

Technische Universität München

Max-Planck-Institut für Biochemie

**Structural and Functional Studies of the Yeast Class II Hda1
Histone Deacetylase Complex**

**X-ray Structure of the Secreted Aspartic Protease Sap5
from *Candida albicans***



Jung-Hoon Lee

2009

Martinsried / München

Technische Universität München

Max-Planck-Institut für Biochemie

**Structural and Functional Studies of the Yeast Class II Hda1
Histone Deacetylase Complex**

**X-ray Structure of the Secreted Aspartic Protease Sap5
from *Candida albicans***

Jung-Hoon Lee

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

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender : Univ.-Prof. Dr. Johannes Buchner
Prüfer der Dissertation: 1. apl. Prof. Dr. Dr. h. c. Robert Huber, i.R.
2. Univ.-Prof. Dr. Michael Groll

Die Dissertation wurde am 14.10.2009 bei der Technischen Universität München
eingereicht und durch die Fakultät für Chemie am 12.11.2009 angenommen.

Most of the work described has been published:

Jung-Hoon Lee, Klaus Maskos, and Robert Huber (2009)

Structural and functional studies of the yeast class II Hda1 histone deacetylase complex

J. Mol. Biol. **391**, 744 ~ 757

Claudia Borelli, Elisabeth Ruge, Jung-Hoon Lee, Martin Schaller, Alexandra Vogelsang, Michel Monod, Hans Christian Korting, Robert Huber, and Klaus Maskos (2008)

X-ray structures of Sap1 and Sap5: Structural comparison of the secreted aspartic proteinases from *Candida albicans*

Proteins **72**, 1308 ~ 1319

Table of Contents

Acknowledgement

Abbreviations

Part I. Structural and functional studies of the yeast class II Hda1 histone deacetylase complex

Summary	1
1. Introduction	2
1.1. Gene regulation by acetylation and deacetylation of chromatin	2
1.2. Sequence-specific and global histone deacetylations in yeast	3
1.3. Classification of histone deacetylases in mammal and yeast	5
1.4. <i>Saccharomyces cerevisiae</i> class II histone deacetylase complex	7
1.5. Helicase and SNF2 family: structural overview	7
1.6. SWI2/SNF2 chromatin-remodeling domain of Rad54 enzymes	9
1.7. Structure determination by X-ray crystallography	9
1.8. Crystal twinning	10
2. Materials and Methods	16
2.1. Materials	16
2.1.1. Chemicals, enzymes, primers, and equipments	16
2.2. Methods I : molecular biology, biochemistry, and biophysics	16
2.2.1. Cloning and protein expression	16
2.2.2. Protein purification	17
2.2.3. Limited proteolysis and N-terminal sequencing	18
2.2.4. Sedimentation velocity	18
2.2.5. Circular dichroism spectroscopy	18
2.2.6. Mutagenesis	18
2.2.7. Deacetylation assay	19

2.2.8. Design of oligonucleotides for DNA-binding assay	19
2.2.9. DNA-binding assay	20
2.3. Methods II : X-ray crystallography	21
2.3.1. Programs, screening solutions, and equipments	21
2.3.2. Crystallization of wild-type scHda3pDBD3 and its triple mutant	21
2.3.3. Preparation of SeMet derivative	21
2.3.4. Data collection and processing	23
2.3.5. Structure determination by MAD phasing	23
3. Results	24
3.1. Cloning and expression of scHda1p, scHda2p, and scHda3p genes	24
3.2. Purification of recombinant Hda1 HDAC complex and its subunits	26
3.3. <i>In vitro</i> deacetylase activity of recombinant Hda1 HDAC complex	27
3.4. Determination of stoichiometry between three subunits in solution	28
3.5. Domain organization of yeast class II Hda1 HDAC complex	29
3.6. Identification of scHda3pDBD3	32
3.7. Crystallization and structure determination of scHda3pDBD3 ^{K168A/Q169A/K170A}	33
3.8. Crystal structure of scHda3pDBD3 ^{K168A/Q169A/K170A}	36
3.9. Structural comparison of scHda3pDBD3 ^{K168A/Q169A/K170A} with SWI2/SNF2 enzymes	38
3.10. DNA-binding activities of Hda1 HDAC complex and it subunits	41
3.11. Identification of the DNA-binding sites of scHda3pDBD3	42
4. Discussion	44
5. References	49

Part II. X-ray structure of the secreted aspartic protease Sap5 from *Candida albicans*

Summary	53
1. Introduction	54
1.1. Proteolytic activity as virulence factor of <i>Candida albicans</i>	54
1.2. Classification of Saps in <i>Candida albicans</i>	54
1.3. Catalytic mechanism of aspartic proteases	55
1.4. Known structures of Sap enzymes in <i>Candida albicans</i>	56
1.5. Aspartic protease inhibitors	57
2. Materials and Methods	58
2.1. Crystallization of <i>Candida albicans</i> Sap5-pepstatin A complex	58
2.2. Data collection and processing	58
2.3. Structure determination by molecular replacement	58
3. Results	59
3.1. Crystallization and structure determination of Sap5-pepstatin A complex	59
3.2. Overall structure of Sap5-pepstatin A complex	60
3.3. The active site of Sap5	61
3.4. Structural comparison of Sap5 with the other Sap enzymes	64
4. Discussion	67
5. References	70
Curriculum vitae	75

Acknowledgement

I am deeply grateful to my supervisor Prof. Dr. Robert Huber for giving me the opportunity to work as a Ph.D student in his laboratory and learn how to set up a new laboratory, organize the research plans, and manage most of things independently for last three years. I would deeply appreciate of his scientific guidance, inspiration, and valuable advices during my Ph.D studies.

Especially, I am indebted to my former mentor Prof. Dr. Yunje Cho in Pohang University of Science and Technology (POSTECH, South Korea) for his previous scientific guidance, inspiration, enthusiasm, and valuable advices when I was scientifically immature. I do not doubt that my attitude towards biological science was established at that time.

Special thanks to the Peter und Traudl Engelhorn-Stiftung, Max-Planck society, and Proteros Biostructures for the financial support.

My thanks to Mrs. Renate Rueller and Mrs. Monika Schneider who friendly helped me for many things and even were happy to see my wife and baby every time.

I would like to extend my thanks to Dr. Klaus Maskos, Dr. Stephan Krapp and Dr. Michael Blasse for helping me many things, Dr. Mekdes Debela, Thomas Steiner, Chang-Uk Lee for the good time together in Munich.

Thanks to all my Korean friends outside the science for providing a non-scientific buffer that is necessary to survive from the drizzly days of research.

I would like to thank my parents, brother, and brother's sister-in-law for the mental support and patience from South Korea.

In particular, thank so much to my wife Kyoung and lovely daughter Hannah who did not complain that I could not spend much time as husband and farther with them so often.

I dedicate this thesis to my family:

Kyoung Youn and Hannah Lee

Abbreviations

Å	angstrom, 1×10^{-10} m
Amp	ampicillin
ARB2	argonaute-binding protein-2
ATP	adenosine-5'-triphosphate
CCD	coiled-coil domain
CD	circular dichroism;
cDNA	complementary deoxyribose nucleic acid
Cm	chloramphenicol
DBD	DNA-binding domain
DBP	DNA-binding protein
DTT	1,4-dithiothreitol
EM	electron microscopy
FL	full-length
FLC	full-length scHda1p-scHda2p-scHda3p complex
FLSC	full-length scHda2p-scHda3p sub-complex
HAT	histone acetyltransferase
HDAC	histone deacetylase
IGR	intergenic region
IPTG	isopropyl- β -D-thiogalactopyranoside
Kan	kanamycin
kDa	kilodalton
MAD	multiwavelength anomalous dispersion
MIR	multiple isomorphous replacement
MR	molecular replacement
MW	molecular weight
NCS	non-crystallographic symmetry
Ni-NTA	nickel-nitriloacetat acid
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulfonylfluoride
PVDF	polyvinylidene fluorid

RMSD	root-mean-square-deviation
SAD	single wavelength anomalous dispersion
SAP	secreted aspartic protease
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SeMet	selenomethionine
SER	surface entropy reduction
SF	superfamily
SIR	single isomorphous replacement
Tris	tris (hydroxymethyl) aminomethane
TSA	trichostatin-A
TEMED	N,N,N',N'-tetramethylethylenediamine
URS	upstream repressive sequence
v/v	volume per volume
w/v	weight per volume
2-ME	2-mercaptoethanol

Part I

Structural and functional studies of the yeast class II Hda1 histone deacetylase (HDAC) complex

Summary

Yeast class II Hda1 histone deacetylase (HDAC) complex is an H2B and H3 specific histone deacetylase in *Saccharomyces cerevisiae* consisting of three previously identified subunits, the catalytic subunit scHda1p and two non-catalytic structural subunits scHda2p and scHda3p. I coexpressed and copurified recombinant yeast class II HDAC complex from bacteria as a functionally active and trichostatin A (TSA)-sensitive heterotetrameric complex. According to an extensive analysis of domain organization and interaction of all subunits (or domains), the N-terminal domain of scHda1p associates through the C-terminal coiled-coil domains (CCDs) of the scHda2p-scHda3p subcomplex, yielding a truncated scHda1pHDAC-scHda2pCCD2-scHda3pCCD3 complex with indistinguishable deacetylase activity compared to the full-length complex *in vitro*.

I characterized the interaction of the HDAC complex with either single-stranded or double-stranded DNA and identified the N-terminal halves of scHda2p and scHda3p as binding modules. A high resolution structure of the scHda3p DNA-binding domain (DBD3) by X-ray crystallography is presented. The crystal structure shows an unanticipated structural homology with the C-terminal helicase lobes of SWI2/SNF2 chromatin-remodeling domains of the Rad54 family enzymes. DNA-binding is unspecific for nucleotide sequence and structure, similar to the Rad54 enzymes *in vitro*.

Structural and functional analyses of the budding yeast class II Hda1 HDAC complex presented here provides insight into DNA recognition and suggests a model for the deacetylation of histones in nucleosomes.

1. Introduction

1.1. Gene regulation by acetylation and deacetylation of chromatin

The DNA packing extensively regulates the gene transcription in eukaryotic cells. The heterochromatin, which is compactly packaged chromatin, is not easily accessible by the transcriptional machinery, so that the gene transcription is inhibited. On the other hand, the euchromatin, which is loosely packaged chromatin, is a suitable site of active gene transcription as transcription factors and RNA polymerase can easily access. The nucleosome is the basic unit of the chromatin and comprises of dsDNA and histone proteins in the manner that 146bp of dsDNA wraps histone octamer of (two copies each of H2A, H2B, H3, and H4) two rounds. Thus, physical tightness between nucleosomes is different between euchromatin and heterochromatin, and chemical modification of nucleosome structure regulates gene transcription. This post-translational modification of nucleosome structure results from the chemical alteration of histone proteins, especially such as acetylation/deacetylation and methylation/demethylation, and these opposing reactions catalyzed by two enzymes set allow for reversible gene regulation (Sengupta, N. *et. al.* 2004, Fig. 1).

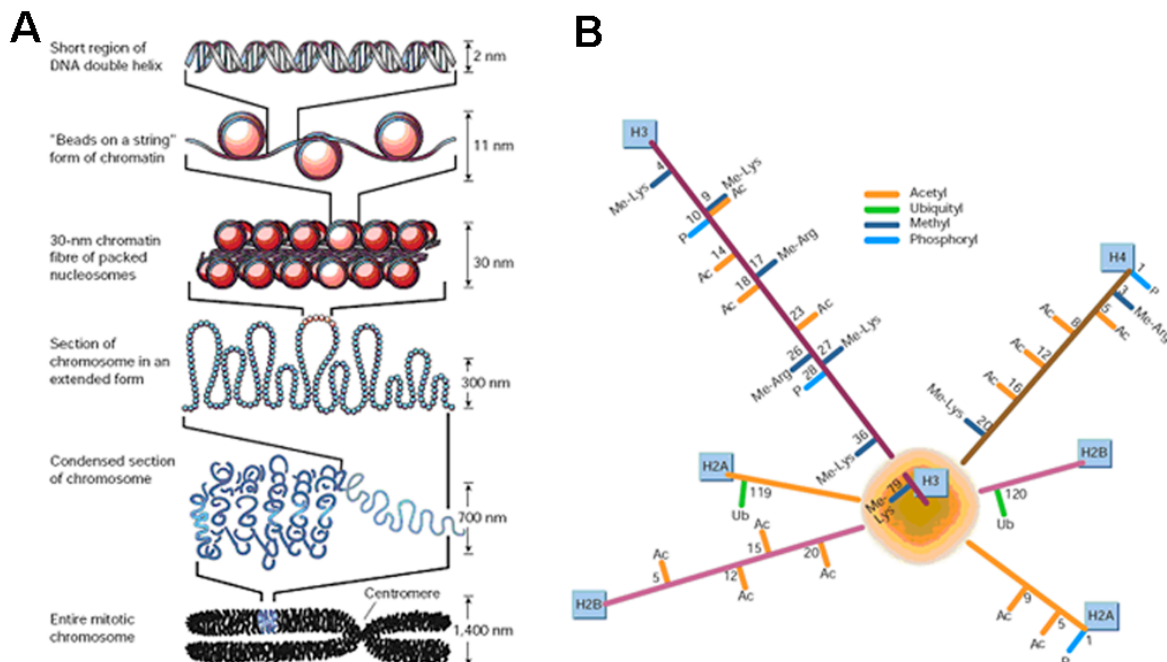


Fig. 1. Packing DNA in chromatin and histone modification. A. The organization of DNA within the chromatin structure. The lowest level of organization is the nucleosome, in which two superhelical turns of DNA (a total of 165bp) are wound around the outside of a histone octamer. Nucleosomes are connected to one another by short stretches of linker DNA. At the next level of organization the string of nucleosomes is folded into a fiber about 30nm in diameter, and these fibers are then further folded into higher-order structures. At levels of structure beyond the nucleosome the details of folding still uncertain. **B.** The chemical modifications of histone N-terminal tails by acetylation/deacetylation (orange), ubiquitination/deubiquitination (green), methylation/demethylation (marine), and phosphorylation/dephosphorylation (cyan) (Felsenfeld, G. *et. al.* 2003).

The histone acetylation and deacetylation regulates gene transcription by changing the compactness of nucleosomes. The histone acetylation occurs at the ϵ -amino groups of lysine residues in N-termini of histone proteins. The acetylation of ϵ -amino group of lysine residues decreases positive charge, which increase repulsive forces against negative charge of phosphates in dsDNA. In this way, histone acetylation helps transcription machinery better approach to dsDNA by widening the gap between nucleosomes and thus promotes gene transcription. In contrast, histone deacetylation results in tight nucleosome structures, inhibits the approach of transcription machinery, and represses gene transcription. Especially, N-terminal histone tails are major acetylation and deacetylation targets. These opposing reactions, acetylation and deacetylation are performed by special key enzymes; histone acetyltransferases (HATs) and histone deacetylases (HDACs).

1.2. Sequence-specific and global histone deacetylations in yeast

Two principle HDACs, class I scRpd3p and class II scHda1p play key roles in nucleosomal repression of transcription, either promoter site-specifically or globally (Rundlett, S. E. *et. al.* 1998, Vogelauer, M. *et. al.* 2000, Wu, J. *et. al.* 2001, Kurdistani, S. K. *et. al.* 2002). These two major HDAC activities in yeast are organized each in large multi-subunit complexes. In general, the class I scRpd3p HDAC enzyme associates with several subunits (e.g., scSin3p, scSap30p, scSde3p, and scPho23p) to form a multi-protein Rpd3 complex, whereas class II scHda1p HDAC assembles with two non-catalytic subunits, scHda2p and scHda3p

to form a hetero-tetrameric Hda1 complex (Kasten, M. M. *et. al.* 1997, Rundlett, S. E. *et. al.* 1996, Wu, J. *et. al.* 2001). Previous studies demonstrated that these two HDAC complexes are recruited to specific promoter loci through DNA-binding proteins for transcriptional repression (Rundlett, S. E. *et. al.* 1998, Wu, J. *et. al.* 2001a, Wu, J. *et. al.* 2001b, Kadosh, D. *et. al.* 1997). For instance, the zinc-finger DNA-binding protein scUme6p specifically recruits class I Rpd3 HDAC complex to the upstream repressive sequence (URS) in the *INO1* promoter through the scSin3p subunit. Likewise, the class II Hda1 HDAC complex is recruited to target promoters (e.g, *ENA1* gene) through the scTup1p repressor which can interact with the sequence-specific DNA-binding protein. Subsequently adjacent nucleosomes are precisely modified at the lysine residues of histone H2B and H3 (but not H2A and H4). Untargeted global deacetylation of most regions of chromatin in *S. cerevisiae* is also primarily conducted by these two HDAC complexes. In the case of the class I enzyme, the Rpd3 HDAC complex binds to non-promoter sequences by a scUme6p-independent mechanism (Kurdistani, S. K. *et. al.* 2002). It was shown that more than 69% of the intergenic regions (IGRs) deacetylated by the HDAC complex are not affected by a loss of *TUP1*, suggesting a scTup1p-independent mechanism of class II Hda1 HDAC recruitment (Robyr, D. *et. al.* 2002).

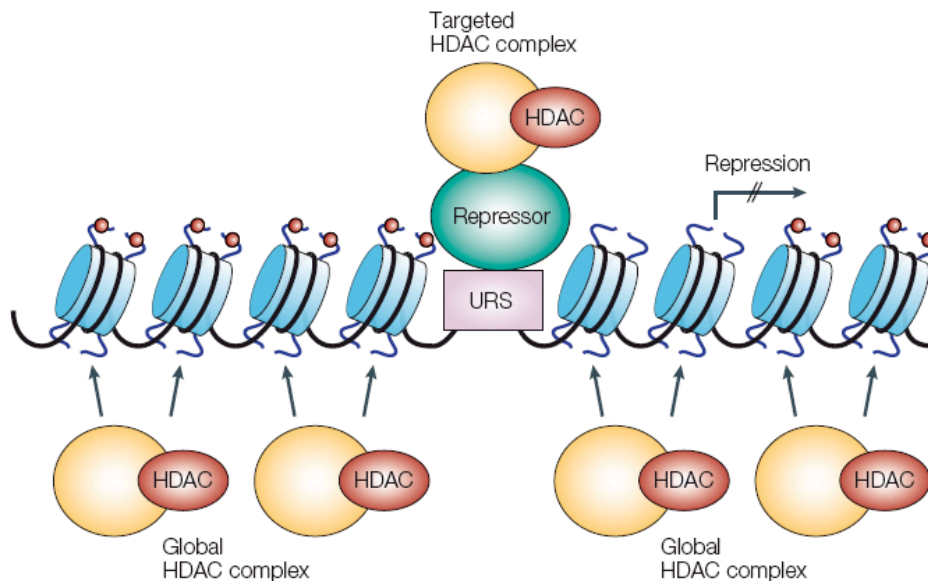


Fig. 2. Histone deacetylation by site-specific and global mechanisms. Promoter-targeted deacetylation occurs in a background of global and untargeted deacetylation.

HDAC complexes generate a highly localized region of deacetylated chromatin (represented by the absence of red circles from chromatin) when tethered to an upstream repressive sequence (URS) in a promoter by a sequence-specific DNA-binding repressor. HDACs also function globally on most regions of chromatin including coding sequences in the absence of any apparent targeting mechanisms. Not shown are the HATs that also function through targeted and global mechanisms. The global, but opposing, actions of HATs and HDACs maintain the steady-state levels of acetylation and allow for reversible gene regulation (Kurdistani, S. K. *et. al.* 2003).

1.3. Classification of histone deacetylases in mammal and yeast

Until now, total 18 HDACs have been discovered, and they are classified into 4 different classes based on the amino acid sequence similarity to yeast HDACs. Phylogenetically, class I, II, and IV HDACs are closely related and share Zn-catalyzed hydrolysis mechanism of acetyl-lysine amide bond in common. These enzyme activities are inhibited by HDAC inhibitors (e.g., TSA and SAHA). However, class III HDACs (Sirtuins) are phylogenetically different with the other classical HDACs. While they are not inhibited by classical HDAC inhibitors, they are NAD-dependent and similar to yeast Sir2 enzyme. Thus, class III enzymes will not be discussed any further in this study.

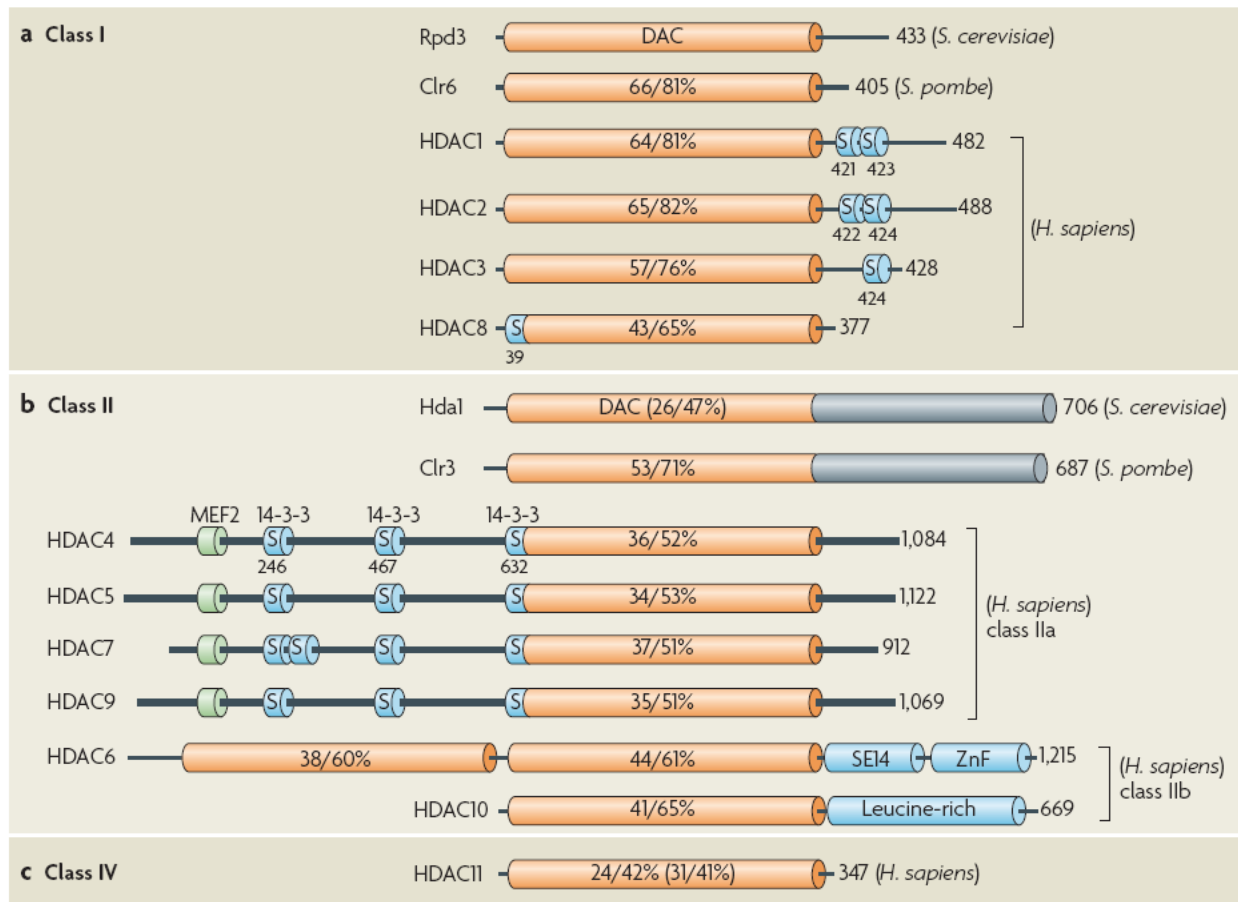


Fig. 3. Domain organization of classical HDACs from yeast and humans. The histone deacetylases (HDACs) are grouped into different classes according to sequence similarity to yeast prototypes. In mammals, class I (Rpd3-like) members include HDAC1, -2, -3 and -8 (a); class II members are HDAC4, -5, -6, -7, -9 and -10 (b), which are related to yeast Hda1. Class IV consists of HDAC11 (c). Class II is further divided into two subclasses: IIa (HDAC4, -5, -7 and -9) and IIb (HDAC6 and -10). The total number of amino acid residues in each deacetylase is shown on the right. Many of the deacetylases have isoforms that result from alternative splicing; for simplicity, the number of amino acids refers to the longest isoform. The deacetylase (DAC) domain is depicted as an orange cylinder, and the percentage amino acid sequence identity/similarity to that of Rpd3 (for class I) or Hda1 (class II/IV) is shown. The sequence identity/similarity of Hda1 and HDAC11 to Rpd3 is given in brackets. The C-terminal domains (grey) of Hda1 and Clr3 are homologous (identity/similarity: 26/57%). Thick black lines represent similar N-terminal domains and C-terminal tails of class IIa HDACs. Myocyte

enhancer factor-2 (MEF2)-binding motifs are depicted as short green cylinders, whereas 14-3-3 binding motifs are shown as short blue cylinders labeled with 'S' (for Ser). Clr3, cryptic locus regulator-3; Hda1, histone deacetylase-1; *H. sapiens*, *Homo sapiens*; Rpd3, reduced potassium dependency-3; *S. cerevisiae*, *Saccharomyces cerevisiae*; SE14, Ser-Glucontaining tetradecapeptide repeats; *S. pombe*, *Schizosaccharomyces pombe*; ZnF, ubiquitin-binding zinc finger (Yang, X. J. *et. al.* 2008).

1.4. *Saccharomyces cerevisiae* class II histone deacetylase complex

The yeast Hda1 HDAC complex is related to the mammalian class II HDACs (e.g., HDAC4, 5, 6, 7, 9, and 10), and consists of a scHda1p homodimer as the catalytic subcomplex and a non-catalytic scHda2p-scHda3p heterodimer, essential for a fully functional HDAC complex *in vivo* and *in vitro* (Wu, J. *et. al.* 2001b). Interestingly, no metazoan homologues of scHda2p and scHda3p were identified so far, suggesting unique structural and functional features of those subunits in fungi. This multi-subunits HDAC complex is known as a transcriptional repressor in yeast together with class I HDAC Rpd3 complex. It was reported that the N-terminal repressor domain of scTup1p interacts with the scHda1p subunit and more weakly with scHda3p to recruit HDAC activity to specific promoter sites in yeast (Wu, J. *et. al.* 2001a). On the other hands, class II Hda1 HDAC complex itself is also responsible for indirect global repression of gene expression by unknown mechanism in yeast *S. cerevisiae*.

1.5. Helicase and SNF2 family: structural overview

1.5.1 core *RecA*-like domains

At the core of all superfamily 1 (SF1) and superfamily 2 (SF2) helicase proteins are a pair of RecA-related domains (Fig. 4), each consisting of a 5-strand beta sheet decorated by alpha helices on either face (Hall, M. C. *et. al.* 1999, Caruthers, J. M. *et. al.* 2002). The ATP hydrolysis site is located at the base of the cleft formed by the inter-domain association, primarily mediated by Walker motifs A and B (helicase motifs I and II) in RecA-like domain 1 and an 'arginine finger' contributed by motif VI of RecA-like domain 2 (Caruthers, J. M. *et.*

al. 2002, Ye, J. *et. al.* 2004).

1.5.2. ATP-driven mechanism

ATP hydrolysis appears to cause a change in the relative orientations of the two domains of the helicase proteins, involving the concerted action of the ATP binding and other 'sensor' residues such as the TxGx motif (conserved block A) and motif III. The distinction between SF1 and SF2 is based on the detailed interactions involved in nucleic acid substrate and ATP binding, and how these are interlinked to direct the mechanical changes in the bound nucleic acid driven by ATP hydrolysis (Korolev, S. *et. al.* 1998, Singleton, M. R. *et. al.* 2002). These distinctions are reflected in subtle differences in the conserved helicase motifs which gave rise to the original classification of superfamilies by Gorbalenya and Koonin (Gorbalenya, A. E. *et. al.* 1993). The Q motif (Fig. 4) is an additional structure found in many SF2 proteins, including the SNF2 family, which makes further interactions with ATP (Tanner, N. K. *et. al.* 2003).

1.5.3. Unique SNF2 family structures

The most remarkable feature of the SNF2 family structures compared to other known SF2 members are several additional structural elements grafted onto the core helicase structure. These comprise antiparallel alpha helical protrusions from both RecA-like domains 1 and 2, a structured linker between the recA-like domains, and a triangular brace packed against the domain 2 alpha helical protrusion (Fig. 4). The two helical protrusions and linker are all encoded within the enlarged span between motifs III and IV. The triangular brace is encoded immediately downstream of motif VI.

1.5.4. Insertion sites

A major insertion site is also located behind protrusion 2 (Fig. 4).

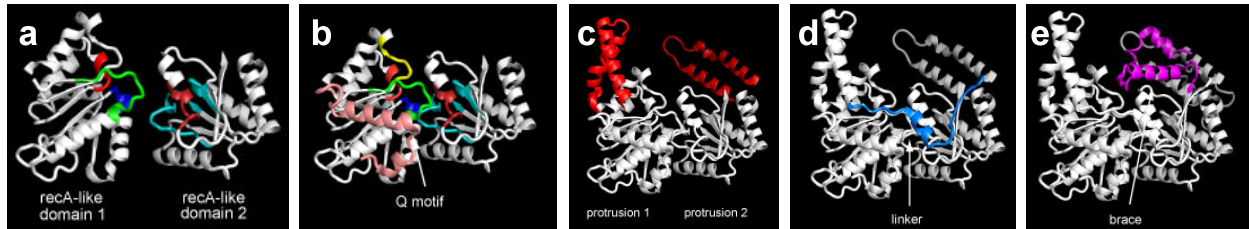


Fig. 4. Structural overview of SF1, SF2, and SNF2 helicase families. (a) Core RecA-like domains, (b) RecA-like domain interaction and Q motif, (c) protrusions from both RecA-like domains, (d) linker spanning between two domains, and (e) brace inside protrusion 2.

1.6. SWI2/SNF2 chromatin-remodeling domain of Rad54 enzymes

The archetype of the Rad54 subfamily is the Rad54 protein from *S. cerevisiae* which was isolated because its inactivation leads to increased sensitivity to ionizing radiation. The scRad54p and its homologues in other organisms play an important but as yet incompletely understood role in homologous recombination by stimulating Rad51-mediated single strand invasion into the target duplex, and subsequent steps in the process (Tan, T. L. *et. al.* 2003, Krogh, B. O. *et. al.* 2004).

Many organisms also contain a second subfamily member, such as *S. cerevisiae* Rdh54p or *S. pombe* tid1p. These are frequently implicated in mitotic repair and meiotic crossover (Klein, H. L. 1997), although the role of the human homologue RAD54B is unclear (Tanaka, K. W. *et. al.* 2002). Rad54 proteins have been extensively studied *in vitro*. They have been shown to be able to generate local changes in DNA topology in supercoiled plasmids to translocate along DNA by biochemical and other methods (Petukhova, G. *et. al.* 1999, Tan, T. L. *et. al.* 1999, Van Komen, S. *et. al.* 2000, Jaskelioff, M. *et. al.* 2003), and to alter the accessibility of nucleosomal DNA (Alexiadis, V. *et. al.* 2002, Alexeev, A. *et. al.* 2003). However, this latter activity appears inefficient compared to purified complexes from the SNF2 and ISWI subfamilies.

1.7. Structure determination by X-ray crystallography

Because protein folds into specific three-dimensional conformation that is intrinsically related to its function, the knowledge of the atomic structure of macromolecules is essential for a fundamental understanding of its function and a prerequisite for structure-based functional studies. Determination of the precise molecular “three-dimensional” structure of protein or nucleic acid has revolutionized the study of its function in many fields of biology. Two principle methods, X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy (solution or solid), are currently the only techniques capable of determining the structures of biological macromolecules at atomic resolution. Electron microscopy (EM) is also valuable to elucidate the molecular architecture of more complicated protein complexes, but the resolution of EM is limited to show the atomic details. The structure determination by NMR, an indirect spectroscopic method, is limited to small molecules (<30 kDa), but provide in addition to structural insights information about intramolecular dynamics or reaction kinetics. In contrast, X-ray crystallography has no size limitation for molecule or complex, but displays only a “snapshot” of the macromolecule and the high accuracy of crystallographic structure has to be paid with high quality crystal.

The x-ray intensity we measure on our detectors with biological macromolecule’s crystal is proportional to the square of the amplitude of the structure factor. Since $F(hkl)$ is a complex number, it's actually the complex conjugate of F :

$$I = F(hkl)F^*(hkl).$$

Because the detectors cannot sample the relative or absolute phase of the incident radiation, for all sorts of perfectly good technical reasons including the extraordinarily high frequency of X-rays, phase information about the structure factor is lost during data collection. This is the so-called phase problem in that we have to leap through flaming hoops in order to retrieve that lost phase information via indirect methods.

Without phases we cannot calculate the electron density corresponding to our molecules, which means we cannot build our structures.

Methods of getting a grip on the phase problem:

- Molecular replacement (MR): using an existing model
- Experimental phase determination: SIR, MIR, SAD, and MAD methods
- Direct methods - probabilistic *ab initio* phase determination

1.8. Crystal twinning

Some crystalline materials form aggregates made up of individual single crystals joined together by one or more definite macroscopic symmetry relations. These aggregate crystals are known as twins. According to Cahn (Cahn, R. W. 1954), two empirical facts are central to the crystallography of twins. First, a twinned piece of material is composed of two or more domains in which the orientations of the domains are related by one or more symmetry elements that are not a part of the space group symmetry of a single crystal of the material. Second, the extra symmetry element(s) must be of the kind encountered in crystal morphology (e.g., a center of symmetry, a mirror plane, or a 2-, 3-, 4-, or 6-fold rotation axis). If the extra symmetry element is a mirror plane, called a twin plane, then this plane must be parallel to a lattice plane of the same d-spacing in both domains. If the extra symmetry is a rotation axis, called a twin axis, then this axis must be parallel to a lattice row common to both domains. If the twin contains domains with more than two sets of orientations, then the sets of domains must be related, pairwise, by an extra symmetry element. It is also possible for two or more crystals to form an aggregate in which the orientations of the crystals are not related by additional symmetry. These pieces are often called twin crystals, even though they do not strictly fit the definition of a twin. If data are collected on these pieces, the data are usually handled as if it was a non-merohedral twin (see below).

Based upon their diffraction patterns, twinned crystals may be grouped into three general categories. *Non-merohedral twins* have two or more crystalline domains with reciprocal lattices that either do not overlap or only partially overlapped. In contrast, *merohedral twins* have domains with diffraction patterns that are completely overlapped. For merohedral twins, the symmetry operations relating the twinned domains are a part of the Laue group of the sample, but are

not a part of the space group. *Pseudo-merohedral* also have domains with completely overlapped diffraction patterns, but the symmetry operation relating the domains is not a part of the Laue group of the sample.

Non-merohedral twins are generally recognized at the time of data collection because of difficulties with indexing or because of obvious multiple sets of reciprocal lattices. Many crystals that are twinned employ this mode of twinning. These twins are sometimes described as having twin lattice quasi-symmetry, TLQS (Giacovazzo, C. 1992). Generally, the twin symmetry element is not a symmetry element of the Laue group. If there are completely overlapped spots in the diffraction pattern, the overlapped spots are usually found in zones or lines that are perpendicular to the twin symmetry element. If most of the crystal is from one domain, it is sometimes possible to collect intensity data from this one domain and continue the structure determination as if the crystal were single (realizing that the final R factor may not be as low as it could be if the crystal were truly single). Point detector instruments have difficulty indexing these samples or generate cells with one or more very large cell edges. Thus, it is rare for data from these types of twins to be collected with point detector instruments. Since area detectors allow data from all parts of reciprocal space to be collected, analysis of crystals with this type of twinning has become more common. Usually the overlapping of spots from the different domains is significant enough that data from all domains must be considered in the final refinement.

Merohedral twins have a twin symmetry element that is not a symmetry element of the space group of the sample but is a symmetry element of the Laue group. Both the direct and reciprocal lattices of the different domains coincide with one another, making the crystal appear to be single until near the end of refinement. This type of twinning is sometimes described as having twin lattice symmetry, TLS, of class I (Giacovazzo, C. 1992). If the volumes of the different domains are nearly equal, then the additional symmetry element often leads to the selection of an incorrect higher symmetry space group. It is sometimes possible to solve and refine the structure in this higher symmetry space group, and obtain a false average of the two orientations of the actual structure. Frequently, in such cases, disorder will occur. Conversely, when extensive "disorder" is encountered, the structure should be checked to make certain that it is not twinned.

As with merohedral twins, Pseudo-merohedral twins have reciprocal lattices that can be indexed on a single lattice and hence appear to be single crystals. Also as with merohedral twins, data sets of these compounds show difficulty with structure solution and especially structure refinement. Often lowering the crystal system symmetry will help with the structure solution. Pseudo-merohedral twins have lattice parameters that suggest higher Laue symmetry than the symmetry appropriate for the point group of the sample and have a twin element that emulates the additional symmetry of a higher symmetry Laue group. For example, consider a monoclinic sample with $\beta = 90^\circ$ that has a twin plane on (100). The twin domain will have peaks that are too close to be distinguished from the first domain. If the volumes of the two domains are nearly equal, the intensities will appear to have mmm symmetry instead of $2/m$, and the structure will not be solvable. Once the nature of the twinning is recognized and included in the model, the structure solution and refinement usually proceed without further problems.

The measured intensities often have observable data from one domain that is superimposed on the systematically absent locations of another domain. In such cases, if twinning is not suspected, a space group may be chosen with few translational symmetry elements, such as $C2221$ or $Pmmm$. When such relatively rare space groups are encountered, and the structure is difficult to solve, it is worth considering the possibility of twinning.

A related type of pseudo-merohedral twinning occurs when the reciprocal cell parameters are such that in every second, third, or n th layer the reciprocal lattices of the two components overlap. Data sets from crystals with these types of twins can easily be collected with either point or area detectors. It is possible to interpret the overlapping reciprocal lattices in terms of a smaller single reciprocal cell, i.e. a larger real cell. The structure produced from this smaller cell is obviously wrong, but the nature of the problem may not be apparent. Often the first hint that something is wrong comes from the presence of implausible systematic absences (Bürger, M. J. 1960). These twins are sometimes classified as having twin lattice symmetry, TLS, of class II (Giacovazzo, C. 1992).

Crystal symmetry	Cell conditions	Apparent symmetry	Twinning operation	Reindexing
Monoclinic <i>P</i>	$a \simeq c$	Orthorhombic <i>C</i>	Twofold \parallel [101]	$hkl > l-kh$
Monoclinic <i>P</i>	$\beta \simeq 90^\circ$	Orthorhombic <i>P</i>	Twofold \parallel <i>a</i>	$hkl > h-k-l$
Rhombohedral <i>R</i>	$\alpha \simeq 60^\circ$	Cubic <i>F</i>	Twofold \parallel [011] or twofold \parallel [101] or twofold \parallel [110]	$hkl > -h, l-h, k-h$ $hkl > l-k, -k, h-k$ $hkl > k-l, h-l, -l$
Rhombohedral <i>H</i>	$c \simeq a\sqrt{3}$	Cubic <i>F</i>	Twofold \parallel [012] or twofold \parallel [-102] or twofold \parallel [1-12]	$hkl > -h, (2h+k+l)/3, (4h+8k-l)/3$ $hkl > (h+2k-l)/3, -k, (-8h, -4k, -l)/3$ $hkl > (-h-2k+l)/3, (-2h-k-l)/3, (4h-4k-l)/3$
Rhombohedral <i>R</i>	$\alpha \simeq 90^\circ$	Cubic <i>P</i>	Twofold \parallel <i>a</i> or twofold \parallel <i>b</i> or twofold \parallel <i>c</i>	$hkl > h-k-l$ $hkl > -hk-l$ $hkl > -h-kl$
Rhombohedral <i>H</i>	$c \simeq a(3/2)^{1/2}$	Cubic <i>P</i>	Twofold \parallel [101] or twofold \parallel [0-11] or twofold \parallel [-111]	$hkl > (h+2k+2l)/3, -k, (4h+2k-l)/3$ $hkl > -h, (2h+k-2l)/3, (-2h-4k-l)/3$ $hkl > (-h-2k-2l)/3, (-2h-k+2l)/3, (-2h+2k-l)/3$
Rhombohedral <i>R</i>	$\alpha \simeq 109.5^\circ$	Cubic <i>I</i>	Twofold \parallel [11-1] or twofold \parallel [1-11] or twofold \parallel [-111]	$hkl > k, h, (-h-k-l)$ $hkl > l, (-h-k-l), h$ $hkl > (-h-k-l), l, k$
Rhombohedral <i>H</i>	$c \simeq a(3/8)^{1/2}$	Cubic <i>I</i>	Twofold \parallel [-201] or twofold \parallel [021] or twofold \parallel [2-21]	$hkl > (h+2k-4l)/3, -k, (-2h-k-l)/3$ $hkl > -h, (2h+k+4l)/3, (h+2k-l)/3$ $hkl > (-h-2k+4l)/3, (-2h-k-4l)/3, (h-k-l)/3$

Table 1. Examples of pseudo-merohedral twinning. In the rhombohedral cell, *R* corresponds to the rhombohedral setting with $a=b=c$, $\alpha=\beta=\gamma$ and *H* to the hexagonal setting with $a=b$ and $\gamma=120^\circ$ (Dauter, Z. 2003).

If the overlapped data are required for structure determination, the relative contributions of the different domains are first estimated and these contributions are applied to "detwin" the overlapped data. If volume fractions of the crystal are known then the relative contributions are calculated from $x = V_1/(V_1 + V_2)$. Initial refinements using the nonoverlapped data can also produce estimates for x . The intensities of the overlapped data may then be estimated using the following formula (Massa, W. *et. al.* 1990).

$$F_{hkl}^2 = F_1^2 [x/(2x - 1)] - F_2^2 [(1 - x)/(2x - 1)]$$

where F_1^2 and F_2^2 are the true intensities for the overlapped hkl peak.

1.8.1. Common indicators of crystal twinning (Sheldrick, G. M. 1997)

- The mean value for $|E^2-1|$ is much lower than the 0.736 expected for a noncentrosymmetric crystal. The intensities from peaks overlapped by more than one twin domain tend to have few very strong or very weak peaks leading to low values for this statistic.

- The data exhibit implausible or unusual systematic absences (a possible partial-merohedral twin).
- The data exhibit no systematic absences for glide planes or screw axes, and the structure is not solvable or solves with difficulty (a possible partial-merohedral twin).
- For the most poorly fit data, the F_o^2 values are consistently larger than the F_c^2 values.
- The space group appears to be either tetragonal or trigonal.
- The model is in a centrosymmetric space group and is highly disordered.
- The metric symmetry of the cell parameters is higher than the Laue symmetry.
- The $R(int)$ for a higher symmetry Laue group is only slightly higher than the $R(int)$ for a lower symmetry Laue group.

Nonmerohedral twins may be indicated by the following.

- The cell has one or more unusually long axes, and many of the data are weak or absent.
- The cell is indexed or refined only with great difficulty.
- Some data are sharp and others appear to be split.

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals, enzymes, primers, and equipments

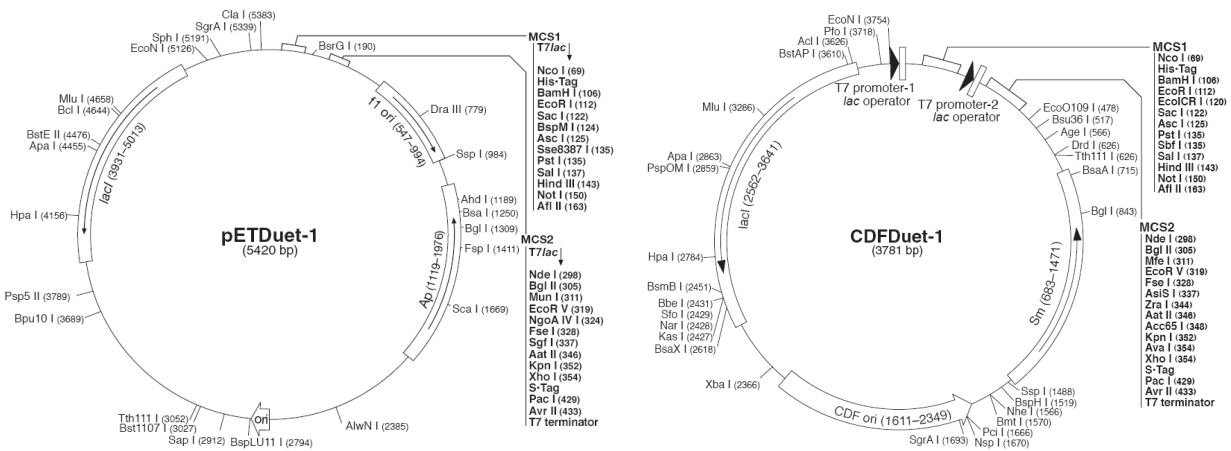
Most of common chemicals were purchased from Merck (Darmstadt, Germany) or Sigma-Aldrich (Munich, Germany). Oligonucleotides for PCR, mutagenesis, and DNA-binding assay were synthesized by Metabion (Martinsried, Germany). Enzymes used for PCR, coning, and mutagenesis were ordered to Fermentas (St. Leon-Rot, Germany), New England Biolabs (Frankfurt am Main, Germany), Stratagene (Waldbronn, Germany), and Takara (Potsdam, Germany). All cloning vectors used here were either purchased from the core facility of MPI-Biochemistry or kindly provided by Dr. Jeong Ho Chang (POSTECH, South Korea). All chromatographic materials and columns were obtained from GE healthcare (Munich, Germany).

2.2. Methods I : molecular biology, biochemistry, and biophysics

2.2.1. Cloning and protein expression

All constructs were generated by using a standard PCR strategy. Full-length (FL) *S. cerevisiae* (sc) Hda1p was subcloned into *Bam*HI/*Sal*I sites of pGEX-4T3 and *Nhe*I/*Xho*I site of pET21b. FL scHda2p was subcloned into *Nde*I/*Xho*I sites of pET28a and pET30a. FL scHda3p was subcloned into *Nde*I/*Xho*I sites of pET28a, pACYCDuet-1 MCS2, and pCDFDuet-1 MCS2. For coexpression of all three subunits, two plasmids of pETDuet-1 MCS1:scHda1p::MCS2:scHda2p and pCDFDuet-1 MCS2:scHda3p were cotransformed into *E. coli* BL21(DE3)CodonPlus-RIL.

	Subunit	Sequence	Forward	Reverse
1	scHda1p	CGGGATCCATGGATTCTGTAATGGTAAAGAAA	BamHI	
2	scHda1p	CCGCTCGAGTCATTCTTCATCACTCCATTCTTC		XhoI
3	scHda2p	GGAATTCCATATGAGTAGGAAAAATTCTAAGAAAC	NdeI	
4	scHda2p	CCGCTCGAGTCATGTGTAATTAGGGGTATTGG		XhoI
5	scHda3p	GGAATTCCATATGGATTTACTACGCATTTTAGAC	NdeI	
6	scHda3p	CCGCTCGAGTTATACATTTCTGGTTTTTTTGC		XhoI

Table 2. Primers used for cloning of FL scHda1p, scHda2p, and scHda3p.**Fig. 5. Bicistronic dual vectors (pETDuet-1 and pCDFDuet-1, Novagen) used for coexpression of all three subunits of the yeast class II Hda1 HDAC complex.**

2.2.2. Protein purification

FL yeast class II Hda1 HDAC complex (FLC) was purified by using a His-affinity chromatography followed by anion-exchange (resource Q) and size-exclusion chromatography (superdex-200). Each domain of scHda1p, scHda2p, and scHda3p subunits was cloned and purified with the same protocols mentioned above. A selenomethionine (SeMet)-substituted scHda3pDBD3 K168A/Q169A/K170A was cultured by growing the expression plasmid in *E. coli* B834(DE3) in minimal medium and purified by the same method used for the native protein.

	Step	Composition
1	His-affinity chromatography (buffer A)	20mM Tris-HCl pH7.5, 200mM NaCl, 4mM 2-ME
2	His-affinity chromatography (buffer B)	Buffer A plus 500mM imidazole
3	Dialysis for His-tag cleavage	20mM Tris-HCl pH8.0, 100mM NaCl, 5mM DTT
4	Anion-exchange chromatography (buffer A)	20mM Tris-HCl pH8.0, 5mM DTT
5	Anion-exchange chromatography (buffer B)	Buffer A plus 1M NaCl
6	Size-exclusion chromatography	20mM BTP-HCl pH7.0, 200mM NaCl, 1mM DTT

Table 3. Buffer composition used for purification of recombinant yeast class II Hda1 complex.**2.2.3. Limited proteolysis and N-terminal sequencing**

To identify domain boundaries and stable fragments of yeast class II Hda1 HDAC complex, limited proteolysis was performed by using two proteases, subtilisin and trypsin. Stable fragments were prepared by limited proteolysis, blotted on PVDF membrane and identified by N-terminal sequencing.

2.2.4. Sedimentation velocity

Sedimentation velocity experiments were carried out in a Beckman XL-A analytical ultracentrifuge equipped with absorbance optics. Protein samples were dialyzed into storage buffer (20mM Tris-HCl pH 7.5, 300mM NaCl) and loaded into cells with two channel Epon center pieces and quartz windows. Data were recorded at 40,000rpm, 20°C, using an An50Ti rotor and an absorbance wavelength of 280nm, with scans taken every 6 min. Solvent density was measured using an Anton Paar DMA 5000 high-precision density meter. Partial specific volumes were calculated using the program SEDNTERP (<http://www.jphilo.mailway.com/>). Data were analyzed using the program SEDFIT (Schuck, P. 2000). Sedimentation coefficient distributions were calculated using the Lamm equation modeling implementing maximum entropy regularization.

2.2.5. Circular dichroism spectroscopy

Structural changes of a native and mutant proteins were monitored by circular dichroism (CD) spectrophotometry (Jasco J-715) at wavelengths between 200-250nm. All samples (10µM) were prepared in the buffer supplemented with 20mM Tris-HCl pH 7.5, 300mM NaCl.

2.2.6. Mutagenesis

All scHda3p mutants used in this study were constructed by PCR-based methods. The mutant proteins were purified by using His-affinity chromatography followed by anion-exchange (resource Q) and gel-filtration (superdex-75).

2.2.7. Deacetylation assay

All protein samples (30nM) were mixed with fluoro-substrate peptide (40 μ M) in assay buffer supplemented with 20mM Tris-HCl pH 8.0, 125mM NaCl, 1% glycerol. Reaction mixtures were incubated at room temperature for 1h and then quenched by adding 50 μ L of buffer containing TSA (2 μ M) and lysyl endpeptidase (1mAU/mL). After further incubation for 30min at room temperature, fluorescence intensities were measure by a fluoro-spectrophotometer with excitation at 355nm and emission at 460nm.

2.2.8. Design of oligonucleotides for DNA-binding assay

	Name	Sequence
1	50m-2	TGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTT
2	50m-2a	AACGTCATAGACGATTACATTGCTAGGACATCTTTGCCACGTTGACCCA
3	50m-5	TGCCGAATTCTACCAGTGCCAGTGATGGACATCTTTGCCACGTTGACCC
4	50m-6	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
5	50m-7	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
6	51m-8	CAACGTCATAGACGATTACATTGCTAATCACTGGCACTGGTAGAATTCGGC
7	24m-10	GGACATCTTTGCCACGTTGACCC
8	26m-15	TGCCGAATTCTACCAGTGCCAGTGAT

Table 4. Oligonucleotide sequences used for band shift assay.

	Oligonucleotides	Final products
1	50bp of DNA	50m-2 + 50m-2a
2	Splayed arm	50m-2 + 50m-8

3	3'-tailed duplex	50m-2 + 24m-10
4	5'-tailed duplex	51m-8 + 26m-15
5	3'-flapped DNA	50m-2 + 51m-8 + 24m-10
6	5'-flapped DNA	50m-2 + 51m-8 + 26m-15
7	3'-PX junction	50m-2 + 50m-6 + 50m-7 + 24m-10
8	5'-PX junction	50m-2 + 50m-6 + 50m-7 + 26m-15
9	X junction	50m-2 + 50m-5 + 50m-6 + 50m-7

Table 5. The assembly of oligonucleotides for band shift assay.

2.2.9. DNA-binding assay

The band shift assay was performed in a buffer containing 20mM Tris-HCl pH 8.0, 50mM NaCl, 5mM MgCl₂, 5mM DTT, with various concentrations of protein samples and various structures and lengths of nucleic acids. For all binding reactions, DNA samples were kept at 5μM, while proteins were varied between 0 and 20μM in 20uL of reaction volume. Protein-DNA mixtures were incubated on ice for 1h and run on 1-5% (w/v) agarose or 3-12% (w/v) non-denaturing native polyacrylamide gels at 4°C. Gels were stained with 1x Gel star solution.

2.3. Methods II : X-ray crystallography

2.3.1. Programs, screening solutions, and equipments

Programs used for crystallography:

XDS (Kabsch, W. 1993)

SHARP/AutoSHARP (Vonrhein, C. *et. al.* 2007)

COOT (Emsley, P. *et. al.* 2004)

CNS (Brünger, A. T. 2007) and PHENIX (Adams, P. D. *et. al.* 2002)

PROCHECK (Laskowski, R. A. *et. al.* 1993)

PYMOL (<http://www.pymol.org/>), MOLSCRIPT (Kraulis, P. J. 1991), and GRASP (Nicholls, A. *et. al.* 1991)

Commercial solutions for screening of initial crystallization:

Index, SaltRX, Crystal screen I, and II (Hampton Research, USA)

Precipitant synergy, Wizard I, and II (Emerald Biostructures, USA)

Equipments for hanging-drop vapor diffusion crystallization:

24-well VDX plate (Hampton Research, USA)

Siliconized square cover slide (Hampton Research, USA)

2.3.2. Crystallization of wild-type scHda3pDBD3 and its triple mutant

All crystallization trials were performed by hanging drop vapor diffusion method at 4°C and room temperature. Among various protein samples, wild-type scHda3pDBD3 I6-K333 and its triple mutant (K168A/Q169A/K170A) crystals were grown against crystallization buffers containing 0.1M HEPES pH 7.5, 0.2M ammonium sulfate or lithium sulfate, 23% (w/v) PEG-3,350, and 0.1M magnesium formate dehydrate and 15% (w/v) PEG-3,350 at 20°C, respectively.

2.3.3. Preparation of SeMet derivative

A selenomethionine (SeMet)-substituted scHda3pDBD3 K168A/Q169A/K170A triple mutant was cultured by growing the expression plasmid in *E.coli* B834(DE3) in minimal

medium and purified by the same protocol used for the native protein.

Table 6. How to prepare the SeMet medium

(1) inorganic compounds (1L):

	g/L		g/L
Na ₂ HPO ₄	6	CaCl ₂	0.01
KH ₂ PO ₄	3	MgCl ₂	0.01
NaCl	0.5	FeCl ₃	0.00001
MgSO ₄	0.12	NH ₄ Cl	1

(2) amino acids (1L):

	g/L		g/L
Gly	0.2	Thr	0.2
Ala	0.2	Lys	0.2
Asn	0.2	Arg	0.2
Leu	0.2	His	0.2
Ile	0.2	Asp	0.2
Pro	0.2	Glu	0.2
Phe	0.2	Gln	0.2
Tyr	0.2	Val	0.2
Cys	0.2	Trp	0.2
Ser	0.2	SeMet	0.04

(3) Dissolve all inorganic chemicals and amino acids in 900mL of DW and autoclave.

(4) Dissolve glucose (4g for 1L culture) in 87.5mL of DW and autoclave (or sterilize by filtering).

(5) vitamins (1L):

	g/L		g/L
Thiamine	0.05	Choline chloride	0.01
Biotin	0.01	Pantothenate	0.01
Pyridoxal	0.01	Riboflavin	0.0001
Nicotinamide	0.01		
Folic acid	0.01		

(6) Dissolve all vitamins in 12.5mL of DW and filter solution through a 0.45um (or 0.22um) filter device.

2.3.4. Data collection and processing

X-ray diffraction data were collected at -170°C with crystals flash-frozen in crystallization buffer supplemented with 20% (v/v) glycerol. A native and multiple-wavelength data set were collected by using native and SeMet crystals on beamlines X06SA at the Swiss Light Source (SLS) and X12 at the Deutsches Elektronen Synchrotron (DESY), respectively. Integration, scaling, and merging of the diffraction data were performed by using the XDS program suites (Kabsch, W. 1993).

2.3.5. Structure determination by MAD phasing

The structure of scHda3pDBD3 triple mutant was determined by multiple anomalous scattering dispersion (MAD) method using SeMet-substituted crystal. Total 32 selenium sites in the asymmetric unit (four molecules / AU) were determined by using the SHARP/AUTOSHARP (Vonrhein, C. *et. al.* 2007). After phase determination, solvent flattening, and non-crystallographic symmetry (NCS) averaging, the experimental electron density map of 3.0\AA resolution was of excellent quality, which allowed me to trace most of $C\alpha$ chains. Successive rounds of model building using COOT (Emsley, P. *et. al.* 2004), refinement using CNS (Brünger, A. T. 2007) and PHENIX (Adams, P. D. *et. al.* 2002) allowed the complete structure determination.

3. Results

3.1. Cloning and expression of scHda1p, scHda2p, and scHda3p genes

All three genes encoding scHda1p, scHda2p, and scHda3p were amplified with genomic DNA of *Saccharomyces cerevisiae* by using a standard PCR method. Each of three genes was subcloned into bacterial expression vectors harboring N- or C-terminal fusion tag (e.g, hexa-histidine or GST) to facilitate the protein purification.

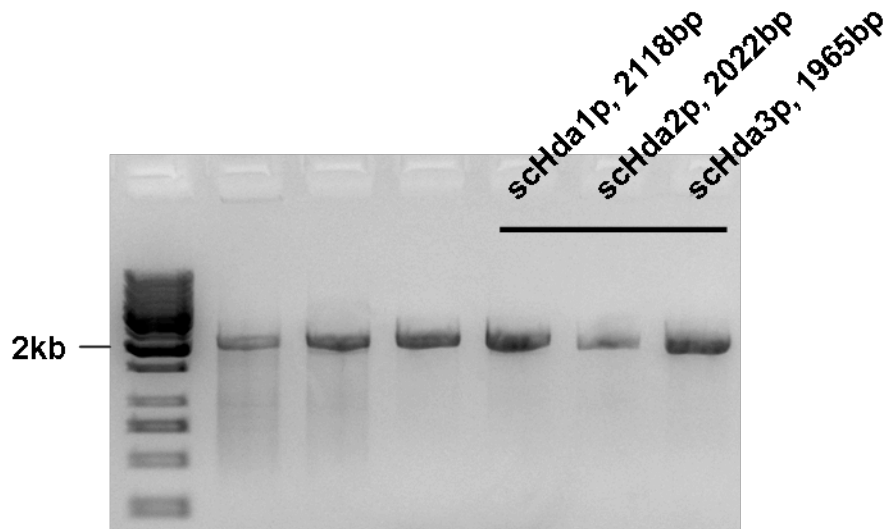


Fig. 6. Amplification of three genes encoding three subunits of yeast class II Hda1 HDAC complex by polymerase-chain reaction (PCR).

Expression plasmids carrying each three subunit were conducted to test the protein expression and solubility in *E.coli*. Full-length GST-scHda1p could be well expressed only in *E.coli* BL21(DE3)CodonPlus-RIL or Rosetta(DE3) harboring episome, which is necessary to overcome the codon bias problem in bacterial expression system. Full-length scHda2p and scHda3p could be well expressed in *E.coli* strain BL21(DE3) without any helper DNA. However, two genes encoding full-length scHda2p and scHda3p were coexpressed in bacteria rather than single expression because these two subunits were largely insoluble when expressed individually.

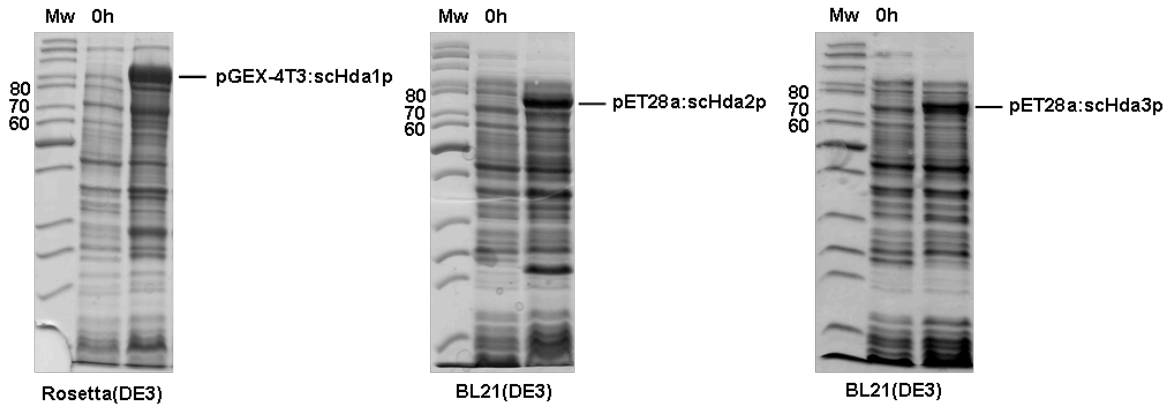


Fig. 7. Expression of each of three subunits comprising yeast class II Hda1 HDAC complex in *E.coli*.

Following combination using two or three expression plasmids lead to successfully coexpress two or all three subunits of yeast class II Hda1 HDAC complex in bacteria.

	scHda1p	scHda2p	scHda3p
scHda2p-scHda3p subcomplex	-	pET28a	pCDFDuet-1 MCS2
class II Hda1 HDAC complex	pETDuet-1 MCS1	pETDuet-1 MCS2	pCDFDuet-1 MCS2
	pET21b	pET30a	pCDFDuet-1 MCS2

Table 7. Coexpression strategy of expression plasmids encoding three subunits of yeast class II Hda1 HDAC complex.

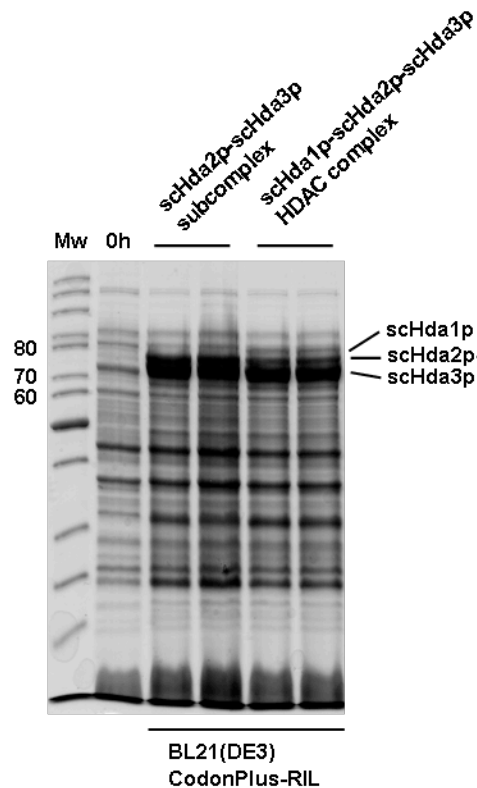


Fig. 8. Coexpression of two or three subunits comprising yeast class II Hda1 HDAC complex in *E.coli*.

3.2. Purification of recombinant Hda1 HDAC complex and its subunits

The three subunits of the yeast class II Hda1 HDAC complex encode more than 650 amino acids and were known to physically interact with each other (Fig. 9). To examine the network of interactions between scHda1p, scHda2p, and scHda3p, I expressed and purified the full-length (FL) scHda1p catalytic subunit, the scHda2p-scHda3p sub-complex (FLSC), and the scHda1p-scHda2p-scHda3p complex (FLC) by affinity, ion-exchange, and size-exclusion chromatography (SEC). FL scHda1p is well expressed and purified in *E.coli*, but scHda2p and scHda3p had to be coexpressed as individual expression led to insoluble material. Purified FL scHda1p showed a retention volume in gel filtration corresponding to an apparent molecular weight of 162kDa, suggesting a homodimer. Unexpectedly, the molecular weight of copurified FLSC was close to a heterotetramer (286kDa). To reconstitute FLC, we further subjected the scHda1p-FLSC mixture to SEC and, in

parallel, copurified all three subunits from an *E. coli* culture co-transformed with a bicistronic plasmid encoding scHda1p and scHda2p and a second compatible plasmid harboring the scHda3p gene. A peak corresponding to a molecular weight of 480kDa migrated on the gel filtration column, suggesting an apparent stoichiometry of 2:2:2 of scHda1p:scHda2p:scHda3p (Fig. 4).

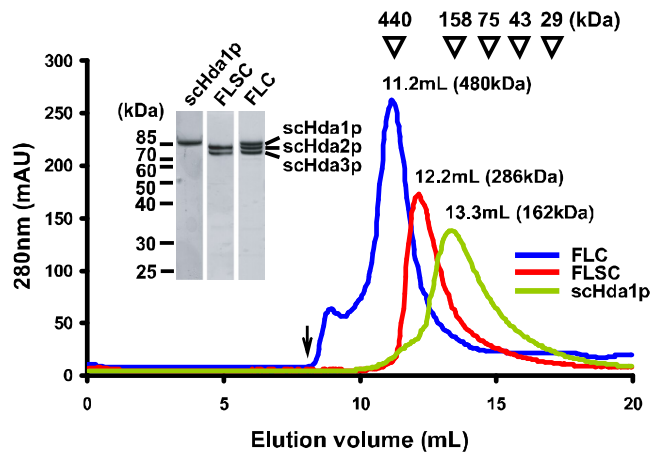


Fig. 9. SEC profiles of scHda1p (green), scHda2p-scHda3p subcomplex (red), and recombinant scHda1p HDAC complex (blue) at ~7.2ug/uL of protein concentration. Left inset shows 15% SDS-PAGE of each peak fraction. The standards for SEC experiments are shown at the top of profile. Arrow indicates the void volume.

3.3. *In vitro* deacetylase activity of recombinant Hda1 HDAC complex

The highly purified recombinant class II Hda1 HDAC complex catalyzed the deacetylation of the fluoro-substrate peptide *in vitro* efficiently but was also inhibited by TSA with an IC_{50} of 3.54nM (Fig. 10).

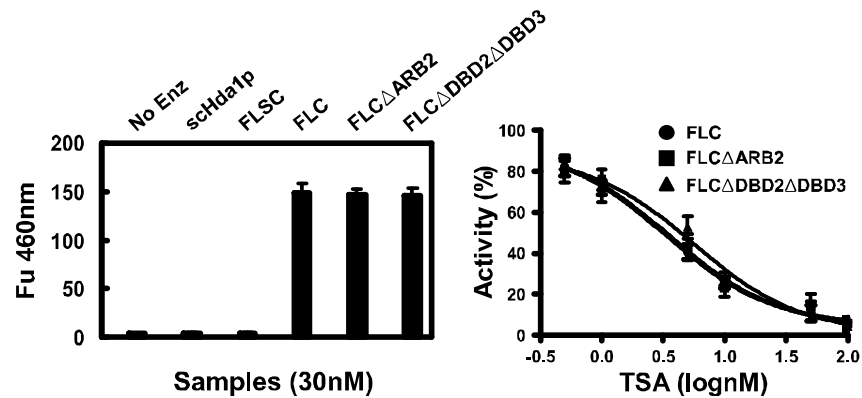


Fig. 10. *In vitro* deacetylase activities of various purified proteins (left), and TSA-mediated inhibitions of recombinant class II Hda1 HDAC complex (FLC) and truncated HDAC complexes (right). All data from two separate measurements were averaged, and the error bars correspond to the standard deviations of the two independent experiments.

3.4. Determination of stoichiometry between three subunits in solution

Interestingly, the estimated molecular weights of either FLC or FLSC are different from those previously reported for the native class II HDAC complex isolated from *S. cerevisiae* (Wu. J. *et. al.* 2001b). Sedimentation velocity experiments of each protein preparation at a concentration of $\sim 1\mu\text{g}/\mu\text{L}$ confirmed that the scHda1p catalytic subunit is a homodimer in solution, whereas scHda2p associates with scHda3p to form a heterodimer of a 1:1 ratio with an experimental molecular mass of 171kDa compared to a calculated molecular weight for a 1:1 of 152kDa (Fig. 11). Because the frictional ratio of FLSC was within the range for globular proteins ($f/f_0=1.20$; 1.05-1.35 for globular proteins and >3 for elongated shapes), an extended conformation of FLSC is unlikely to cause the apparent molecular mass differences by SEC and ultracentrifugation. Instead, FLSC showed the protein concentration-dependent dynamic assembly, indicating a dimer-tetramer equilibrium at low and high protein concentrations (Fig. 11). The experimental molecular mass of FLC by SEC at low (341kDa at $\sim 1\mu\text{g}/\mu\text{L}$) and high (480kDa at $\sim 7.2\mu\text{g}/\mu\text{L}$) protein concentrations or sedimentation velocity analysis (332kDa at $\sim 1\mu\text{g}/\mu\text{L}$) also indicated dynamic assemblies of 2:1:1 (homodimeric scHda1p and heterodimeric scHda2p-scHda3p) or 2:2:2 (homodimeric

scHda1p and heterotetrameric scHda2p-scHda3p) forms depending on protein concentrations (Fig. 9 and 11).

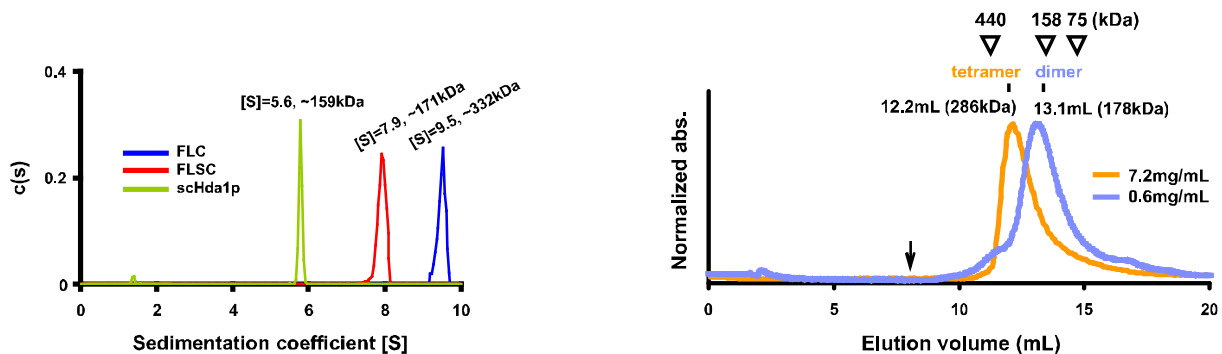


Fig. 11. Sedimentation velocity analysis (left) of scHda1p (green), scHda2p-scHda3p subcomplex (FLSC, red), and recombinant class II Hda1 HDAC complex (FLC, blue) at ~1 μ g/ μ L of protein concentration. Experimental molecular mass of each sample is indicated. SEC profiles (right) of protein concentration-dependent dynamic assembly of scHda2p-scHda3p subcomplex at high (~7.2 μ g/ μ L) and low (~0.6 μ g/ μ L) concentrations. Arrow indicates the void volume.

3.5. Domain organization of yeast class II Hda1 HDAC complex

The scHda1p is divided into an N-terminal half, which harbors the catalytic activity (Fig. 12) and a C-terminal half containing an ARB2 (Argonaute-Binding Protein-2) domain which was recently identified in fission yeast spArb2p, essential for histone H3 Lys9 (H3-K9) methylation, heterochromatin assembly, and siRNA generation (Buker, S. M. *et. al.* 2007). The Pfam database analysis with ARB2-containing sequences revealed that at least 33 proteins have related domain architecture (data not shown). To map the interaction domain of scHda1p with FLSC, two fragments, corresponding to N-terminal half (M1-A436) harboring the catalytic HDAC domain (HDAC) and the C-terminal ARB2 domain (N437-E706) based on secondary structure prediction and amino sequence alignment between class I and II HDACs, were individually expressed in *E.coli* and purified to homogeneity. According to pull-down assays, the N-terminal half of scHda1pHDAC spanning residues M1-A436 is sufficient to interact with FLSC, while the C-terminal scHda1pARB2 does not interact with the scHda2p-scHda3p subcomplex (Fig. 13).

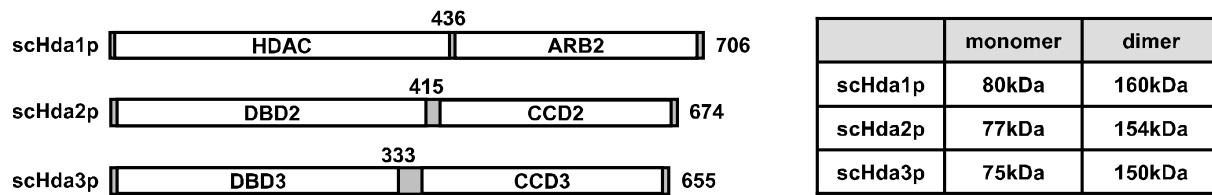


Fig. 12. Summary of domain organization of three subunits comprising of yeast class II Hda1 complex. Calculated molecular weights correspond to monomeric and dimeric conformations are summarized in table.

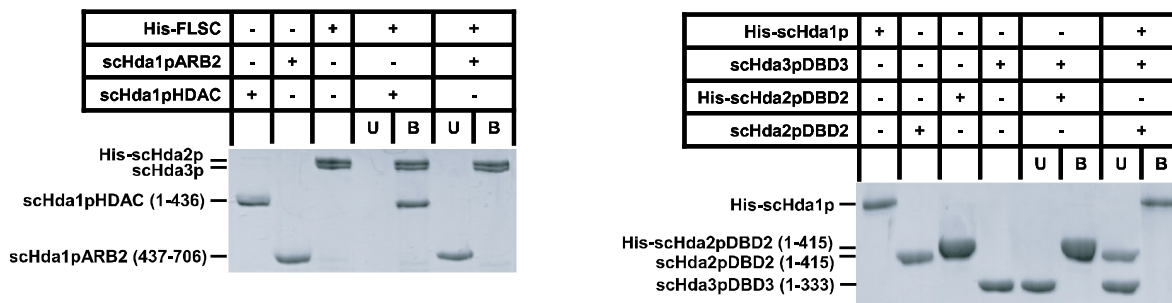


Fig. 13. His-pull down assay for identification of interaction between three subunits. Reactions were precipitated with Ni-NTA resin, and the unbound (U) and bound (B) fractions were analyzed with 15% SDS-PAGE and Coomassie staining.

Extensive amino acid sequence alignment, secondary structure prediction, and 3D-fold recognition of scHda2p and scHda3p revealed that their N-terminal halves (DBD2 and DBD3 in Fig. 12 and 14) may be related to helicases especially found in SWI2/SNF2 chromatin-remodeling domains. The C-terminal halves showed the characteristics of long coiled-coil domains (CCD2 and CCD3 in Fig. 12 and 14) spanning residues of approximately 200 amino acids. Although coiled-coil domains of many proteins are known to be crucial for protein-protein interaction (Strauss, H. M. *et. al.* 2008), we have asked which domain(s) of scHda2p and scHda3p contribute to assemble the subcomplex, and finally interact with scHda1p. Based on the secondary structure prediction and limited proteolysis of FLSC with two proteases, subtilisin and trypsin, we have identified two N-terminal fragments of scHda2p and scHda3p spanning residues N5-K415 (DBD2) and I6-K333 (DBD3), respectively. The two CCDs were largely degraded by proteolytic digestion in these experiments (data not shown). To examine the interaction between scHda2p and scHda3p, we assessed the binding of

the His-tagged scHda2p N-terminal domain to that of untagged scHda3p using a pull down assay. Interestingly, the two N-terminal domains neither bound to each other nor to the catalytic subunit of scHda1p (Fig. 12). However, coexpression and copurification of several truncated complexes revealed that the two CCDs of the non-catalytic subunits are essential to reconstitute the class II Hda1 HDAC complex with three subunits (Fig. 14). Furthermore, the purified scHda1pARB2 and the two N-terminal domains of scHda2p and scHda3p migrated as monomers in gel filtration. A truncated complex consisting of scHda1p and two CCDs of the structural subunits migrated as a heterocomplex (either heterotetramer or heterohexamer depending on protein concentration) on the gel filtration column (Fig. 14). *In vitro* fluorometric deacetylase assay with fluoro-peptide of the truncated complex lacking scHda1pARB2, scHda2pDBD2, and scHda3pDBD3 showed indistinguishable deacetylase activity compared with that of FLC (Fig. 10), suggesting that the truncated scHda2pCCD2-scHda3pCCD3 sub-complex serves as a scaffold for an enzymatically competent yeast class II Hda1 HDAC complex (Fig. 14).

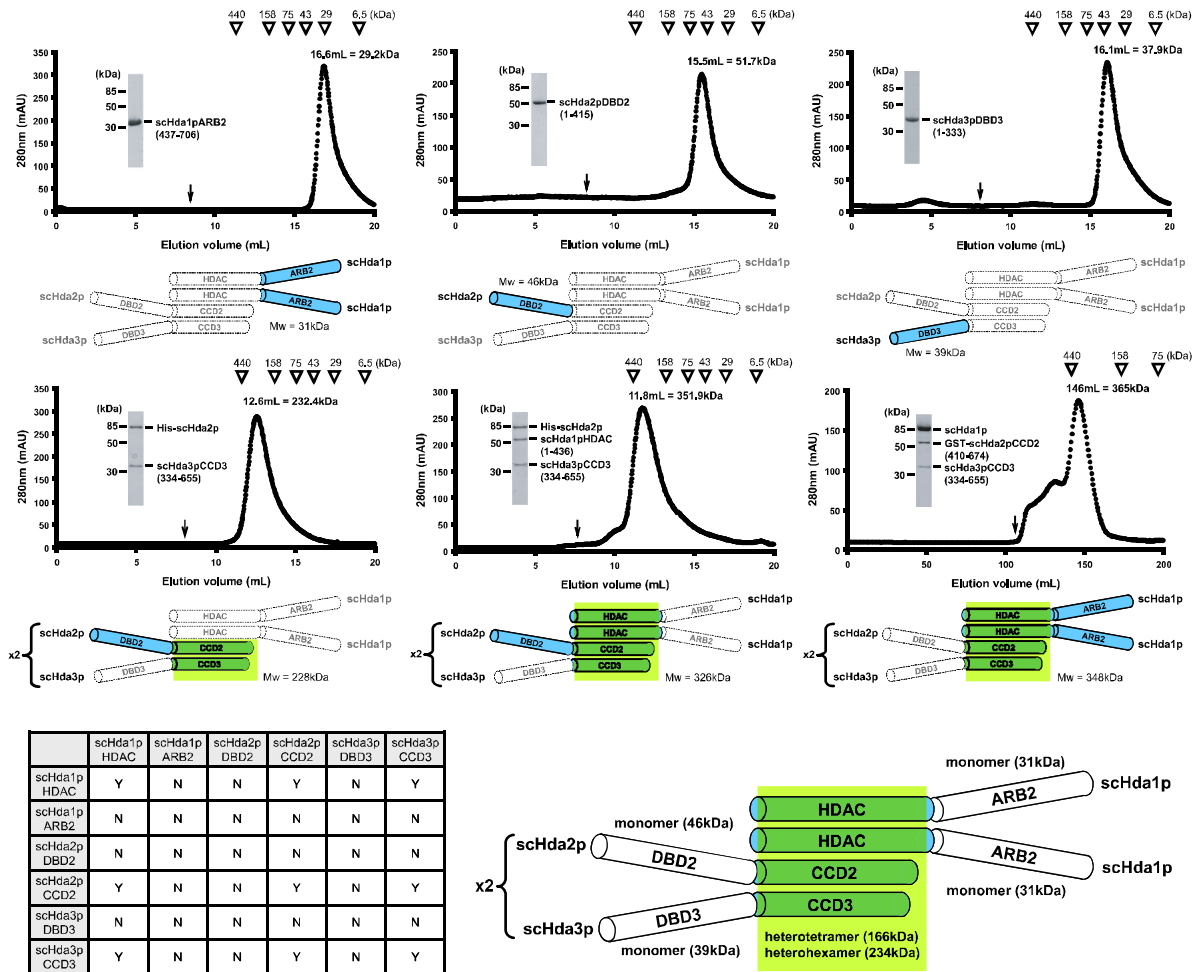


Fig. 14. SEC profiles of individual domains, truncated sub-complexes, and truncated HDAC complexes at $\sim 7.9\mu\text{g}/\mu\text{L}$ of protein concentration. Peak fractions after SEC are shown by 15% SDS-PAGE and Coomassie staining. Estimated molecular weights measured from retention volume of each peak fraction are indicated. Schematic diagrams together with calculated molecular weights corresponding to purified proteins (blue color) are shown below each profile. Domains which are crucial for interaction between three subunits are highlighted in green color. Void volumes are indicated by arrows. Summary of interaction network between scHda1p, scHda2p, and scHda3p are highlighted.

3.6. Identification of scHda3pDBD3

Primary aim to crystallize an intact full-length HDAC complex expressed in *E.coli*

was unsuccessful. Therefore, I performed limited proteolytic digestion of purified FLSC and/or FLC to identify stable proteolytic fragments for crystallization (Fig. 15). Two fragments were identified corresponding to scHda2pDBD2 (N5-K415) and scHda3pDBD3 (I6-K333). The N-terminal domain of scHda2p could be further cleaved to two sub-fragments (approximately 28kDa and 20kDa) with trypsin to protein ratio of 1:10, whereas scHda3pDBD3 was resistant to excess amount of trypsin.

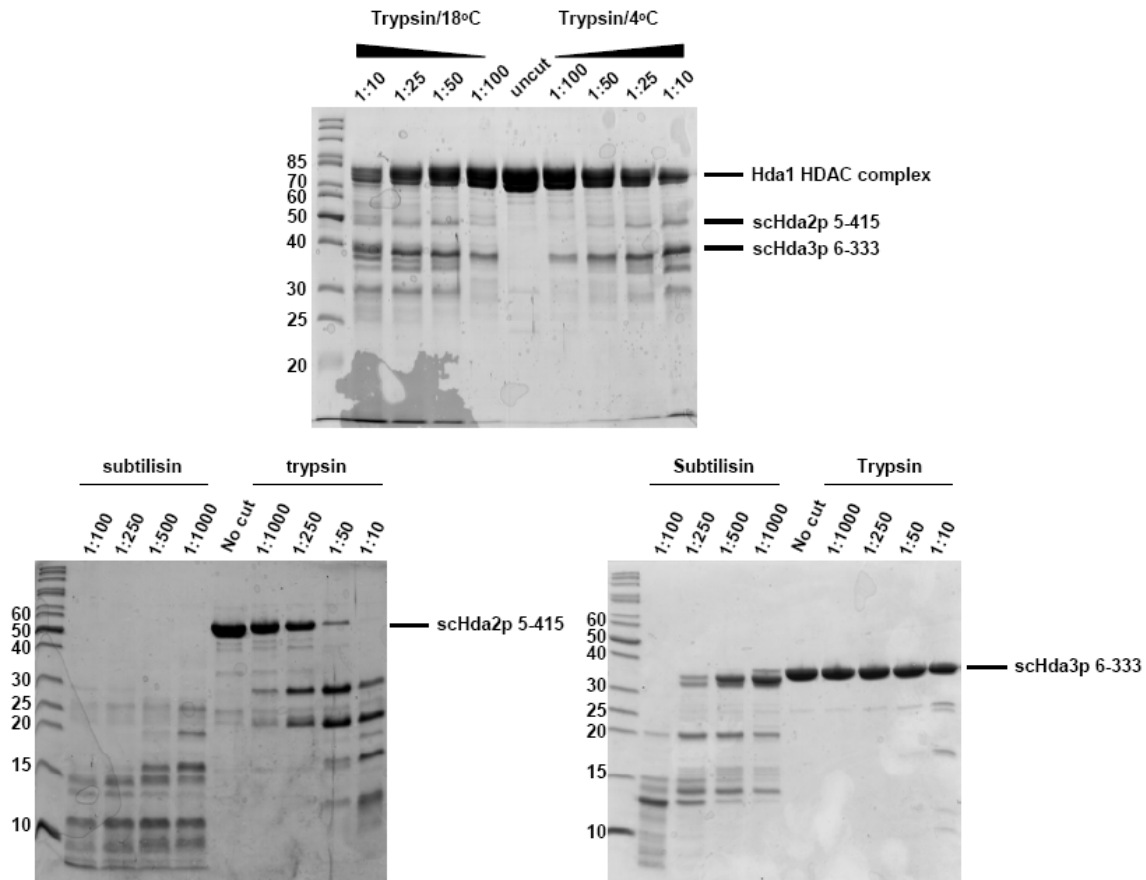


Fig. 15. Limited proteolysis of recombinant class II Hda1 complex, scHda2p 5-415, and scHda3p 6-333 using subtilisin and trypsin.

3.7. Crystallization and structure determination of scHda3pDBD3^{K168A/Q169A/K170A}

Initially, I obtained crystals of wild-type scHda3pDBD3 encompassing residues I6-K333 by the hanging drop method. However, the crystal quality was unsuitable

for data collection and structure determination because of crystal twinning (Fig. 16). Neither extensive screening of crystallization conditions with additives and detergents nor micro-seeding was successful to improve the crystal quality. Therefore, we decided to engineer the protein by lysine to alanine mutation of several positively charged residues which were predicted to be solvent exposed. Four possible clusters for mutation (K168A/Q169A/K170A, K253A/K254A, K223A/R224A/E225A/K227A, and K66A/E67A) were screened on SER server (<http://www.doe-mbi.ucla.edu/Services/SER>) according to the SERp score. Among them, one of the scHda3pDBD3 variants (scHda3pDBD3^{K168A/Q169A/K170A}) with highest SER score was successfully crystallized in single crystals which belong to space group $P2_1$ with $a=55.18$, $b=116.57$, $c=55.19\text{\AA}$ and contain two molecules in the asymmetric unit (Fig. 16 and Table 8). The overall data quality of native scHda3pDBD3^{K168A/Q169A/K170A} crystal was assessed with the PHENIX program suite prior to the preparation of the SeMet derivative for MAD phasing. Unfortunately, however, a possible pseudo-merohedral twinning problem was indicated by the nearly identical a - and c -axis lengths and Wilson statistics for acentric reflections ($\langle I^2 \rangle / \langle I \rangle^2$ of 1.8). The twin operator was determined as $I, -k, h$, and the twin fraction was 0.296. To avoid the undesirable problems from twinned data during structure determination, a wide variety of crystallization conditions for untwined SeMet scHda3pDBD3^{K168A/Q169A/K170A} crystals were explored. Finally, we identified an untwined crystal of the SeMet derivative and successfully determined the crystal structure by MAD phasing and refined it. After successful search of this model in the twinned native crystal data set, it was refined further with R_{work} of 19.8% and R_{free} of 23.7% at 2.2Å resolution (Table 8).

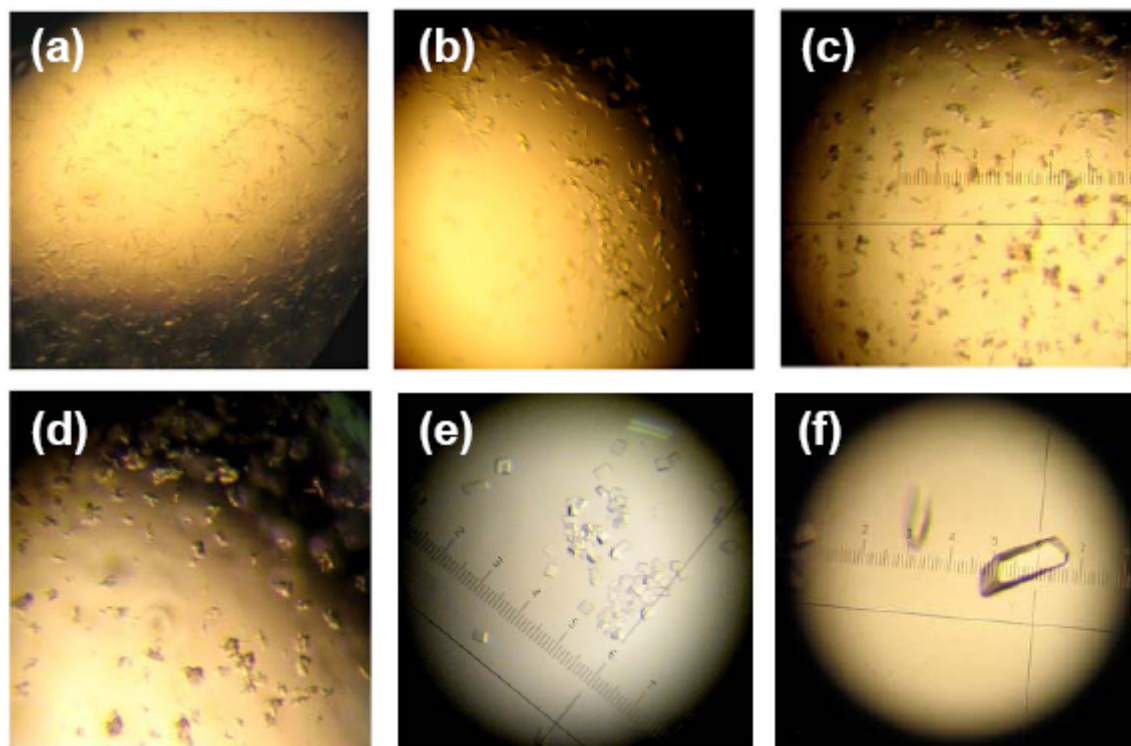


Fig. 16. Crystallization of scHda3p DNA-binding domain and its variant. Needle-cluster or sea urchin-like crystals of wild-type scHda3pDBD3 were grown against (a, b) 0.15M K_2HPO_4/NaH_2PO_4 pH6.6, 17%(w/v) PEG1000, 12%(w/v) PEG400, (c, d) 0.1M Tris-HCl pH8.0, 0.2M $(NH_4)_2HPO_4$, 22%(w/v) PEG3350. Crystals of scHda3pDBD3^{K168A/Q169A/K170A} 6-333 were grown against (e) 0.1M HEPES pH7.5, 0.2M $(NH_4)_2SO_4$, 23%(w/v) PEG3350 for native and (f) 0.1M magnesium formate dehydrate, 15%(w/v) PEG-3350 for SeMet.

	Native	SeMet peak	SeMet inflection	SeMet high remote
Data collection				
Beamline	SLS-X06SA	DESY-X12	DESY-X12	DESY-X12
Wavelength (Å)	1.0007	0.97815	0.97876	0.97469
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions (Å / °)	55.18, 116.57, 55.19	60.08, 132.58, 90.17	60.08, 132.58, 90.17	60.08, 132.58, 90.17
(a, b, c / α, β, γ)	90, 97.42, 90	90.00, 92.38, 90.00	90.00, 92.38, 90.00	90.00, 92.38, 90.00
Twin law / fraction	l, -k, h / 0.296			
Resolution (Å)	30 – 2.2 (2.34 – 2.2)	30 – 3.0 (3.12 – 3.0)	30 – 3.0 (3.12 – 3.0)	30 – 3.0 (3.12 – 3.0)
R_{sym} (%)	4.3 (11.0)	12.6 (54.5)	12.9 (56.8)	13.0 (59.3)
Completeness (%)	95.4 (88.7)	99.9 (100)	99.9 (99.9)	99.9 (99.9)

I/sigma (I)	21.40 (10.27)	11.38 (2.74)	11.21 (2.63)	11.13 (2.54)
Reflections (total / unique)	128877 / 33474	219245 / 55625	219191 / 55617	219181 / 55620
Redundancy	3.9	3.9	3.9	3.9
Phasing (MAD)				
Resolution (Å)		30 – 3.0		
Phasing power		0.771		
Figure of merit		0.443		
Refinement				
Resolution (Å)	20 – 2.2	20 – 3.0		
Number of reflections (total / test)	33430 / 1741	28181 / 1410		
$R_{\text{work}} / R_{\text{free}}$ (%)	19.8 / 23.7	22.7 / 28.0		
Number of protein / water atoms	4666 / 149	9555 / 0		
RMSD bond length / angle (Å / °)	0.007 / 1.080	0.005 / 0.904		
Mean B value (protein / solvent) (Å)	39.47 / 34.56	40.11 / 0		

Table 8. Data collection, MAD phasing, and refinement statistics. Values in parentheses are for the highest resolution shell. $R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - I_h| / \sum_h \sum_i I_{h,i}$, where I_h is the mean intensity of the i observations of symmetry-related reflections of h . $R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, where $F_{\text{obs}} = F_p$, and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections).

3.8. Crystal structure of scHda3pDBD3^{K168A/Q169A/K170A}

The refined structures against native and SeMet peak data of scHda3pDBD3^{K168A/Q169A/K170A} contained two and four molecules in the asymmetric unit, respectively. I observed identical homodimeric assemblies of two molecules (buried surface area = 767.5Å²/protomer, corresponding to 5.3% of the total surface area of one protomer) related by two-fold non-crystallographic symmetry. All six protomers (two in the native and four in the SeMet derivative) superimpose well with RMS deviations of ~0.42Å for 286 matched Cα atoms. Two loops spanning residues K166-N172 (between β3 and β4) and R224-Y232 (between α7 and β6) are invisible in the native structure and are probably flexible. The first of these disordered segments harbors the mutations

(K168A/Q169A/K170A). One of these loops, out of four copies, is visible in the SeMet-substituted structure (data not shown).

The scHda3pDBD3^{K168A/Q169A/K170A} monomer forms a α/β -mixed structure consisting of 12 helices and 6 strands, and has a dimension of 51.6 x 43.1 x 40.0Å (Fig. 17). The overall structure may be divided into two subdomains, a RecA-like subdomain with a core of seven parallel β -strands decorated by six α -helices, and a helical domain with a six helix bundle inserted between β 6 and α 10 of the RecA-like domain. A RecA-like domain of scHda3pDBD3^{K168A/Q169A/K170A} is topologically similar to that of superfamily 2 (SF2) helicases (Caruthers, J. M. *et. al.* 2002, Singleton, M. R. *et. al.* 2007). The helical domain bulge is stabilized by extensive hydrophobic interactions between its six α -helices (α 1, α 2, α 3, α 8, α 9, and α 12) and by packing interactions with the RecA-like core. The dimer observed in the crystals is mediated by four hydrogen bonds and seventeen salt bridges from α 1, α 3, α 4, α 9, and α 10.

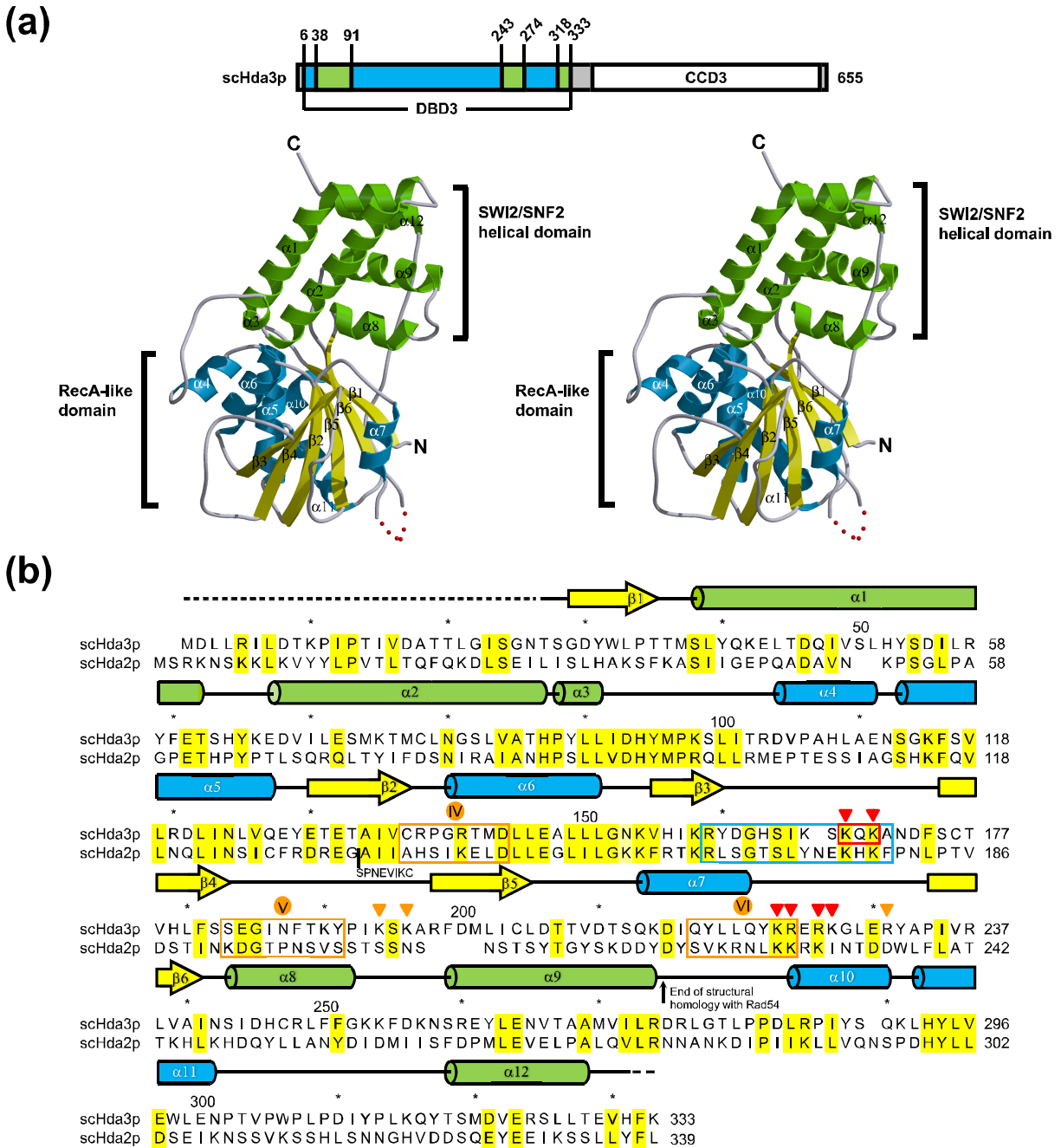


Fig. 17. Crystal structure and sequence conservation of scHda3p DNA-binding domain. (a) Overall structure of scHda3pDBD3^{K168A/Q169A/K170A} is shown as ribbon stereo-diagrams. The scHda3pDBD3 comprises two subdomains, a RecA-like domain (blue and yellow) and a SWI2/SNF2 helical domain (green). Flexible loop (R224-L229) which is not visible in electron density map is shown as red dot. (b) Structural alignment and sequence conservation in DBDs of two structural subunits. Conserved residues are shown in yellow, and mutation points

to facilitate the crystallization are highlighted in red box. Three helicase motifs of scHda3pDBD3 equivalent to those of SF1 and SF2 helicases are indicated in orange boxes. A SWI2/SNF2-specific motif (Snf-D) found in zebrafish drRad54 structure is shown in blue box. Positively charged residues of scHda3pDBD3 which are critical for DNA recognition are highlight in red (highly effective) and orange (less effective) triangles. Every 10th residue is marked with an asterisk.

3.9. Structural comparison of scHda3pDBD3^{K168A/Q169A/K170A} with SWI2/SNF2 enzymes

BLAST search with the amino acid sequence of scHda3pDBD3 did not show any significant homology with any known protein except scHda3p orthologs in several fungi. Position-specific iterative BLAST search gave only moderate results (amino acid sequence identities of 13-16%) between scHda3pDBD3 and several chromo-domain helicase DNA-binding proteins. However, database search in DALI revealed that scHda3pDBD3^{K168A/Q169A/K170A} structure most closely resembles that of SWI2/SNF2 chromatin-remodeling enzymes such as Rad54 (Fig. 18, Z-scores = 16.6 with ssRad54 and 15.7 with drRad54). Generally, two tandem lobes of Rad54 which consist of RecA-like α/β domains with a helical protrusion are structurally related (Thomae, N. H. *et. al.* 2005, Duerr, H. *et. al.* 2005). The ATP-binding site is located in the first RecA-like domain where a Walker A motif (G/DXXXXGKS/T) is found. The scHda3pDBD3^{K168A/Q169A/K170A} structure and the second helicase lobe of the SWI2/SNF2 ATPase core of two Rad54 family enzymes share a common structural feature of six parallel β -strands encompassed by six α -helices (RecA-like subdomain) and six α -helical bundle (SWI2/SNF2-type helical subdomain). The RMS deviations of these proteins were 2.9-3.7Å for T27–R276 C α atoms of scHda3pDBD3^{K168A/Q169A/K170A}. Furthermore, three helicase motifs VI, V, and IV of scHda3pDBD3^{K168A/Q169A/K170A} are structurally equivalent to those of the two closest homologues (Fig. 18). Similar to the ‘ATP-less’ second helicase lobe of the SWI2/SNF2 family enzyme Rad54, scHda3pDBD3^{K168A/Q169A/K170A} also does not have any obvious Walker motifs, hallmarks for ATP-binding.

However, scHda3pDBD3^{K168A/Q169A/K170A} and the second helicase lobes of these archaeal and zebrafish Rad54 enzymes have noticeable structural differences. First,

the loop between $\beta 3$ and $\beta 4$ of scHda3pDBD3^{K168A/Q169A/K170A}, where three residues (K168, Q169, and K170) were simultaneously mutated to alanine for crystallization, is largely shifted by $\sim 15\text{\AA}$ toward to $\alpha 6$ of scHda3pDBD3^{K168A/Q169A/K170A} compared to the corresponding the α -helices ($\alpha 5$ of ssRad54-lobe 2 and $\alpha 7$ of drRad54-lobe 2) of two structural homologues. Likewise, a second loop between $\beta 4$ and $\beta 5$ of scHda3pDBD3^{K168A/Q169A/K170A} which corresponds to helicase motif V is also shifted into the direction of $\alpha 6$ in scHda3pDBD3^{K168A/Q169A/K170A} by $\sim 5\text{\AA}$ compared to equivalent elements of two Rad54 structures. The other difference between the three structures is observed at the C-terminal extension of scHda3pDBD3^{K168A/Q169A/K170A}. Structural homology of scHda3pDBD3^{K168A/Q169A/K170A} with ssRad54 and drRad54 ends with R276. The scHda3pDBD3 continues with two α -helices ($\alpha 11$ and $\alpha 12$) connected by a Pro-rich (five Pro residues between P302 and T318) loop (Fig. 17b, 18). The C-terminal extension of $\alpha 12$ and the Pro-rich loop have no equivalent elements in the two Rad54 structures. They extend the contact area and stabilize the SWI2/SNF2-type helical domain by salt-bridges and hydrophobic interactions (W306, L308, I311, V322, and L326) with $\beta 1$, $\alpha 1$, $\alpha 8$, and $\alpha 9$.

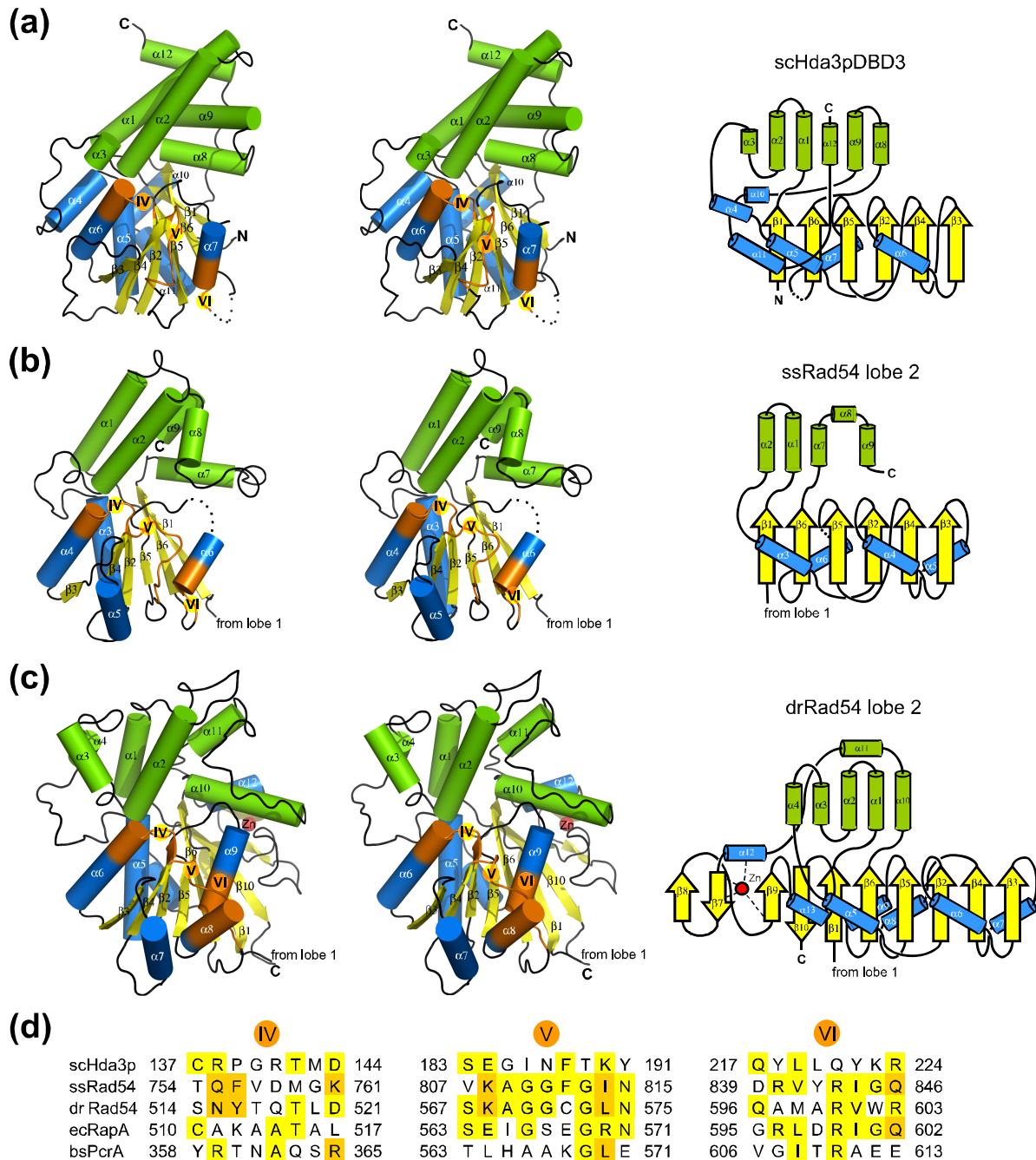


Fig. 18. Structural comparison of scHda3pDBD3^{K168A/Q169A/K170A} with its two closest structural homologues. Stereo-drawings of (a) *Saccharomyces cerevisiae* scHda3pDBD3^{K168A/Q169A/K170A}, (b) *Sulfolobus solfataricus* ssRad54 (PDB ID code 1Z6A), and (c) *Danio rerio* drRad54 (PDB ID code 1Z3I) are illustrated on the basis of superposition. Color code of all structures is same with those of Fig. 11, and topology diagrams of each structure are shown in right side of each stereo-diagram. Only the second lobes (lobe 2) of two Rad54 structures are represented for clarity.

helicase motifs (IV, V, and VI) predicted on the basis of helicase structures are shown as orange. (d) Structure-based sequence alignment for three functional motifs between scHda3pDBD3 and four structural homologues.

3.10. DNA-binding activities of Hda1 HDAC complex and its subunits

Recently it was demonstrated that the class II HDAC complex interacts with centromere DNA in *S. cerevisiae* (Kanta, H. *et. al.* 2006). Nucleic acid binding of the HDAC enzyme may be direct or mediated by cofactors. Extensive database search, identification of the domain architecture of class II HDAC complex and crystal structure of scHda3pDBD3^{K168A/Q169A/K170A} suggested that the two N-terminal domains of the non-catalytic subunits scHda2p and scHda3p may function as the nucleic acid binding modules in yeast. Therefore, I first examined the DNA-binding properties of yeast class II Hda1 HDAC complex and its subunits using *in vitro* band shift assays (Fig. 19). Both FLC and FLSC directly bound to 50bp of duplex DNA. In contrast, the truncated complex with both N-terminal domains of the structural subunits (FLC Δ DBD2 Δ DBD3) failed to interact with oligo-nucleotide, suggesting that scHda2pDBD2 and scHda3pDBD3 are responsible for direct recognition of dsDNA (Fig. 19a). In general, SWI2/SNF2 chromatin-remodeling enzymes bind to dsDNA in a nucleic acid sequence and structure-independent manner. To further characterize the binding specificities of the non-catalytic subunits, I performed band shift assays with 50mers of ssDNA, 50bp of dsDNA, and oligonucleotides of various shapes (Fig. 19b). Interestingly, FLSC bound not only duplex DNA but also single-stranded Poly(A) chain which does not possess secondary structure. FLSC also efficiently bound to a splayed arm which has 25bp of a stem and 25mers of a single-stranded arm as well as a Holliday junction which has four each of 25bp duplex DNA, suggesting a complete lack of DNA sequence and structure-specificity of the scHda2p-scHda3p subcomplex. I also observed the direct interaction of various oligonucleotides with highly purified scHda3pDBD2 and scHda3pDBD3 even though apparent binding affinities of individual DBDs were \sim 10-fold weaker than those of intact FLSC (Fig. 19c and 19d), which may act as a bi-dentate ligand with strongly enhanced affinity.

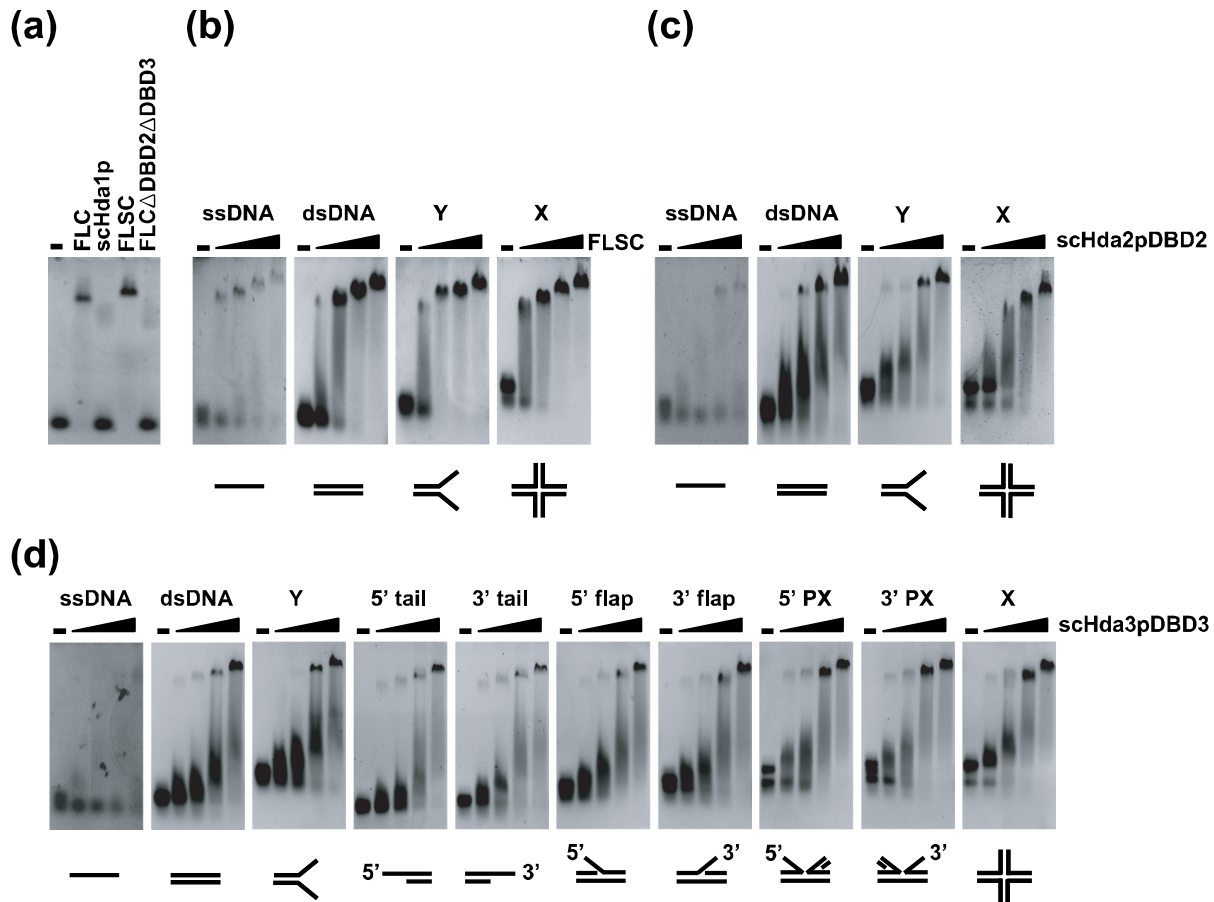


Fig. 19. DNA-binding activities of yeast class II Hda1 HDAC complex and its two DBDs. (a) Functional HDAC complex recognizes the DNA molecule via both N-terminal domains of scHda2p and scHda3p. A 50bp of dsDNA (5 μ M) was mixed with 4 μ M of each protein sample, and DNA-binding was analyzed by 1% agarose gel electrophoresis at 4 $^{\circ}$ C. (b) Four DNA samples (5 μ M) were incubated with FLSC (0.5-4 μ M) and (c) scHda2pDBD2 M1-K415 (2-16 μ M) at 4 $^{\circ}$ C for 1h, and all reactions were analyzed by 1% agarose or 3-12% polyacrylamide gel electrophoresis at 4 $^{\circ}$ C. Each gel was stained with 1x Gel star solution. (d) Various DNA substrates (5 μ M) were incubated with scHda3pDBD3 M1-K333 (2-16M) at 4 $^{\circ}$ C for 1h, and all mixtures were visualized as 1x Gel star staining after 3-12% polyacrylamide gel electrophoresis at 4 $^{\circ}$ C.

3.11. Identification of the DNA-binding sites of scHda3pDBD3

As mentioned above, the scHda3pDBD3^{K168A/Q169A/K170A} structure closely resembles

the second helicase lobe of SWI2/SNF2 family of the ssRad54 ATPase core, whereas the DNA-binding site of the archaeal enzyme is predominantly located on the first lobe of the bi-lobal structure where the Walker A motif is found. Only few residues from the second lobe (e.g., R728 and K781) contribute to the DNA contact. To address a possible DNA interaction area of scHda3pDBD3, sixteen K/R→A mutants chosen from electrostatic surface calculations were constructed (Fig. 20). I first selected eight mutation sites by homology with the DNA-binding site of ssRad54 (Fig. 20a, left). DNA-binding activities of these eight variants were analyzed with 50bp of dsDNA by *in vitro* band shift assay, and I confirmed that none of these mutations disrupted the DNA interaction (Fig. 20b). The other mutation area of scHda3pDBD3 was deduced from the crystal structure and DNA-binding activity of scHda3pDBD3^{K168A/Q169A/K170A}. In fact, two of four variants used for crystallization showed very low DNA-binding activities (Fig. 20b, triple: K168A/Q169A/K170A and quadruple: K223A/R224A/E225A/K227A). Coincidentally, two loops containing multiple mutations of two critical variants were solvent exposed and disordered showing no defined electron density in the six protomers of native and SeMet scHda3pDBD3^{K168A/Q169A/K170A} structures, except for one.

Positions of the three helicase motifs (IV, V, and VI) in scHda3pDBD3^{K168A/Q169A/K170A} structure identified by structural superposition with three SWI2/SNF2 enzyme structures (ssRad54, drRad54, and ecRapA) and one superfamily 1 (SF1) structure (bsPcrA) were mapped on the electrostatic surface of scHda3pDBD3^{K168A/Q169A/K170A} structure (Fig. 20a, right). Any K→A mutation (e.g., R138A and/or R141A) in the helicase motif IV caused insolubility when expressed in *E.coli*. However, motif V mutants (K190A and K194A/K196A) and VI mutant (K223A/R224A) were successfully purified to homogeneity and characterized for DNA-binding by band shift assays (Fig. 20b). The double mutation introduced in motif V (K194A/K196A) which is close to the K168A/K170A mutation area showed moderately lower DNA-binding affinity to duplex DNA, but multiple K/R→A mutations of K168/K170 loop, and motif VI enriched with many positively charged residues (e.g., K223, R224, and R226) abolished DNA-binding. Therefore, I concluded that these two areas are crucial for direct DNA recognition. These areas are different from the SWI2/SNF2 family of ssRad54 enzyme.

catalytic scHda2p-scHda3p subcomplex (Wu, J. *et. al.* 2001b). Extending these studies, I found that scHda2p and scHda3p dimerize by their C-terminal coiled-coil domains (CCD2 and CCD3) and directly bind to the scHda1p homo-dimer through the scHda1pHDAC domain as a truncated heterocomplex, scHda1pHDAC-scHda2pCCD2-scHda3pCCD3 (stoichiometry of 2:1:1 or 2:2:2 dependent on the protein concentration) which has full deacetylase activity *in vitro*. According to the previously studies of class II Hda1 complex, direct interaction between GST-scHda1p and ³⁵S-labeled scHda3p was suggested (Wu, J. *et. al.* 2001b). However, I failed to observe a significant interaction between these two subunits since coexpression of scHda1p and scHda3p (or scHda2p) in bacteria did not indicate the existence of a stable scHda1p-scHda3p (or scHda2p) binary complex (data not shown). The domain characterization of the yeast class II Hda1 HDAC complex presented here clearly showed that only scHda1pHDAC and the CCDs of the two structural subunits oligomerize and participate in the reconstitution of a functional deacetylase complex. I identified the N-terminal domains of scHda2p and scHda3p (DBD2 and DBD3) as DNA recognition modules with the potential to target the deacetylation activity of scHda1pHDAC to histone tails in nucleosomes *in vivo*.

Structural details of DNA-binding domain of yeast class II Hda1 HDAC complex were revealed by the determination of crystal structure of scHda3pDBD3^{K168A/Q169A/K170A}. Surprisingly, the molecular structure of scHda3pDBD3^{K168A/Q169A/K170A} resembles the helicase fold as found in the SWI2/SNF2-type chromatin-remodeling ATPase such as ssRad54. The amino acid sequence identity between the two proteins was insufficient to be identified by BLAST search (e.g., identity of 12% between scHda3pDBD3 6-333 and ssRad54 664-906). Moreover, scHda3pDBD3^{K168A/Q169A/K170A} had a structural similarity with only the C-terminal second lobe of known Rad54 structures. In general, SF1 helicases (e.g., PcrA) and SF2 enzymes (e.g., HCV NS3 and Rad54) possess a bi-lobal structure which consists of two tandem RecA-like protomers with related folds even though ATP-binding and ATPase activity is confined to the N-terminal lobes.

As there is only one helicase domain in scHda3p, I considered whether the N-terminal domain of the second structural subunit, scHda2p, may play the role of a second helicase domain (Fig. 17b). Indeed, domain architecture and amino acid sequences are similar in scHda2p and scHda3p, with 29.1% identical residues in the DBDs (333 residues) and 24.9% in the CCDs (261 residues) indicating strongly a close structural and probably also functional relationship between the two structural subunits of the complex which both bind DNA. In two different crystals of

scHda3pDBD^{K168A/Q169A/K170A}, identical dimeric assemblies are found, probably a consequence of the very high protein concentration under crystallization conditions, because in solution, at moderate concentrations, the protein is monomeric. The observation, nevertheless, designates a sticky surface area, possibly used in the heterodimeric and heterotetrameric scHda2p-scHda3p subcomplexes. Although these are held together predominantly by their CCD2 and CCD3 domains, the DBD interactions may add to their stability.

Interestingly, an other class II HDAC enzyme, spClr3p in fission yeast *S. pombe*, forms a quaternary complex (SHREC complex) with three subunits of which one, spMit1p, has a SNF2/helicase domain inserted in the amino acid sequence (Sugiyama, T. *et. al.* 2007), suggesting that the two activities (deacetylation and DNA-binding/chromatin remodeling) may be integrated in one complex to control the gene expression in fungi as seen in the Hda1 HDAC complex.

Recently, it was reported that all three subunits of the Hda1 HDAC complex associate exclusively with centromeric DNA (Kanta, H. *et. al.* 2006). Data presented here confirm the DNA-binding activity of the HDAC complex, but disagree with the claim of DNA specificity. I show by *in vitro* band shift assays and mutational studies that DNA-binding is to the N-terminal DBDs of scHda3p, and probably also to scHda2p, in a sequence and structure-independent mode as seen with the SWI2/SNF2 chromatin-remodeling factors even though DNA-binding areas between two molecules are completely different. Based on the identification of DNA contact areas of scHda3pDBD3 by mutagenesis studies, I propose a possible DNA-binding mode of scHda3pDBD3 by modeling with 20bp of dsDNA (Fig. 15d). Two crucial loops (K168/K170 and K223/R224/R226/K227) as well as K194/K196 residues of scHda3pDBD3 protruded from RecA-like subdomain are well positioned into two minor grooves and one major groove of dsDNA and the potential CCD3 extension from the end of scHda3pDBD3 is located in the opposite space of protein-DNA contact. Furthermore, minimal length of dsDNA (17bp) is deduced from the model of protein-DNA complex and verified with FLSC and scHda3pDBD3 by *in vitro* band shift assay (data not shown).

Microarray deacetylation maps of the yeast chromosome (Robyr, D. *et. al.* 2002) had suggested earlier global and untargeted genome-wide deacetylation by class I Rpd3 and class II Hda1 HDAC complexes. Especially, more than 69% of the intergenic regions (IGRs) affected by the class II HDAC complex in *S. cerevisiae* were not

regulated by a *TUP1* deletion, suggesting scTup1p-independent mechanism of genome-wide deacetylation in yeast chromatin. On the basis of our structural and functional studies we propose that both N-terminal DBDs of the two structural subunits scHda2p and scHda3p serve as unspecific DNA-binding modules and as anchors for scHda1p in proximity of the H2B and H3 histone tails of neighboring nucleosomes for deacetylation (Fig. 21).

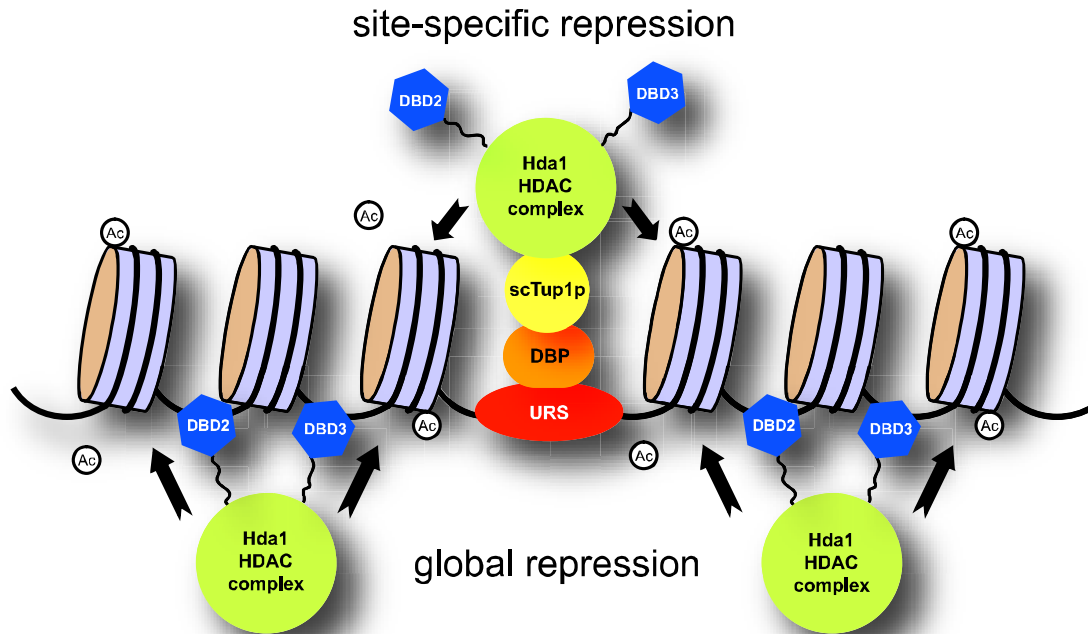


Fig. 21. Possible mechanisms for sequence-specific and untargeted deacetylation of nucleosomes by class II Hda1 HDAC complex in yeast. Promoter-targeted deacetylation takes place by recruitment of Hda1 HDAC complex to specific site (URS) through scTup1 repressor and sequence-specific DNA-binding protein (DBP). On the other hand, untargeted global deacetylation could be mediated by direct DNA recognition of class II Hda1 HDAC complex via DBD2 and DBD3 of two structural subunits.

Despite the relationship of the molecular structure and fold of scHda3pDBD3 with the SWI2/SNF2 chromatin-remodeling enzymes including the accessory domain (e.g., SWI2/SNF2-specific helical domain), the recombinant Hda1 HDAC complex lacks translocase/supercoiling activity (data not shown) and has lost ATPase activity. It shares the DNA-binding property albeit using different contact areas and thus

represents another example of extensive structural remodeling resulting in gain or loss of function of an 'ancestor' protein module during evolution.

5. References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. -W., Loerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta. Crystallogr. D.* **58**, 1948-1954.

Brünger, A. T. (2007) Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728-2733.

Buker, S. M., Iida, T., Bühler, M., Villén, J., Gygi, S. P., Nakayama, J. & Moazed, D. (2007) Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nat. Struct. Mol. Biol.* **14**, 200-207.

Bürger, M. J. (1960) *Crystal Structure Analysis*, Wiley, New York, 53-68

Cahn, R. W. (1954) Twinned crystal. *Advances in Physics* **3**, 363-445

Caruthers, J. M. & McKay, D. B. (2002) Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123-133.

Dauter, Z. (2003) Twinned crystals and anomalous phasing. *Acta Crystallogr.*, D59, 2004-2016

Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K. P. (2005) X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363-373.

Emsley, P. & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D.* **60**, 2126-2132.

Felsenfeld, G. & Groudine, M. (2003) Controlling the double helix. *Nature* **421**, 448-453

Giacovazzo, C. (1992) *Fundamentals of Crystallography*, Oxford Univ. Press, New York, 83-87.

Gorbalenya, A. E. & E. V. Koonin (1993) Helicases: amino acid sequence comparisons and structure-function relationship. *Curr. Opin. Struct. Biol.* **3**, 419-429

Hall, M. C. & S. W. Matson (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* **34**, 867-77

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795-800.

Kadosh, D. & Struhl, K. (1997) Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**, 365-371.

Kanta, H., Laprade, L., Almutairi, A. & Pinto, I. (2006) Suppressor analysis of a histone defect identifies a new function for the Hda1 complex in chromosome segregation. *Genetics* **173**, 435-450.

Kasten, M. M., Dorland, S. & Stillman, D. J. (1997) A large protein complex containing the yeast Sin3p and Rpd3p transcriptional regulators. *Mol. Cell. Biol.* **17**, 4852-4858.

Korolev, S., N. Yao, *et al.* (1998) Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. *Protein Sci.* **7**, 605-10

Kurdistani, S. K. & Grunstein, M. (2003) Histone Acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell. Biol.* **4**, 276-284.

Kurdistani, S. K., Robyr, D., Tavazoie, S. & Grunstein, M. (2002) Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat. Genet.* **31**, 248-254.

Massa, W., Wocadlo, S., Lotz, S., & Dehnicke, K. (1990) *Z. anorg. allg. Chem.* **587**, 79-88

Millar, C. B. & Grunstein, M. (2006) Genome-wide patterns of histone modifications in yeast. *Nat. Rev. Mol. Cell. Biol.* **7**, 657-666.

Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S. K., Wang, A., Suka, N. & Grunstein, M. (2002) Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* **109**, 437-446.

Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M. & Grunstein, M. (1996) Hda1 and Rpd3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl. Acad. Sci. USA* **93**, 14503-14508.

Rundlett, S. E., Carmen, A. A., Suka, N., Turner, B. M. & Grunstein, M. (1998) Transcriptional repression by Ume6 involves deacetylation of lysine 5 of histone H4 by Rpd3. *Nature* **392**, 831-835.

Sengupta, N. & Seto, E. (2004) Regulation of histone deacetylase activities. *J. Cell Biochem.* **93**, 57-67

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606-1619.

Sheldrick, G. M. (1997) *SHELXTL Reference Manual*, Bruker-AXS, Inc., Madison, WI

Singleton, M. R., Dillingham, M. S. & Wigley, D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23-50.

Singleton, M. R. & Wigley, D. B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* **184**, 1819-26

Strauss, H. M. & Keller, S. (2008) Pharmacological interference with protein-protein interactions mediated by coiled-coil motifs. *Handb. Exp. Pharmacol.* **186**, 461-482.

Sugiyama, T., Cam, H. P., Sugiyama, R., Noma, K.-I., Zofall, M., Kobayashi, R. & Grewal, S. I. S. (2007) SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell* **128**, 491-504.

Tanner, N. K., O. Cordin, *et al.* (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol. Cell* **11**, 127-38

Thomä, N. H., Czyzewski, B. K., Alexeev, A. A., Mazin, A. V., Kowalczykowski, S. C. & Pavletich, N. P. (2005) Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nat. Struct. Mol. Biol.* **12**, 350-356.

Vogelauer, M., Wu, J., Suka, N. & Grunstein, M. (2000) Global histone acetylation and deacetylation in yeast. *Nature* **408**, 495-498.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215-230.

Wu, J., Carmen, A. A., Kobayashi, R., Suka, N. & Grunstein, M. (2001a) Hda2 and Hda3 are related proteins that interact with and are essential for the activity of the yeast histone deacetylase Hda1. *Proc. Natl. Acad. Sci. USA* **98**, 4391-4396.

Wu, J., Suka, N., Carlson, M. & Grunstein, M. (2001b) Tup1 utilizes histone H3/H2B-specific Hda1 deacetylase to repress gene activity in yeast. *Mol. Cell* **7**, 117-126.

Yang, X. J. & Seto, E. (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat. Rev. Mol. Cell. Biol.* **9**, 206-218.

Ye, J., A. R. Osborne, *et al.* (2004) RecA-like motor ATPases--lessons from structures. *Biochim Biophys Acta* **1659**, 1-18

Part II

X-ray structure of the secreted aspartic protease Sap5 from *Candida albicans*

Summary

Proteolytic activity is an important virulence factor for *Candida albicans*. It is attributed to the family of the secreted aspartic proteases (Saps) from *C. albicans* with a minimum of 10 members. Saps show controlled expression and regulation for the individual stages of the infection process. Distinct isoenzymes can be responsible for adherence and tissue damage of local infections, while others cause systemic diseases.

Previously, two crystal structures of Sap2 and Sap3, and very recently, high resolution structure of Sap1 were determined by X-ray crystallography. In this study, I have succeeded in solving the crystal structure of Sap5 in complex with pepstatin A at 2.5Å resolution. Of subgroup Sap4-6, the structure of the enzyme Sap5 is the first structure that has been described up to now. This facilitates comparison of structural details as well as inhibitor binding modes among the different subgroup members. Structural analysis reveals a highly conserved overall secondary structure of Sap1-3 and Sap5. However, Sap5 clearly differs from Sap1-3 by its electrostatic overall charge as well as through structural conformation of its entrance to the active site cleft. Design of inhibitors specific for Sap5 should concentrate on the S4 and S3 pockets, which significantly differ from Sap1-3 in size and electrostatic charge. Both Sap1 and Sap5 seem to play a major part in superficial *Candida* infections. Determination of the structures of isoenzymes can contribute to the development of new Sap-specific inhibitors for the treatment of superficial infections with a structure-based drug design.

1. Introduction

1.1. Proteolytic activity as virulence factor of *Candida albicans*

The key process of virulence is the production of hydrolytic enzyme, which is known to play an important role in the pathogenicity of bacteria (Finlay, B. B., and Falkow. S. 1989), protozoa (McKerrow, J. H. et. al. 1993), and pathogenic yeasts (Ogrydziak, D. M. 1993). Among a variety of hydrolytic enzymes, proteases are the most commonly coupled with virulence. The three most significant extracellular hydrolytic enzymes produced by *Candida albicans* are the secreted aspartyl proteases (Sap), phospholipase B, and lipases. The Sap proteins have been the most intensively studied as key virulence factors of *C. albicans*. This yeast strain is not the only *Candida* species known to produce extracellular proteases. Many of the pathogenic *Candida* species have been known to carry Sap genes, including *C. dubliniensis* (Gilfillan, G. D. et. al. 1998), *C. tropicalis* (Monod, M. et. al. 1994, Togni, G. et. al. 1991, Zaugg, C. et. al. 2001), and *C. parapsilosis* (de Viragh, P. A. et. al. 1993, Monod, M. et. al. 1994), all of which produce the active extracellular proteases *in vitro* (Gilfillan, G. D. et. al. 1998). Less or non-pathogenic *Candida* species do not show to produce significant amounts of protease, even though they probably possess aspartyl protease genes. We should note that all secreted proteases from *Candida* belong to the same class of the aspartyl proteases. Neither extracellular serine nor cysteine proteases nor metalloproteases have been identified in pathogenic *Candida* species.

1.2. Classification of Saps in *Candida albicans*

Among 10 Sap family enzymes in *C. albicans*, three distinct groups are clustered within the family. Sap1 to Sap3 are up to 67% identical (family I), and Sap4 to Sap6 are up to 89% identical (family II), while Sap7 is only 20-27% identical to other Sap proteases (family III). Sap9 and Sap10 both have the C-terminal consensus sequences typical for GPI proteins and constitute the third distinct group (family IV).

Gene	Prepropeptide size (aa) ^f	No. of KR and KK sites ^b	ORF size (bp) ^c	Mature enzyme size (aa) ^e	No. of N-glycosylation (propeptide) sites ^d	Chromosome ^d
<i>SAP1</i>	50	2 KR	1,173	341	1 (0)	6
<i>SAP2</i>	56	2 KR	1,194	342	0 (2)	R
<i>SAP3</i>	58	1 KR	1,194	340	1 (1)	3
<i>SAP4</i>	75	4 KR	1,254	342	0 (1)	6
<i>SAP5</i>	76	4 KR	1,254	342	0 (0)	6
<i>SAP6</i>	76	4 KR	1,254	342	0 (1)	6
<i>SAP7</i>	211	1 KK	1,764	377	1 (4)	1
<i>SAP8</i>	73	2 KR	1,215	405	1 (1)	3
<i>SAP9</i>	50	1 KR, 1 KK	1,632	544	4 (0)	3
<i>SAP10</i>	38	1 KR	1,326	403	8 (0)	4

Table 1. *C. albicans* Sap genes and proteins. All Sap proteins contain Lys/Arg (KR)- or Lys/Lys (KK)-processing sites and four conserved residues (b). The sizes of the prepropeptide (signal peptide and propeptide) and the mature enzyme are shown in amino acids, and the size of the open reading frame (ORF) is shown in base pairs (c). The number of potential N-glycosylation sites, including those located in the propeptide (in parentheses), and the chromosomal location of the gene are also shown. Sap9 and 10 have structural elements typical of GPI proteins.

1.3. Catalytic mechanism of aspartic proteases

Aspartic proteases are a family of protease enzymes that use an aspartate residue for catalysis of their peptide substrates. In general, they have two highly conserved aspartates in the active site and are optimally active at acidic pH. Nearly all known aspartyl protease are inhibited by pepstatin. Eukaryotic aspartic proteases include pepsins, cathepsins, and renins. They have a two-domain structure, probably arising from ancestral duplication. Retroviral and retrotransposon proteases (Pfam PF00077) are much smaller and appear to be homologous to a single domain of the eukaryotic aspartyl proteases.

While a number of different mechanisms for aspartyl proteases have been proposed, the most widely accepted is a general acid-base mechanism involving coordination of a water molecule between the two highly-conserved aspartate residues (Brik, A. & Wong, C. H. 2003, Suguna, K. et. al. 1987). One aspartate activates the water by abstracting a proton, enabling the water to attack the carbonyl carbon of the substrate scissile bond, generating a tetrahedral oxyanion intermediate. Rearrangement of this intermediate leads to protonation of the scissile amide.

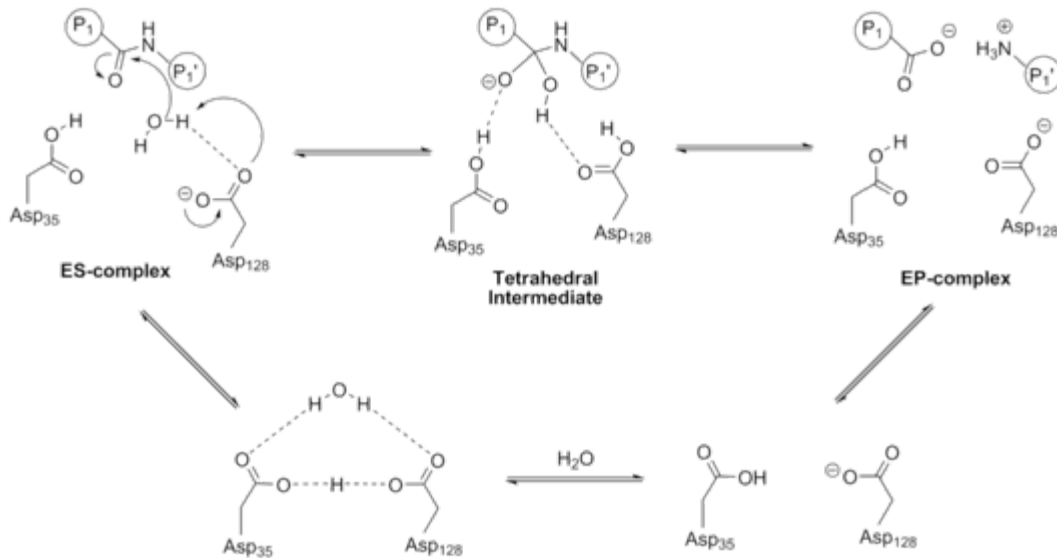


Fig. 1. General acid-base catalytic mechanism of peptide cleavage by aspartyl proteases (porcine pepsin numbering).

1.4. Known structures of Sap enzymes in *Candida albicans*

All Sap enzymes of *C. albicans* are known as pre-proenzymes approximately 60 amino acids longer than the mature form, which are then processed when transported via the secretion system. The mature forms contain typical sequence motifs for all aspartyl proteases, including the two conserved aspartate residues of the active site and conserved cysteine residues implicated in the maintenance of the three-dimensional structure. Most Sap enzymes contain putative N-glycosylation sites, but it remains to be determined which Sap proteins are glycosylated. Unlike Sap1-Sap8, Sap9 and Sap10 enzymes have the C-terminal consensus sequences for glycosylphosphatidylinositol (GPI) proteins (Felk, A. et. al. 2000, Monod, M. et. al. 1998).

Structural studies of the *C. albicans* Sap protease family have concentrated on Sap2 and its complex with inhibitors (Abad-Zapatero, C. et. al. 1996, Cutfield, S. M. et. al. 1995), which is the most abundant secreted protein *in vitro* when grown in the presence of protein as the sole source of nitrogen (Hube, B. et. al. 1994, White, T. C. & Agabian, N. 1995). The overall structure of Sap2 has the classical aspartic protease fold typified by pepsin. Very recently, two additional structures of apo-Sap1 and Sap3-pepstatin A complex were also determined by using X-ray crystallography (Borelli, C. et. al. 2007,

Borelli, C. et. al. 2008).

1.5. Aspartic protease inhibitors

Proteases were once considered to function primarily as "enzymes of digestion", but are now recognized as the largest and most diverse class of enzymes. The scientific interest in protease inhibitors has been enormous both in academia and in private industry in recent years. Numerous protease inhibitors are currently in development in the pharmaceutical industry, and at least 50 different proteases are considered as potential targets. A number of these belong to the group of aspartic proteases which are involved in key processes that are related to several human disease states. Different aspartic proteases are currently targeted for the development of new medicines for use in the treatments of a range of conditions including hypertension (renin), malaria (plasmeprins), HIV/AIDS (HIV-1-protease) and Alzheimer's disease (β -secretase).

The function of aspartic proteases depends on the presence of two aspartic acid residues in the active site (Fig. 1). A water molecule is bound between the two catalytic Asp residues, and through a "push-pull" general acid-base mechanism the water molecule acts as a nucleophile that attacks the carbonyl of the peptide bond to be broken (scissile peptide bond). Most aspartic proteases recognize and bind a 6-10 amino-acid stretch of their substrates, but the enzymatic cleavage of the substrate depends on several key events to occur. The enzyme active site is covered by a flap region, and to allow for substrate entry into the active site, this flap must first be opened, and then closed again over the substrate in the catalytic cleft during enzyme catalysis. To allow for diffusion of products from the active site the flap region needs to re-open after the substrate hydrolysis has occurred.

Early protease inhibitors were often found to have poor oral absorption and low bioavailability, due to their peptidic nature, and further development of these compounds into non-peptide peptidomimetic drugs has been necessary. In the search for tight binding inhibitors, structure-based drug design has proved to be a powerful tool. Access to protease-inhibitor crystal structures, has opened the way for reduction of the molecular weight of inhibitors as well as reducing the peptidic

features, and allowed for designed selectivities in lead inhibitors.

2. Materials and Methods

2.1. Crystallization of *Candida albicans* Sap5-pepstatin A complex

A 1.5-fold excess amount of pepstatin A was added to Sap5 solution (10mg/uL). The Sap5 in complex with pepstatin A crystallized against the crystallization buffer containing 10% (w/v) PEG-8,000, 8% (v/v) ethylene glycol, 0.1M HEPES pH7.5 by sitting drop vapor diffusion method at 20°C for overnight.

2.2. Data collection and processing

X-ray data set in the resolution range of 30-2.5Å were collected by using a native crystal of the Sap5-pepstatin A complex on the beamline BW6 at the DESY (Deutsches Elektronen Synchrotron, Hamburg). Integration, scaling, and truncation of the diffraction data were performed by using the DENZO and SCALEPACK (Otwinowski, Z. & Minor, W. 1997).

2.3. Structure determination by molecular replacement

The structure of Sap5-pepstatin A complex was determined by a molecular replacement method with PHASER (McCoy, A. J. et. al. 2005) using the structure of Sap2 (PDB ID code 1EAG) as a search model. The solution consisted of two molecules in the asymmetric unit. The molecular replacement model was refined by a rigid body protocol followed by energy minimization and simulated annealing methods and the R factor dropped to 43%. Successive rounds of rebuilding using COOT (Emsley, P. & Cowtan, K. 2004), simulated annealing refinement with CNS (Bruenger, A.T. et. al. 1998) allowed the complete interpretation of the model and clearly showed the density for the inhibitor pepstatin A in the active site of Sap5. In each round of refinement, the R_{free} value was monitored, and the omit maps were used to verify every part of the molecule at the end of the refinement. The final structure was assessed with PROCHECK (Laskowski, R. A. et. al. 1993), and 85.5% of the main chain ϕ/ψ angles were found in the most favored regions on the Ramachandran plot.

3. Results

3.1. Crystallization and structure determination of Sap5-pepstatin A complex

The Sap5 in complex with pepstatin A was crystallized as tetragonal form which belongs to space group $P4_12_12$ with $a=92.20$, $b=92.20$, and $c=182.01\text{\AA}$ and contain two molecules in the asymmetric unit. To determine the crystal structure of Sap5-pepstatin A complex, the Sap2 monomer (PDB ID code 1EAG) was used as a search model for molecular replacement using program PHASER. The first solution (Z-score of rotation function = 8.3; Z-score of translation function = 22.4; LLG = 358) after the first molecular replacement was readily improved as the model which contains two molecules in the asymmetric unit (Z-score of rotation function = 9.1; Z-score of translation function = 50.3; LLG = 1622). Successive rounds of manual model building using COOT and refinement using CNS allowed the complete structure determination with R_{work} of 22.4% and R_{free} of 27.5% at 2.5Å resolution (Table 1).

Data collection	Sap5-pepstatin A complex
Parameter	
Space group	$P4_12_12$
Wavelength (Å)	1.05
Resolution (Å)	30.0 – 2.5
Total observations	628,654
Unique observations	27,970
Data coverage (%) (2.59 – 2.50)	100 (100)
R_{sym} (2.59 – 2.50)	0.083 (0.424)
Refinement	
Resolution range (Å)	30.0 – 2.5
Number of reflections	26,393
$R_{\text{work}} / R_{\text{free}}$ (%)	22.4 / 27.5
Number of protein atoms	5,246
Number of water atoms	370
Number of inhibitor atoms	96
RMSD bond length (Å)	0.007
RMSD bond angle (°)	1.38

Table 2. Data collection and refinement statistics. Values in parentheses are for the highest resolution shell. $R_{\text{sym}} = \frac{\sum_h \sum_i |I_{h,i} - I_h|}{\sum_h \sum_i I_{h,i}}$, where I_h is the mean intensity of the i observations of symmetry-related reflections of h . $R = \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum F_{\text{obs}}}$, where $F_{\text{obs}} = F_p$, and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections). RMSD in bond lengths and angles are the deviations from ideal values.

3.2. Overall structure of Sap5-pepstatin A complex

Within the asymmetric unit of crystal, there are two complexes of Sap5 bound to pepstatin A (Fig. 1A). Two complexes in the asymmetric unit have similar conformations; the root-mean-squared deviations (RMSDs) of the $C\alpha$ atoms between two molecules are 0.64Å for Sap5 and 0.44Å for the pepstatin A. The final structure includes Sap5, pepstatin A bound in the active site of Sap5, and 370 water molecules. 85.5% of the non-glycine residues were found in the most favored regions and 14.5% of the residues were found in additionally allowed regions on the Ramachandran plot.

The Sap5 folds into mainly β -structure which can be divided into two lobes related by a pseudo two-fold axis, as found in the structures of aspartic proteases (Fig 1B). The N-lobe consists of four helices (α_1 , α_2 , α_3 , and α_4) and 15 β -strands (β_1 - β_{12} , β_{24} , β_{25} , and β_{26}). The C-lobe is comprised of two helices (α_5 and α_6) and 11 β -strands (β_{13} - β_{23}). Topology diagram with the secondary structure of Sap5 is shown in Fig. 1C. The active site is located in the deep cleft between two lobes which contain two catalytic aspartic acids, Asp32 and Asp218 on the N and C-lobes, respectively. Two loops spanning β_4 and α_1 , and β_5 and β_6 protruded from the N-lobe and located around the active site. These flap structures play a role in trapping substrate.

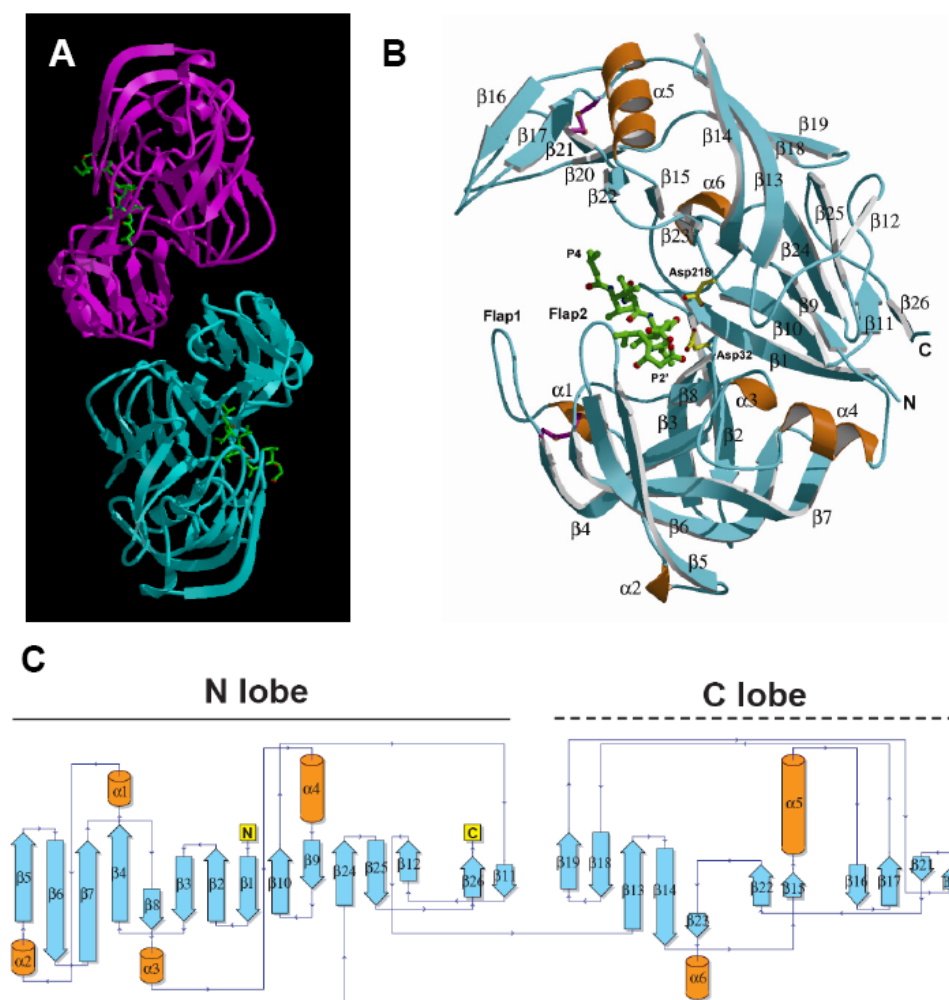


Fig. 2. Overview of the structure of *Candida albicans* Sap5-pepstatin A complex. (A) Structural orientation of two Sap5-pepstatin A complexes (Sap5: magenta and blue, pepstatin A: green) in the asymmetric unit. (B) Overall structure of Sap5-pepstatin A complex. Inhibitor pepstatin A (green) is located in the deep cleft. Catalytic aspartate residues, Asp32 and Asp218 are colored in yellow. Two disulfide bonds formed by Cys47-Cys59 and Cys256-Cys294 are colored in magenta. N, O and S atoms are colored in blue, red and orange, respectively. (C) Topology diagram of Sap5 structure.

3.3. The Active site in Sap5

The residues forming the active site cleft in Sap5 are highly conserved across the members of the Sap family of enzymes from *C. albicans*. These residues include Asp32,

Gly34, Ile82, Tyr84, Gly85, Asp86, Ile123, Asp218, Gly220, Thr221, Thr222, and Ile223 (Fig. 2 and Fig. 3). In the course of model building, I have observed strong density for the bound pepstatin A in the active site cleft. The $2Fo-Fc$ omit map for the bound pepstatin A in the active site of Sap5 is shown in Fig. 3B, where all atoms are clearly assigned. The detailed structure of the active site cleft and the bound pepstatin A molecule are shown in Fig. 3C. Overall, the bound pepstatin A is recognized by Sap5 protease through multiple hydrogen bonds and van der Waals interactions.

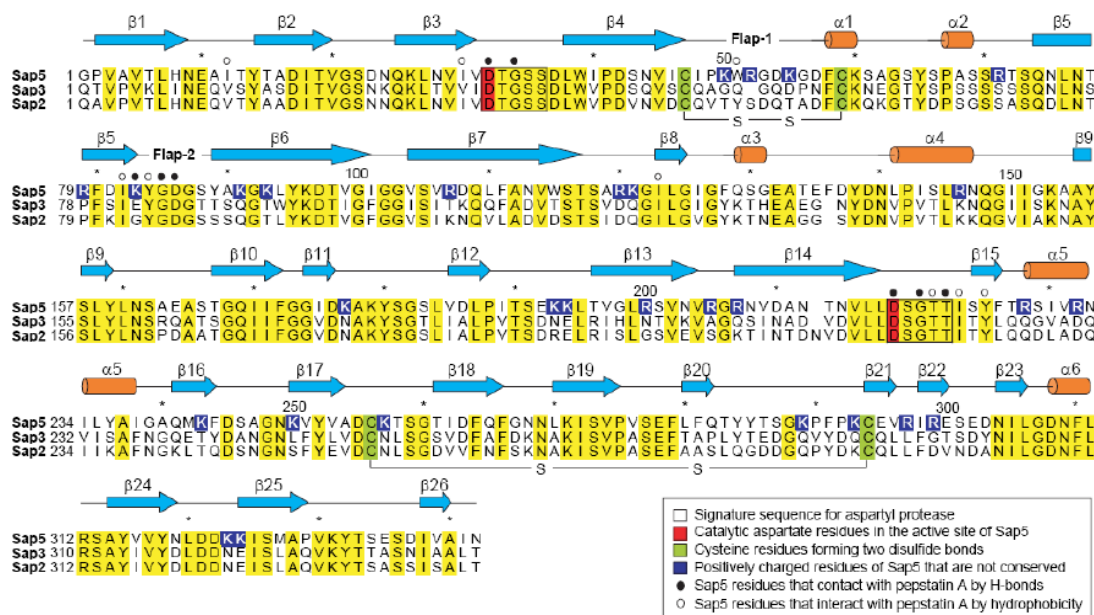


Fig. 3. Sequence alignment and conservation in Sap isoenzymes of *C. albicans*. Conserved residues in Sap family are highlighted in yellow. The secondary structure elements are indicated above the alignment. Every 10th residue is marked with an asterisk.

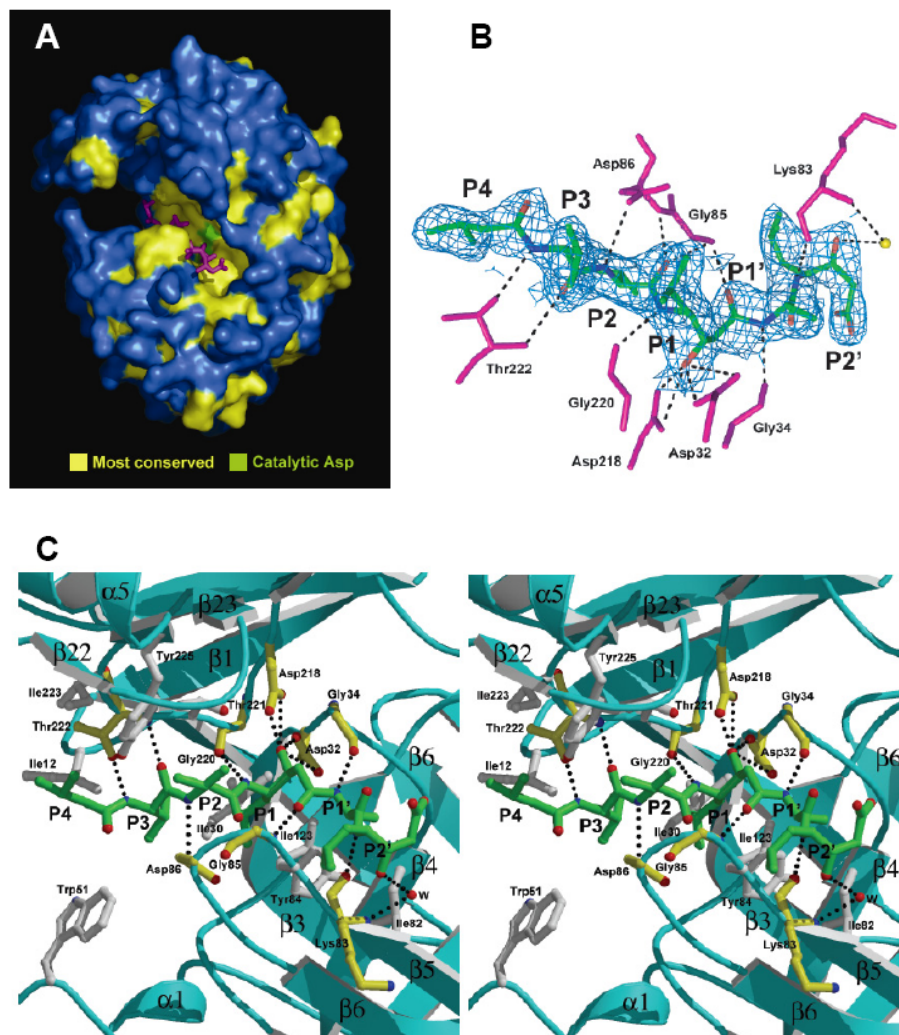


Fig. 4. Surface representation and active site of Sap5-pepstatin A complex. (A) Molecular surface of Sap5-pepstatin A complex. The surface of Sap5 residue is 100% conserved in the Sap family (Sap2, 3, and 5) and colored in yellow. The surface of two catalytic Asp residues is colored in green. The bound pepstatin A is colored in magenta. (B) Hydrogen bonding networks between the bound pepstatin A and key residues (magenta) in the active site of Sap5. The $2F_o - F_c$ omit map (blue) contoured at 1.0σ and was calculated by simulated annealing using CNS. Water molecule is shown in yellow. (C) Stereoview of Sap5-pepstatin A interaction. Residues of Sap5 that are related to the intermolecular hydrogen bonds and hydrophobic interactions in the active site are colored in yellow and white, respectively. N and O atoms are shown in blue and red, respectively. The dotted lines indicate the intermolecular hydrogen bonds between Sap5 and pepstatin A.

3.3. Structural comparison of Sap5 with the other Sap enzymes

Sequence alignment and structural analysis suggested that Sap5 protease contains the typical structure with two catalytic aspartate residues, Asp32 and Asp218 as found in the aspartic protease enzyme family. Structural superposition between Sap5-pepstatin A complex and two recently determined Sap-inhibitor complexes revealed that the overall fold of Sap5 is closely similar (Fig. 5A). The RMSD values of these proteins are between about 1.6 and 2.4Å for all C α atoms. The Sap5 and these isoenzymes share the common secondary structural elements. However, Sap5 and the other Sap isoenzymes, Sap2 and Sap3 are markedly different in terms of other structural features. Major difference between Sap5 and two Sap isoenzymes is the charge distribution and the electrostatic potential (Fig. 4). Although three Sap proteases share the common acidic nature in the active site cleft, the electrostatic charges, especially around the active site of Sap5 are remarkably different from the “negative” Sap2 and Sap3. Reasons for the conservation of the charge distribution at the active site and the variation in the surrounding area in the Sap enzyme family are unclear. It may be related to a particular tissue localization of the individual isoenzymes.

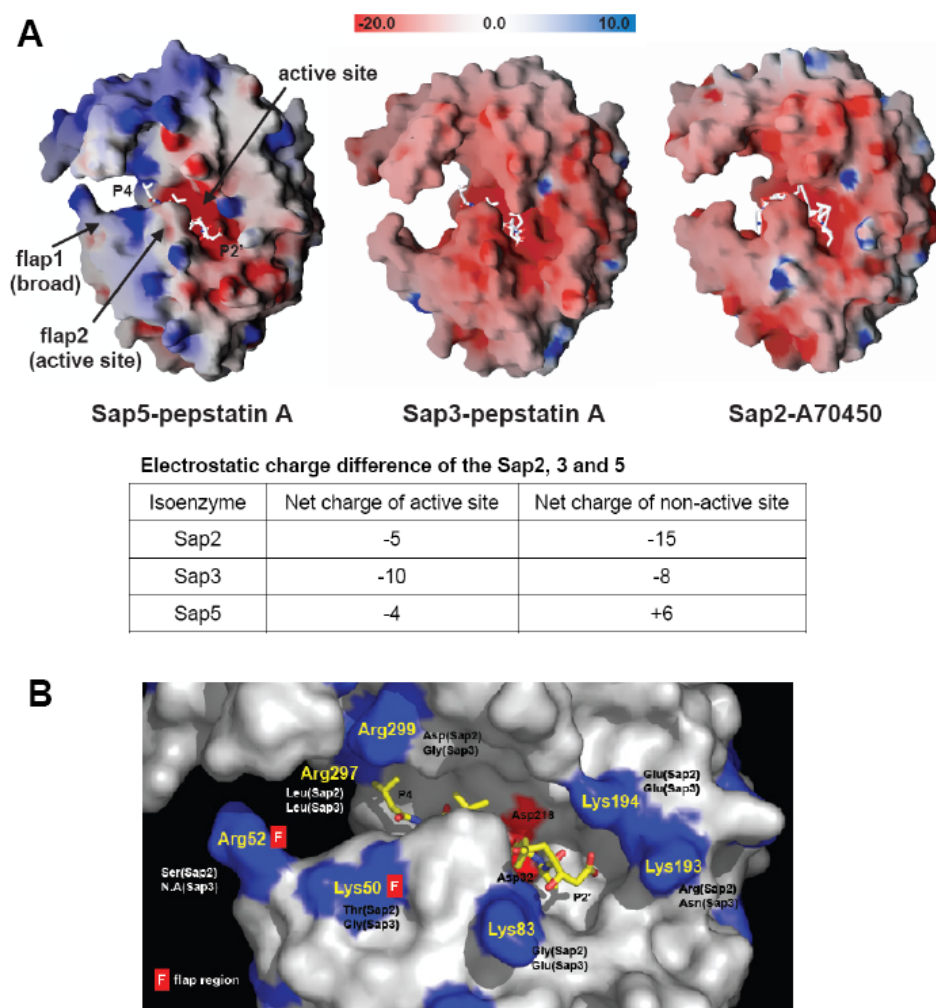


Fig. 5. Structural comparison of three Sap-inhibitor complexes. (A) Comparison of the electrostatic surface potential of three Sap-inhibitor complexes. Basic and acidic residues are colored in red and blue, respectively. (B) The variation of electrostatic surface potential between Sap5 and the other Sap isoenzymes. Positively charged residues around the active site cleft that are not conserved in Sap family are highlighted in blue.

The other structural difference between Sap5 and two Sap isoenzymes is the movement of flap region spanning $\beta 4$ and $\alpha 1$ (flap 1), and amino acid substitutions at the subsites of the active site among three Sap proteases (Fig. 5.). The flap 1 region of Sap5 enzyme bound pepstatin A has moved into the active site more than in Sap2 and Sap3 protease complexes. In the case of the active site cleft, the S3 subsite is a site of significant variation within the active sites of Sap2, 3, and 5 proteases. Due to the larger

aromatic residue in Sap5 (Trp51), the volume of S3 subsite is smaller than in Sap2 and Sap3. The S4 subsite is extremely different between three Sap structures, both in shape and polarity: Sap2 and Sap3 have neutral and hydrophobic residues comprising the S4 subsite, such as Ala281, Gln295, Leu297, and Gly299 (or Asp299 for Sap2). However, the S4 pocket of Sap5 consists of aromatic Phe281, charged Glu295, Arg297 and Arg299. Because of bulky shape and sidechain orientation of residues comprising the S4 subsite of Sap5, the S4 pocket of Sap5 bound pepstatin A is more tightly packed than those of Sap2 and Sap3 structures.

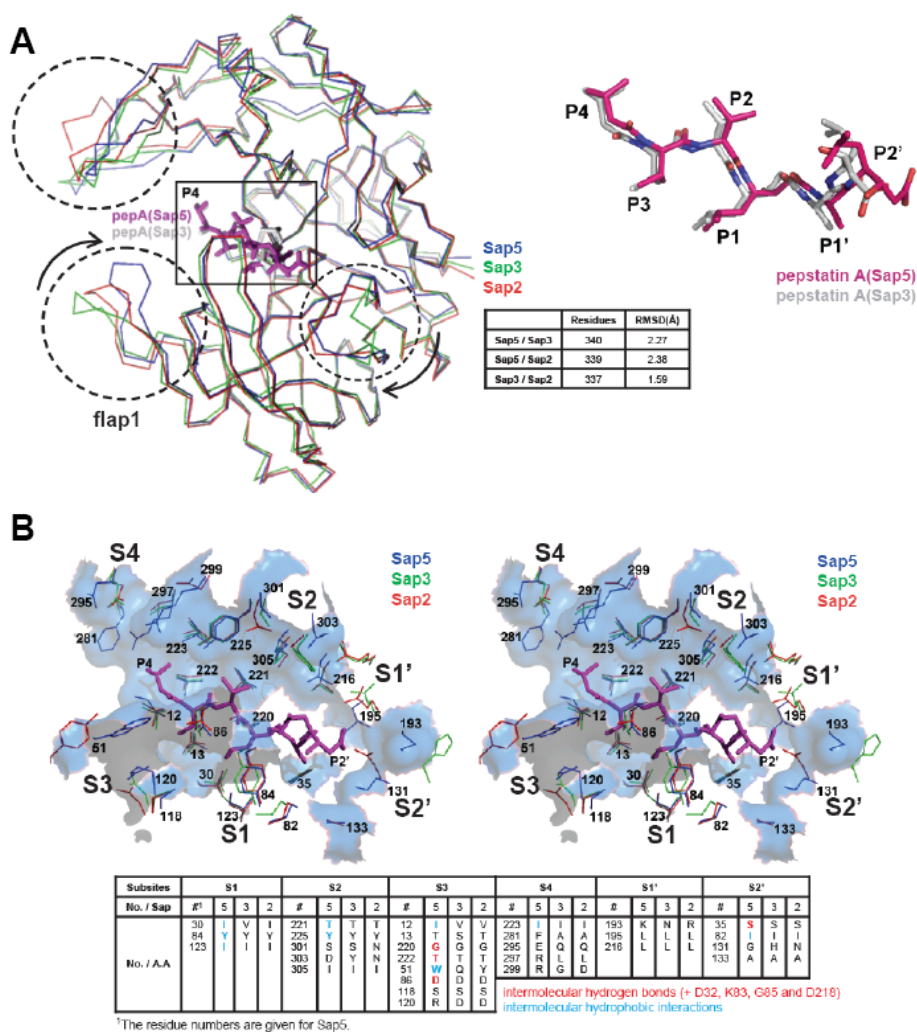


Fig. 6. Structural comparison of three Sap-inhibitor complexes. (A) Superposition and local difference between three Sap-inhibitor complexes. The structures are superimposed on C α atoms using LSQKAB and RMSD values are calculated using LSQMAN. The curved arrows indicate the local movement of

Sap5 as compared with Sap2 and Sap3 structures. The A70450 coordinate from Sap2-inhibitor complex (PDB ID code 1EAG) was not presented for clarity. (B) Superposition of the residues in three Sap structures involved in the inhibitor interaction and the proposed pocket formation. The residues are superimposed on C α atoms using LSQKAB. Amino acid substitutions at the different active site subsites among three Sap isoenzymes (Sap2, Sap3, and Sap5) and the key residues in Sap5 involved in the intermolecular interaction with pepstatin A are summarized.

4. Discussion

Previous structural studies of two Sap isoenzymes, *C. albicans* Sap2 and Sap2X have allowed us to understand the structural and functional importance of Sap enzymes as virulence factor of *C. albicans* (Abad-Zapatero, C. et. al. 1996, Cutfield, S. M., et. al. 1995). All four Sap enzymes (Sap1, Sap2, Sap3, and Sap5) which were determined the 3D structures by X-ray crystallography have showed similar features that identified them as a unique family within pepsin-like aspartic proteases. The disulfide linkage between Cys47 and Cys59, which constitutes the flap-1, is a highly important structural difference. Compared to Sap1-3, the flap-1 of Sap5 has voluminous side-chains such as Lys50, Trp51, and Arg52, which reach down to the substrate binding cleft and are probably the additional barrier for entering substrates to the active site.

The Sap5 clearly differs from Sap1-3 by its electrostatic charge distribution (Fig. 5). The positively charged residues surrounding the active site of Sap5 make the optimal pH for the action rise to pH 5.0 with a still noticeable activity at neutral pH. The Sap2 (Sap3) and Sap1, however, are most active at pH 3.5 and 4.5, respectively (Borg-von Zepelin, M. et. al. 1998). The difference in electrostatic charges of Saps and thus their pH optima can be due to the adaptation of the isoenzymes to their specific host environment. This is in accordance to the finding that Saps are differently expressed in *C. albicans* during superficial and systemic infections depending on the environmental conditions (Hube, B. et. al. 1994, White, T. C. & Agabian, N. 1995).

The active site cleft itself remains negatively charged with all four isoenzymes

presumably due to the highly conserved catalytic reaction. The accepted mechanism of pepsin-like enzyme function assumes that one of the catalytic aspartic acid residues has to be charged, whereas the other one has to be protonated (Davies, D. R. 1990, James, M. N., et. al. 1992, Parris, K. D., et. al. 1992, Suguna, K., et. al. 1987, Veerapandian, B., et. al. 1992). Andreeva and Rumsh (Andreeva, N. S. & Rumsh, L. D. 2001) have shown the crucial role of amino acid residues that border the catalytic center in stabilizing this balanced protonated/deprotonated state. Accordingly, the catalytic mechanism demands a highly conserved site adjacent to the catalytic center.

Generally, the aspartic proteases prefer substrates with hydrophobic residues at positions P1 and P10 for cleavage (Fruton, J. S. 1976). However, fungal aspartic proteases differ from mammalian enzymes in that they also cleave substrates with polar residues such as histidine or lysine in the P1 position (Lowther, W. T., et. al. 1995). Fig. 6B shows a high conservation of substrate binding site residues of Sap2-5, which suggests very similar substrate specificities for all members of this isoenzyme subgroup. Unfortunately, only substrate specificities at the P1 and P10 sites for Sap isoenzymes Sap1-3 and Sap6 have been examined so far (Koelsch, G., et. al. 2000). P10 specificities are generally broader than those observed for P1. However, substrate specificities specific to either subgroup Sap1-3 or Sap4-6 in particular could not be defined. Because of the considerable changes regarding character and size of the S4 and S3 pockets in Sap5, different individual substrate specificities might also be expected for Sap1-3 and Sap4-6. The Sap4-6 enzymes, displaying two Arg in the S4 pocket and one Arg (Sap5) or His (Sap4/6) in the S3 pocket, are likely to prefer hydrophobic and/or negatively charged residues.

The activity of Sap5 was strongly inhibited by pepstatin A with an IC_{50} value of 6nM (data not shown). The binding affinities for pepstatin A to Sap1, Sap2, and Sap3 indicated by K_i values being 15, 17 (and IC_{50} 27nM), and 60nM have been reported (Capobianco, J. O., et. al. 1992, Koelsch, G., et. al. 2000). The hydrogen bonding network found in Sap5 which strongly contributes to inhibitor anchoring is a highly conserved not only for the Sap isoenzymes but also for other protein complex inhibited by pepstatin A (Metcalf. P. & Fusek. M. 1993, James, M. N., et. al. 1982, Suguna, K., et. al. 1992). Hydrophobic contacts slightly differ within the pepstatin A-Sap complexes but a sufficient amount of

contacts results in the similar binding affinities for pepstatin A to Sap2, Sap3, and Sap5.

Previous studies have demonstrated that the HIV protease inhibitors (e.g., ritonavir, saquinavir, and indinavir) inhibit Sap1-3 isoenzymes in the micromolar range. Especially, ritonavir was stated as the most potent inhibitor (Korting, H. C. et. al. 1999, Pichova, I. et. al. 2001). For Sap4-6, a marginal inhibitory effect from HIV protease inhibitors has been detected ($IC_{50} > 300$ IM). HIV-1 protease, a member of the aspartic protease family, displays two catalytic aspartic residues that are provided by the subunits of the symmetrical dimer (Miller, M., et. al. 1989, Weber, I. T. et. al. 1989). To explore the potential interactions between the HIV protease inhibitor ritonavir and the Saps, we superimposed the structure of HIV-1 bound to ritonavir (PDB Code: 1HXW) with the structures of Sap1-3 and Sap5 (Fig. 7).

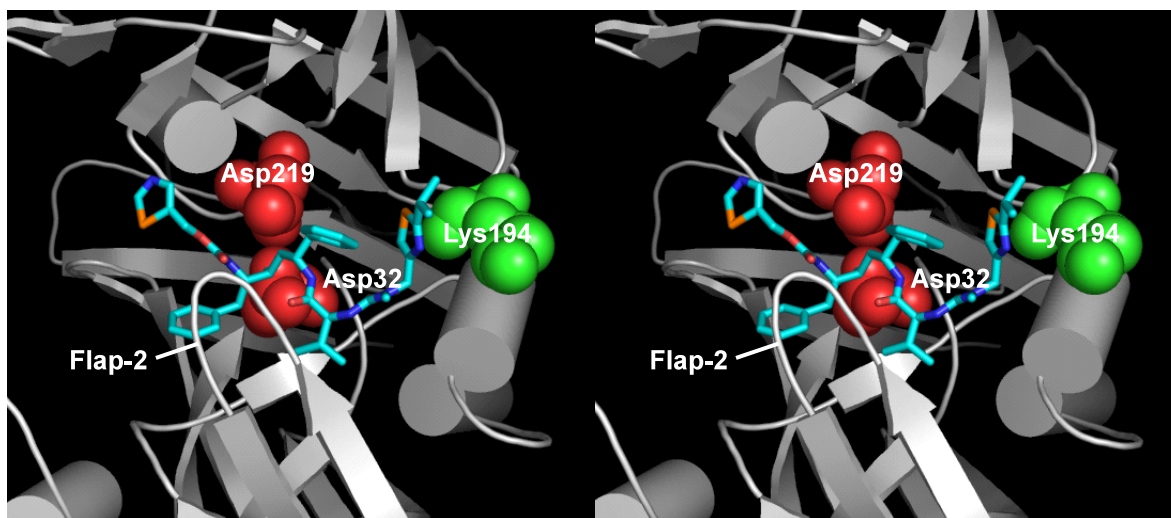


Fig. 7. Structural model of the interaction between Sap enzymes and ritonavir. For clarity, stereo-diagram only for Sap5 structure bound with ritonavir is shown.

Ritonavir was modeled into the active site of Sap1-3 and Sap5 with its central phenylstatine residue (Phe-CH-CHOH-) directed toward the active site residues Asp32 and Asp219. The most significant difference between Sap1-3 and Sap5 was found at the position 194 of the isoenzymes. In Sap5, the long side chain of Lys194 represents a significant barrier for ritonavir's isopropylthiazolyl group when

binding to the enzyme, whereas the smaller glutamic acid found in Sap1-3 at that position is of no hindrance. This could also explain the reduced inhibitory effect of ritonavir on Sap5. However, a conformational change of the respective side chain of ritonavir after binding to Sap5 is conceivable.

5. References

Abad-Zapatero, C., Goldman, R., Muchmore, S. W., Hutchins, C., Stewart, K., Navaza, J., Payne, C. D. & Ray, T. L. (1996) Structure of a secreted aspartic protease from *C. albicans* complexed with a potent inhibitor: implications for the design of antifungal agents. *Protein Sci.* **5**, 640–652.

Andreeva, N. S. & Rumsh, L. D. (2001) Analysis of crystal structures of aspartic proteinases: on the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci* **10**, 2439–2450.

Borelli, C., Ruge, E., Schaller, M., Monod, M., Korting, H. C., Huber, R. & Maskos, K. (2007) The crystal structure of the secreted aspartic proteinase (Sap) 3 from *Candida albicans* and its complex with pepstatin A. *Proteins* **68**, 738–748.

Borelli, C., Ruge, E., Lee, J. -H., Schaller, M., Vogelsang, A., Monod, M., Korting, H. C., Huber, R. & Maskos, K. (2008) X-ray structures of Sap1 and Sap5: Structural comparison of the secreted aspartic proteinases from *Candida albicans*. *Proteins* **72**, 1308-1319.

Borg-von Zepelin, M., Beggah, S., Boggian, K., Sanglard, D. & Monod, M. (1998) The expression of the secreted aspartyl proteinases Sap4 to Sap6 from *Candida albicans* in murine macrophages. *Mol Microbiol* **28**, 543–554.

Brik, A. & Wong, C. H. (2003) HIV-1 protease: mechanism and drug discovery *Org. Biomol. Chem.* **1**, 5–14.

Bruenger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.

Capobianco, J. O., Lerner, C. G. & Goldman, R. C. (1992) Application of a fluorogenic substrate in the assay of proteolytic activity and in the discovery of a potent inhibitor of *Candida albicans* aspartic proteinase. *Anal Biochem* **204**, 96–102.

Cutfield, S. M., Dodson, E. J., Anderson, B. F., Moody, P. C., Marshall, C. J., Sullivan, P. A. & Cutfield, J. F. (1995) The crystal structure of a major secreted aspartic proteinase from *Candida albicans* in complexes with two inhibitors. *Structure* **3**, 1261–1271.

Davies, D. R. (1990) The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* **19**, 189–215.

de Viragh, P. A., Sanglard, D., Togni, G., Falchetto, R. & Monod, M. (1993) Cloning and sequencing of two *Candida parapsilosis* genes encoding acid proteases. *J. Gen. Microbiol.* **139**, 335–342.

Emsley, P. & Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

Felk, A., Schafer, W. & Hube, B. (2000) *Candida albicans* secretory aspartic proteinase (SAP10) gene. Accession number AF146440.

Finlay, B. B. & Falkow, S. (1989) Common themes in microbial pathogenicity. *Microbiol. Rev.* **53**, 210–230.

Fruton, J. S. (1976) The mechanism of the catalytic action of pepsin and related acid proteinases. *Adv Enzymol Relat Areas Mol Biol* **44**, 1–36.

Gilfillan, G. D., Sullivan, D. J., Haynes, K., Parkinson, T., Coleman, D. C. & Gow, N. A. R. (1998) *Candida dubliniensis*: phylogeny and putative virulence factors. *Microbiology* **144**, 829–838.

Hube, B., Monod, M., Schofield, D. A., Brown, A. J. P. & Gow, N. A. R. (1994) Expression of seven members of the gene family encoding secretory aspartyl proteinases in *Candida albicans*. *Mol. Microbiol.* **14**, 87–99.

James, M. N., Sielecki, A. R., Hayakawa, K. & Gelb, M. H. (1992) Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **31**, 3872–3886.

James, M. N., Sielecki, A., Salituro, F., Rich, D. H. & Hofmann, T. (1982) Conformational flexibility in the active sites of aspartyl proteinases revealed by a pepstatin fragment binding to penicillopepsin. *Proc Natl Acad Sci USA* **79**, 6137–6141.

Koelsch, G., Tang, J., Loy, J. A., Monod, M., Jackson, K., Foundling, S. I. & Lin, X. (2000) Enzymic characteristics of secreted aspartic proteases of *Candida albicans*. *Biochim Biophys Acta* **1480**, 117–131.

Korting, H. C., Schaller, M., Eder, G., Hamm, G., Bohmer, U. & Hube, B. (1999) Effects of the human immunodeficiency virus (HIV) proteinase inhibitors saquinavir and indinavir on in vitro activities of secreted aspartyl proteinases of *Candida albicans* isolates from HIV-infected patients. *Antimicrob Agents Chemother* **43**, 2038–2042.

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

Lowther, W. T., Majer, P. & Dunn, B. M. (1995) Engineering the substrate specificity of rhizopuspepsin: the role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. *Protein Sci* **4**, 689–702.

Metcalf, P. & Fusek, M. (1993) Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. *EMBO J* **12**, 1293–1302.

McKerrow, J. H., Sun, E., Rosenthal, P. J. & Bouvier, J. (1993) The proteases and pathogenicity of parasitic protozoa. *Annu. Rev. Microbiol.* **47**, 821–853.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464.

Miller, M., Jaskolski, M., Rao, J. K., Leis, J. & Wlodawer, A. (1989) Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* **337**, 576–579.

Monod, M., Hube, B., Hess, D. & Sanglard, D. (1998) Differential regulation of *SAP8* and *SAP9*, which encode two new members of the secreted aspartic proteinase family in *Candida albicans*. *Microbiology* **144**, 2731–2737.

Monod, M., Togni, G., Hube, B. & Sanglard, D. (1994) Multiplicity of genes encoding secreted aspartic proteinases in *Candida* species. *Mol. Microbiol.* **13**, 357–368.

Ogrydziak, D. M. (1993) Yeast extracellular proteases. *Crit. Rev. Biotechnol.* **13**, 1–55.

Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

Parris, K. D., Hoover, D. J., Damon, D. B. & Davies, D. R. (1992) Synthesis and crystallographic analysis of two rhizopuspepsin inhibitor complexes. *Biochemistry* **31**, 8125–8141.

Pichova, I., Pavlickova, L., Dostal, J., Dolejsi, E., Hruskova-Heidingsfeldova, O., Weber, J., Ruml, T. & Soucek, M. (2001) Secreted aspartic proteases of *Candida albicans*, *Candida tropicalis*, *Candida parapsilosis* and *Candida lusitanae*. Inhibition with peptidomimetic inhibitors. *Eur J Biochem* **268**, 2669–2677.

Suguna, K., Padlan, E. A., Bott, R., Boger, J., Parris, K. D. & Davies, D. R. (1992) Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins* **13**, 195–205.

Suguna, K., Padlan, E. A., Smith, C. W., Carlson, W. D. & Davies, D. R. (1987) Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7009–7013.

Togni, G., Sanglard, D., Falchetto, R. & Monod, M. (1991) Isolation and nucleotide sequence of the extracellular acid protease gene (ACP) from the yeast *Candida tropicalis*. *FEBS Lett.* **286**, 181–185.

Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B. & Hoover, D. J. (1992) Direct observation by X-ray analysis of the tetrahedral “intermediate” of aspartic proteinases. *Protein Sci* **1**, 322–328.

Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M. & Wlodawer, A. (1989) Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* **243**, 928–931.

White, T. C. & Agabian, N. (1995) *Candida albicans* secreted aspartyl proteinases: isoenzyme pattern is determined by cell type, and levels are determined by environmental factors. *J. Bacteriol.* **177**, 5215–5221.

Zaugg, C., Borg-von Zepelin, M., Reichard, D., Sanglard, D. & Monod, M. (2001) Secreted aspartic proteinase family of *Candida tropicalis*. *Infect. Immun.* **69**, 405–412.

Curriculum Vitae

Jung-Hoon Lee

Max-Planck-Institute of Biochemistry,
Am Klopferspitz 18, 82152 Martinsried, Germany
Tel: 49-89-3859-0354, email: lee@biochem.mpg.de

Personal information

Date of birth: 12.11.1974
Place of birth: Seoul, South Korea
Gender: Male (married)
Nationality: South Korea

Education

PhD candidate, Max-Planck-Institute of Biochemistry, Martinsried, Germany, 02.2006 ~
MS, Dept. of Biochemistry, Yonsei University, Seoul, South Korea, 09.1997 ~ 08.1999
BS, Dept. of Biology, Konkuk University, Seoul, South Korea, 03.1993 ~ 02.1997
Obligatory military service (Korean Army), 12.1999 ~ 02.2002

Research experience

PhD (candidate) (X-ray crystallography)
Max-Planck-Institute of Biochemistry, Martinsried, Germany, 02.2006 ~
Thesis advisor: Prof. Robert Huber
 Structural and functional studies of yeast class II Hda1 HDAC complex
 Structural studies of secreted aspartic protease Sap5 of *C. albicans*
 Structural studies of ovarian tumor suppressor Ovca1/Dph1-Dph2-Dph3 complex

Research assistant (X-ray crystallography)
Dept. of Life Science, POSTECH, Pohang, South Korea, 03.2004 ~ 01.2006
Advisor: Prof. Yunje Cho
 Structural studies of the peptide:N-glycanase Png1 and its complex with DNA repair
 protein Rad23 using X-ray crystallography
 Structural studies of Mus81-Eme1-DNA complex using X-ray crystallography

Research assistant (Biochemistry)

Samsung Biomedical Research Institute, Suwon, South Korea, 03.2002 ~ 02.2004

Advisor: Prof. Kyeong Kyu Kim

Cloning, purification, and crystallization of the adenosine to inosine editing enzyme ADAR1 and its complex with the left-handed Z-DNA

MS (NMR spectroscopy)

Dept. of Biochemistry, Yonsei University, Seoul, South Korea, 09.1997 ~ 08.1999

Thesis advisor: Prof. Weontae Lee

Structural studies of the melanocyte-stimulating hormone and its analogues using NMR

Structural studies of a new hypothalamic neuropeptide orexin-B using NMR

Structural studies of a sweet protein single-chain monellin using NMR

Structural studies of the substrates and inhibitor of HCV NS3 protease using NMR

Structural studies of the platelet aggregation inhibitor salmosin by NMR

Award

Prospective Graduate Student Award, Yonsei University, Seoul, South Korea, 05.1999

Scholarship

Max Planck society, 01.2009 ~

Peter und Traudl Engelhorn Stiftung, Stuttgart, Germany, 01.2006 ~ 12.2008

Korea Science and Engineering Foundation, Seoul, South Korea, 09.1997 ~ 08.1999

Publication

Jung-Hoon Lee*, Klaus Maskos, and Robert Huber (2009) Structural and functional studies of the yeast class II Hda1 histone deacetylase complex, *J. Mol. Biol.* 391, 744 ~ 757 (*corresponding author)

Claudia Borelli, Elisabeth Ruge, **Jung-Hoon Lee**, Martin Schaller, Alexandra Vogelsang, Michel Monod, Hans Christian Korting, Robert Huber, and Klaus Maskos (2008) X-ray structures of Sap1 and Sap5: Structural comparison of the secreted aspartic proteinases from *Candida albicans*, *Proteins* 72, 1308 ~ 1319

Jeong Ho Chang, Jeong Joo Kim, Jung Min Choi, **Jung-Hoon Lee** and Yunje Cho (2008) Crystal structure of the Mus81-Eme1 complex, *Genes & Dev.* 22, 1093 ~ 1106

Jung-Hoon Lee, Jung Min Choi, Changwook Lee, Ki Joung Yi, and Yunje Cho (2005) Structure of a peptide:*N*-glycanase-Rad23 complex: Insight into the deglycosylation for denatured glycoproteins. *Proc. Natl. Acad. Sci. USA* 102, 1029144 ~ 9149

Jung-Hoon Lee, Eunjung Bang, Kyeong-Jun Chae, Jin-Young Kim, Dai Woon Lee, and Weontae Lee (1999) Solution structure of a new hypothalamic neuropeptide, human hypocretin-2/orexin-B. *Eur. J. Biochem.* 266, 831 ~ 839

Jung-Hoon Lee, Keunsoo Bang, Jin-Won Jung, In-Ae Ahn, Seonggu Ro, and Weontae Lee (1999) Solution conformations of the substrates and inhibitor of Hepatitis C virus NS3 protease, *Bull. Korean Chem. Soc.* 20, 301 ~ 306

Seok-Yong Lee, **Jung-Hoon Lee**, Ho-Jin Chang, Joong Myung Cho, Jin-Won Jung, and Weontae Lee (1999) Solution structure of a sweet protein single-chain monellin determined by nuclear magnetic resonance and dynamical simulated annealing calculations. *Biochemistry (USA)* 38, 2340 ~ 2346

Song-Zhe Li, **Jung-Hoon Lee**, Weontae Lee, Cang-Ju Yoon, Ja-Hyun Baik, and Sung-Kil Lim (1999) Type I beta-turn conformation is important for biological activity of the melanocyte-stimulating hormone analogues. *Eur. J. Biochem.* 265, 430 ~ 440

Jung-Hoon Lee, Sung-Kil Lim, Sung-Ho Huh, Donghan Lee, and Weontae Lee (1998) Solution structures of the melanocyte-stimulating hormones by two-dimensional NMR spectroscopy and dynamical simulated-annealing calculations. *Eur. J. Biochem.* 257, 31 ~ 40