "Histone acetylation and chromatin assembly: a single escort, multiple dances?" Cell **87**: 5-8.

Ruiz-Garcia, A. B., Sendra, R., Galiana, M., Pamblanco, M, Perez-Ortin, J. E., Tordera, V. (1998). "HAT1 and HAT2 proteins are components of a yeast nuclear histone acetyltransferase enzyme specific for free histone H4." J Biol Chem **273**(20): 12599-605.

Ruiz-Garcia A, B., S. R., Pamblanco M, Tordera V. (1997). "Gcn5p is involved in the acetylation of histone H3 in nucleosomes." Febs Lett. **403**(2): 186-90.

Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M. and Grunstein, M. (1996). "HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription." PNAS **93**: 14503-14508.

Sakaguchi, K., Herrera, J. E., Saito, S., Miki, T., Bustin, M., Vassilev, A., Anderson, C. W., Appella, E. (1998). "DNA damage activates p53 through a phosphorylation-acetylation cascade." Genes Dev. **12**(18): 2831-2841.

Sánchez-Puig N, F. A. (2006). "Characterization of the native and fibrillar conformation of the human Nalpha-acetyltransferase -ARD1." Protein Sci. **15**(8): 1968-1976.

Seemayer, C., Kuchen, S., Neidhart, M., Kuenzler, P., Rihoskova, V., Neumann, E., Pruschy, M., Aicher, W., Muller-Ladner, U., Gay, R., Michel, B., Firestein, G. and Gay, S. (2003). "p53 in rheumatoid arthritis synovial fibroblasts at sites of invasion." Ann. Rheum. Dis. **62**(12): 1139-1144.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., Van der Oost, J. (2001). "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2." Proc Natl Acad Sci U S A. **98**(14): 7835-7840.

Shikama, N., Lyon, J. and La Thangue, N. B. (1997). "The p300/CBP family: integrating signals with transcription factors and chromatin." Trends. Cell Biol. **7**: 230-236.

Smith, E. R., Eisen, A., Gu, W., Sattah, M., Pannuti, A., Zhou, J., Cook, R. G., Lucchesi, J. C. and Allis, C. D. (1998). "ESA1 is a histone acetyltransferase that is essential for growth in yeast." Proc. Natl. Acad. Sci. USA **95**(7): 3561-3565.

Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D., and Lucchesi, J. C. (2000). "The Drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation." Mol. Cell Biol. **20**(get paper): 312-318.

Song, O. K., Wang, X., Waterborg, J. H. and Sternglanz, R. (2003). "An N(alpha)-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A." J Biol Chem **278**(40): 38109-38112.

Spiegelberg, B. D., Hamm, H. E. (2005). "G betagamma binds histone deacetylase 5 (HDAC5) and inhibits its transcriptional co-repression activity." J Biol Chem. **280**(50): 41769-41776.

Sreerama, N., Venyaminov, S. Y. and Woody, R. W. (1999). "Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. " Protein Sci. **8**(22): 370-380.

Stein, R. W., Corrigan, M., Yaciuk, P., Whelan, J., Moran, E. (1990). "Analysis of E1A-mediated growth regulation functions: binding of the 300-kilodalton cellular product correlates with E1A enhancer repression function and DNA synthesis-inducing activity." J Virol. **64**(9): 4421-4427.

Sterner, D. E. and Berger., S. L. (2000). "Acetylation of histones and transcription-related factors." Microbiol Mol Biol Rev. **64**(2): 435-59.

Sternglanz, R. (1996). "Histone acetylation: a gateway to transcriptional activation." Trends Biochem Sci **21**(10): 357-8.

Sternglanz, R. and Schindelin, H. (1999). "Structure and mechanism of action of the histone acetyltransferase Gcn5 and similarity to other N-acetyltransferases." Proc Natl Acad Sci USA **96**(16): 8807-8.

Struhl, K. (1998). "Histone acetylation and transcriptional regulatory mechanisms." Genes Dev. **12**: 599-606.

Sugiura, N., Adams, S. M., Corriveau, R. A. (2003). "An evolutionarily conserved N-terminal acetyltransferase complex associated with neuronal development." J Biol Chem. **278**(41): 40113-40120.

Syntichaki, P. and Thireos. G. (1998). "The Gcn5.Ada complex potentiates the histone acetyltransferase activity of Gcn5." J Biol Chem **273**(38): 24414-9.

Tak, P. P., Zvaifler, N. J, Green, D. R and Firestein, G. S (2000). "Rheumatoid arthritis and p53: how oxidative stress might alter the course of inflammatory diseases." Immunol. Today **21**: 78-82.

Takei, Y., Swietlik, M., Tanoue, A., Tsujimoto, G., Kouzarides, T. and Laskey, R. (2001). "MCM3AP, a novel acetyltransferase that acetylates replication protein MCM3." EMBO reports **2**(2): 119-123.

Takemura, R., Okabe, S., Umeyama, T., Kanai, Y., Cowan, N. J. and Hirokawa, N. (1992). "Increased microtubule stability and tubulin acetylation in cells transfected with microtubule-associated proteins MAP1B, MAP2 or tau". J Cell Sci, 103:953–964.

Tanner, K. G., Langer, M. R., Kim, Y., Denu, J. M. (2000). "Kinetic mechanism of the histone acetyltransferase GCN5 from yeast." J Biol Chem **275**(29): 22048-55.

Tanner, K. G., Trievel, R. C., Kuo, M. H., Howard, R. M., Berger, S. L., Allis, C. D., Marmorstein, R., Denu, J. M. (1999). "Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator." J Biol Chem **274**(26): 18157-60.

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

Taylor, J. P., Taye, A. A., Campbell, C., Kazemi-Esfarjani, P., Fischbeck, K. H., Min, K. T. (2003). "Aberrant histone acetylation, altered transcription, and retinal degeneration in a Drosophila model of polyglutamine disease are rescued by CREB-binding protein." Genes Dev. **17**(12): 1463-1468.

Teng, T. Y. and Moffat., K. (1998). "Cooling Rates During Flash Cooling." J. Appl. Cryst. **31**: 252-257.

Teng, Y., Yachuan. Y., Ferreiro, J. A. and Waters, R. (2005). "Histone acetylation, chromatin remodeling, transcription and nucleotide excision repair in S. cerevisiae: studies with two model genes." DNA Repair (Amst) **4**(8): 870-83.

Tercero, J. C., Dinman. J. D., Wickner, R. B. (1993). "Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus." J Bacteriol **175**(10): 3192-4.

Tercero, J. C., Riles, L. E, Wickner, R. B. (1992). "Localized mutagenesis and evidence for post-transcriptional regulation of MAK3. A putative N-acetyltransferase required for double-stranded RNA virus propagation in Saccharomyces cerevisiae." J Biol Chem **267**(28): 20270-6.

Timmermann, S., Lehrmann, H., Polesskaya, A., Harel-Bellan, A. (2001). "Histone acetylation and disease." Cell. and Mol. life sci. **58**(5-6): 728-736.

Trievel, R. C., Rojas. J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L., Marmorstein, R. (1999). "Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator." Proc Natl Acad Sci USA **96**(16): 8931-6.

Trueblood, C. E., Wright, R. M. and Poyton, R. O. (1998). "Differential Regulation of the Two Genes Encoding *Saccharomyces cerevisiae* Cytochrome C Oxidase Subunit V by Heme and the HAP2 and REOJ Genes." Mol. and Cell. Biol. **8**(10): 4537-4540.

Tse, C., Georgieva. E. I., Ruiz-Garcia, A. B., Sendra, R., Hansen, J. C. (1998).

"Gcn5p, a transcription-related histone acetyltransferase, acetylates nucleosomes and folded nucleosomal arrays in the absence of other protein subunits." J Biol Chem **273**(49): 32388-92.

Tse, C., Sera, T., Wolffe, A. P. and Hansen, J. C. (1998). "Disruption of Higher-Order Folding by Core Histone Acetylation Dramatically Enhances Transcription of Nucleosomal Arrays by RNA Polymerase III." Mol Cell Biol. **18**(8): 4629-4638.

Tulinsky, A. (1999). "The Protein Structure Project, 1950-1959: First Concerted Effort Of a Protein Structure Determination In the U.S." The Rigaku Journal. **16**(1): 1950-1959.

Utley, R. T., Ikeda, K., Grant, P. A., Cote, J., Steger, D. J., Eberharter, A., John, S., Workman, J. L. (1998). "Transcriptional activators direct histone acetyltransferase complexes to nucleosomes." Nature **394**(6692): 498-502.

Van Lint, C., Emiliani, S. and Verdin, E. (1996). "The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation." Gene Expr. **5**: 245-253.

van't Hoff, J. H. (1884). Etudes de Dynamique Chimique, Amsterdam, F. Muller & Co. pg 112.

Vatsis, K. P., Weber, M. K. (1991). "Diverse point mutations in the human gene for polymorphic N-acetyltransferase." Proc Natl Acad Sci U S A. **88**(44): 633306337.

Verdel, A. and Khochbin, S. (1999). "Identification of a new family of higher eukaryotic histone deacetylases. Coordinate expression of differentiation-dependent chromatin modifiers." J. Biol. Chem. **274**: 2440-2445.

Verreault, A., Kaufman, P. D., Kobayashi, R. and Stillman, B. (1996). "Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4." Cell **87**(Get this paper!): 95-104.

Vetting, M., Magnet, E., Nieves, S., Roderick, S., Blanchard, J. (2004). "A bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones. Chem Biol 11(4):565-73." Chem. & Biol. **11**(4): 565-573.

Vetting, M. W., de Carvalho, L. P., Roderick, S. L., Blanchard, J. S. (2005). "A novel dimeric structure of the RimL Nalpha-acetyltransferase from Salmonella typhimurium." J Biol Chem. **280**(23): 22108-22114.

Visuri, K., Kaipainen, E., Kivimäki, J., Niemi, H., Leisola, M., Palosaari, S. (1990). "A new method for protein crystallization using high pressure." Biotechnology (N Y). **8**(6): 547-549.

Vogelauer, M. R. L., Lucas, I., Brewer, B. J., Grunstein, M. (2002). "Histone acetylation regulates the time of replication origin firing." Mol Cell **10**(5): 1223-33.

Vogelauer, M., W.. J., Suka, N., Grunstein, M. (2000). "Global histone acetylation and deacetylation in yeast." Nature **408**(6811): 495-8.

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

Walia, H., Chen, H. Y., Sun, J-M., Holth, L. T. and Davie, J. R. (1998). "Histone Acetylation Is Required to Maintain the Unfolded Nucleosome Structure Associated with Transcribing DNA." J Biol Chem **273**(23): 14516-14522.

Walsh, C. T. (2007). "Posttranslational Modification of Proteins. Expanding Nature's Inventory." 152.

Waltregny D, G., Tran, S. L., North, B. J., Verdin, E., Colige, A., Castronovo, V.. (2005). "Histone deacetylase HDAC8 associates with smooth muscle alpha-actin and is essential for smooth muscle cell contractility." FASEB J. **19**(8): 966-968.

Wang, A. H., Bertos, N.R., Vezmar, M., Pelletier, N., Crosato, M., Heng, H.H., Th'ng, J., Han, J. and Yang, X.J. (1999). "HDAC4, a human histone deacetylase related to yeast HDA1, is a transcriptional corepressor." Mol. Cell. Biol. **19**: 7816-7827.

Wang, H., Huang, Z-Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B. D., Briggs, S. D., Allis, C. D., Wong, J., Tempst, P. and Zhang, Y. (2001). "Methylation of Histone H4 at Arginine 3 Facilitating Transcriptional Activation by Nuclear Hormone Receptor." Science **293**(5531): 853-857.

Wang, L., Berger, S. L. (1998). "Critical residues for histone acetylation by Gcn5, functioning in Ada and SAGA complexes, are also required for transcriptional function in vivo." Genes Dev **12**(5): 640-53.

Wang, L. M., Ying, C., Candau, R., Barlev, N., Brownell, J., Allis, C. D., Berger, S. L. (1997). "Histone acetyltransferase activity is conserved between yeast and human GCN5 and is required for complementation of growth and transcriptional activation." Mol Cell Biol. **17**(1): 519-27.

Wardleworth, B. N., Russell, R.J.M., Bell, S.D., Taylor, G.L. and White, M.F. (2002). "Structure of Alba: and archaeal chromatin protein modulated by acetylation." EMBO J **21**(17): 4654-4662.

Watanabe, M., Sofuni, T., Nohmi, T. (1992). "Involvement of Cys69 residue in the catalytic mechanism of N-hydroxyarylamine O-acetyltransferase of Salmonella typhimurium. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for S. typhimurium and higher organisms." J Biol Chem. **267**(12): 8429-8436.

Westendorp, M. O., Shatrov, V. A., Schulze-Osthof, K., Frank, R., Kraft, M., Los, M., Krammer, P. H., Droge, W. and Lehmann, V. (1995). "HIV-1 Tat potentiates TNF-induced NF-kappa B activation and cytotoxicity by alter- alteringing the cellular redox state." EMBO J **14**(3): 546-554.

White, M. F., and Bell, S.D. (2002). "Holding it together: chromatin in the archaea." Trends Genet. **18**: 621-626.

Whiteway, M., Freedman, R., Van Arsdell, S., Szostak, J. W. and Thorner, J. (1987). "The yeast ARD1 gene product is required for repression of cryptic mating-type information at the HML locus." Mol Cell Biol. **7**(10): 3713-22.

Whitman, W. B., Pfeifer, F., Blum, P., Klein, A. (1999). "What archaea have to tell biologists." Genetics. **152**(4): 1245-1248.

Whitmore, L. and Wallace, B. A. (2004). "DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data." Nucleic Acids Research **32**: W668-673.

Winkler, G. S., Kristjuhan. A., Erdjument-Bromage, H., Tempst, P., Svejstrup, J. Q. (2002). "Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels *in vivo*." Proc Natl Acad Sci USA **99**(6): 3517-22.

Wisniewski, J. R., Zougman, A., Krueger, S. and Mann, M. (2006). "Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations and phosphorylations as well as differences between cell culture and tissue." Molecular and Cellular Proteomics **6**(1): 72-87.

Wittschieben, B. O., Otero, G., de Bizemont T, Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C. D., Tempst, P., Svejstrup, J. Q. (1999). "A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme." Mol Cell **4**(1): 123-8.

Wolf, E., Vassilev, A., Makino, Y., Sali, A., Nakatani, Y., Burley, S.K. (1998). "Crystal structure of a GCN5-related N-acetyltransferase: Serratia marcescens aminoglycoside 3-N-acetyltransferase." Cell **94**(4): 439-49.

Xu, W., D. G. Edmondson, and S. Y. Roth. (1998). "Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. Mol. Cell. Biol. 18: 5659–5669." Mol Cell Biol. **18**(10): 5659-5669.

Xue, H., Guo, R., Wen, Y., Liu, D., Huang, L. (2000). "An Abundant DNA Binding Protein from the Hyperthermophilic Archaeon Sulfolobus shibatae Affects DNA Supercoiling in a Temperature-Dependent Fashion." J Bacteriol **182**(14): 3929-3933.

Yamamoto, T., and M. Horikoshi. (1997). "Novel substrate specificity of the histone acetyltransferase activity of HIV-1-Tat interactive protein Tip60." J. Biol. Chem. **272**(49): 30595-30598.

Yamanishi, Y., Boyle, D.L., Pinkoski, M.J., Mahboubi, A., Lin, T., Han, Z., Zvaifler, N.J., Green, D.R., Firestein, G.S. (2002). "Regulation of joint destruction and inflammation by p53 in collagen-induced arthritis." Am J Pathol. **160**(1): 123-130.

Yan Y, B. N., Haley RH, Berger SL, Marmorstein R (2000). "Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases." Mol Cell **6**(5): 1195-205.

Yang, X.-J., Ogryzko, V. V., Nishikawa, J-I., Howard, B.H. and Nakatani, Y. (1996). "A p300/CBP-associated factor that competes with the adenoviral E1A oncoprotein." Nature **382**: 319-324.

Yu Y, T. Y., Liu H, Reed SH, Waters (2005). "UV irradiation stimulates histone acetylation and chromatin remodeling at a repressed yeast locus." Proc Natl Acad Sci USA **102**(24): 8650-5.

Zhang W, B. J., Edmondson DG, Turner BM, Roth SY (1998). "Essential and redundant functions of histone acetylation revealed by mutation of target lysines and loss of the Gcn5p acetyltransferase." EMBO J **17**(11): 3155-67.

Zhao, K., Chai, X., Marmorstein, R. (2003). "Structure of a Sir2 Substrate, Alba, Reveals a Mechanism for Deacetylation-induced Enhancement of DNA Binding*." J Biol Chem **278**(28): 26071-26077.

Zhao, X., Ito, A., Kane, C.D., Liao, T.S., Bolger, T.A., Lemrow, S.M., Means, A.R., Yao, T.P.. (2001). "The modular nature of histone deacetylase HDAC4 confers phosphorylation-dependent intracellular trafficking." J Biol Chem. **276**(37): 35042-35048.

Zhou, X., Richon, V.M., Wang, A.H., Yang, X.J., Rifkind, R.A. and Marks, P.A. (2000). "Histone deacetylase 4 associates with extracellular signal-regulated kinases 1 and 2 and its cellular localization is regulated by oncogenic Ras." Proc. Natl Acad. Sci. USA. **97**: 14329-14333.

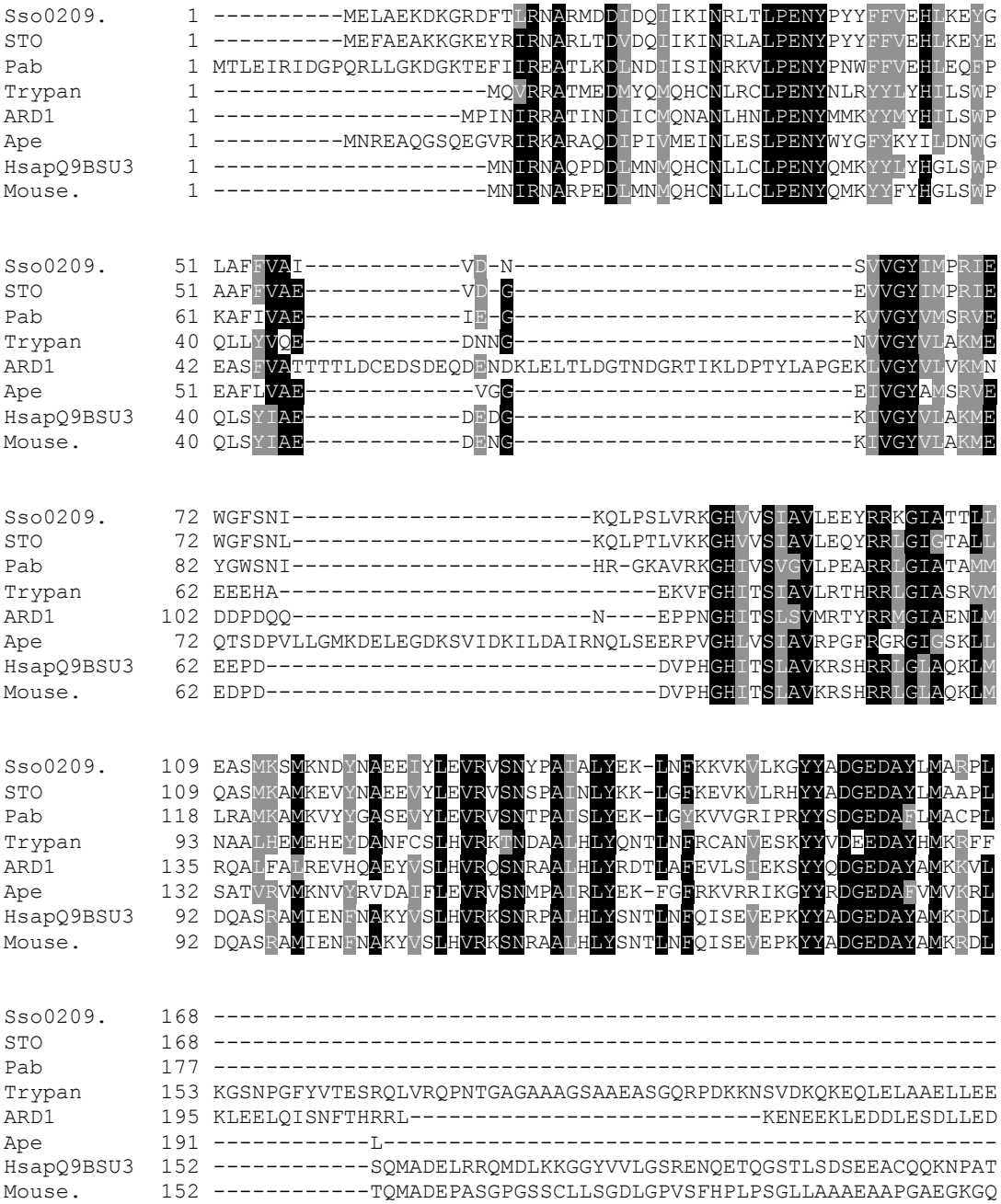
Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H. and Scholz, I. (1980). "The Sulfolobus-Caldariella group - taxonomy on the basis of the structure of DNA-dependent RNA-polymerases." Archives of microbiology **125**(3): 259-269.

Zuccotti, S., Rosano, C., Bemporad, F., Stefani, M. and Bolognesi, M. (2005). "Preliminary characterization of two different crystal forms of acylphosphatase from the hyperthermophile archaeon *Sulfolobus solfataricus*." *Acta Cryst.* F61(1):144-146

Zwiebel, J. A. (2000). "New agents for acute myelogenous leukaemia." Leukaemia **14**(3): 488-490.

Appendix 1

Sso0209 blast search. Black shaded residues indicate absolute identity, grey shaded residues indicate similarity.



```
Sso0209.          ------------------------------
STO               ------------------------------
Pab               ------------------------------
Trypan       213 DLKSSGKGRRQPHQQHQKGGKHGKGKK----
ARD1         228 IIKQGVNDIIV--------------------
Ape               ------------------------------
HsapQ9BSU3   200 EESGSDSKE-PKESVESTNVQDSSESSDSTS
Mouse.       200 AHGSGGLGE-QSGEQRQRAFELRRGLS----
```

*Sso*0209: *Sulfolobus solfataricus ss*Ard1 N-acetyltransferase, STO: *Sulfolobus tokodaii* 167aa long hypothetical N-terminal acetyltransferase, Pab: *Pyrobaculum aerophilum* str. IM2 176aa N-acyltransferase, Trypan: *Trypanosoma brucei* putative N-acetyltransferase subunit Ard1, Ard1: *Saccharomyces cerivisiae* N-terminal acetyltransferase complex, Ape: Aeropyrum pernix (strain K1) N-terminal acetyltransferase complex subunit APE1954, HsapQ9BSU3: *Homo sapiens* MGC10646 protein, Mouse: *Mus musculus* Ard1 protein.

Appendix II

*ss*Ard1 alignment with other Ard1 homologues.

```
CLUSTAL W (1.83) multiple sequence alignment


ssArd1                                  ---------------------------------MELAEKDKGRD-FTLRNA 17
gi|15920446|ref|NP_376115.1|            ---------------------------------MEFAEAKKGKE-YRIRNA 17
gi|70606292|ref|YP_255162.1|            ---------------------------------MEITEDSKRKINYQIRLA 18
gi|14601741|ref|NP_148282.1|            ---------------------------------MN-REAQGSQEGVRIRKA 17
gi|67466743|ref|XP_649513.1|            -----------------------------------------MFTIRKA 7
gi|51593357|gb|AAH80651.1|              PSTRFSPVVLAPWTRFLRQTVPLKVGDSLFWVHCSPTGSLPSAIMNIRNA 50
gi|149701444|ref|XP_001492559.          ---------------------------------------------MNIRNA 6
gi|20071196|gb|AAH27219.1|              ---------------------------------------------MNIRNA 6
gi|10835057|ref|NP_003482.1|            ---------------------------------------------MNIRNA 6
gi|149029890|gb|EDL85002.1|             ---------------------------------------------MNIRNA 6
gi|47086985|ref|NP_998499.1|            ---------------------------------------------MNIRNA 6
gi|58267770|ref|XP_571041.1|            ---------------------------------------------MDIRQA 6
gi|71657918|ref|XP_817467.1|            ---------------------------------------------MQIRRA 6
gi|15240705|ref|NP_196882.1|            ----------------------------------------MVCIRRA 7
gi|7649677|emb|CAB89123.1|              ---------------------------------------------MQVRRA 6
gi|94468972|gb|ABF18335.1|              ---------------------------------------------MNIRCA 6
gi|126276259|ref|XP_001386974.         ----------------------------------------MGITIRQA 8
                                                                                    :* *


ssArd1                                  RMDDIDQIIKINRLTLPENYPYYFFVEHLKEYGLAFFVAI---------- 57
gi|15920446|ref|NP_376115.1|            RLTDVDQIIKINRLALPENYPYYFFVEHLKEYEAAFFVAE---------- 57
gi|70606292|ref|YP_255162.1|            TLSDIDQIIRINRSALPENYPYYFFVEHLKEYGQAFYVAD---------- 58
gi|14601741|ref|NP_148282.1|            RAQDIPIVMEINLESLPENYWYGFYKYILDNWGEAFLVAE---------- 57
gi|67466743|ref|XP_649513.1|            TPADLPAIQNVNLTNLPENYNLQLYYYHLILYP-TSFVAV---------- 46
gi|51593357|gb|AAH80651.1|              QPDDLMNMQHCNLLCLPENYQMKYYLYHGLSWPQLSYIAE---------- 90
gi|149701444|ref|XP_001492559.          RPDDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE---------- 46
gi|20071196|gb|AAH27219.1|              RPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE---------- 46
gi|10835057|ref|NP_003482.1|            RPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE---------- 46
gi|149029890|gb|EDL85002.1|             RPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE---------- 46
gi|47086985|ref|NP_998499.1|            RPEDLMNMQHCNLLCLPENYQMKYYFYHGLSWPQLSYIAE---------- 46
gi|58267770|ref|XP_571041.1|            TIDDLLGMQNANLLNLPENYTFKYYLYHALTWPELSYVAV---------- 46
gi|71657918|ref|XP_817467.1|            TMEDMYQMQHCNLRCLPENYNLRYYFYHILSWPQLLYVQE---------- 46
gi|15240705|ref|NP_196882.1|            TVDDLLAMQACNLMCLPENYQMKYYLYHILSWPQLLYVAE---------- 47
gi|7649677|emb|CAB89123.1|              TMEDMYQMQHCNLRCLPENYNLRYYLYHILSWPQLLYVQE---------- 46
gi|94468972|gb|ABF18335.1|              KPEDLMNMQHCNLLCLPENYQMKYYFYHGLTWPQLSYVAE---------- 46
gi|126276259|ref|XP_001386974.         TINDIQAMQNANLHNLPENYQLKYYMYHILSWPQASFVATTYEDIVESND 58
                                           *:  :   *   *****    :      :     :


ssArd1                                  -------------VDN---------SVVGYIMPRIEWG------------ 73
gi|15920446|ref|NP_376115.1|            -------------VDG---------EVVGYIMPRIEWG------------ 73
gi|70606292|ref|YP_255162.1|            -------------LEG---------EVVGYVMPRIEWG------------ 74
gi|14601741|ref|NP_148282.1|            -------------VGG---------EIVGYAMSRVEQTSDPVLLGMKDEL 85
gi|67466743|ref|XP_649513.1|            ------------TPDN---------KVVGYCLTKIEDDDP---------- 65
gi|51593357|gb|AAH80651.1|              ------------DEDG---------KIVGYVLAKMEE-EP---------- 108
gi|149701444|ref|XP_001492559.          ------------DEDG---------KIVGYVLAKMEE-DP---------- 64
gi|20071196|gb|AAH27219.1|              ------------DENG---------KIVGYVLAKMEE-DP---------- 64
gi|10835057|ref|NP_003482.1|            ------------DENG---------KIVGYVLAKMEE-DP---------- 64
gi|149029890|gb|EDL85002.1|             ------------DENG---------KIVGYVLAKMEE-DP---------- 64
gi|47086985|ref|NP_998499.1|            ------------DENG---------KIVGYVLAKMEE-DP---------- 64
gi|58267770|ref|XP_571041.1|            ------------DPKG---------RIVGYILAKMEEEP---------- 64
gi|71657918|ref|XP_817467.1|            ------------DYNR---------NIVGYVLAKMED-EEK--------- 65
gi|15240705|ref|NP_196882.1|            ------------DYNG---------RIVGYVLAKMEE-ES--------- 65
gi|7649677|emb|CAB89123.1|              ------------DNNG---------NVVGYVLAKMEE-EEH--------- 65
gi|94468972|gb|ABF18335.1|              ------------DDKG---------NIVGYVLAKMEEPEPG--------- 66
gi|126276259|ref|XP_001386974.         TEAGEEFGGEKEDPKGDSAYINRGEKIVGYVLGKMEDDPEAE-------- 100
                                                                  :*** : ::*
```

```
ssArd1                          ---FSNIKQLPSLVR--------KGHVVSIAVLEEYRRKGIATTLLEASM 112
gi|15920446|ref|NP_376115.1|    ---FSNLKQLPTLVK--------KGHVVSIAVLEQYRRLGIGTALLQASM 112
gi|70606292|ref|YP_255162.1|    ---FSNLKHIPSLVR--------KGHIVSIAVLEPFRKIGVGTSLLQNSL 113
gi|14601741|ref|NP_148282.1|    EGDKSVIDKILDAIRNQLSEERPVGHLVSIAVRPGFRGRGIGSKLLSATV 135
gi|67466743|ref|XP_649513.1|    -------------------HPVVTGQVTSISVLRTYRRLGIATKLIRAAE 96
gi|51593357|gb|AAH80651.1|      -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 139
gi|149701444|ref|XP_001492559.  -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 95
gi|20071196|gb|AAH27219.1|      -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 95
gi|10835057|ref|NP_003482.1|    -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 95
gi|149029890|gb|EDL85002.1|     -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 95
gi|47086985|ref|NP_998499.1|    -------------------DDVPHGHITSLAVKRSHRRLGLAQKLMDQAS 95
gi|58267770|ref|XP_571041.1|    -------------------SDAPSGHVTSISVLRPYRRLGLANKLMKQSQ 95
gi|71657918|ref|XP_817467.1|    -------------------PANIHGHITSIAVLRTHRRLGIASRVMRASM 96
gi|15240705|ref|NP_196882.1|    --------------------NECHGHITSLAVLRTHRKLGLATKLMTAAQ 95
gi|7649677|emb|CAB89123.1|      -------------------AEKVFGHITSIAVLRTHRRLGIASRVMNAAL 96
gi|94468972|gb|ABF18335.1|      -------------------EESTHGHITSLAVKRSYRRLGLAQKLMNQAS 97
gi|126276259|ref|XP_001386974.  -------------------DKTPHGHITSLSVMRTYRRMGIAEKLMRQSL 131
                                                    *::.*::*   .*  *:.  ::  :

ssArd1                          KSMKNDYNAEEIYLEVRVSNYPAIALYEK-LNFKKVKVLKGYYADGEDAY 161
gi|15920446|ref|NP_376115.1|    KAMKEVYNAEEVYLEVRVSNSPAINLYKK-LGFKEVKVLRHYYADGEDAY 161
gi|70606292|ref|YP_255162.1|    KAMKDTYNAEEVYLEVRVTNYPAISLYKK-FNFREVKLLKHYYADGEDAY 162
gi|14601741|ref|NP_148282.1|    RVMKNVYRVDAIFLEVRVSNMPAIRLYEK-FGFRKVRRIKGYYRDGEDAF 184
gi|67466743|ref|XP_649513.1|    NSMIEVFGARAMMLQVRVSNQPALHLYEKTIGFTVTKVSKHYYLDGEDAL 146
gi|51593357|gb|AAH80651.1|      RAMIENFNAKYVSLHVRKSNRPALHLYSNTLNFQISEVEPKYYADGEDAY 189
gi|149701444|ref|XP_001492559.  RAMIENFSAQYVSLHVRKSNRAALHLYSNTLNFQVSEVEPKYYADGEDAY 145
gi|20071196|gb|AAH27219.1|      RAMIENFNAKYVSLHVRKSNRAALHLYSNTLNFQISEVEPKYYADGEDAY 145
gi|10835057|ref|NP_003482.1|    RAMIENFNAKYVSLHVRKSNRAALHLYSNTLNFQISEVEPKYYADGEDAY 145
gi|149029890|gb|EDL85002.1|     RAMIENFNAKYVSLHVRKSNRAALHLYSNTLNFQISEVEPKYYADGEDAY 145
gi|47086985|ref|NP_998499.1|    RAMIENFNAKYVSLHVRKSNRAALHLYSNTLKFQISEVEPKYYADGEDAY 145
gi|58267770|ref|XP_571041.1|    EAMVAHYDAHHITLHVRKSNRAAISLYRDTLGFEVHGMEKSYYADGEDAY 145
gi|71657918|ref|XP_817467.1|    KEMEKEYDANYCSLHVRKTNDAALHLYQETLGFRCANVEKGYYVDEEDAF 146
gi|15240705|ref|NP_196882.1|    AAMEQVYEAEYVSLHVRRSNRAAFNLYTETLGYKINDVEAKYYADGEDAY 145
gi|7649677|emb|CAB89123.1|      HEMEHEYDANFCSLHVRKTNDAALHLYQNTLNFRCANVESKYYVDEEDAY 146
gi|94468972|gb|ABF18335.1|      HAMVECFNAQYVSLHVRKSNRAALNLYTNSLGFRILEIEPKYYADGEDAY 147
gi|126276259|ref|XP_001386974.  YAMCESFDAQYVSLHVRKSNRAALHLYRDSLNFEVTSIEKSYYQDGEDAY 181
                                  *   : .    *.** :* .*: ** . : :        ** * ***

ssArd1                          LMARPL------------------------------------------- 167
gi|15920446|ref|NP_376115.1|    LMAAPL------------------------------------------- 167
gi|70606292|ref|YP_255162.1|    LMAAPL------------------------------------------- 168
gi|14601741|ref|NP_148282.1|    VMVKRLL------------------------------------------ 191
gi|67466743|ref|XP_649513.1|    ILTHNFTLDTLIN---------------------------------- 159
gi|51593357|gb|AAH80651.1|      AMKRDLSQMADELRRQMD-LKKGGYVVLGSRENQETQG---------- 226
gi|149701444|ref|XP_001492559.  AMKRDLAQMADELRRQLE-QRKGGYVV-GSRENQETQG---------- 181
gi|20071196|gb|AAH27219.1|      AMKRDLTQMADEPAS----GPGSSCLLSGDLGPVSFHPL---------- 180
gi|10835057|ref|NP_003482.1|    AMKRDLTQMADELRRHLELKEKGRHVVLGAIENKVESKG---------- 184
gi|149029890|gb|EDL85002.1|     AMKRDLTQ------------------------PQGLAPL---------- 160
gi|47086985|ref|NP_998499.1|    AMKRNLTQMADELQ-------KPGVRLWGSEAPPSQDTS---------- 177
gi|58267770|ref|XP_571041.1|    GMRYVFKKPE------------------------------------- 155
gi|71657918|ref|XP_817467.1|    HMKKFFKGPNPGLYVAAN----RQLVRQQNSAAAAVTGMAALQRNGGTAP 192
gi|15240705|ref|NP_196882.1|    DMRKNLKGKQN------------------------------------- 156
gi|7649677|emb|CAB89123.1|      HMKRFFKGSNPGFYVTES----RQLVRQPNTGAGAAAG---------- 180
gi|94468972|gb|ABF18335.1|      SMRRDLAELAELYG-------------------------------- 161
gi|126276259|ref|XP_001386974.  AMRLTLELDEL------------------------------------ 192
                                  :   :

ssArd1                          -------------------------------------------------
gi|15920446|ref|NP_376115.1|    -------------------------------------------------
gi|70606292|ref|YP_255162.1|    -------------------------------------------------
gi|14601741|ref|NP_148282.1|    -------------------------------------------------
gi|67466743|ref|XP_649513.1|    -------KDCSSVVVDEWKRVMQEQENKQKE------------------ 183
gi|51593357|gb|AAH80651.1|      ---STLSDSEEACQ-QKNPATEESGSDSKEPKE---SVESTNVQDSSESS 269
gi|149701444|ref|XP_001492559.  ---STHPGSEGACQEQKISAADDSGSDSKEPSE---SPESTDVQDSSEDS 225
gi|20071196|gb|AAH27219.1|      ---PSGLLAAAEAAPG-AEGKGQAHG-SGGLGEQSGEQRQRAFELRRGLS 225
gi|10835057|ref|NP_003482.1|    ---NSPPSSGEACREEKGLAAEDSGGDSKDLSEVSETTESTDVKDSSEAS 231
gi|149029890|gb|EDL85002.1|     ---VSCLET------------------------------------- 166
gi|47086985|ref|NP_998499.1|    ---VTGLVEKLTVQDGEKEGDGDSGGESKEMSEVSEATESTDVKDSS--S 222
gi|58267770|ref|XP_571041.1|    -----------ESLKE-------------------------------- 160
gi|71657918|ref|XP_817467.1|    PPPPPSSSSAGGQR-GENDTADLEKEQLKLVAELLEESNKESGKGRQQKQ 241
gi|15240705|ref|NP_196882.1|    --------HHHAHGHHHHHGGGCCSGDAKVVET----AQAVDGKAVSK-- 192
gi|7649677|emb|CAB89123.1|      -----SAAEASGQRPDKKNSVDKQKEQLELAAELLEEDLKSSGKGRRQPH 225
gi|94468972|gb|ABF18335.1|      -----AGREHSDAVISNHTGGSHAGHDGHCC------------------- 187
gi|126276259|ref|XP_001386974.  ----------LPSLAQKGERDDLSKDLLESL------------------- 213
```

```
ssArd1                              -----------------
gi|15920446|ref|NP_376115.1|        -----------------
gi|70606292|ref|YP_255162.1|        -----------------
gi|14601741|ref|NP_148282.1|        -----------------
gi|67466743|ref|XP_649513.1|        -----------------
gi|51593357|gb|AAH80651.1|          DSTS------------- 273
gi|149701444|ref|XP_001492559.      DSAS------------- 229
gi|20071196|gb|AAH27219.1|          -----------------
gi|10835057|ref|NP_003482.1|        DSAS------------- 235
gi|149029890|gb|EDL85002.1|         -----------------
gi|47086985|ref|NP_998499.1|        DS--------------- 224
gi|58267770|ref|XP_571041.1|        -----------------
gi|71657918|ref|XP_817467.1|        KQQQQSKGGNKRGKTKK 258
gi|15240705|ref|NP_196882.1|        -----------------
gi|7649677|emb|CAB89123.1|          QQHQ--KGG-KHGKGKK 239
gi|94468972|gb|ABF18335.1|          -----------------
gi|126276259|ref|XP_001386974.     -----------------
```

Identities of proteins [and their host organisms] from alignment (in order of occurrance):

hypothetical N-terminal acetyltransferase [*Sulfolobus tokodaii* str. 7]
gi|15920446|ref|NP_376115.1|[15920446]
N-acetyltransferase [*Sulfolobus acidocaldarius* DSM 639]
gi|70606292|ref|YP_255162.1|[70606292]
N-terminal acetyltransferase [*Aeropyrum pernix* K1]
gi|14601741|ref|NP_148282.1|[14601741]
N-terminal acetyltransferase [*Entamoeba histolytica* HM-1:IMSS]
gi|67466743|ref|XP_649513.1|[67466743]
ARD1B protein [*Homo sapiens*]
gi|51593357|gb|AAH80651.1|[51593357]
PREDICTED: similar to ARD1 homolog B protein [*Equus caballus*]
gi|149701444|ref|XP_001492559.1|[149701444]
Ard1 protein [*Mus musculus*]
gi|20071196|gb|AAH27219.1|[20071196]
N-acetyltransferase, homolog of S. cerevisiae ARD1 [*Homo sapiens*]
gi|10835057|ref|NP_003482.1|[10835057]
N-acetyltransferase ARD1 homolog (*S. cerevisiae*) (predicted), isoform CRA_b [*Rattus norvegicus*]
gi|149029890|gb|EDL85002.1|[149029890]
ARD1 homolog a, N-acetyltransferase [*Danio rerio*]
gi|47086985|ref|NP_998499.1|[47086985]
ard1 family protein [*Cryptococcus neoformans* var. *neoformans* JEC21]
gi|58267770|ref|XP_571041.1|[58267770]
N-acetyltransferase complex ARD1 subunit [*Trypanosoma* cruzi strain CL Brener]
gi|71657918|ref|XP_817467.1|[71657918]
GCN5-related N-acetyltransferase, putative [*Arabidopsis thaliana*]
gi|15240705|ref|NP_196882.1|[15240705]
putative N-acetyltransferase subunit ARD1 [*Trypanosoma brucei*]
gi|7649677|emb|CAB89123.1|[7649677]
major N alpha-acetyltransferase subunit [*Aedes aegypti*]
gi|94468972|gb|ABF18335.1|[94468972]
N-terminal acetyltransferase complex ARD1 subunit (Arrest-defective protein 1) [*Pichia stipitis* CBS 6054]
gi|126276259|ref|XP_001386974.1|[126276259]