Structural Analysis of a Transactivation Domain-Cofactor Complex

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Venkatesh Ramakrishnan
aus
Theni, Indien

Göttingen 2005

Referent: Prof. Dr. Hans Joachim Fritz
Korreferent: Prof. Dr. Christian Griesinger
Tag der mündlichen Prüfung:

To My Parents

Acknowledgements

This thesis work is based on my research in the Department of NMR Based Structural Biology, Max Planck Institute for Biophysical Chemistry, Goettingen. I would like to thank all the people who made this possible:
Prof. Dr. Christian Griesinger for providing me the opportunity to work in a state-of-the art, world class facility and the constant support he extended in terms of academic input. His lectures on NMR were a source of inspiration. Prof. Dr. H.J. Fritz for being a committed doctoral advisor.
Dr. Markus Zweckstetter for introducing me to experimental NMR and patiently helping me through the analysis of data. Stefan Becker, Karin Giller and Melanie Wegstroth for educating me in the techniques of protein expression and purification.
Prof. Dr.Thomas Heimburg for teaching me titration calorimetry. Kerstin Overkampf for providing me all the peptides I needed for my work. Dr. Adelia Razeto for her role in the project as a crystallographer. Vinesh Vijayan for assiting me at the spectrometer. Jochen Junker for his expertise in computation. Wolfgang Hoyer for CD experiments.
Pierre Montaville, Konark Mukherjee, Christof Fares, Jonathan Farjon, Karel Kubicek, Nils Lakomek, Ovidio Andronesi, Young Sang Jung, Fernando Rodriguez and Devanathan Raghunathan for their thoughtful insights, useful discussions, instantaneous help and words of encouragement.
All the long term and short term members of the Griesinger group, for a friendly atmosphere and cheerful coffee time discussions. The people in the Studentenwohnheim Rosenbachweg for the relaxing dinner time chats and weekend breakfasts.
The Max Planck Gesselschaft for the financial support.
My parents for their lasting support and love.
And Almighty for His eternal mercy.

Abstract

Signal Transducers and Activators of Transcription (STAT) proteins were discovered as targets for interferon activation. These proteins are dormant in the cytoplasm until they are activated by one or more cytokines, that interact with their cognate cell surface receptors. These extracellular signaling proteins can activate various tyrosine kinases inside the cell that phosphorylate the STAT proteins. The activated STAT proteins migrate to the nucleus and drive transcription. Since STAT activity is modulated by several post-translational modifications and protein-protein interactions, these transcription factors are capable of integrating inputs from multiple signaling networks. The carboxy-terminal region of the STAT proteins functions as a transcriptional activation domain. Seven mammalian STATs have been discovered so far. Inspite of the progress that has been made in identifying the interaction targets of the activation domains and the recognition of the importance of particular residues for their function, the structural basis for the ability of transactivation domains to stimulate transcription is not understood well. The structural understanding of the transactivation domains is barely developed, in contrast to the well known DNA-binding domains. This thesis work involved the biochemical and structural characterization of the interaction between the transactivation domain of STAT6 and its co-factor, the Nuclear Receptor Co-activator -1(NCoA-1). The amino acid sequences in both STAT6 and NCoA-1 involved in this interaction had been identified previously.

The sequence spanning amino acids 794-814 in the STAT6 C- terminal region was identified to be essential in mediating the interaction with NCoA-1 and comprised of a LXXLL motif. In the case of NCoA-1, the region between the amino acids 213-462 had been identified to be involved in the interaction. The complex was reconstituted *in vitro*. Biochemical characterization of this complex using techniques like gel filtration,

limited proteolysis, mass spectroscopy and protein sequencing suggested that the residues 257-385 of the NCoA-1 is the minimal domain required for the interaction with the STAT6 and that the ratio of the binding partners in the complex was 1:1. Secondary structure prediction showed that the region C- terminal to residue 366 in the NCoA-1 was unstructured. Two shorter constructs of NCoA-1, one spanning residues 257-385 and other comprising of residues 257-370 were cloned for further experiments.

Titration calorimetry showed that the dissociation constant of the binding was in the order of 10^{-7}M. The crystal structure of the complex between the NCoA-$1_{(257-385)}$ and STAT$6_{(794-814)}$ peptide revealed that the NCoA-$1_{(257-385)}$ fragment in complex with STAT$6_{(794-814)}$ peptide forms a canonical Per-Arnt-Sim (PAS) domain. The free forms of both the NCoA-1 constructs failed to crystallize, therefore NMR was used for the study of these constructs. Isotope labeled NCoA-$1_{(257-385)}$ and NCoA-$1_{(257-370)}$ fragments were used for multidimensional NMR experiments. Backbone resonance assignment proved that the residues C- terminal to amino acid 366 were unstructured. Further structural analysis using Residual Dipolar Couplings (RDCs) was not successful, as the couplings obtained were not suitable for such studies. The reason was that the alignment of the protein in liquid crystalline medium which is a pre-requisite for the measurement of dipolar couplings was not optimal.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Man is distinguished from mouse in not having very different proteins or different numbers of them, but rather in having a common set of proteins, which are temporally and spatially separated in the developing organism[1, 2]. This separation is achieved by transcribing or repressing specific genes, degrading or stabilizing particular proteins, splicing RNA transcripts in one way or the other etc.,[3, 4, 5]. The distinction in the phenotype among various kinds of cells in higher eukaryotes are largely due to the differences in expression of genes encoding different proteins[6, 7]. Therefore, in spite of the prodigious numbers of coding sequences in their genomes and the apparent enormity of the task of genetic control, only a limited set of genes is used to generate diversity in functions and complexity in living organisms[8]. Most eukaryotic cells, except the haploid sex cells, contain the same set of genes. Yet, each specialised cell type expresses only a small fraction of the genome[9]. Selective gene expression allows cells to be metabolically strict, synthesizing only those gene products that are of immediate use under the prevailing environmental conditions. In multicellular organisms selective gene expression allows cells to fulfill specialized roles[10]. This process is described as cellular regulation, in simplest terms and its study calls for expertise from different branches of science. A simple mechanism called regulated recruitment forms the basis of many of the regulatory decisions taken by the cells, whether they constitute unicellular or multicellular organisms. The same regulatory processes not only generate the intricacy involved in the development of a complex organism from a single fertilized egg or the evolution of different forms of life, but also enable the cells to respond to their surroundings[11]. Proliferation of the cells and differentiation processes, that

underlie development are controlled by programs of regulated gene expression within the embryo. Higher life forms must continue to respond to various endogenous and exogenous stimuli to maintain the functioning of the adult organism, in the course of post-natal development. These stimuli induce specific adaptive responses that involve quantitative and qualitative changes in gene expression.

The ability of the cells to respond to essential signals in their environment constitutes one of the most fundamental features of life. These signals may have a variety of forms from simple organic molecules to large peptides, temperature, light or pressure. They may trigger an immediate change in the metabolism of the cell, change the electrical charge distribution across the plasma membrane or change the gene expression by initiating transcription inside the cell's nucleus. The increased knowledge of cellular signaling offers the opportunity to develop novel substances that target specific pathway molecules. Eukaryotic gene expression is regulated by multiple control mechanisms acting on at least several different levels - (i) the formation and maintenance of inactive and active chromatin, (ii) disruption of the nucleosome to form a productive transcription unit, (iii) processing of transcription (iv) transport of the mRNA to the cytoplasm, (v) translation of the mRNA, (vi) mRNA stability (vii) protein activation and stability[6, 3]. As eukaryotic DNA is highly folded, transcription is regulated to some extent by the uncoiling of appropriate chromosomal regions. However, these structural changes in the active region of the genome do not answer the question how the DNA sequences are selected for activation. The answer is to be found in the phenomenon of transcriptional control. Most intracellular signaling proteins function as modular switches[12]. In eukaryotes, many of these proteins seem to have a component-based architecture[13] and some of them have catalytic functions whereas others mediate protein-protein or protein-lipid interactions[14].

Prokaryotic RNA polymerase can initiate transcription on its own, but eukaryotic polymerases (RNA polymerases I, II and III) require formation of an assembly of additional transcription factors called the general transcription factors before the initiation of transcription[15]. This assembly provides a platform for transcriptional regulation, as many eukaryotic regulatory proteins are thought to work by aiding or inhibiting this process. Some of them activate while others repress transcription. These proteins bind to regulatory sequences arranged in modules in the DNA sequence, and the proteins themselves are called *trans*-acting proteins whereas the DNA sequences that participate in this interaction are called *cis*-acting elements[16]. The transcription machinery

consists of the core RNA polymerase and these transcription factors, 40 of them have been discovered so far, They were found to be assembled stepwise, during trials to assemble this machinery *in vitro*[17].

Most of the transcription factors have several modules in common, [18], they are :

1. DNA binding domain 2.Activation domains 3.Dimerization domain 4.Ligand binding domains

Some regulatory transcription factors do not have direct contacts with the basal transcription machinery[1], but instead bind to co-activators that contact the basal apparatus[19, 20]. General transcription factors, essential for the transcription of all genes, are required for initiating RNA synthesis at all promoters[20]. They assemble with the RNA polymerase at the core promoter[21], a DNA region consisting of a short initiator sequence at the transcriptional start point[15], and $\tilde{2}5$ nucleotides upstream, a TATA box,[2][22, 23] to form a complex surrounding the site of initiation of transcription. The interaction of the general factors with RNA polymerase at the core promoter initiates a low basal transcription rate. In addition to their own core promoter, most genes have short DNA sequences upstream or downstream to which other transcription factors can bind, dramatically improving the efficiency of the transcription of the core promoter[24]. These sequences called response elements, allow transcription to be tuned up or down, in response to various types of signals. Those which increase the rate of transcription are called enhancers and those that decrease it are called silencers[25, 26]. Placement of the same response element adjacent to genes residing at different chromosomal loci, allows a group of genes to be coordinately transcribed and their expression is regulated by the same regulatory protein. Transcription factors that selectively bind to one of the response elements, or to other DNA control elements located outside the core promoter, are called regulatory transcription factors. They increase or decrease the rate of transcription initiation, by interacting with the components of the basal transcription apparatus. This stimulation of transcription by a transcription factor binding to the DNA and activating adjacent proteins is known as transactivation[27, 28]. There are two types of transcription factors.

[1]The RNA polymerase and its associated proteins

[2]A eukaryotic DNA element upstream of the transcription start site.

1. Upstream factors are DNA binding proteins that recognize specific short sequences located upstream of the initiation start point[18]. The activity of these factors is not regulated[29]. They are ubiquitous[26], and act upon any promoter that contains the appropriate binding region on DNA, increasing the efficiency of initiation.
2. Inducible factors function in the same general way, but have a regulatory role. They are synthesized and activated at specific times or in specific tissues.

Both the upstream and inducible factors function by interacting with certain general factors in the basal transcription apparatus. A gene is expressed at a maximum level only when the set of transcription factors present includes all the regulatory transcription factors that bind to the gene's positive control elements[3, 30, 31]. Transcription factors contain at least two major domains : A DNA binding domain, responsible for tethering the transcription factor to the DNA helix and an activation domain (responsible for interaction with the general transcription apparatus to promote transcription)[32]. The activity of the transcription factors may be controlled by changing their DNA binding affinity in response to extracellular signals.

An ever increasing body of information suggests that proteins involved in the regulation of cellular events such gene expression, signal transduction, the cell cycle, protein trafficking, targeted proteolysis and cytoskeletal organization are constructed in a modular fashion from a combination of interacting and catalyzing domains[33, 12]. Interaction domains direct signaling polypeptides into specific multi-protein complexes, and link cell surface receptors to intracellular biochemical pathways that control cellular responses to external signals. The pathways and networks that connect receptors to their ultimate targets frequently involve a series of protein-protein interactions. As a consequence of these interactions signaling proteins recruited and confined to appropriate subcellular locations. These interactions determine the specificity with which enzymes interact with their targets, e.g. association of protein kinases and their substrates[34]. In an era of whole genome sequencing, hundreds of proteins essential for cellular regulation have been identified[35]. However, this list reveals neither the exact function of these proteins, nor their assembling into the molecular machinery and functional networks that control the cellular behavior[11]. The regulation of the cellular processes requires the interaction of protein domains, with one another and with phospholipids, small molecules or nucleic acids. Evolution of cellular signaling pathways is likely to have been facilitated by the modular nature of these domains and

their flexible binding properties[36].

1.0.1 Modular Interaction Domains

Interaction domains usually consist of 35-150 amino acids and can fold independently with their N- and C- termini abutted to each other in space. They can be expressed in isolation from their host proteins while retaining their intrinsic ability to bind their physiological partners. They can be incorporated into a larger protein such that their ligand-binding surface is exposed through which they recognize exposed surfaces on their partner molecules. A protein interaction typically recognizes a core determinant, where additional contacts and an element of selectivity are provided by flanking or non-contiguous residues[37]. The affinity of a single domain for a polypeptide motif alone is sufficient for a particular interaction in cells, in some cases. Selectivity in signaling can also be attributed to tertiary interactions, the structural organization of the interacting proteins and subcellular localization, domain competition, and multidomain interactions, *in vivo*. Consensus amino acid sequences allow signaling domains to be identified, which permit the prediction of binding properties and biological functions of a protein, based on the domain composition[38]. In the human genome, several interaction domains are found in hundreds of copies, and they are used repeatedly to regulate distinct aspects of cellular organization[35]. For example, about 115 SH2 domains and 253 SH3 domains are encoded by the human genome[39, 40, 41]. Some domains have specific functions whereas others can bind motifs from a bigger set of proteins and display a wider range of biological activity. SH3 domains, for example, regulate processes such as signal transduction, protein and vesicular trafficking, cytoskeletal organization, cell polarization and organelle biogenesis[39, 41]. Versatility in binding properties is a characteristic of interacting domains. Several distinct ligands can be engaged by an individual domain, either simultaneously or at successive stages of signaling[42, 43, 44, 45]. Different members of the same domain type can bind to various motifs. Another way of building interaction surfaces is through the joining of repeated copies of a small peptide motif, forming a much larger structure with multifaceted binding properties. Many modular domains undergo homo- or hetero typic domain-domain interactions. Such domains frequently identify proteins in a common signaling process and then direct their coassembly into functional oligomeric complexes[46]. There is no absolute distinction between the domains that bind peptide

motifs and those that interact with other folded domains[47].

1.0.2 Modular Signaling Systems

Interaction domains couple various cell surface receptors with their targets, and mediate the formation of signaling complexes in the cytoplasm and the nucleus[48]. They also play a role in regulating targeted proteolysis, endocytosis, vesicle and protein trafficking, cell polarity, cytoskeletal organization and gene expression. Strikingly common strategies are used by these different regulatory systems to assemble functional complexes, to compartmentalize molecular components and to direct enzymes to their targets[11]. In principle, by analyzing the *in vitro* binding specificities of interaction domains and by directly analyzing protein complexes by techniques such as mass spectrometry and yeast two hybrid analysis, it is possible to assemble a circuit for cellular protein interactions. The repeated use of signaling domains might have led to the evolution of new cellular functions, because the domains may be joined in new combinations to create new connections and pathways inside the cell. The fusion of separate domains can also create a new composite protein entity with more complex properties than either of the domain alone. Thus interaction domains provide a way to increase the connectivity of existing proteins, and impart new functions to these proteins. Recruitment of positive or negative regulators, that exhibit feed-forward or feedback control is an example for protein-protein interactions that can influence the signaling kinetics. In addition to this, if a protein-protein interaction is dependent on secondary events like phosphorylation at multiple sites, this may create a switch like response, as the activity of the relevant kinase rises above a certain threshold.

1.0.3 Transactivation Domains

The domains of the regulatory proteins that bind to other proteins are called transactivation domains, and are functional modules which enable sequence specific DNA binding to stimulate transcription. Activation domains are classified according to the preponderance of amino acid residues such as glutamine, proline and those bearing acidic side chains. Among these classes, the third class, which are called acidic activators, is the most extensively studied[49].

A consistent theme has emerged from the studies of the functional domains that bind to DNA sites distant from the RNA polymerase II initiation sites. These distant binding sites require one or more proteins that integrate the activating potential of the distant binding factor with the RNA polymerase II and other general transcription factors. These integrating factors are called co-activators, mediators, or TATA box binding protein-associated factors. When contacted by a distant DNA binding factor, the integrating proteins enhance the probability that RNA polymerase II will begin transcription at a particular site.

1.1 Topic for the thesis

The topic chosen for the thesis is the structural analysis of the interaction between a particular transactivation domain, that of the STAT6 and its coactivator protein, the NCoA-1.

1.2 STAT (Signal Transducers and Activators of Transcription) Proteins

The STAT (Signal Transducers and Activators of Transcription) are named after their dual role as signal transducers and activators of transcription[50]. The STAT proteins were discovered as targets for interferon activation[51]. These proteins are dormant in the cytoplasm until they are activated by one or more cytokines, that interact with their cognate cell surface receptors. These extracellular signaling proteins can activate various tyrosine kinases inside the cell that phosphorylate the STAT proteins[52]. The activated STAT proteins migrate to the nucleus and drive transcription. These proteins need to cross the cell membrane to get linked with the cytokine-responsive genes. Since STAT activity is modulated by several post-translational modifications and protein-protein interactions, these transcription factors are capable of integrating inputs from multiple signaling networks.

The carboxy-terminal region of the STAT proteins functions as a transcriptional activation domain. STATs 1,3,4,5a and 5b have relatively short carboxy terminal regions of about 40 to 50 amino acids whereas STATs 2 and 6 have extended C-

STAT protein	*Phenotype of null mice*
STAT1	Impaired responses to interferons; increased susceptiblity to tumors; impaired growth control
STAT2	Impaired responses to interferons
STAT3	Embryonic lethality; impaired response to pathogens
STAT4	Impaired T_H1 differentiation owing to loss of Il-12 responsiveness
STAT5A	Impaired mammary gland development owing to loss of prolactin responsiveness
STAT5B	Impaired growth owing to loss of growth hormone responsiveness
STAT6	Impaired T_H1 differentiation owing to loss of Il-4 responsiveness

Table 1.1: Role of STAT proteins as revealed by gene targeting in mice (adapted from Darnell *et al.*, 2002)

terminal domains[53]. STAT6 has been shown to be the most common transducer of the interleukin (IL-4) receptor α chain and mediates IL-4- and IL-13- initiated responses[54]. IL-4 and IL-13 are considered to be the crucial cytokines involved in the development of type II allergic diseases because IL-4 can promote immunoglobulin class switching to the IgE isotype in B cells[55] and IL-4 and IL-13 mediate the induction of Th2 differentiation[56]. Mice which are deficient in STAT6 genes are defective in IL-4 and IL-13 induced Th2 cell differentiation and IL-4 induced class switching of B cells to IgE and IgG1 isotypes[57]. In addition, STAT6 knock-out mice are protected from antigen-induced respiratory tract hyper-responsiveness and mucus production, and also from IL-13 induced respiratory tract hyper-responsiveness[51].

Seven mammalian STATs have been identified[53]which range in size from 750-850 amino acids. Chromosomal mapping of the mouse and human genes showed that, in both species, the STAT genes are localized in clusters, STAT1 and STAT4 in one group, STAT2 and STAT6 in the second group, STAT3 and STAT5a, STAT5b in the third set[58]. The STAT family of genes is evolutionarily conserved, and a STAT gene has been found in the lower eukaryote *Dictyostelium discoidium*[59]. The finding that STAT genes are not present in yeast suggests that the emergence of this protein family was an early step in the evolution of multicellular organisms. A STAT gene has also been identified in *Drosophila*[60]. The similarity of *Dictyostelium* and *Drosophila* STATs to STAT5 suggests that it is the ancestral gene. Table 1.1 shows the effect of the deficiency of different STATs in mouse models.

1.2.1 Transcriptional coregulators of STATs

The activity of STAT 3 and 5 has been shown to be upregulated by their interaction with transcription factors/co-regulators like Cyclic AMP Response Element Binding Protein (CREB) binding protein (CBP), Centrosomal P4.1 Associated Protein (CPAP), *c-jun*, Silencing Mediator for Retinoid Thyroid Hormone Receptors (SMRT), the glucocorticoid receptor and histone deacetylases[61]. STAT1 synergically acts with Sp1[62], while STAT6 interacts with CBP at a composite Interleukin-4 (IL-4) response element site[63]. The association of CBP with the STAT family proteins is thought to be rate limiting to the process of transactivation. CBP has been shown to interact with STAT1 and STAT2 and STAT6. CBP binds to serine- phosphorylated transcription factors and RNA polymerase. Some STATs have been known to be associated physically and functionally with coactivator proteins. It is possible that more than one co-operating factor can associate with the same STAT, providing a mechanism by which that same STAT can be involved in the activation of different cell types or activation sites, depending on which co-operating factors are expressed.

1.2.2 Functional architecture of the STAT proteins

All the STATs share the same functional organization as shown in the Figure 1.1, with six conserved domains, the N-terminal domain, the linker domain, the coiled coiled domain, the DNA binding domain, the *Src-homology* 2 (SH2) domain and the C- terminal transcriptional activation domain[63].

The N- terminal domain comprising of approximately 130 amino acids (Figure 1.1), is responsible for co-operative binding to multiple DNA sites[64]. It is an independently folded and stable moiety which can be cleaved from the full-length molecule by limited proteolysis. The crystal structure of the N-terminus of STAT4 reveals a dimer[65]. Studies suggest that this N- terminal dimerization promotes co-operativity of binding[64, 66, 65]. The N-terminal STAT domain is involved in the interaction with the transcriptional coactivator, CREB [3] binding protein (CBP)/p300[67], the Protein Inhibitor of Activated STAT (PIAS) family, receptor domains, and it regulates nuclear translocation[68]. A flexible polypeptide chain links the N- terminal domain to the coiled-coil domain which consists of four α helices (approximately amino acids 135-

[3] cyclic AMP response element binding protein

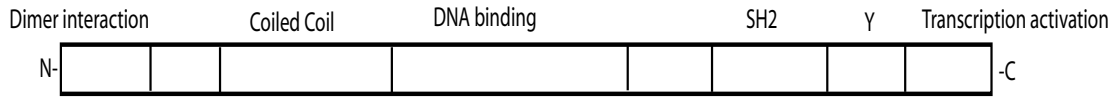

Figure 1.1: Organization of the functional domains in STAT family proteins
The SH2 domains mediate receptor binding and dimerization. The DNA-binding domain is located in the center of the molecule. The amino-terminal domain confers dimer-dimer interaction involved in the co-operative DNA binding. The transcriptional activation domains of the STAT proteins in located at the carboxy-terminal end.

315). The crystal structures of STAT1 and STAT3 reveal that this domain protrudes about 80 $\mathring{A}$ A laterally from the core structure[69, 70]. This domain forms a large, predominantly hydrophilic surface that is available for specific interactions with other helical proteins. Studies have also implicated the coiled-coil domain in receptor binding, tyrosine phosphorylation and nuclear export[71, 72]. Figure 1.2a shows the structure of a STAT3β homodimer. Figure 1.2b shows the domain boundaries in the monomer. The structure lacks the both the extreme N- and C- terminal regions, namely the cooperative DNA binding and the transactivation domains. A bundle of four antiparallel helices constitutes the N-terminal domain, which is followed by an eight stranded β-barrel. The β-barrel domain is linked to the SH2 domain by a small helical domain, formed by two helix-loop-helix modules. Apart from the STAT3 β homodimer, other STAT molecules that have been structurally characterized are STAT1 and STAT4.

The DNA binding domain (approximately amino acids 320 - 480) is a β, barrel with an immunoglobulin fold and lies C-terminal to the coiled-coil domain[70]. The number of direct contact sites between amino acid residues and DNA is modest, accounting for a dissociation constant in the nanomolar range. The DNA binding domain is located at the center of the STATs[73]. All the STATs have similar DNA binding specificities *in vitro*[73]. A palindromic sequence TTCnnnGAA (where n is any amino acid) is the optimal binding site for STATs 1,3,4 and 5[74]. STAT6 binds to the same palindromic sequence but with a spacing of four nucleotides.

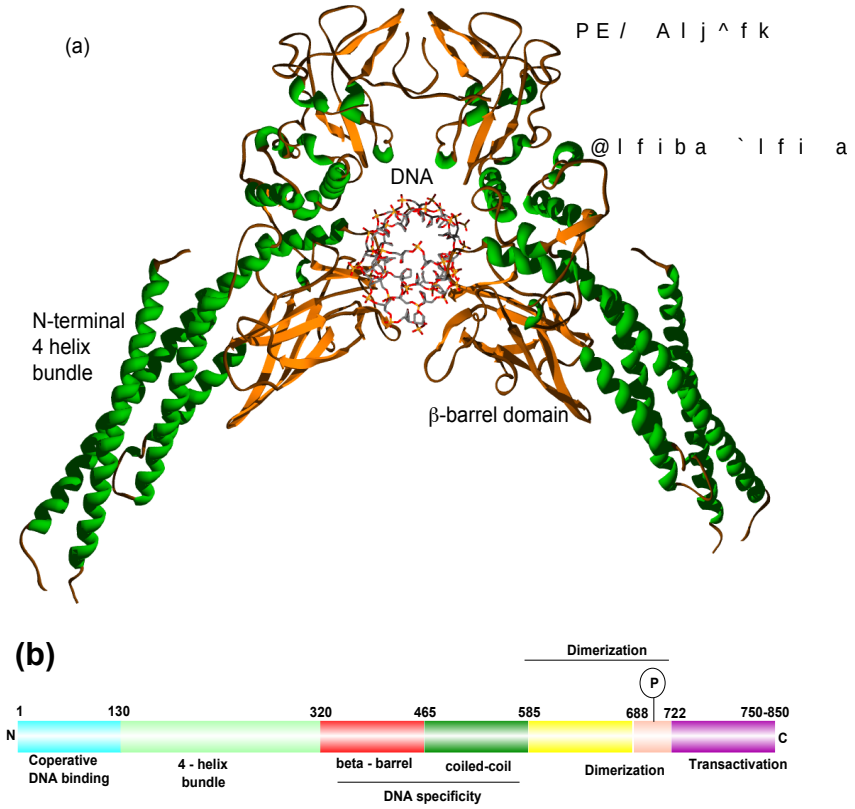

Figure 1.2: (a) Crystal structure of STAT3β homo dimer in complex with DNA (from Becker *et al*, [69]). (b) Diagrammatic representation of the domain arrangement in STAT3 β monomer.

The linker domain connects the DNA-binding domain with the SH2/dimerization domain. Mutational studies have also implicated the linker domain of STAT1 in transcriptional regulation[75].

SH2 domains play an important role in signaling through their capacity to bind to specific phosphotyrosine motifs. It is the most highly conserved STAT domain. The SH2 domain of the Dictyostelium STAT appears to represent one of the earliest identified SH2 domains [59]. Although the sequence of the STAT SH2 domain (residues 580-680) is quite divergent from other SH2 domains, its structure is well conserved. It consists of an anti-parallel β-sheet flanked by two α-helices, which form a pocket. An absolutely conserved arginine, which mediates the interaction with phosphate, lies at the base of this pocket (Arg-602 for STAT1, Arg-609 for STAT3). The ability of this SH2

domain to recognize specific phosphotyrosine motifs plays an essential role in three STAT signaling events: (1) recruitment to the cytokine receptor through recognition of specific receptor phosphotyrosine motifs. (2) association with the activating JAK [76]. (3) STAT homo- or heterodimerization [77]. DNA binding capacity can be regulated by structural changes in the SH2 domain as it binds phosphotyrosine[70]. As seen in the crystal structures, STAT dimerization depends on the interaction between the SH2 domain of one STAT monomer and the tyrosine phosphorylated tail segment of the other monomer. Residues most involved in defining the specificity of the interaction between the SH2 domain and tyrosine motif are located at positions +1, +3 and +5, +6, +7 C-terminal to the phosphotyrosine [70]. Closely positioned amino acids of the SH2 domain (e.g. Ala-641, Val-642) appear to participate in this interaction. All STATs except STAT2 have been shown to form stable homodimers *in vitro* and *in vivo*. Additionally, many STATs, including STAT2, can heterodimerize with other STATs through this reciprocal SH2-phosphotyrosine interaction [78].

Consistent with its ability to regulate unique transcriptional responses, the carboxy-terminal domain is poorly conserved among the STATs. The first evidence that the carboxy-terminus encodes transcriptional activation domain (TAD) came from a comparative analysis of the full-length STAT1 and an alternatively spliced isoform STAT1β, which lacks the last 38 carboxy-terminal amino acids [79]. Well characterized C-terminally truncated isoforms have also been identified for STAT3, STAT4 and STAT5 [80]. They appear to function as dominant-negative regulators.

Although a detailed understanding of how the STAT carboxy-terminus regulates transcription remains to be determined, important progress has been made. It has been determined that the transcriptional activity of several STATs can be modulated through serine phosphorylation [81]. Serine phosphorylation appears to enhance the transcription of some, but not all target genes. The highly conserved amino-terminal region of the STAT family is also involved in the dimer-dimer interactions leading to co-operative DNA binding[66].

1.2.3 JAK-STAT Pathway

Figure 1.3 shows a model of Jak-STAT pathway. Signaling through the JAK/STAT pathway is initiated when a cytokine binds to its corresponding receptor. This binding leads to conformational changes in the cytoplasmic part of the receptor, which lead

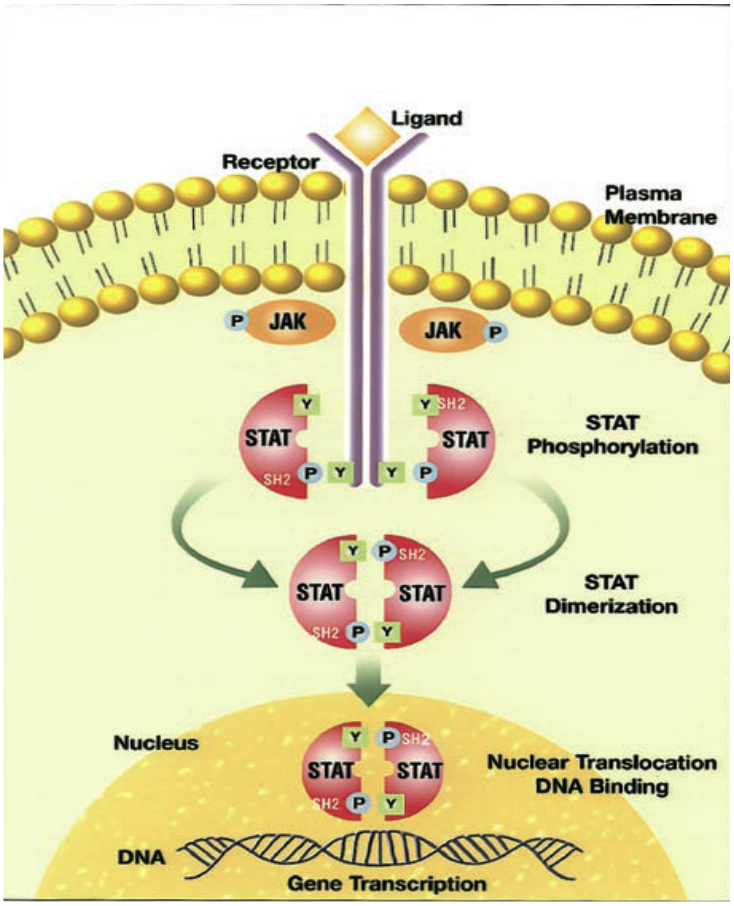

Figure 1.3: A schematic representation of the Jak-STAT pathway

to the recruitment of receptor associated members of the receptor associated Janus kinase (JAK) family of kinases. The JAKs in turn mediate the phosphorylation of specific tyrosine residues in the cytoplasmic tail of the cytokine receptor. These phosphorylated tyrosines then serve as docking sites for specific STATs. Once docked to the receptor the STAT molecule also becomes phosphorylated by JAKs, on a single tyrosine residue around amino acid 700 in the STAT molecule. These activated STATs then dissociate from the receptors and dimerize. Thereafter, they translocate to the nucleus and bind to DNA regions that belong to Gamma Activated Site (GAS) family of gene enhancers.

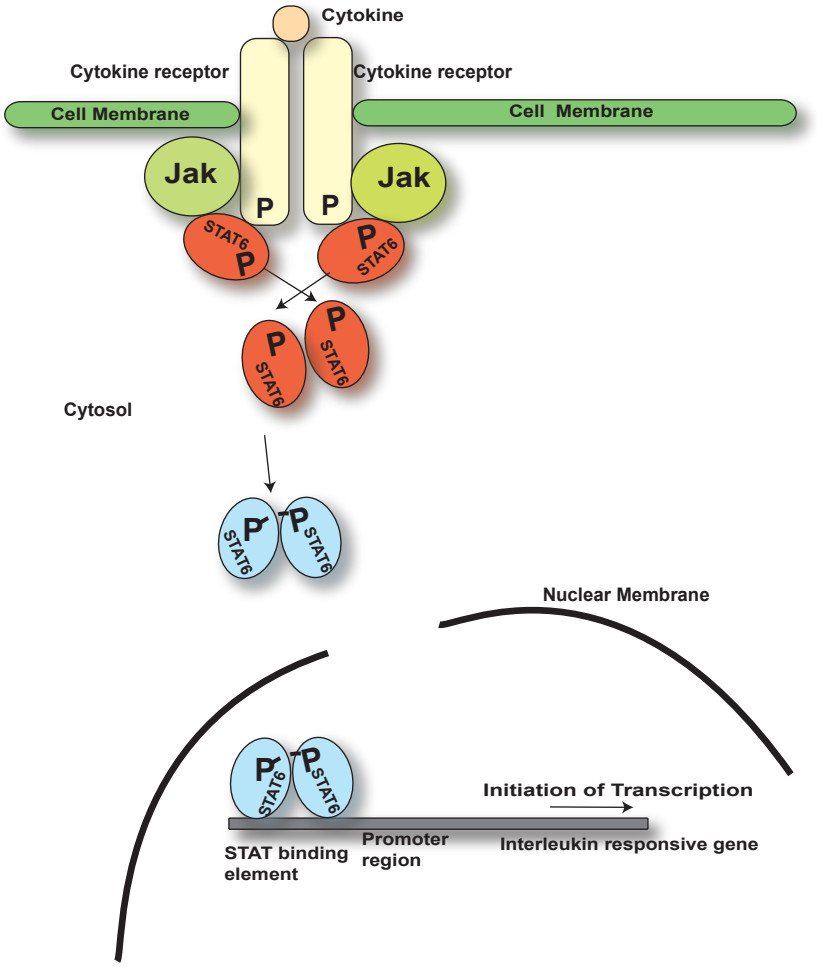

Figure 1.4: A schematic representation of the Jak-STAT6 pathway

JAK-STAT Pathway-with respect to STAT6

Figure 1.4 represents the JAK-STAT pathway. With STAT6 as the STAT component, IL-4 stimulation results in activation of Janus kinase 1 and 3, which in turn phosphorylate tyrosine residues in the cytoplasmic domain of the IL-4 receptor α. Thereafter, the STAT6 is recruited to the cytosolic domain of the receptor, which interacts with the phosphotyrosine residues via its SH2 domain. Subsequently, STAT6 gets phosphorylated

at Tyr694 by the Janus kinase, dimerizes and translocates to the nucleus where it modulates transcription through binding to STAT6 response elements. STAT6-binding elements were identified in the I_ϵ promoter, the CD23 promoter, and the IL-4 promoter. It has been shown by Arimura *et al*[82]., that STAT6 acts both as a positive and negative regulator of the gene expression induced by IL-4. Upstream of the phosphorylated tyrosine residue is the SH2 domain which is involved in the recognition of the activated receptors and in the mediation of the STAT dimerization. The DNA binding domain (DBD) of STAT6 lies between amino acid residues 268-448[73].

At least two different coactivators, p300/CBP and the Nuclear receptor Coactivator -1 (NCoA-1) are required for the transcriptional induction by IL-4 and they interact with the STAT6 transactivation domain (TAD)[83, 84]. p300 and CBP are functionally related multifunctional coactivators that participate in transcriptional activation by multiple transcription factors[20]. They possess an intrinsic histone deacetyl transferase (HAT) activity. Distinct domains for binding to several transcription factors, other coactivators and components of the basal transcription machinery have been characterized. p300/CBP was shown to increase IL-4 induced transcription of STAT6 reporter genes[83]. The region of CBP that interacts with STAT6 was distinct from interaction domain of CBP with other STAT family members. However it also contains the interaction site with the coactivators of the NCoA family[85, 86]. Nuclear Receptor Coactivators (NCoAs), also called the p160- or SRC (steroid receptor coactivators) were originally identified as nuclear receptor binding proteins, which enhance transcriptional activation by these ligand induced transcription factors.

1.2.4 Identification of a transactivation domain in STAT6

The carboxy-terminal region of both human and mouse STAT6 are proline rich, a common feature of some classes of transactivation domains[87]. To determine whether STAT6 contains a region capable of activating transcription, the modular property of transcription factors was exploited to generate fusion proteins, in which parts of the STAT6 molecule were fused to the DNA binding region of a heterologous protein, Gal4. The level of expression of these Gal4-STAT6 fusion proteins is not affected by the addition of IL-4 to these cultures. These experiments demonstrated that STAT6 contains a modular transcription activation domain that does not need IL-4 for its

activation function. Similar experiments with different lengths of the carboxy-terminal of the STAT6 fused with Gal4 protein were used to map the exact location of this trans-activation domain. The carboxy-terminal region of the STAT6 is required for IL-4 inducible transcription[87].

1.2.5 Nuclear Receptor Coactivators (NCoAs)

Coactivators are defined as proteins that can interact with DNA-bound nuclear receptors and enhance their transcriptional activation function[88]. Numerous nuclear receptor coactivators have been identified[89], particularly the steroid receptor coactivator (SRC) family has been extensively studied, in recent years[90]. The steroid and thyroid nuclear receptor superfamily is a large class of ligand-dependent transcription factors involved in the regulation of genes that play crucial roles in a wide range of biological processes like reproduction, development and homeostasis[91]. The steroid receptor coactivator 1 (SRC-1) was the first nuclear receptor coactivator to be cloned[88].

The members of the SRC-1 family of coactivators also share a common structure[92]. Figure 1.5 shows the domain arrangement in SRC family proteins. The N- terminal bHLH (Helix-Loop-Helix)-Per-Arnt-Sim(PAS)[4]region is the most conserved region in all the members of this family. The bHLH and the PAS motifs play a role in protein-protein interactions and dimerization[93]. The bHLH-PAS

[4]Per-Arnt-Sim domain : All the three domains had similar folds and were the first domains of this structural family to be identified

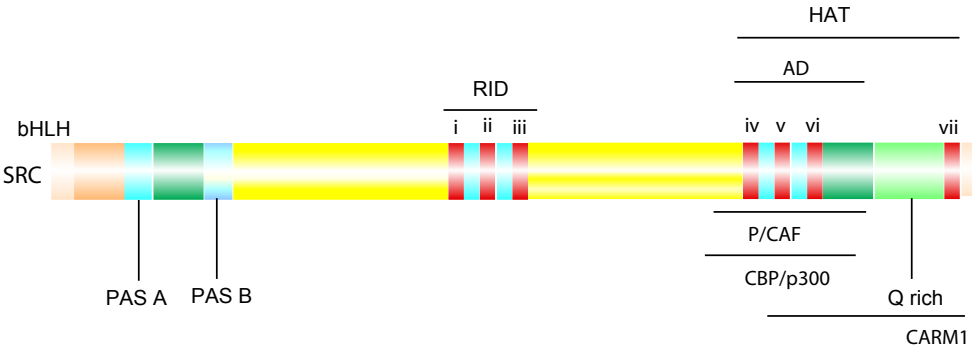

Figure 1.5: Domain arrangement in SRC family proteins

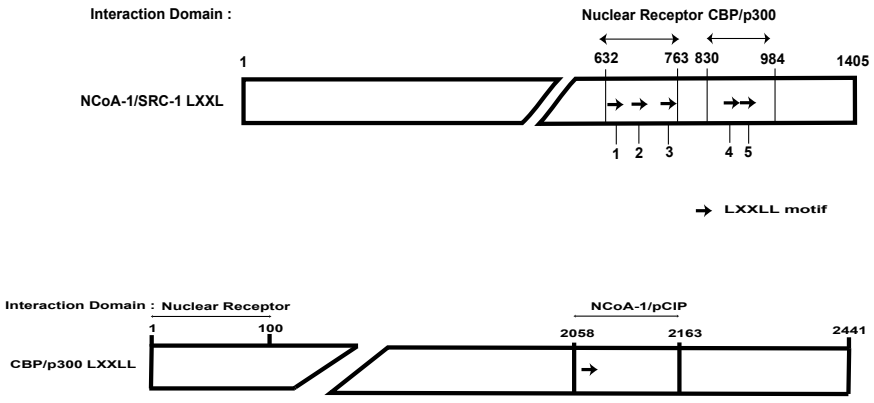

Figure 1.6: Location of LXXLL motifs in NCoA-1 and CBP

domain is followed by a central receptor interaction domain (RID) and a C-terminal transcriptional activation domain (AD). Analysis of the sequence of the RID revealed a conserved LXXLL motif, where L is leucine and X is any amino acid, that is termed the 'NR'[5]box[94]. This sequence has been identified in a number of proteins. Three such motifs were found in the RID of SRC coactivators. These motifs are required for mediating interactions between coactivators and liganded nuclear receptors. Figure 1.6 shows the location of LXXLL motifs in NCoA and CBP molecules. There are several types of recognition interfaces between the nuclear receptors and their coactivators. But a large number of these interactions are mediated by the highly conserved NR box[94]. Crystallographic and protein structure prediction analyses have shown that these motifs form amphipathic helices, with the leucine residues forming a hydrophobic surface on one side of the helix. Interestingly, different nuclear receptors prefer different NR boxes of the RID for interaction with coactivators, indicating the existence of a receptor-specific code. However, mutation of a single motif does not completely abolish coactivator interaction with nuclear receptors, since multiple NR

[5]NR-Nuclear Receptor

boxes contribute to the overall, high affinity binding to the receptor. The LXXLL motif is essential to mediate the binding of these coactivators to the ligand bound NRs, amino acids flanking the core motif are important for the recognition of NRs[95]. NCoA coactivators are also involved in transcriptional activation by several transcriptional activators, like AP-1, p53, Serum Response Factor (SRF), Nuclear Factor κB (NFκB) and STAT1[96]. Only one member of NCoA family proteins, the NCoA-1 interacts with STAT6[97].

1.2.6 Transcriptional activation by STAT6 requires direct interaction with NCoA-1

Transfection assays using a luciferase reporter construct with multimerized STAT6 response elements were carried out in IL-4 responsive liver HepG2 cells[84]. Vectors containing the coding sequences for NCoA-1, NCoA-2 and NCoA-3 were cotransfected. After transfection, the cells were treated with IL-4 and untreated cells were used as control. In comparison to the cells untreated with IL-4, induction with IL-4 led to a 6-fold increase in the reporter gene expression. The reporter gene expression was enhanced up to 20 times, when NCoA-1 was present, whereas cotransfection with NCoA-2 and NCoA-3 did not have any effect. In a further study, two Gal4 fusion constructs with STAT6 amino acids (677-791) and STAT6 amino acids 792-847 representing the N- and C- terminal regions of the STAT6 transactivation domains respectively, were used to assay gene expression. The results showed that NCoA-1 effected a stronger coactivation on the C-terminal part of the STAT6 TAD than the N-terminal part. Litterst *et al*. [97], found that the N-terminal part of the STAT6 TAD recruited CBP, but the far C-terminal part of the TAD directly contacted the NCoA-1.

They suggested a model (see Figure 1.7) where NCoA-1 binds to the C-terminal part of the STAT6-TAD and recruits CBP which interacts directly with the N-terminal part of the STAT6 TAD. Multiple interactions of CBP, NCoA-1, and STAT6 would stabilize binding to the promoter.

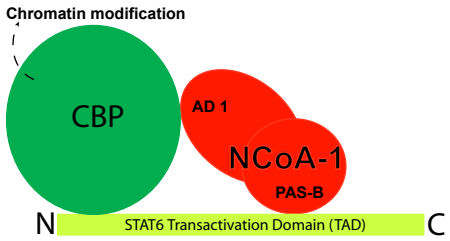

Figure 1.7: A schematic illustration showing simultaneous interaction between CBP, NCoA-1 and STAT transactivation domain

1.2.7 Preliminary studies of the NCoA-1 domain that interacts with the STAT6 transactivation domain

Based on GST-pull down assays Litterst *et al*[84]identified that the region comprising amino acids 213-462 that contains the predicted PAS-B domain and serine/threonine rich region in the NCoA-1 was identified to be necessary for its interaction with the STAT6 TAD. The existing model of transcriptional activation proposes that coactivators function as bridging factors, to recruit additional cofactors, and the basal transcriptional machinery to the DNA bound transcriptional factors. Some of them also function chromatin remodeling enzymes. NCoA-1 through its bHLH-PAS domain coactivates the transactivation of STAT6 and through its C- terminus also binds to the CBP/p300. The bound CBP could possibly stabilize the NCoA-1/STAT6 interaction because it binds to STAT6 TAD at a region close to the NCoA-1 binding site. CBP possesses histone acetylase activity required for chromatin remodeling. This concerted interaction between CBP/p300, NCoA-1 and STAT6 TAD leads to the transcription of the IL-4 responsive germ line (GL) ϵ gene. It is not yet known, if the expression of other STAT6 activated genes like GL γ1, CD23, and the Major Histocompatibility Complex (MHC) are dependent on this mechanism.

1.2.8 An LXXLL motif in the Transactivation domain of STAT6 mediates recruitment of NCoA-1/SRC-1

Litterst *et al*[97] identified that a 56-amino acid sequence, in the STAT6-TAD from amino acids 792-847 mediates the recruitment of NCoA-1. They showed that a peptide

containing the LXXLL signature motif of STAT6 from amino acids 794-814, and the antibodies for this peptide were strong inhibitors of the interaction of the STAT6 domain with NCoA-1 *in vitro*. They also showed that the mutations (L802A and L805A) of the STAT6-LXXLL motif, abolish the NCoA-1 binding, both *in vitro* and *in vivo*. It was found that the STAT6 LXXLL motif has sequence homology to the known coactivator motifs, but does not interfere with NR/coactivator interactions, suggesting that this motif might represent a specific target for STAT6 inhibition. Further, the LXXLL motif of the STAT6 is required for transactivation, but not involved in DNA-binding.

The NCoA coactivators were identified as nuclear receptor-binding proteins, which enhance transcriptional activation by ligand-induced transcription factors. Three homologous factors, termed NCoA-1, also called SRC-1 [98, 99]; NCoA-2, also called TIF2 or Glutamate Receptor Interacting Protein 1 (GRIP1)[100]; and NCoA-3, also called p/CIP, Activin Receptor Protein (ACTR), or Amplified in Breast Cancer 1 (AIB1) [101, 88, 102], have been identified by several laboratories. Subsequently a number of transcription factors have been shown to recruit these coactivators. It has been proposed that distinct classes of transcription factors selectively use specific coactivators and histone acetyltransferase activities for their function [20, 103]. In accordance with this model, a specific role of NCoA-1 was observed, but not of NCoA-2 or NCoA-3, for STAT6-mediated transactivation[84]. Only NCoA-1 acts as a coactivator of STAT6. STAT6 directly contacts NCoA-1 via a small carboxyl-terminal part of its transactivation domain. Overexpression of the STAT6-interacting domain of NCoA-1 inhibits transactivation by STAT6 in a transdominant manner, demonstrating the importance of NCoA-1 for STAT6 transactivation. An LXXLL motif (where L is leucine and X is any amino acid) with a predicted alpha -helical structure was indentified in the shortest NCoA-1 interacting fragment of the STAT6-TAD. The alpha -helical LXXLL motif was originally found in a variety of coactivators, e.g. NCoA/p160/SRC members, p300/CBP, RIP-140. It is described as a signature motif, which mediates the recruitment of these proteins by the nuclear hormone receptors [94]. Specific LXXLL motifs of NCoAs mediate the ligand-dependent interaction with nuclear receptors, as well as interaction with p300/CBP. STAT6 recruits NCoA-1 via an LXXLL motif in its transactivation domain. A peptide containing the LXXLL motif of STAT6, as well as antibodies raised against this peptide, are potent inhibitors of the interaction with NCoA-1 *in vitro*. Mutagenesis of the STAT6-LXXLL motif abolishes the binding of NCoA-1 *in vitro* and *in vivo*. A STAT6 variant, in which the LXXLL motif is mutated, is less

active in induction of a STAT6-dependent reporter gene. The LXXLL mutant of STAT6 shows impaired induction of eotaxin-3 expression, indicating that recruitment of NCoA-1 by the STAT6-LXXLL motif is essential for full transactivation of this endogenous target gene. The LXXLL motif of STAT6 shows no sequence homology to the known coactivator motifs. NCoA-1 does not only use its own LXXLL motifs to contact DNA-bound factors and coactivators, but can also be recruited itself by an LXXLL motif of DNA-bound transcription factors, such as STAT6[97]. Based on the above studies, a peptide containing amino acids 794-814 of the STAT6 transactivation domain and a construct containing the amino acids 257-420 of the NCoA-1 were selected for structural analysis.

1.3 Justification of the scientific question

Transcriptional activation is mediated by large protein complexes assembled on target gene promoter regions. These complexes contain activators and coactivators of transcription as well as elements of basal transcription machinery. The specificity, timing, and degree of transcriptional activation depend not only on the proteins forming the complex, but also on the way they interact with each other. Coactivators are part of chromatin remodeling complexes, and serve as bridges to the basal transcription machinery [104]. Inspite of the progress that has been made in identifying the interaction targets of the acidic activation domains and the recognition of the importance of particular residues for their function, the structural basis for the ability of transactivation domains to stimulate transcription is not understood well. The structural understanding of the transactivation domains is barely developed, in contrast to the well known DNA-binding domains. The reason is partly because transactivation domains share little sequence similarity. They have a poor intrinsic propensity to form secondary structure, although they apparently need to make specific interactions with several different target factors[53]. The structural analysis of the transactivation domains is highly hampered by the highly flexible conformation of such domains in the unbound state. Most transactivation domains adopt a regular structure only in complex with a coactivator. This functional conformation is brought about by an induced fit. This corroborates the idea that the minimal requirement for partner recognition is defined by very short stretches of amino acids, which become structured only

when bound to target proteins. While the LXXLL and similar motifs form structurally homologous amphipathic helices, their target proteins share no structural similarity. Structural analysis, in particular solving the three dimensional structure of the complex between a transactivation domain and its cofactor would make a very important contribution to this very active research field.

The aim of this thesis work is to structurally characterize the complex between the transactivation domain of STAT6 and the transactivation domain interacting region of the NCoA-1 as well as the free form of the NCoA-1, using NMR (Nuclear Magnetic Resonance), crystallography and biochemical methods. Thereby the implications of this structure for the specificity of the STAT6 trans-activation domain and NCoA-1 and in general the mechanism of a transactivation domain-cofactor interaction would be better understood.

Chapter 2

Materials and Methods

2.1 Materials

Chemicals and reagents were purchased from Roth, Merck, and Sigma, unless specified otherwise. $^{15}N\,H_4Cl$, ^{13}C - glucose were obtained from *Spectral Isotopes*. Enzymes and buffers were from New England Biolabs (Beverly, MA). Plasmid purification kits were from Qiagen (Hildesheim, Germany).

2.1.1 Buffers and Solutions

Table 2.1 gives the composition of the buffers and reagents used in this thesis work.

2.1.2 Media

Luria Bertani (LB) medium - for expressing unlabeled proteins (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.4)

Minimal medium - for ^{15}N and $^{15}N/^{13}C$ labeled proteins:200 ml 5$\times$ M9 salts, 20 ml 20% glucose , 2 ml 1M $MgSO_4$, 50μl, 2M $CaCl_2$, 6 ml 1mg/ml Thiamin,adjusted to 1 litre with double distilled H_2O.

OD2 rich medium (*Silantes*®) for ^{15}N and^{15}N / ^{13}C labeling of proteins. A certain chemolithoautotrophic bacterial lysate supplied with M9 salts. The OD level gives an

Experiment	*Buffer/Solution*	*Components*
His_6 purification	Lysis Buffer	20 mM Tris-HCl, 500 mM NaCl,
on Ni-NTA agarose		20 mM Imidazole, 0.5 mM PMSF, pH 7.9
		Complete®[1] without EDTA
	Wash buffer	20 mM Tris-HCl, 500 mM NaCl
		40 mM Imidazole, 0.5 mM PMSF, pH 7.9
	Elution buffer A	20 mM Tris-HCl, 500 mM NaCl
		100 mM Imidazole, 0.5 mM PMSF, pH 7.9
	Elution buffer B	20 mM Tris-HCl, 500 mM NaCl
		100 mM Imidazole, 0.5 mM PMSF, pH 7.9
TEV digestion	TEV digestion buffer	50 mM Tris-HCl, 300 mM NaCl, pH 8.0
NMR	NMR buffer	50 mM NaH_2PO_4, 150 mM NaCl,
		1 mM DTT, pH 6.5
Crystal screening	HEPES buffer	50 mM HEPES, 150 mM NaCl,
		1 mM DTT, pH 7.0
5 × M9 salts		33.9 g of Na_2HPO_4, 15 mg KH_2PO_4,
		2.5 g NaCl, 5 gm NH_4Cl, in 1 litre of dd H_2O
		adjusted to pH 7.4 and autoclaved.
Polyvinylidene fluoride		
	10× CAPS buffer	221.13 gms of CAPS in 900 ml of ddH_2O
(PVDF) Blotting		pH 11.0
	Transfer Buffer	100 ml of 10× CAPS buffer, 100 ml methanol, 800 ml dd H_2O

Table 2.1: Table showing various buffers

indication of the amount of enriched compounds present in the medium[2]. For example, in the OD2 medium, uninduced bacteria of common strains like BL21DE3 can grow up to an OD of 2.0 at 600 nm.

2× Yeast Tryptone (YT) medium : for transformation purpose (16g bacto-tryptone, bacto-yeast extract 10g, NaCl 5g)

2.1.3 Bacterial strains

E. coli XL2Blue for cloning.
E. coli BL21DE3 for expression.

2.1.4 Expression vectors

Table 2.2: Expression vectors used for expression

Vector	*Antibiotic marker*	*Affinity tag*	*Protease*
pET16b (*NEB*)	Ampicillin	N-His_6	TEV
pTWIN1	Ampicillin	Chitin binding domain	Intein based splicing

2.1.5 Oligonucleotide primers

Primer	*Sequence*	*Restriction Sites*
NCoA-$1_{(257-370)}$	EP-7 : 5'-GCTAAGCTTCATATGACGGGTGTAGAATCCTTTATGACC - 3' EP-9neu : 5'-GCTCTCGAGGGATCCCTACCCACTGTGCTGCCTGTCG-3'	NdeI and Bam H1
NCoA-$1_{(257-385)}$	EP-7 : 5'-GCTAAGCTTCATATGACGGGTGTAGAATCCTTTATGACC - 3' EP-14 : 5'-GGTCTCGAGGGATCCTTATCGGGGAATTGACATTCCAGAATTAG-3'	NdeI and Bam H1
NCoA-$1_{(257-420)}$	EP-7: 5'-GCTAAGCTTCATATGACGGGTGTAGAATCCTTTATGACC - 3' EP-8 : 5' -GCTCTCGAGGGATCCCTAGCGGTTTATTCTGGTGGATACC-3'	NdeI and Bam H1

Table 2.3: Primers and restriction enzymes used for cloning NCoA-1 constructs

[2]data from Silantes media manual

2.2 Biochemical Methods

2.2.1 Cloning of NCoA-1 fragments

Fragments containing 257-385, 257-370, 257-420 of the NCoA-1 PAS-B region were cloned into a modified pET16B (*Novagen*) vector containing a Tobacco-Etch-Virus (TEV) protease cleavage site. A full length NCoA-1 clone provided by Dr. Edith Pfitzner, Georg Speyer Haus, University of Frankfurt. All NCoA-1 constructs were derived from the full length clone using appropriate primers as shown in Table 2.3.

Polymerase Chain Reaction (PCR) was performed with the NCoA-1 full length fragment in pET30C as the template. The primers used for different constructs are listed in Table 2.3. After an initial incubation at 95 °C, 25 cycles were performed, with each cycle consisting of an incubation for 3 seconds at 95 °C and annealing at 55 °C for a minute followed by extension for 3 minutes at 68 °C. The amplified fragments were digested with the appropriate restriction enzymes (refer Table 2.3) for 3 hours at 37 °C. The pET16b vector was linearized using appropriate restriction enzymes mentioned in Table 2.3 based on the domain to be cloned. The digested vectors were treated calf intestinal phosphatase (Boehringer Mannheim) with to prevent self-ligation in the following step. These fragments were ligated into the previously linearized pET16b expression vector using T4 DNA ligase (MBI Fermentas) by incubation at 14 °C overnight. All the restriction enzymes were purchased from MBI Fermentas.

2.2.2 Expression and Purification of NCoA-1

Transformation

About 1μl of a miniprep of the respective expression construct was added to 50μl of calcium-competent BL21DE3 (*Stratagene)* cells. After an incubation on ice for 30 minutes, the cells were subjected to a heat shock of 42 °C for 30 seconds, and incubated again in ice for 2 minutes. The cells were transferred to 1 ml of prewarmed YT medium, and incubated at 37 °C for 1 hour after which they were pelleted down. About 850 μl of the supernatant was discarded and the remaining medium was used to resuspend the cells, and they were plated on LB agar plates with ampicillin as selection marker. The plates were incubated overnight at 37 °C, and colonies developed on them. The plates were stored at 4 °C.

Expression of NCoA-1

A single colony was picked from a fresh plate, inoculated into 2 ml of LB medium with ampicillin and incubated at 37 °C for 6 hours. When the culture turned slightly turbid, 10 μl of this culture was used to inoculate 50 ml of LB medium which was shaken overnight in a baffled conical flask. The next day, the culture was diluted into 1 litre of LB medium. The cells were grown up to an optical density (O.D.$_{600nm}$) of 0.45 at 600 nm and then transferred to 28 °C, grown to an O.D.$_{600nm}$ of 0.75. At this point, the cells were induced with 0.4 mM isopropyl β-thiogalactoside (IPTG), for 7 hours. Thereafter, they were harvested by centrifuging at 6000 r.p.m. for 15 minutes at 4 °C, and stored at -80 °C.

Purification of NCoA-1 fragments

The frozen cells were thawed in ice and resuspended in lysis buffer. This suspension was kept in ice and sonicated 8 times for 20 seconds, with the maximum power at 80 %, each time. The sample was cooled for 4 minutes on ice between the sonication steps. After sonication, this suspension was centrifuged using a Beckman centrifuge for 45 minutes, 20,000 r.p.m., at 4 °C. The supernatant was incubated for 60 minutes at 4° C with pre-equilibrated Ni-NTA resin (*Qiagen*), on a tumbling shaker. The pellet was stored for further analysis. After incubation, the supernatant and resin mixture was taken in a 10 ml polypropylene column (*Pierce*) and the flow through was collected. The resin was washed with 20 ml of lysis buffer and 20 ml wash buffer. The NCoA-1 fragment was eluted along with the histidine tag within 10 $\times$ 1.5 ml fractions, first with elution buffer A and followed by elution buffer B (see Table 2.1). Fifteen microlitres of the flow through, the first wash, the second wash, and each of the 20 elution fractions and a scoop of the resin were taken for analysis with a 15% SDS-PAGE gel.

The fractions containing the protein were pooled together and dialyzed against TEV digestion buffer, overnight at 4 °C, using Spectrapor®, 3500 MW cutoff dialysis membrane. Proteolysis was performed using a TEV concentration of 1.0 mg/100 mg of the protein. The extent of digestion was periodically checked using 15% SDS-Polyacrylamide (PAGE) gels. After the complete digestion, the protein was dialyzed overnight against a buffer containing 20 mM Tris-HCl, pH 7.9, 500 mM NaCl and 25 mM imidazole at 4 °C. The dialyzed NCoA-1 was incubated with 1 ml of Ni-NTA resin

and then loaded on a 10 ml column. The flow through which contained the NCoA-1 alone was collected, whereas the histidine tag remained bound to the column. The resin was washed with 6 $\times$ 1 ml of the same dialysis buffer. The flowthrough and the wash fractions containing the digested protein were pooled together and dialyzed against an appropriate buffer for further studies.

2.2.3 Media switch

The commercially available isotope labeled medium and minimal medium enriched with isotopes are relatively expensive, especially when the expression level from these media is low. In these cases, the amount of labeled medium required can be reduced by growing the cells in LB medium up to the point of induction with IPTG, and after induction the cells are washed with M9 salts and transferred to a lower volume of labelling medium. The protein of interest is expressed only after induction by IPTG, where the cells are introduced into a medium in which the nitrogen and carbon are isotope-labeled. During translation, most of the protein would have ^{15}N and ^{13}C isotopes incorporated into their amino acids. This media switch protocol was used in the case of producing isotope labeled NCoA-1$_{(257-385)}$ and NCoA-1$_{(257-370)}$ fragments, where the cells were grown at 37 $^{\circ}C$ up to an O.D.$_{600nm}$ = 0.75 at 37 °C and washed with M9 salts twice before being resuspended in a labeled medium and induced with IPTG. The ratio of shrinkage from unlabeled medium to labeled medium was usually 3 : 1.

2.2.4 *Protocol for the expression and purification of the NCoA-1$_{(257-385)}$ fragment*

The general protocol mentioned in the section 2.2.2. was also employed in the expression and purification of the NCoA-1$_{(257-385)}$ and NCoA-1$_{(257-370)}$ fragments. But a few changes had to be made to the protocol when expressing NCoA-1$_{(257-385)}$ fragment and while purifying the NCoA-1$_{(257-370)}$ fragment. The yield of NCoA-1$_{(257-385)}$ fragment was comparatively lower than that of the NCoA-1$_{(257-420)}$ and NCoA-1$_{(257-370)}$ using the same expression strategy. Therefore, a different protocol was employed in the case of the expression of NCoA-1$_{(257-385)}$. Cells were grown for 4 hours in 3 ml of LB preculture, and transferred to 20 ml LB medium with ampicillin. The next morning the cells were

transferred to 1 litre of LB medium and grown until they reached an O. D. $_{600}$ = 0.75, centrifuged at 4000 r.p.m for 4 minutes and washed twice with 1 $\times$ M9 salts. After the wash, the cells were resuspended in 300 ml of doubly labeled *Silantes* medium and induced with 0.025 mM IPTG, doubling the IPTG added, every hour for the next 3 hours, at 37°C. After 3 hours, the cells were transferred to a temperature of 18° C and grown for about 20 hours. NCoA-1$_{(257-385)}$ was purified using the same protocol that was used for other constructs of NCoA-1.

2.2.5 *The problem of impurity in the purification of NCoA-1*$_{257-420}$ *fragment*

The NCoA-1$_{(257-370)}$ purified through the Ni-NTA process had a strong absorption at 260 nm, suggesting the possibility of DNA bound to the protein. This absorption was removed by increasing NaCl from 500 mM to 1 M and adding DNAse 1 to the lysis buffer which was used to resuspend the thawed cell pellet before sonication.

2.2.6 Electrophoresis

Polyacrylamide gel electrophoresis with Sodium Dodecyl Sulphate (SDS) is a powerful tool to separate polypeptides in complex biological solutions[105]. Gel electrophoresis was done using SDS buffer at 25 mA, using a *Biorad* gel electrophoresis appartus. The protein was stained with Coomassie and the gel was destained using 10% acetic acid[106].

2.2.7 Gel Filtration

The NCoA-1$_{(257-385)}$ protein fragment was dialyzed against 50 mM HEPES buffer, pH 7.0, with 150 mM NaCl. STAT6$_{(794-814)}$ peptide was added in 5 times molar excess and incubated at room temperature, for about 1 hour and the sample was concentrated upto 0.9 mM using a Vivaspin® (*Sartorius*) 10 kDa molecular weight cut-off concentrator at a speed of 6000$\times$g. The concentrated sample was passed through a pre-equilibrated Superdex HR 75® (*Amersham*) column. The fractions corresponding to the UV absorption peak at 280 nm were checked for eluted protein.

2.2.8 Limited Proteolysis

Many studies have demonstrated the utility of limited proteolysis as a very useful tool to probe conformational states of proteins[107, 108]. A protein substrate can be proteolysed only if the polypeptide chain in question can bind and adapt to the specific stereochemistry of the protease's active site[109]. Since the active sites of proteases do not fit a stretch of a globular protein at least 8-10 amino acids in length, adaptation of a globular protein substrate to the active site of the polypeptide is not possible[110]. In the case of locally disordered or unfolded domains, there is enhanced segmental mobility and conformational flexibility, which makes these regions suitable for attack by the protease enzyme[111, 112]. To analyse the minimal fragment of NCoA-1 that interacts with the STAT6$_{(794-814)}$ fragment, NCoA-1$_{(257-420)}$ and the complex of NCoA-1$_{(257-420)}$ with STAT6$_{794-814}$ were subjected to a limited proteolysis experiment using trypsin as the protease. Different ratios of the protease to its substrate were used at varying lengths time. The reaction was stopped by quick freezing the mixture in liquid nitrogen. Samples were also taken for analysis with a 15$\%$ SDS polyacrylamide gel. The resuting bands of protein were blotted on a Polyvinylidine Fluoride (PVDF) membrane. The protein transferred to the membrane was N-terminally sequenced by Edman degradation.

2.2.9 Polyvinylidine Fluoride (PVDF) Blotting

The advantage of PVDF blotting is that it removes all free SDS and buffer salts and the blotted protein is suitable for direct NH2-terminal sequencing[113]. The membrane was prepared by immersing it in 100 % methanol for 1 minute and three times in 1 $\times$ transfer buffer for 15 minutes, changing the buffer each time. The gel was incubated in the transfer buffer for 5 minutes. The membrane and the gel were assembled in a Biorad blot transfer apparatus, along with strips of Whatman filter paper soaked in transfer buffer. Transfer of the protein from the gel to protein was done by applying a voltage difference of 60 V across the gel to the membrane, for 35 minutes. After the transfer, the membrane was removed, rinsed with water twice, and washed with methanol. The transferred protein was stained by Coomassie and destained by washing the membrane three times in 50 % methanol.

2.2.10 Mass Spectrometry

Fractions from the corresponding samples used for PVDF blotting were taken for mass spectroscopic analysis. The measurements were performed in the laboratory of Dr. Henning Urlaub, MPI for Biophysical Chemistry, using a Bruker (Karlsruhe, Germany) Matrix Assisted Laser Desorption and Ionization(MALDI) [114, 115, 116] mass spectrometer.

2.2.11 Circular Dichroism(CD)

CD spectra of different NCoA-1 fragments were measured at a concentration of 50 μM, with the same buffer conditions used for NMR measurements (see NMR Buffer Table 2.1). The sample for the CD spectroscopy consisted of 50 μM of the NCoA-1 protein taken in a quartz cuvette of 0.2 cm pathlength. The spectra were measured using a *JASCO* CD spectrometer at 20 °C. Phosphate buffer, containing 50 mM Na_2HPO_4, and 150 mM NaCl, 1 mM DTT and pH 6.5 was used throughout the measurements. STAT6$_{(794-814)}$ peptide was also dissolved in the above buffer such that the concentration was 50 μM and a CD spectrum was measured. A complex between the NCoA-1$_{(257-370)}$ fragment and STAT6$_{(794-814)}$ peptide was constituted in the above buffer, and the final concentration of the complex that was used for CD measurements was 50 μM. The spectra were analyzed using *SpectralView* software supplied by JASCO Corporation, Osaka, Japan. All the measurements were performed using the same set of parameters.

2.2.12 Fluorescence spectroscopy

Thiol reactive dyes

A 10 mM stock solution of Texas Red® dye (*Molecular Probes*) was prepared in the same buffer as the protein. The solution was protected from light as much as possible by wrapping the container in aluminium foil. In order to attach fluorescent dyes for anisotropy studies, STAT6 peptides with N- terminal and C- terminal cysteines were synthesized using solid phase peptide synthesis[117] and purified using High Pressure Liquid Chromatography(HPLC). Texas Red was added in 20 times molar excess and incubated at 4 °C overnight. The sample was checked for dye labeled peptide using Electron Spray Ionization (ESI)-Mass spectrometer(*Waters*).

Synthesis of a fluorescent labeled probe

Fluorescein is a popular probe for biophysical studies, because of it biocompatibility. Carboxyfluorescein, especially has remained the molecule of choice for the preparation of hydrolytically stable fluorescent peptides, and conjugated proteins. Fluorescein-labeled peptide derivatives have been employed as fluorescent markers in many biophysical and cell biological applications. STAT6 peptides (amino acids 801-814) were synthesized separately with N- and C- terminal cysteines with a carboxyfluorescein dye attached to the cysteins. These peptides were used as probes for detecting conformational changes in the NCoA-$1_{(257-370)}$ fragment upon titration. The fluoresence emission was measured during the process of stepwise titration of the peptides into NCoA-$1_{(257-370)}$ sample of 0.01 mM concentration inside a quartz cuvette fitted with a magnetic stirrer. The starting concentration of the peptide was 0.001 mM and it was increased in 0.001 mM steps.

2.2.13 Isothermal Titration Calorimetry

Experimental Setup

The samples were degassed for 5-10 minutes with care to avoid evaporation. It was ensured that the samples were devoid of any precipitation, since it would influence the heat of the reaction. The NCoA-$1_{(257-370)}$ sample solution was carefully filled into the reaction cell of the ITC unit of a Microcal MCS-2 calorimeter, by using a Hamilton syringe, without introducing any air bubbles, as the presence of even tiny air bubbles might lead to serious artifacts in the signal. Prior to the experiment, the cell and its contents were equilibrated with its surroundings to have the same temperature, so that neither heat is lost nor gained.

Design of an ITC experiment

From the crystal structure of the complex between NCoA-$1_{(257-370)}$ and $STAT6_{794-814}$ fragment, it was evident that only one peptide molecule binds to a single NCoA-1 molecule. In the case where there is only one binding site on a macromolecule *M* to which a ligand *X* binds, then,

according to the law of mass action, the binding constant,K_A, is given by,

$$K_A = \frac{[boundsites]}{[unboundsites][X]} \tag{2.1}$$

and,

$$\triangle G^0 = -RTlnK_A = RTlnK_D = \Delta\, H - T\Delta\, S \tag{2.2}$$

where,
R = gas constant (8.314 J K^{-1} mol^{-1})
T= absolute temperature (in K, degrees Kelvin)
K_A=Association constant in units of concentration^{-1} (expressed as Moles^{-1})
K_D=Dissociation constant in units of concentration(expressed as Moles)
Δ H = Enthalpy (kcal/Mole) Δ S = Entropy (kcal/Mole)

In biochemistry, dissociation constants are used, as the units are easier to follow.

$$\Delta\, H = \frac{\Delta\, Q}{\Delta\, X} \tag{2.3}$$

Δ H is expressed in kcal/mol. where,
$\Delta\, Q$ - total heat released under each titration point
$\Delta\, X$- number of moles of the ligand

The power released/second (μJ/sec.) was plotted and a characteristic pattern of peaks was obtained. The area under each peak was integrated to obtain the total heat released at a titration point (Δ Q). This value is divided by the diluted concentration of the ligand at each titration point and the values were plotted. In the case of NCoA-$1_{(257-385)}$/STAT6(794-814) interaction, a molar ratio of binding at equilibrium was expected as indicated from the biochemical studies. Therefore, Δ H was calculated for this ratio of the protein/peptide.

By using a non-linear squares fit of calorimetric titration data K_D is determined in a single experiment.

Titration

The number of steps of the titration depends on the following factors:[118]
(i) sensitivity of the equipment to detect signals, (ii) energy released or absorbed during

the reaction, (iii) time taken by the process to reach the equilibrium, (iv) additional processes like conformational changes, (v) concentrations of the interacting partners.

In order to get accurate results, the number of titration steps were increased as much as possible.

A series of samples of NCoA-$1_{(257-370)}$ were made, such that there were 25 nanomoles, 35 nanomoles, 45 nanomoles, 60 nanomoles, 75 nanomoles in the reaction cell (volume of the reaction cell = 1.33 ml). The samples were filled in the cell such that it was flowing over the brim, to avoid air bubbles in the cell of the calorimeter.

The STAT6 peptide was taken in exactly the same buffer as NCoA-$1_{(257-370)}$. The pH was the same in both NCoA-$1_{(257-370)}$ and STAT6 peptide solutions to avoid artifacts. Since the STAT6 peptide solution could be concentrated much farther than the NCoA-$1_{(257-370)}$, $>>$ 5mM, it was used as a titrant against the NCoA1 (257-370) in the cell. The amount of STAT6 peptide was 20 times higher than that of NCoA-$1_{(257-370)}$ in the cell, so that after the completion of the titration, there was twice as much STAT6 peptide than NCoA-$1_{(257-370)}$ in the cell.

2.3 Methods in NMR spectroscopy

In most of the experiments the sample concentration was around 0.3 mM. 50 mM PO_4, 150 mM NaCl, and 1 mM DTT, pH 6.5 was the buffer condition for all measurements, except for the cases where salt titration was done during alignment with the Pf1 phages. The temperature of measurement for the NCoA-$1_{(257-420)}$ fragment was 293 K. For all other experiments, the measurement temperature was 300 K. In the case of measurement of samples containing bicelles however the temperature for measurement was increased to 309 K. Table 2.4 shows the summary of all types of experiments measured.

2.4 NMR experiments

Over the last 55 years NMR has developed from a description of a physical phenomenon to a powerful tool to elucidate three dimensional structures of biomolecules in solution and solid state, their dynamics and the kinetics of their interaction with other

Table 2.4: Summary of the NMR spectra measured on NCoA1

Protein	*Experiment*	*Field strength in MHz*	*Temp in degree Kelvin*
NCoA-1$_{(257-420)}$	HNCA	600	293
	HNCACB	600	293
	CB(CA)CONH	600	293
	HNCO	600	293
NCoA-1$_{(257-385)}$	HNCACB	700	300
NCoA-1$_{(257-370)}$	HNCA	600	300
	HN(CA)CO	600	300
	HNCACB	600	300
	CBCA(CO)NH	600	300
	HNCO	600	300
	HNCO(TROSY)	600	300
	IPAP	600, 800, 900	300
	ECOSY-HNCO	800	300

biomolecules. This development of NMR has been possible due to three important concomitant factors : (i) Availability of very high magnetic fields up to 24 T with a spatial and temporal homogeneity of the order of 10^{-9}. (ii) Development of powerful computers that can control the radiofrequency preparation and excitation, with the added ability to vary phases and amplitudes in any conceivable way. (iii) Development of highly sensitive probes that facilitate the measurement of ample signals from samples as low as 10^{-7} M in concentration.

2.5 Theoretical Basis of NMR Phenomenon

Four parameters are mainly used in NMR based structural elucidation of proteins and nucleic acids [119] :

2.5.1 Chemical shift

Chemical shift is the property of nuclei to resonate in an external magnetic field at a frequency that depends on the chemical environment. Thus, physically identical nuclei such as the ^{1}H , resonate at different frequencies in different chemical environments . The chemical shift of nuclei also depends on the orientation of the molecule in

the magnetic field. In isotropic conditions, where there is equal probability for the molecules to take all possible orientations with respect to the magnetic field, only the orientation-independent part is observed. There is a strong relationship between α-CH ^{1}H chemical shifts and the protein secondary structure for all the 20 amino acids[120, 121]. It is a simple method which can be used in conjunction with other methods to determine protein secondary structure. The method is almost entirely empirical. Early studies by Clayden and Williams[121] suggested that strong conformation dependent tendencies exist in chemical shifts in the case of α helices and β sheets. Figure 2.1

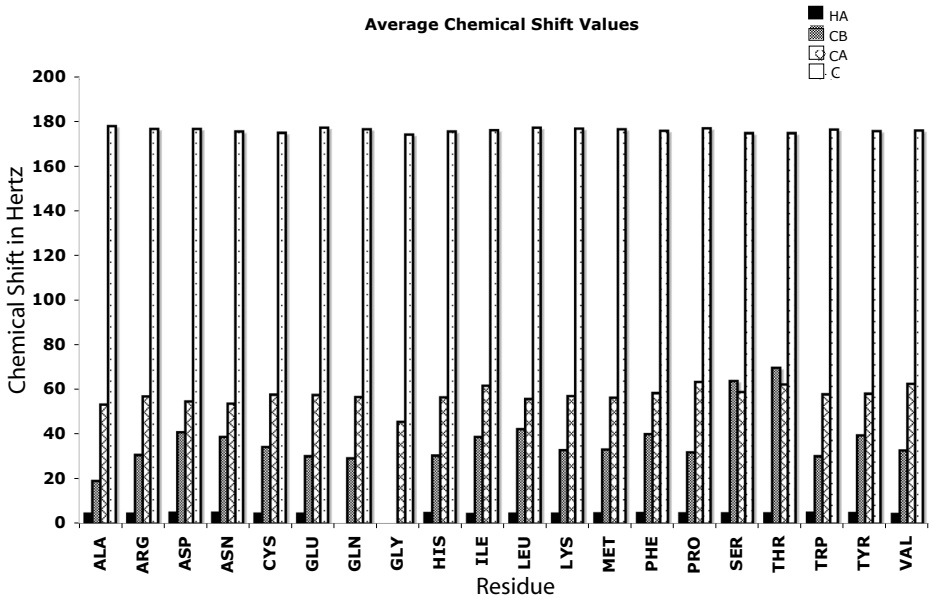

Figure 2.1: Average chemical shifts for different nuclei in proteins, from Biological Magnetic Resonance Bank (BMRB) database

shows the average chemical shift values for the commonly used chemical shifts of Cα, Cβ and Hα and CO (carbonyl carbon) values.

2.5.2 J coupling

The J coupling is a scalar quantity, which is a measure of the interaction of two nuclei. It is isotropic in nature and hence is independent of the orientation of the molecule with respect to the external field. J couplings exist between covalently bonded nuclei. A J

coupling is reflected as an internal splitting of a resonance line in the NMR spectrum. The size of the J couplings depend on the involved nuclei, as well as the nature of the bonds that connect the nuclei. When the nuclei are connected through only one bond, the J sizes are predictable. The J coupling across single bonds is an unambiguous way to infer constitution from connectivities. J couplings of nuclei connected through two bonds and three bonds are represented as ^{2}J and ^{3}J and are also called as *geminal* and *vicinal* couplings respectively. For proteins and nucleic acids with repetitive elements correlations *via* single bonds are also the established approach. Completely replacing the normal carbon nuclei with ^{13}C and the nitrogen with ^{15}N is required for this method. For example, the whole protein molecule becomes a chain of NMR- active nuclei in which each of the nuclei are connected by one bond at least to two others. The correlation information is inferred by essentially relating the chemical shift information of each nucleus within an amino acid to the amide ^{15}N and ^{1}H chemical shifts. The coupling topology is represented for a segment of a protein chain, in the Figure 2.2.

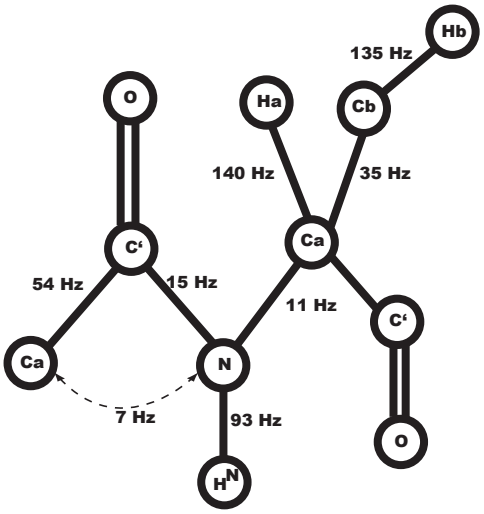

Figure 2.2: Coupling topology in polypeptides

J couplings as a structural tool

^{3}J coupling constants could be used as a direct source of information on torsional angles, by the virtue of Karplus relation which allows the calibration of Karplus curves for different molecules. There is almost no angle in biomolecules that cannot be defined by coupling constants quite accurately. The magnitude of a ^{3}J scalar coupling constant is a function of the dihedral angle, θ, formed by the three covalent bonds :

$$^3J = A\cos^2\theta + B\cos\theta + C. \tag{2.4}$$

where the constants A, B, and C depend on the particular nuclei involved in the covalent bonds[122].

2.5.3 Nuclear Overhauser Effect (NOE)

The NOE is the conventional parameter for structure elucidation. It describes the flow of z-magnetization of a nucleus to a neighbouring one with a cross-correlated relaxation rate σ. This rate depends on the distance between protons *via* the inverse 6th power, as well as the diffusional rotation correlation time of the molecule.

$$\sigma_{HH} = \frac{\gamma_H^4}{10r_{H,H}^6}(\frac{\hbar\mu_0^2}{4\pi})(-\tau_c + 6\frac{\tau_c}{1+4\omega_H^2\tau_c^2}) \tag{2.5}$$

where, γ_H is the gyromagnetic ratio of the proton, $r_{H,H}$ is the distance between the interacting protons, μ_0 is the permeability constant, and τ_c is the correlation time.

2.6 Basic Principles of Multidimensional NMR

NMR spectra of biological macromolecules contain hundreds and thousands of resonance lines which are impossible to be resolved in a 1D experiment. For a detailed analysis, accurately resolved lines are a prerequisite. Moreover, the interpretation of NMR data requires correlations between different nuclei, which are difficult to extract from 1D spectra. Multidimensional NMR spectra provide both increased resolution and correlations, which are much easier to analyze. The important step in increasing the dimensionality of NMR experiments lies in extending the spectrum from one to two

dimensions. Spectra of higher dimensionality consist of a straightforward combination of 2-D experiments.

Multidimensional NMR offers the following distinct advantages[123]: 1. Overcrowding of resonances is avoided, as the spectral information is spread out in a plane or a cube rather than along a single frequency dimension as in the 1D, 2. Correlation of pairs of resonances. A 2D spectrum, S(ω_1, ω_2) is obtained from a two-dimensional Fourier transformation of a time domain signal, *s(t_1, t_2)*, that is a function of two independent time variables, t_1 and t_2. 2D spectra are commonly presented as contour plots, where contours are drawn through the points of equal intensity, analogous to elevation lines on a geographic map.

A n-dimensional Fourier transform of S(t_1,t_2,.....t_n) provides a nD spectrum that depends on n frequency variables S(ω_1,ω_2,.....,ω_n). If two nuclei suitably interact within this mixing time this interaction will manifest as a cross peak, at a position characterized by the resonances of the interacting nuclei. In principle, it is possible to correlate additional physical parameters and to display 3D data in a cube, and 4D NMR spectra can be displayed by spreading the data in each plane of the 3D plot into a separate set of planes.

2.6.1 The basic 2D experimental procedure

In a typical 2D experiment, four time intervals can be distinguished[124]. 1. A preparation period (τ_p) 2. An evolution period (t_1) 3. A mixing period (τ_M) 4. Detection period (t_2). In many experiments, during the preparation period, the nuclear spin system is allowed to come to equilibrium, while in some others, elaborate pulse sequences may be used to ”prepare” the system for the rest of the 2D experiment. The final step in the preparation period causes the magnetization **M** to be rotated to the *xy* plane, by a simple 90° pulse, and occasionally by a more complex pulse sequence. During t_1, the magnetization evolves freely, so that M_{xy} precesses at its Larmor frequency, and each nuclear magnetization is ”tagged” according to its Larmor frequency, which depends on the particular Hamiltonian that is operative in the spin system. The mixing period τ_M gives nuclear magnetizations in the *xy* plane a chance to interact or ”mix” their wavefunctions. Finally, t_2 is the usual data acquisition period in which an FID is acquired, just as in an one dimensional experiment. This procedure is repeated a large number of times, with different durations of the evolution period,

keeping all other things a constant. For each value of t_1, the signal recorded during t_2 is digitized and stored in computer memory. Upon completion of the entire 2D experiment, a data matrix representing the two-dimensional time domain signal *s(t_1, t_2)* is available in the computer memory. Two Fourier transforms, with respect to t_1 and t_2 plus other data processing similar to that in one-dimensional NMR then provide the desired 2D spectrum. A large number of 2D and multidimensional NMR techniques are available and new variations are continuously being developed and improved. A summary of the commonly used 2D experiments is provided.

2.6.2 Heteronuclear Single Quantum Coherence(HSQC)

In the heteronuclear single quantum coherence (HSQC) experiment[125], the INEPT[3] (Insensitive Nuclei Enhanced by Polarization Transfer) sequence is used to transfer I spin polarization into antiphase heteronuclear single quantum coherence. The antiphase heteronuclear coherence evolves during the subsequent t_1 evolution period. A second INEPT sequence is used to transfer the frequency-labeled heteronuclear coherence back to proton magnetization for detection. This is one of the most sensitive of the 2D experiments.

2.6.3 Measurement of J and D couplings

IPAP-HSQC (In Phase Anti Phase HSQC)

In this method[126] two spectra are recorded in an interleaved manner, one with in-phase doublets and another with anti-phase doublets. Upon addition of the two spectra, a $^{1}H^{N}$ -^{15}N correlation spectrum showing only the upfield or downfield component of the ^{15}N doublet is generated. There are lesser resonance overlaps relative that of the decoupled HSQC spectrum.

2.6.4 Novel sensitivity enhanced experiments

In addition to the large one bond ^{15}N -^{1}H dipolar couplings, backbone one bond ^{15}N -^{13}CO and two bond $^{1}H^{N}$ -^{13}CO dipolar couplings play vital roles in protein NMR, inspite

[3]A basic unit of a complex NMR pulse sequence; explanation beyond the scope of the thesis. A very useful scheme for obtaining better S/N ratio with samples of very low concentration

of their small values. In order to measure the one-bond NCO couplings and the more difficult to obtain, two bond HCO couplings from NCoA1, a recently developed E-COSY-HSQC type experiment [127] was used. It is difficult to implement J-correlation type experiments for the measurement of two bond $^{1}H^{N}$ -^{13}CO couplings. In a classical E-COSY experiment, both couplings can be measured in a single spectrum, but the E.COSY introduced splitting causes a decrease in the resolution and the sensitivity of the spectrum. In the novel E-COSY type HSQC experiment, one bond ^{15}N -^{13}CO and two bond$^{1}H^{N}$ -^{13}CO couplings are obtained from two spectra by a complex combination of splittings. Manipulation of the E-COSY split pair in a echo-antiecho fashion correlates the frequencies of the low field and high field peaks of the E.COSY doublets with the proton and nitrogen frequencies. This manipulation results in twice the number of scans leading to a 1.4 times increase in sensitivity.

2.6.5 Three Dimensional NMR

A 3D pulse sequence can be obtained from two separate 2D experiments by deleting the detection period, of one experiment and the preparation period of the other, to obtain two evolution periods (t_1 and t_2), and one detection period t_3. Conventional 3D assignment is by NOE connectivities between adjacent residues involving NH, $C^{\alpha}H$ and $C^{\beta}H$ protons. The same approach can be used in 3D NMR.

Sequential Assignment Via One-Bond J Couplings

Triple resonance experiments can be used to sequentially assign the backbone NH, ^{15}N , $^{13}C^{\alpha}$,^{13}CO and $C^{\alpha}H$ resonances by means of heteronuclear one-bond couplings without any knowledge of the spin systems involved. This involves the measurement of 6 different kinds of spectra namely the HNCA, HN(CA)CO, HNCO, HN(CO)CA, HNCACB, CBCA(CO)NH. The nomenclature for triple-resonance experiments is systematic. The frequency labeled spins during acquisition in the indirect evolution periods or the acquisition periods are listed using HN, N, HA, CA, CO, HB, and CB to represent the ^{1}H , ^{15}N , $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, ^{13}CO , $^{1}H^{\beta}$ and $^{13}C^{\beta}$ spins respectively.

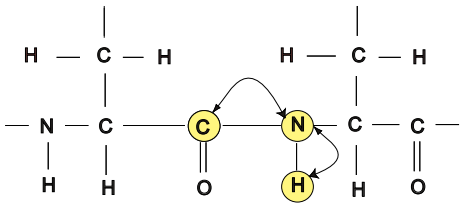

Figure 2.3: The HNCO experiment

HNCO

The 3D HNCO experiment[128] correlates the NH and ^{15}N chemical shifts of residue *i* with the ^{13}CO shift of the previous residue through the one-bond $^{1}J_{NCO}$ coupling of 15 Hz, thereby providing sequential connectivity information. In this experiment, magnetization originating from the amide protons is transferred to the directly bonded ^{15}N spins using an INEPT sequence, following which the ^{15}N magnetization evolves exclusively under the influence of the ^{15}N as a result of ^{1}H -^{13}CO and $^{13}C^{\alpha}$ decoupling through the application of the 180° pulses at the midpoint of t_1 interval. During the delay δ ^{15}N magnetization becomes antiphase with respect to the polarization of the carbonyl spin of the preceding residue via the $^{1}J_{NCO}$ coupling and the subsequent 90° ^{13}CO pulse converts this magnetization into ^{15}N - ^{13}CO two spin coherence. Evolution of the ^{13}CO chemical shift then occurs during the period t_2, and the contributions of ^{15}N chemical shift and ^{1}H -^{15}N ^{1}J coupling on the one hand, and of the $^{13}C^{\alpha}$ -^{13}CO ^{1}J coupling on the other, are removed by the application of 180° ^{15}N and $^{13}C^{\alpha}$ pulses, respectively, at the t_2 midpoint. Magnetization is finally transferred to back to the NH protons by reversing the transfer steps and detected during t_3.

HNCA

The HNCA experiment[128] correlates the ^{1}H and ^{15}N chemical shifts with the $^{13}C^{\alpha}$ of the same residue. The one bond ^{15}N -$^{13}C^{\alpha}$ J coupling is exploited to correlate ^{15}N and $^{13}C^{\alpha}$ spins. The magnetization is also transferred from the ^{15}N spins to the $^{13}C^{\alpha}$ of the preceding residue through the interresidue two bond ^{15}N -$^{13}C^{\alpha}$ J coupling, which can be as large as 9 Hz[128, 129].

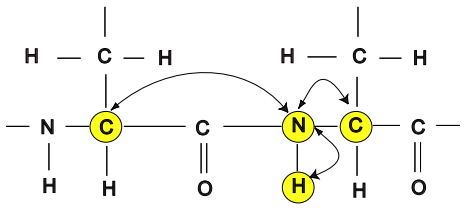

Figure 2.4: The HNCA experiment

HN(CA)CO

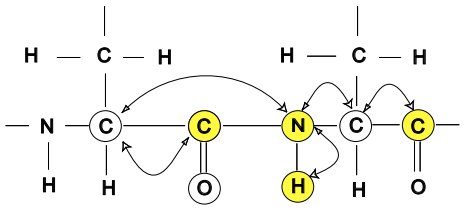

Figure 2.5: The HNCACO experiment

The HN(CA)CO experiment[130] provides correlations between the amide ^{1}H , ^{15}N and ^{13}CO chemical shifts by using the one-bond ^{15}N -^{13}CO J couplings for coherence transfer[130]. This experiment provides sequential connectivities from the ^{15}N spins to the ^{13}CO of the previous amino acid through the interresidue two-bond ^{15}N - $^{13}C^{\alpha}$ J coupling.

HNCACB

The HNCACB experiment[129] correlates the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances with the amide ^{1}H and ^{15}N resonances of the same residue and the amide ^{1}H and ^{15}N resonances of the succeeding residue through $^{1}J_{CN}$ and $^{2}J_{CN}$ couplings, respectively. In addition to the sequential connectivities the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts provide information on the amino acid type. The $^{13}C^{\beta}$ is not present in the strip containing the Glycine and its subsequent residue as the Glycine does not have a $^{13}C^{\beta}$ spin. The HNCAB spectrum is very helpful in resolving ambiguities, as it provides sequential resonance information

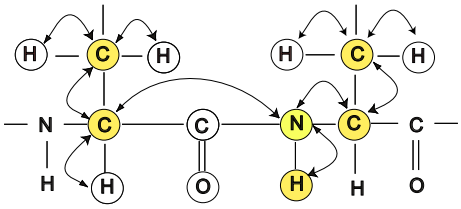

Figure 2.6: The HNCACB experiment

using two possible sequential connectivities. Some residues are very conspicuous in a HNCACB spectrum as they have characteristic resonances.

CBCA(CO)NH

The CBCA(CO)NH experiment[129] correlates both the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances of an amino acid residue with the amide ^{1}H and ^{15}N resonances of the following residue. These correlations are highly useful if significant $^{13}C^{\alpha}$ -$^{1}H^{\alpha}$ chemical shift degeneracy exists. The interresidue correlations are established by transferring coherence through the intervening ^{13}CO spin.

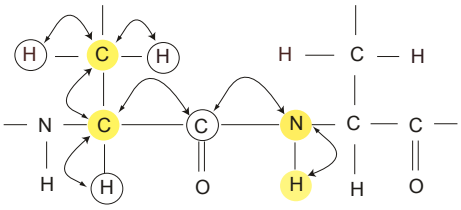

Figure 2.7: The CBCACONH experiment

In case of a $^{13}C^{\alpha}$ - $^{1}H^{\alpha}$ degeneracy, this correlation is useful. $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances provide information on the type of amino acid preceding the amide. The interresidue correlations are established by transferring the coherence through the ^{13}CO spin.

Chemical Shift Index

The chemical shift index calculation is a simple method of estimating the secondary structure, based on NMR resonance assignment data[120]. Typically,

$$CSI = \delta CB - \delta CA, \tag{2.6}$$

where δ CA = C$\alpha_{randomcoil}$ - C$\alpha_{expt.}$ and δ CB = C$\beta_{randomcoil}$ - C$\beta_{expt.}$.

HNCO-TROSY

Two main problems limit the study the large molecules using NMR. First, the large number of signals that arise causes increased resonance overlap, which makes the intrepretation of spectra very difficult. Second, large molecules relax faster, which leads to increased line broadening and subsequently poor resolution of the spectra. In theory, the resonance overlap problems can be reduced by a suitable isotope labeling scheme [131], but the limitation due to faster relaxation is a more severe challenge. The hydrogen atoms, which constitute a major source of relaxation also provide the most sensitive NMR signal and therefore valuable structural information. Therefore, their replacement by deuterium ions to substantially reduce transverse relaxation, is an option. Even this fails when the size of the molecule extends beyond 50 kDa. Transverse Relaxation Optimized Spectroscopy (TROSY)[132], uses spectroscopic means to reduce relaxation, and is suitable for large molecules as it reduces the relaxation to such an extent that satisfactory linewidths could be obtained and sensitivity enhanced.

Theory of TROSY

^{1}H spins are scalar coupled to ^{15}N nuclei which gives rise to a proton spectrum where there are two peaks - one from those protons which are spin up with respect to ^{15}N and another from those protons which are in spin down status. In a large molecule, the two lines have very different line widths. The two lines are collapsed by decoupling in conventional NMR, with the compromise of averaging the relaxation rates. The signal will be much more attenuated in the case of larger molecules because one of the peaks will arise due to faster relaxation. The TROSY technique is used to exclusively select the slower relaxing peak neglecting the faster relaxing peak. The loss of half of the

potential signal is compensated for, during acquisition.

2.6.6 Sequential Resonance Assignment

All spectra were processed by NMRPipe software from the NIH[133]. The processed spectra were analyzed using Felix (Accelrys) or Sparky (Computer Graphics Lab, University of San Francisco) NMR data analysis programs. Analysis consisted of initial peak picking and cleaning noise from the peak lists of different spectra. The peak list was verified and was used for manual assignment of the resonances. Sequential resonance assignment was also performed using the program MARS developed in the department. Resonance assignment results were output by the program in a Sparky compatible format. The output was once again read into Sparky program and cross checked.

2.7 Residual Dipolar Couplings(RDCs)

Dipolar couplings are the dominant interactions in solid state NMR of spin nuclei, they are averaged to zero for isotropically reorienting molecules in the liquid state[134, 135]. This makes it possible to achieve high resolution in liquid state NMR[136]. On the other hand, a wealth of structural information is lost, if the couplings disappear[137, 135]. Even in liquid state NMR, molecules can be weakly aligned, e.g., by external fields, or by anisotropic solvents [138, 139, 140, 141, 134, 135]. Dipolar couplings are potentially large interactions, caused by the magnetic flux lines of one nucleus affecting the magnetic field at the site of another nucleus[135].

In the Figure 2.8, nuclei P and Q are part of a molecule in solution and $\vec{r}$ is the internuclear distance separating them. In the lab frame the vector representing the magnetic field, $\vec{B_0}$, is constant whereas the internuclear vector is time dependent as a result of molecular tumbling. For simplicity it is assumed that the molecule is rigid and hence there is no internal dynamics. Therefore, the magnitude of r is also a constant. The component parallel to the magnetic field $\vec{B_0}$ is the only component of concern, since the components perpendicular to the $\vec{B_0}$ magnetic field have a negligible effect on the total magnitude of the vector sum of the external and the dipolar field. The z component of the dipolar magnetic field of the nucleus P will change the resonance

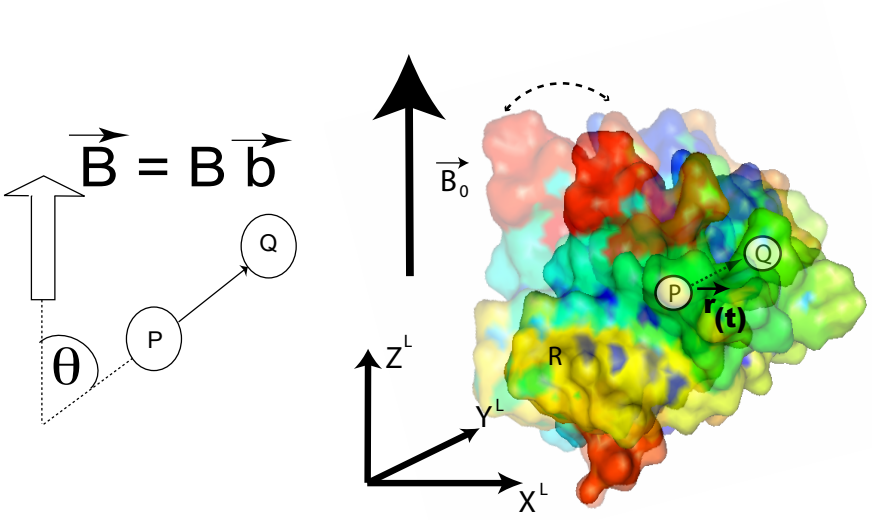

Figure 2.8: In the lab frame, the magnetic field vector $\vec{B}$ is constant (pointing along the z^L axis) but the internuclear vector is time dependent

frequency of nucleus Q by an amount that is dependent on the internuclear distance and on the orientation of the internuclear vector relative to $\vec{B_0}$. For a fixed orientation of the vector, which is parallel to $\vec{B_0}$, nuclear spin P can increase or decrease the total magnetic field at nucleus Q, depending on whether P is parallel or anti-parallel at $\vec{B_0}$. In a molecular ensemble, half of the nuclei will be parallel to $\vec{B_0}$, and other half will be anti-parallel. Spin Q will show two resonances (doublet), separated in frequency by

$$D^{PQ} = D^{PQ}_{max}\langle(3cos^2\theta - 1)/2\rangle \tag{2.7}$$

where θ is the angle between the internuclear vector and B_0, and the $\langle\rangle$ brackets denote ensemble averaging over time. Equation 2.7 shows that the dipolar splitting, D^{PQ} provides direct information on the angle θ, i.e., on the orientation of the internuclear vector, and that it is proportional to the inverse of the cubed internuclear distance. Under isotropic conditions, rotational Brownian diffusion rapidly averages the internuclear dipolar interaction to exactly zero. As a result, the solution NMR spectrum shows narrow resonances, which can be assigned to individual nuclei in the protein, but which no longer contain valuable orientation information. In solids, such an averaging does not take place, and each nucleus couples with a very large number of other nuclei

(each of which can point parallel or anti-parallel to the magnetic field), resulting in unresolvable, very broad resonances. Mechanical rotation of the sample around an axis that makes an angle of 54.7° with the B_0 magnetic field, together with a radiofrequency irradiation, can be used to average the dipolar interaction to zero and reintroduce sharp features in the NMR spectrum [142]. In an intermediate case where, the protein is dissolved in a slightly anisotropic aqueous medium, where not all orientations of the protein have an equal chance of occurrence, the alignment of the protein can be described by an alignment tensor, '**A**'. This tensor is a real symmetric matrix, which can be diagonalized. In the corresponding molecular frame for the tensor 'A', its principal components A'$_{xx}$,A'$_{yy}$, A'$_{zz}$ reflect the probabilities for the x,y, and z axes to be parallel to B_0(Figure 2.9(a) and (b)). The difference among A'$_{xx}$,A'$_{yy}$, and A'$_{zz}$ contributes to dipolar coupling. Therefore the alignment matrix is used in its traceless form, **A**. When A_{zz}, A_{yy}, and A_{xx} are defined, the dipolar coupling depends on the polar coordinates of the P-Q vector in the frame of the diagonalized alignment tensor in the following way :

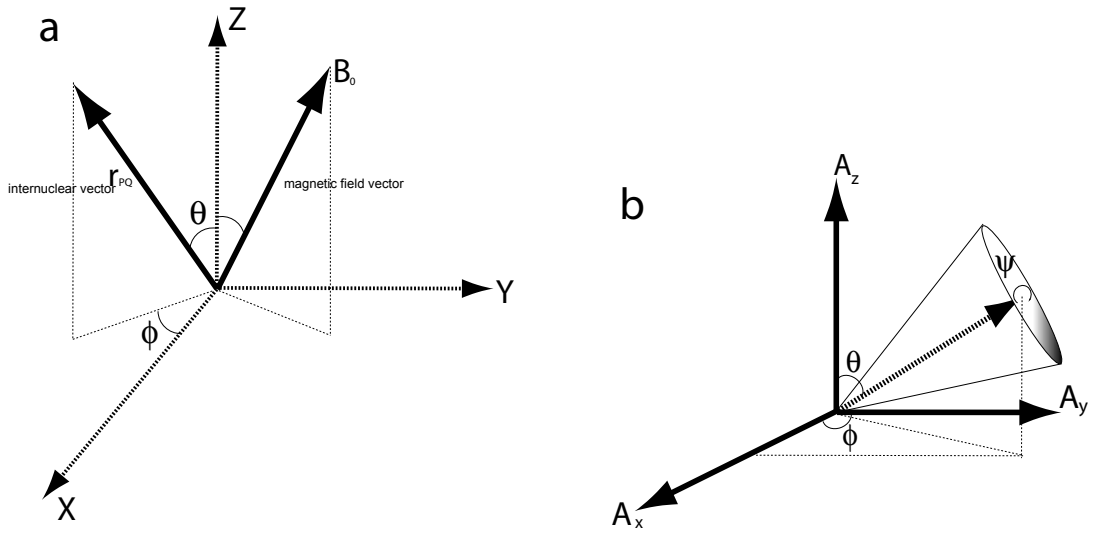

Figure 2.9: (a) Depiction of the orientation of the internuclear vector and the magnetic field relative to a molecule fixed frame. The time dependance of the magnetic field vector arises due to molecular reorientation. (b) The three angles θ, ϕ, ψ called the Euler angles describe the overall molecular alignment (A_x, A_y, A_z)relative to a principal axis system.

$$D^{PQ}(\theta,\phi) = 3/4 D^{PQ}_{max}[(3\cos^2\theta - 1)A_{zz} + \sin^2\theta\cos 2\phi(A_{xx} - A_{yy})] \tag{2.8}$$

where, θ

This equation can also be written as,

$$D^{PQ}(\theta,\phi) = D_a[(3\cos^2\theta - 1) + 3/2R\sin^2\theta\cos 2\phi] \tag{2.9}$$

where $D_a = 3/4D^{PQ}_{max}A_{zz}$ is referred to as the magnitude of the dipolar coupling tensor, usually normalized to the N-H dipolar interaction, and $R = 2/3(A_{xx} - A_{yy})$ is called the rhombicity.

Equation 6.3 indicates that for a given value of D^{PQ} there can be an entire cone of (θ,ϕ) solutions that correspond to this dipolar coupling. So, the dipolar coupling does not define an unique orientation for the internuclear vector but restricts it to be on the surface of a distorted cone. The inverted vector orientation QP also gives rise to the same coupling, and an inverted cone is also included in to the solutions to equation 6.3.

2.8 Alignment in liquid crystals

Intermolecular complexes, mutidomain proteins and nucleic acids that are more difficult to study than globular proteins, because of comparatively low intensities of meaningful Nuclear Overhauser Effects (NOEs), are now far more accessible to experimental characterization in solution. The measurement of relatively small sets of RDCs can allow pH and/or temperature dependent shifts in conformation to be detected, binding interfaces to be characterized and rapid establishment of relative orientation in solution [143]. The wide timescale sensitivity of RDCs to molecular motions ($<10^{-3}$ seconds) allows experimental accessibility, at atomic resolution, to functionally important motions occurring in the nanosecond to microsecond time scale. Apart from the necessity of finding an alignment medium that is compatible with the molecule under relevant solution conditions, the nature of molecular alignment can vary according to the physical properties of the medium used. In the past few years, there is a rapid development in RDC techniques to quickly establish the average solution state conformations of intermolecular complexes or the relative orientation of domains. RDCs can be used to probe the conformations of these systems in solution without carrying out a complete structure determination, in the assumption that the distinct structural elements estimated *a priori* are rigid entities. NMR plays a complementary

role to crystallography, as the RDC constraints are used to refine global arrangement of these structural elements, which originate mostly from crystallographic studies. Several procedures have been described in order to make such a refinement, such as order tensor analysis [144] or rigid body minimization [145] but they all have a fundamental similarity in that the orientations of the partners involved are optimized by a best fit to the RDC data. A whole array of liquid crystalline solvent systems useful for inducing molecular alignment of macromolecules have been introduced. The rapid translational diffusion eliminates intermolecular dipolar interactions, but large intermolecular dipolar couplings can be seen.

2.9 Application to structure calculation and structure validation

The quality factor Q[146] is a convenient criterion to validate the compatibility between structures and RDCs measured in solution.

$$Q = rms(D^{calc} - D^{obs})/rms(D^{obs}) \tag{2.10}$$

where D^{obs} and D^{calc} are observed and calculated one-bond dipolar couplings. It is important that the couplings used in the evaluation are not used in the structure calculation process, which would make the validation meaningless. In order to predict the dipolar couplings using equation 6.3, the magnitude and orientation of the alignment tensor need to be known. There are two ways to obtain this : 1. It can be predicted from the molecular shape[147] 2. Using singular value decomposition[144], the alignment tensor that best fits with the experimental data can be found. There are five independent parameters in such a fit, and it is hence important in Q factor evaluations that the number of fitted dipolar couplings is very much larger than five.

Structure determination based on existing structures

The conventional method of structure determination by NMR relies on short-range interproton distances (NOEs). When applied to larger proteins or more complex systems this method encounters severe limitations. RDC data have been used to

determine relative orientation of multidomain macromolecules using rigid body modeling assuming a common alignment tensor and have been shown to decrease the disorder present in NMR structural ensembles when combined with NOEs. RDC measurements have been successfully used to predict the backbone conformation of protein in solution. The molecular fragment replacement (MFR) method [148] is analogous to a method proposed by Kraulis and Jones for determining local structures of protein fragments and is similar to the database approach for fitting the main chain electron density of protein X - ray structures. This method has been demonstrated in ubiquitin. Four backbone couplings namely the N-H, the C'-N, the C'-H^N and C^α - H^α were measured. A fragment size of 7 residues was used and for each fragment the best fit between its set of measured dipolar couplings and each 7 residue fragment found in the Brookhaven Protein Data Bank (PDB) is determined by using a linear least squares method.

2.10 Alignment of NCoA-1$_{(257-370)}$ Using Pf1 Phages

Much effort has been directed towards development of alignment media that are stable over a wide range of pH and temperature and that have a minimal adverse effect on rotational diffusion. [149]. Availability of multiple media is useful as it allows for measurement of residual dipolar couplings in different alignment frames [150, 151]. The combination of disc shaped bicelles [152] and rod-shaped virus particles is particularly useful in this regard [153]. Two different mechanisms of interaction, steric for neutral bicelles and mainly electrostatic for phage, generally result in linearly independent alignment tensors [154]. But, bicelle liquid crystalline phase is unstable in the presence of certain proteins and offers only a limited temperature range [126]. Phage particles are highly negatively charged and can bind non-specifically to macromolecules that have a significant positively charged surface patch. Pf1 phage from Profos Biolabs (Munich) was used for alignment. The phage was added to the wall of the NMR sample tube with the sample. The top of the tube was closed using parafilm and the tube was inverted up and down many times, so that phage and the protein sample mixed well. The protein sample with the phages was centrifuged slightly to remove the air bubbles.

2.11 Alignment of NCoA1$_{(257-370)}$ Using Bicelles

A few years ago, application of liquid crystal NMR techniques were extended to some larger, biologically relevant molecules with the aid of isotopic labeling and a medium comprised of an aqueous dispersion of lipid bilayer disks (most commonly a mixture of dimyrystoylphosphatidylcholine and dihexanoylphosphatidylcholine, DMPC and DHPC) [155]. However, the use of bicelle solutions to orient different molecules is not problem free. At low concentrations of solutes the samples often phase separate or become isotropic in times short compared to normal data acquisition times. Instabilities seem to be aggravated by the presence of buffers and electrolytes needed to maintain native structures of biomolecules [156]. Addition of small amounts of charged amphiphiles to normal bicelle preparation can avoid some of these instabilities and extend the range over which these preparations maintain a useful degree of orientation and homogeneity [157]. Water in the medium provides a probe of orientation if some water protons are replaced with deuteriums. Deuterium has a quadrupolar nucleus that interacts with both the laboratory magnetic field and the local electrical field gradient. The electric field gradient interactions add an anisotropic perturbation that splits the normally degenerate resonances of a deuterium nucleus. Under isotropic tumbling, these splittings average to zero. In the presence of bicelles, a small quadrupolar coupling results from the fast exchange of bulk solvent, that is isotropic and the solvent associated with the bilayer, which is partially oriented. Spectra as a function of temperature were measured. The starting temperature was 25 °C and it was raised by two degrees at each step. The sample was allowed to equilibrate at the desired temperature for at least 10 minutes prior to acquisition of the spectra. A DMPC:DHPC::3.2:1 ratio was used at 36°C. The sample was allowed to equilibrate at 306 K for 10 minutes.

2.11.1 Preparation of bicelles

A premixed powder containing DMPC and DHPC in the ratio 3.2:1 was obtained from *Avanti Polar Lipids, Alabama*. A 15% w/v solution was made using NMR buffer at pH 6.5. The lipid mixture was hydrated by many cycles of freezing and thawing on ice and 38° C respectively, with vigorous vortexing. The sample was incubated on ice for 10 minutes and then vortexed at room temperature, heated at 38° C for 10 minutes

and vortexed at room temperature. This cycle was repeated six times until the solution reached a stage where it was a clear transparent fluid at lower temperature, milky at about 25°C and transparent but viscous at 38°C. There was a slight bluish tint. At this stage, 2 times protein solution was mixed with 1 times bicelle solution to get a 5 % w/v bicelle sample for NMR. The protein solution contained 15% D_2O so that the final content of D_2O would be 10% of the total volume. 200 μl of the sample that contained 5 % bicelle solution, 15% D_2O was added to a Shigemi (*Shigemi Corporation, Tokyo*). The probe was preheated to 308 K and the splitting of D_2O was observed in a few minutes. At 308 K, a splitting of 7 Hertz was observed. A HSQC spectrum was measured at this temperature. A comparison with the HSQC spectrum measured at 300 K (the temperature at which all other types of experiments were measured), did not show any shift of peak position due to a shift in temperature. However, the signal from the modified HNCO-TROSY[158] type experiment failed to show a S/N ratio comparable to the HSQC. It was assumed that this difference in signal could be because of the interaction between the protein and the bicelles. After a few hours, when the sample was removed out of the magnet, there was phase separation inside the NMR tube, where the bicelles settled as white fluffy mass in the lower part of the NMR tube. This condition was reversible upon vigorous remixing, but the behaviour repeated in a few hours again. This condition is not suitable for measurement of spectra over prolonged periods of time. Therefore a stabilizing agent Cetylpyridinium Tetraethyl Ammonium Bromide (CTAB) was added at 0.7 mM final concentration.

2.12 Alignment using polyethylene glycol/hexanol based medium

A non-ionic liquid crystalline medium, [149] was used for alignment. This medium consists of an n-alkyl poly(ethylene glycol)/n-alkyl alcohol mixture and is known as the Otting Phase. $C_{12}E_5$ is the polyethylene glycol component and hexanol is the n-alkyl alcohol [159, 160, 161]. This medium belongs to a class of liquid crystals which are uncharged, insensitive to pH, fairly insensitive to salt, feature little or no binding capacity to macromolecules, and can be used at temperatures below 0° C up to 40° C, covering the temperature range of interest for most NMR studies of macromolecules. Under certain conditions, this system forms a lyotropic liquid crystalline phase, where

the hydrophobic n-alkyl chains aggregate into planar bilayers with the hydrophilic polyethylene glycol headgroups directed into the water phase. This phase, called the Lα phase, is optically clear, with slight opalescence, has a lamellar-like super structure. The spacing of the stacked bilayers and hence the arrangement of protein molecules in the water rich layers can be tuned by adjusting the surfactant to water ratio. The spacing can range from several hundred nanometers (1 wt %) to a few nanometers in the more concentrated regime (>5 wt%). In the magnetic field, the bilayer surfaces are oriented parallel with respect to the direction of the magnetic field [155] possibly with a superstructure in which the bilayers bend into a set of concentric tubes with the axis aligned along the field. At high surfactant concentrations (>50 %) and temperature (>50 $^\circ$ C), the liquid crystalline phase is already formed by binary mixture with alcohol. The addition of n-alkyl alcohols lowers the temperature range of stability for the liquid crystalline phase and makes the lyotropic phase accessible even at low surfactant concentrations [159].

2.12.1 Preparation of the Otting Phase

A 5$\times$ stock solution (25 % w/v of the polyethylene glycol $C_{12}E_5$) of the Otting Phase was made by taking 130 μl of $C_{12}E_5$ in a clean Eppendorf tube and votesting well with 370 μl of 90 % H_2O/10 % D_2O solution. The resulting mixture was clear and viscous. To this, hexanol was added in microlitre steps. After addition of 15 μl of hexanol, the solution turned turbid and the addition of hexanol was continued in one micro-liter steps. When the total amount of hexanol reached 18 μl the solution turned clear with a bluish tint. Two more mico-liters of hexanol were added to this in order to compensate for the loss of hexanol by evaporation. The final ratio of $C_{12}E_5$/ hexanol was 1:0.96 (w/w). The phase remained in this state, even after vigorous vortexing and centrifugation. This stock solution was diluted to 5% $C_{12}E_5$ by adding protein solution containing 10% D_2O. There was a slight disturbance in the phase, but the mixture became clear by addition of just one μl of hexanol. But this was not stable for more than 10 minutes as a white precipitate was observed after this time.

2.13 Polyacrylamide Gels in Alignment

Polyacrylamide gels provide an alternative approach to the preparation of weakly aligned samples of membrane proteins. This method is called strain induced alignment in gel (SAG)[162, 163]. The method is based on the fact that either radial or vertical compression of the gel alters the pore shape and induces preferential alignment of the protein[164]. Polyacrylamide gels are chemically inert with respect to the solute proteins, the samples are stable over a wide range of temperature, ionic strength and pH and even allow the study of proteins in partially or fully denatured conditions [165]. Several gels were cast using the apparatus as mentioned in a previous work[164]. The polymerized gels were washed several times using double distilled H_2O, cut into rods of 2 mm length and dried. Part of the drying was done in a desiccator and the rest was carried out by storing the gels overnight at 8 °C. A completely dried gel without a mechanical flaw was chosen and transferred to a shigemi tube. Five hundred microlitres of protein solution of 0.4 mM concentration was added to the gel and closed with the plunger. The protein was allowed to diffuse into the gel overnight. The swollen gel was measured for D_2O splitting. There was a splitting of 10 Hz, but the proton spectrum showed very broad lines from the protein. A HSQC spectrum of the sample showed only a few peaks from the backbone amide residues and the side chain peaks.

2.14 Alignment trials using paramagnetic tags

Protein molecules which are intrinsically diamagnetic do not show enough alignment in a strong magnetic field, and it was one of the reasons for the introduction of external alignment media. However, paramagnetic molecules do not require external alignment media[141]. Structural restraints such as paramagnetic relaxation as well pseudocontact shifts can also be obtained from such molecules in addition to dipolar couplings when they are placed in high magnetic fields. Diamagnetic proteins can be converted to paramagnetic proteins by attaching a paramagnetic tag, which is a small organic molecule that specifically binds a lanthanide [166]. Certain derivatives of EDTA can be used for this purpose, as they have sub-picomolar dissociation constants, for lanthanides in neutral pH. The tag was added in 3 times in molar excess to 0.1 mM of NCoA-$1_{(257-370)}$ in 50 mM MOPS buffer, pH 6.5, 150 mM NaCl and incubated overnight

at room temperature.

2.15 Titration experiments

Chemical shift perturbation is the most widely used NMR method to map protein interfaces. The ^{15}N-HSQC spectrum of one protein is monitored when the unlabeled interaction partner is titrated in, and the perturbations of the chemical shifts are recorded. The interaction causes environmental changes on the surface of the protein interfaces, and therefore the chemical shifts of the nuclei in this area. In some cases, the entire protein may shift conformation and all the shifts may be affected. Then the chemical shift perturbation fails as a mapping technique, but is a very good indicator of allosteric processes [167, 168]. Chemical shift perturbations are sensitive to very subtle effects. To quantify the shift perturbation, most workers use the length of the vector that connects the two-dimensional end points, normalized by typical shift ranges [169]. The precise values of these normalization factors are not important, because of the complicated relationship between structure and chemical shift [170]. The results are significant when a contiguous surface patch of shifting resonances are obtained. Titrations allow a good estimation of affinity, specificity of binding as well as the kinetics of binding. The kinetics of the interaction determine how the chemical shifts of the labeled protein change during the titration. If the dissociation of the complex is very fast, there is only a single set of resonances even during the titration. The chemical shifts are the weighted average of the chemical shifts of the free and bound form. The resonances of the nuclei at the interface move in a continuous manner during the titration. This regime is referred to as fast chemical exchange and is often observed for weaker interactions. If the complex dissociation is very slow, there are two sets of resonances, one for the bound form of the protein and the other for the free form. The set of resonances corresponding to the free form will disappear during titration and will be replaced by those from the bound form. Most of the resonances of the two sets will overlap with each other but the differences will mark the interaction interface. In the case of slow exchange it is not known to which new location the resonance has moved unless there is an independent assignment of the bound form. Therefore it is not easy to quantitate the degree of change. Some workers have approached this problem by assuming that the new resonance which appears closest to the one from the free

form corresponds to bound state [171, 172]. To account for the intrinsic differences in chemical shift dispersion of ^{1}HN, ^{15}N, and ^{13}C nuclei, was defined thus:

$$\delta = (\Delta\delta 1HN/3.5) + (\Delta\delta/30) \tag{2.11}$$

Titration experiments were performed with all the three constructs of the NCoA-1 namely the NCoA-1$_{(257-420)}$, the NCoA-1$_{(257-385)}$ and the NCoA-1$_{(257-370)}$ fragments with the STAT6$_{(794-814)}$ peptide. In the case of NCoA-1$_{(257-420)}$, STAT6$_{(794-814)}$ peptide was added to the protein sample such that the molar ratio of the protein to the STAT6$_{(794-814)}$ peptide was 1:1. With the NCoA-1$_{(257-385)}$ fragment, the peptide was titrated such that molar ratio of the protein to the peptide was 5:1, 2:1, 1:1 and 1.5:1. STAT6$_{(794-814)}$ peptide was titrated into the NCoA-1$_{(257-370)}$ sample such that molar ratio of the protein to the peptide was 1:1, 1:2, 1:3 and 1:4. The conditions of measurement were maintained constant throughout the titration process.

Chapter 3

Results

3.1 Results from Biochemical and Biophysical Experiments

3.1.1 Protein Expression and Purification

The expressions were carried out in different media, namely the Luria Bertani Medium (LB), minimal medium with M9 salts, and a commercial medium, *Silantes*. Table 3.1 gives the yield from different media. The yield of NCoA-$1_{(257-420)}$ in LB medium was around 12 mg, after the final step as calculated from UV_{280nm} absorption from one litre of the medium. In order to make an isotope labeled sample, however, the bacteria were grown in minimal medium, with M9 salts (see Table 2.1). But the yield in this medium was negligible when compared to that from LB medium. Since the loss during concentration of the protein has to be taken into account, sufficient quantity of protein is required before concentration. Therefore, the expression level was checked in a rich, isotope labeled medium, Silantes (O.D. 2)®, marketed by Silantes GmbH, Munich, Germany. As seen from the Figure 3.1, the solubility of the NCoA-$1_{(257-420)}$ when expressed using this medium was comparable to the results from LB medium. The reason for poor solubility in minimal medium expression is not clear.

Figure 3.1 shows the expression profile of the NCoA-$1_{(257-420)}$ fragment in *Silantes* medium. Figure 3.3(a) shows the elution profile of NCoA-$1_{(257-370)}$ after the histidine tag was removed. The level of purity of the protein samples is seen in the gel. It is the purity level of samples used for NMR measurements. Figure 3.3(b) shows the

Type of Medium	LB	Minimal medium	Silantes
final protein yield (mg/liter)	12	1	14

Table 3.1: Comparison of yield of NCoA-$1_{(257-420)}$from expressions in different media

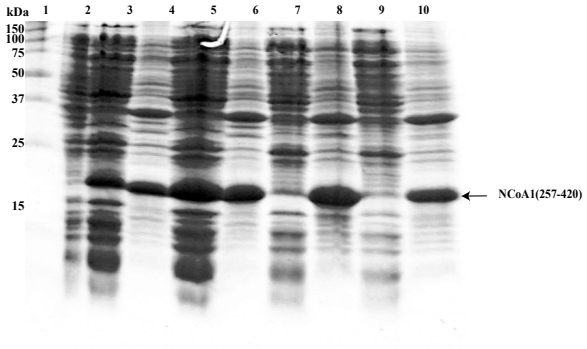

Figure 3.1: A comparison of the expression of NCoA-$1_{(257-420)}$ in different media. 1. Broad range protein marker. 2. Control - cell extract from uninduced cells (45 μl of the cells were taken from the uninduced culture and boiled with 15 μl of 4 $\times$ SDS buffer, out of which 10 μl was loaded. 3. cell extract from cells 6 hours after induction. 4 and 5. insoluble fraction and soluble fractions of the NCoA-$1_{(257-420)}$ from 22 °C induction in minimal medium with 1 $\times$ m9 salts. 6 and 7. insoluble fraction and soluble fractions of the NCoA-$1_{(257-420)}$ from 28 °C induction in minimal medium with 1 $\times$ m9 salts. 7 and 8. insoluble fraction and soluble fractions of the NCoA-$1_{(257-420)}$ from 28 °C induction in LB medium. 9 and 10. insoluble fraction and soluble fractions of the NCoA-$1_{(257-420)}$ from 28 °C induction in Silantes medium. Note that the protein is completely soluble when LB and Silantes were used, and the expression levels are comparable. Amount of medium (100 ml) and time of induction were equal in all cases.

sample purified after gel filtration. The second step of purification after TEV proteolysis removes many other impurities in protein samples which can be seen in the gel Figure 3.3 (a) (Lane 3. resin sample).

3.1.2 Gel Filtration

Figure 3.4 shows the results of the gel filtration experiment of the NCoA-$1_{(257-420)}$ in complex with the STAT6$(794-814)$ peptide. The fractions under absorption peak at 280 nm were collected and analyzed. N-terminal sequencing by Edman degradation and MALDI analysis revealed that the STAT$6_{(794-814)}$ peptide, which is of a much smaller

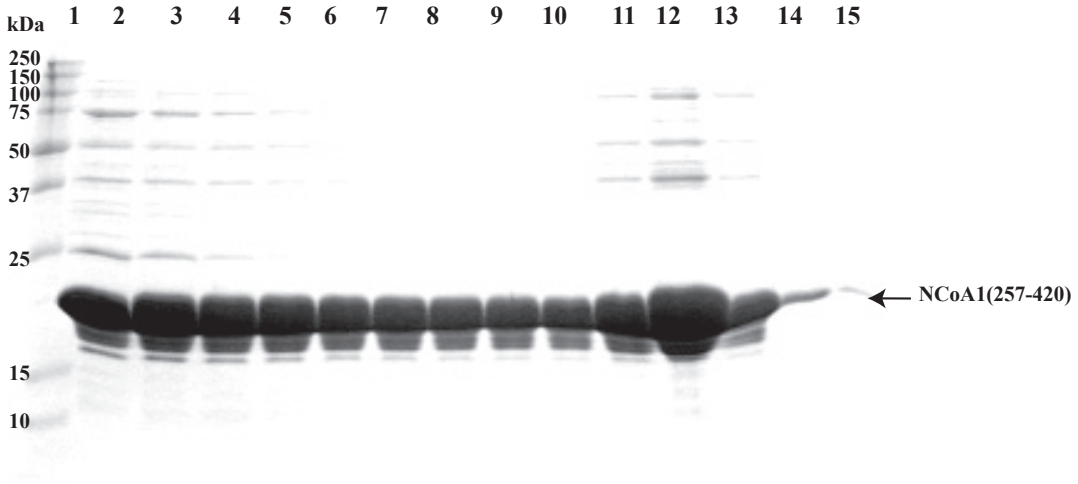

Figure 3.2: Large Scale purification of NCoA-$1_{(257-420)}$ in Silantes medium. 1. Broad Range Marker. 2-10. Elution with elution buffer A. 11-15. Elution with elution buffer B.

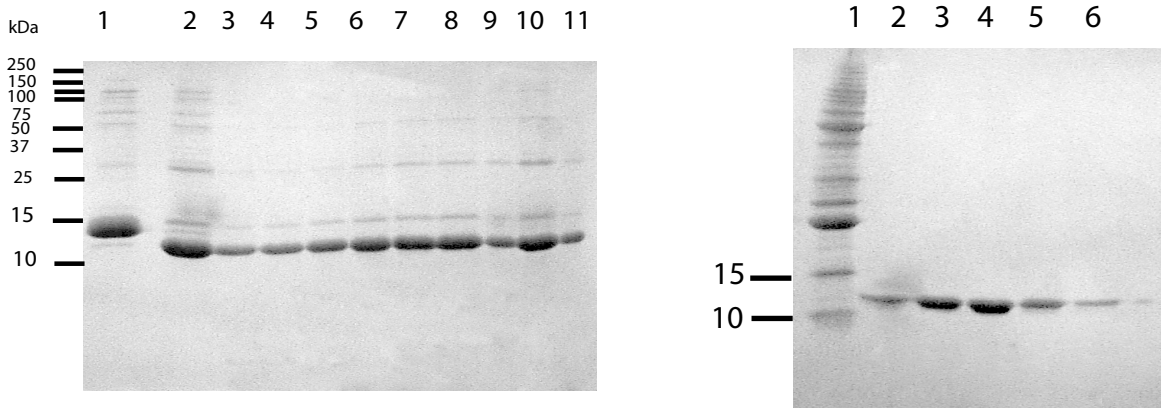

Figure 3.3: (a) Removal of the histidine tag after passing the sample through the Ni-NTA column. Lane 1. Undigested control. Lane 2. Resin sample. Lane 3. Flowthrough. Lanes 4-11. Wash with 1 ml of dialysis buffer. (b) Purity of the NCoA-1 (257-370) sample after gel filtration. Lane 1. Broad range marker. Lanes 2-5 Sample after gel Filtration.

molecular weight than the protein co-migrated with the NCoA-$1_{(257-420)}$ fragment and the peptide and the protein were in a 1:1 molar ratio. Macromolecules are separated during gel filtration because of the fact that bigger molecules elute faster than smaller

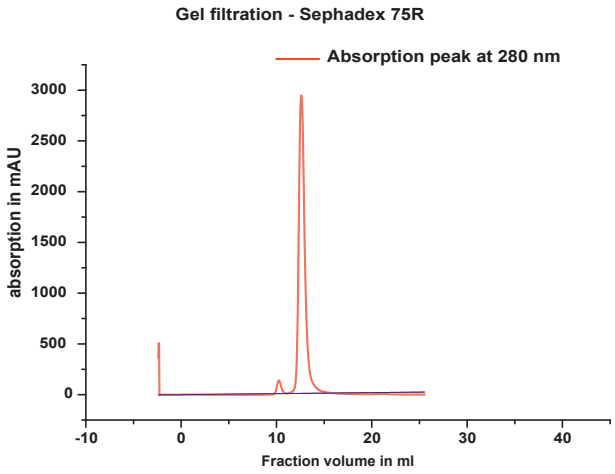

Figure 3.4: Chromatograph of a gel Filtration experiment showing the elution profile of the NCoA-$1_{(257-385)}$ in complex with the STAT6 peptide.

molecules. Since the peptide co-eluted with the protein it is clear that the peptide was in a stable complex with the protein. This experiment showed that a stable complex between the NCoA-1 and STAT$6_{(794-814)}$ could be reconstituted *in vitro* and the binding was strong, as it could withstand the gel filtration step.

3.1.3 Secondary Structure Prediction

Secondary structure prediction was done by using the program SOPMA[173]. The prediction showed that amino acids C- terminal of Asp366 formed a random coil.

3.1.4 Limited Proteolysis

Trypsin digestion revealed that the NCoA-$1_{(257-385)}$ was the most stable fragment as expected from secondary structure prediction.

Figures 3.5 and 3.6 show the SDS polyacrylamide gel analysis of the samples taken from different trials of limited proteolysis on the complex of NCoA-$1_{(257-420)}$ and STAT6 peptide and free NCoA-$1_{(257-420)}$, respectively. The NCoA-$1_{(257-420)}$ complexed with STAT$6_{(794-814)}$ peptide, and the protein blot corresponding to the band in lane 10 was cut out for sequencing by Edman degradation. Corresponding protein sample was also analyzed by mass spectrometry. The fragment had a mass of 14,795 daltons.

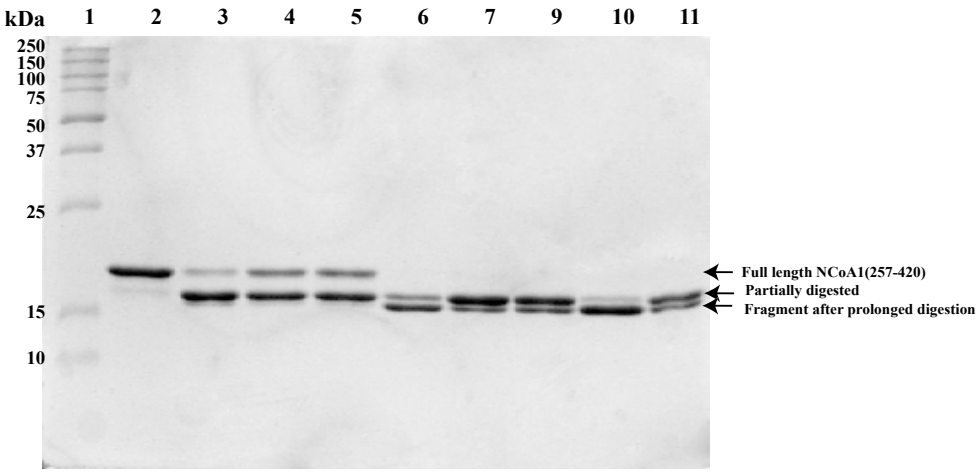

Figure 3.5: Limited proteolysis of the NCoA-$1_{(257-420)}$ fragment in complex with the STAT6 peptide. 1. Broad Range Marker. 2. Undigested control. Lanes 3,6,9 and 13 show proteolysis carried out for 10, 30, 60 and 120 minutes respectively with a protease/protein ratio of 0.004(w/w). Lanes 4, 7, 10, and 14 show proteolysis carried out for 10, 30, 60 and 120 minutes respectively with a protease/protein ratio of 0.002(w/w). Lanes 5,8,12,15 show proteolysis carried out for 10, 30, 60 and 120 minutes respectively with a protease/protein ratio of 0.001(w/w). The digestion was stopped by immersing the samples in liquid nitrogen. While part of the samples were used for mass spectroscopy, the bands on the gel were blotted on a PVDF membrane and the bands were cut out for protein sequencing.

Sequencing results showed that the N- terminus of this fragment was the same as that of the NCoA-$1_{(257-420)}$ fragment, thus verifying the identity of the protein. Based on the secondary structure prediction which showed that the amino acids C-terminal to Asp366 as disordered and limited proteolysis results where it was found that the NCoA-$1_{(257-385)}$ was the most stable fragment, two new constructs, NCoA-$1_{(257-370)}$ and NCoA-$1_{(257-385)}$ were designed and cloned into a pET16b vector.

3.1.5 Circular Dichroism

Both the free state NCoA-1 as well as the domain bound to STAT$6_{(794-814)}$ peptide were analyzed with circular dichroism. The CD spectrum of STAT6 peptide alone was also measured. While there was no significant difference between the bound and the free

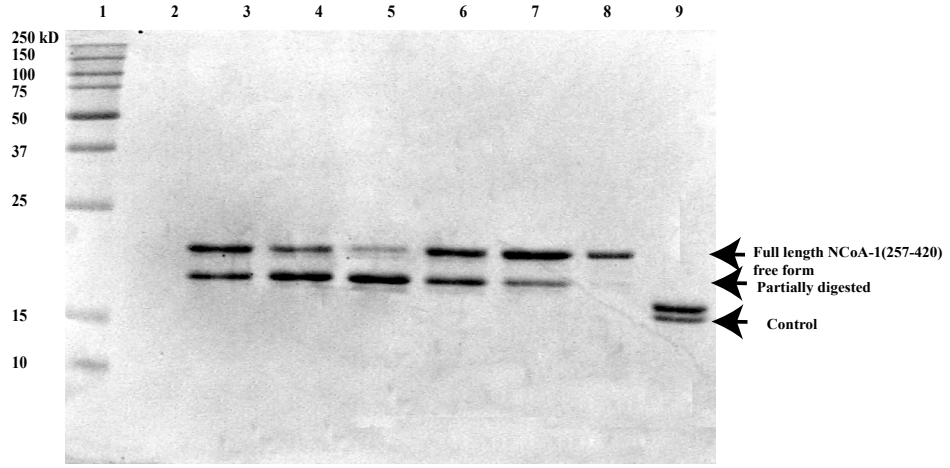

Figure 3.6: Limited proteolysis the free NCoA-$1_{(257-420)}$ fragment. Lane 1. Broad Range Marker. Lanes 3,4,5, and 6 show proteolysis samples with a protease/ protein ratio of 0.004 where the digestion was carried out for 10, 30, 60, and 120 minutes respectively. Lane 7. Proteolysis sample with protease/protein ratio of 0.002 where the digestion was carried out for 10 minutes. Lane 8. Undigested control. The digestion was arrested by immersing the samples in liquid nitrogen. While part of the samples were used for mass spectroscopy, the bands on the gel were blotted on a PVDF membrane and were cut out for protein sequencing.

NCoA-1 in the case of all the three fragments. Figure 3.7 shows the overlay of the CD spectrum of the bound and free form of NCoA-$1_{(257-370)}$ fragment. The STAT$6_{(794-814)}$ peptide alone showed a CD spectrum characteristic of an unfolded peptide(Figure 3.8). This is in agreement with the notion that the transactivation domains are unfolded before the interaction with their coactivator partners.

3.1.6 Isothermal Titration Calorimetry (ITC)

Characterization of affinity and stoichiometry for binding interaction between biomolecules is a key area in analytical biochemistry[174]. Isothermal Titration Calorimetry (ITC) is a method that uses the amount of heat released or absorbed during a binding reaction (i.e. the binding enthalpy change) to measure the parameters associated with binding[175]. Accuracy is one of the main advantages of ITC over other general-use methods[175]. The strongest affinity that can be measured directly

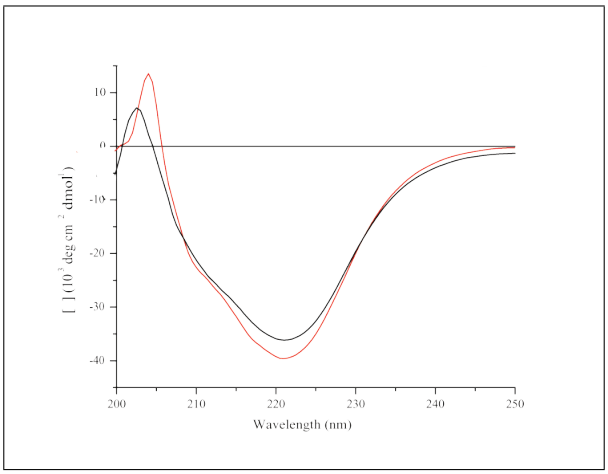

Figure 3.7: CD spectrum of the complex
Far UV CD spectra of the NCoA-1$_{(257-370)}$ with and without the peptide. The spectrum colored black is of the complex and the red one is of the free form. The concentration of the NCoA-1 was 50 μM in 25 mM phosphate buffer, pH 6.5 and the complex was reconstituted in the CD cuvette by adding an equimolar amount of the STAT6 peptide and spectra were measured after a delay of 1 hour.

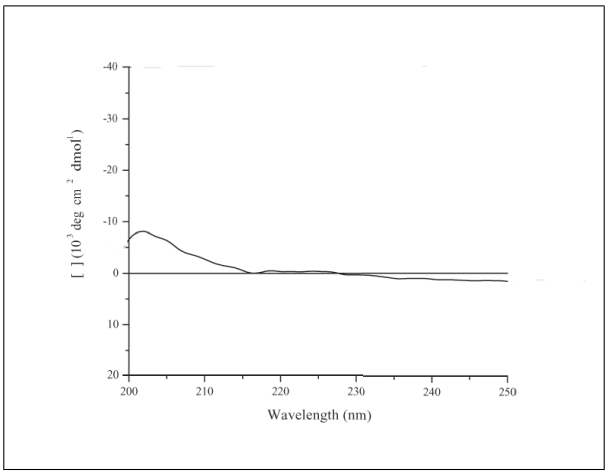

Figure 3.8: CD spectrum of the STAT6$_{794-814}$
Far UV circular dichroism spectrum the the transactivation domain of the 21 amino acid STAT6 peptide, was measured with 25 mM concentration of the STAT6 peptide in 25 mM phosphate buffer, 150 mM NaCl. Since there is no representative absorption at the usual range, it can be presumed that the fragment containing amino acids 794-814 of the STAT6 transactivation domain is unstructured.

by ITC is in the nanomolar range[176].

Analysis of the results

Figure 3.9 shows results of the ITC experiment where NCoA-$1_{(257-370)}$ was incubated with the concentration of the STAT$6_{(794-814)}$ peptide increased in incremental steps. The

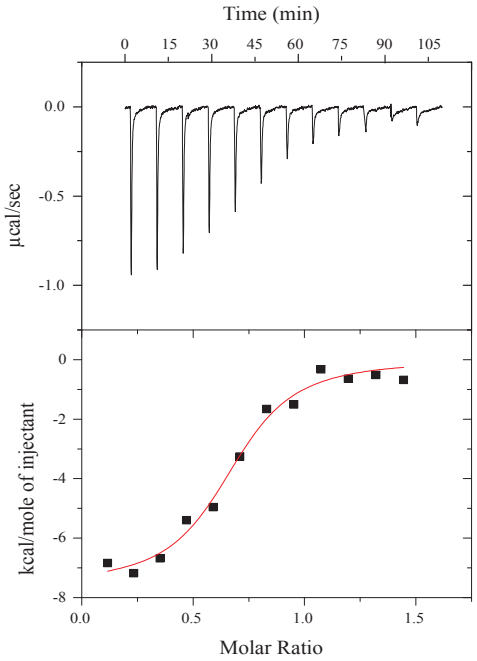

Figure 3.9: High sensitivity isothermal calorimetric titration of STAT$6_(794 - 814)$ peptide with NCoA-$1_{(257-420)}$ in 50 mM HEPES buffer, pH 7.0, 150 mM NaCl and 1 mM DTT. The concentration of the protein was 60 μM. The upper panel shows the power output in units of $\mu cal.s^{-1}$ as a function time for successive injections of the peptide. Eight microlitres of the peptide was injected each time. The lower panel shows the binding isotherm corresponding to the experiment shown in the upper panel. The enthalpy released per mole of the injectant versus the molar ratio of the peptide (injectant) to protein was plotted. Using this plot, the binding constant was calculated by making a fit using the non-linear least squares fit method.

results were analyzed using ORIGIN software provided by Microcal. The dissociation constant (K_D) calculated from isothermal calorimetry experiments was 0.8 $\pm 0.4\mu$ M.

3.1.7 Fluorescence Spectroscopy

Electron Spray Ionization (ESI), analysis showed that the trials with peptides with both N- and C- terminal cysteins failed to react with the dye. Next, the synthetic carboxyfluorescein-tagged peptide was titrated into the NCoA-$1_{(257-370)}$ sample and the emission spectrum at 480 nm was monitored using a spectrofluorimeter. The spectrum turned out to be uninterpretable, possibly due to non-specific interaction of the tagged peptide with the NCoA-$1_{(257-370)}$ fragment.

3.2 Results from NMR experiments

NMR studies of the NCoA-$1_{257-420}$ fragment

The NCoA-$1_{(257-420)}$ fragment was concentrated up to 0.4 mM and a HSQC spectrum was measured. This construct had the 10-Histidine tag attached to it. The spectrum (Figure 3.10 (A)) was indicative of an unfolded protein or a protein arrested in the, "molten globule state", during its folding process. When the peptide was titrated into this sample in a molar ratio of 1:1, the characteristic signature of an unfolded or partially folded protein changed to that of a folded spectrum (Figure 3.10 B).

The HSQC spectrum of the NCoA-$1_{(257-420)}$ fragment without the histidine tag complexed with STAT6$_{(794-814)}$ peptide in a 1:1 molar ratio, showed a number of poorly dispersed N-H correlation peaks, which are rather sharp and indicative of amino acids from the disordered region (Figures 3.11(a) and (b). Many other peaks were so broad that they were hard to detect at all. A preliminary evaluation of the spectrum suggested that about 60% of the NCoA-$1_{(257-420)}$ fragment had a defined structure. A series of 3D heteronuclear experiments (see Table 2.4 were measured to facilitate sequential resonance assignment in the case of the NCoA-$1_{(257-420)}$ fragment in complex with STAT6 peptide. Further analysis and resonance assignment trials indicated that the region from the residue Leu371 in C- terminus onwards could be unstructured as their resonances were very sharp in the ^{1}H , ^{15}N as well as the ^{13}C$'$ dimensions. Since the rest of the residues showed very broad resonances assignment was very difficult. Table 3.2 shows the number of residues in each of the three constructs studied by NMR and their assignment details. Only fifteen amino acids could be assigned in the case of NCoA-$1_{(257-420)}$ fragment and they were the peaks that had very sharp

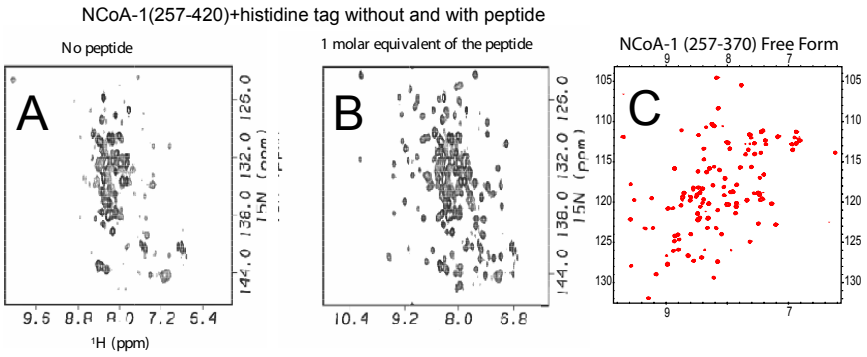

Figure 3.10: Comparison of HSQC spectra of the NCoA-1 (257-420) and NCoA-1 (257-370) fragments
In the left is a HSQC spectrum of the NCoA-$1_{(257-420)}$ with the histidine tag (A), and the right spectrum is of the NCoA-$1_{(257-420)}$, when 1 molar equivalent of the STAT6 peptide was titrated into the NMR sample tube(B). While the left HSQC is characteristic of an unfolded or partially folded protein that resembles the spectrum of a "molten globule", e.g. ubiquitin in 8 M urea, the right HSQC shows well dispersed peaks indicating that the protein starts folding once the peptide was added. Spectra were measured in 50 mM phosphate buffer, pH 6.5, with 150 mM NaCl. The lyophilized peptide was also taken in the same buffer at a stock conc. of 5 mM. (C) An HSQC spectrum of the NCoA-1 (257-370) fragment free form. The spectra was measured at 600 Mhz, with a sweepwidth of 2500 Hz and 2400 Hz in the 15N and 1H dimensions respectively. Number of scans = 4.

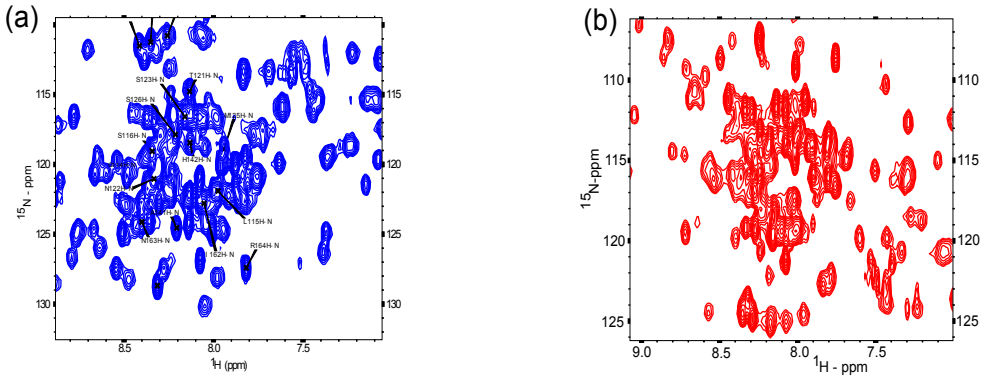

Figure 3.11: A comparison of HSQC spectra of NCoA-1 (257-420) with and without peptide
A comparison of HSQC spectra of the NCoA-$1_{(257-420)}$ with and without STAT6 peptide. The NCoA-$1_{(257-420)}$ is without the Histidine tag in this case. The left spectrum which is of the NCoA-$1_{(257-420)}$ with the STAT6 (794-814) peptide, and the one in the right is of NCoA-$1_{(257-420)}$ without the peptide.

lineshapes. From the assignment it was readily seen that they were from the C-terminus. From this the conclusion is that the peaks in the extreme C- terminus numbering about fifty are disordered, as all of them give rise to peaks characteristic of the unfolded domain. Similar spectral quality was also found in the case of the NCoA-$1_{(257-385)}$, where comparison with the NCoA-$1_{(257-370)}$ HSQC spectrum showed that the 15 amino acids from Lys371 could be in a disordered coil. Figure 3.12 shows the overlay of the NCoA-$1_{(257-370)}$ and the NCoA-$1_{(257-385)}$ HSQC spectra. This is in agreement with the crystal structure of the complex, where residues after amino acid 366 were not visible due to their unstructured and highly flexible nature. Only the NCoA-$1_{(257-370)}$ construct showed a better quality spectrum amenable for three dimensional heteronuclear experiments and subsequent resonance assignment. Nine residues could not be assigned with the measured set of triple resonance spectra as they had highly ambiguous resonances in the $^{13}C^{\alpha}$ dimension and lacked $^{13}C^{\beta}$ chemical shift information. Figure 3.13 shows the HSQC peaks of the free NCoA-$1_{(257-370)}$ fragment and the resonance assignments.

The nine ambiguous residues are as follows : 1. Thr277 2. Ser279 3. Asp290 4. Lys291 5. Ile296 6. Phe299 7. Arg311 8. Leu313 9. Tyr348

It is notable that among these residues that could not be assigned only Ile296 has close interaction with the STAT6 peptide.

Peak details	NCoA-$1_{(257-420)}$	NCoA-$1_{(257-385)}$	NCoA-$1_{(257-370)}$
No. of residues	164	139	114
Total no. of assignable peaks	140	131	107
Peaks assigned by MARS	-	-	78
Peaks assigned manually	15	-	20
Ambiguous peaks	-	-	-9

Table 3.2: Statistics of resonance assignment of the three NCoA-1 constructs

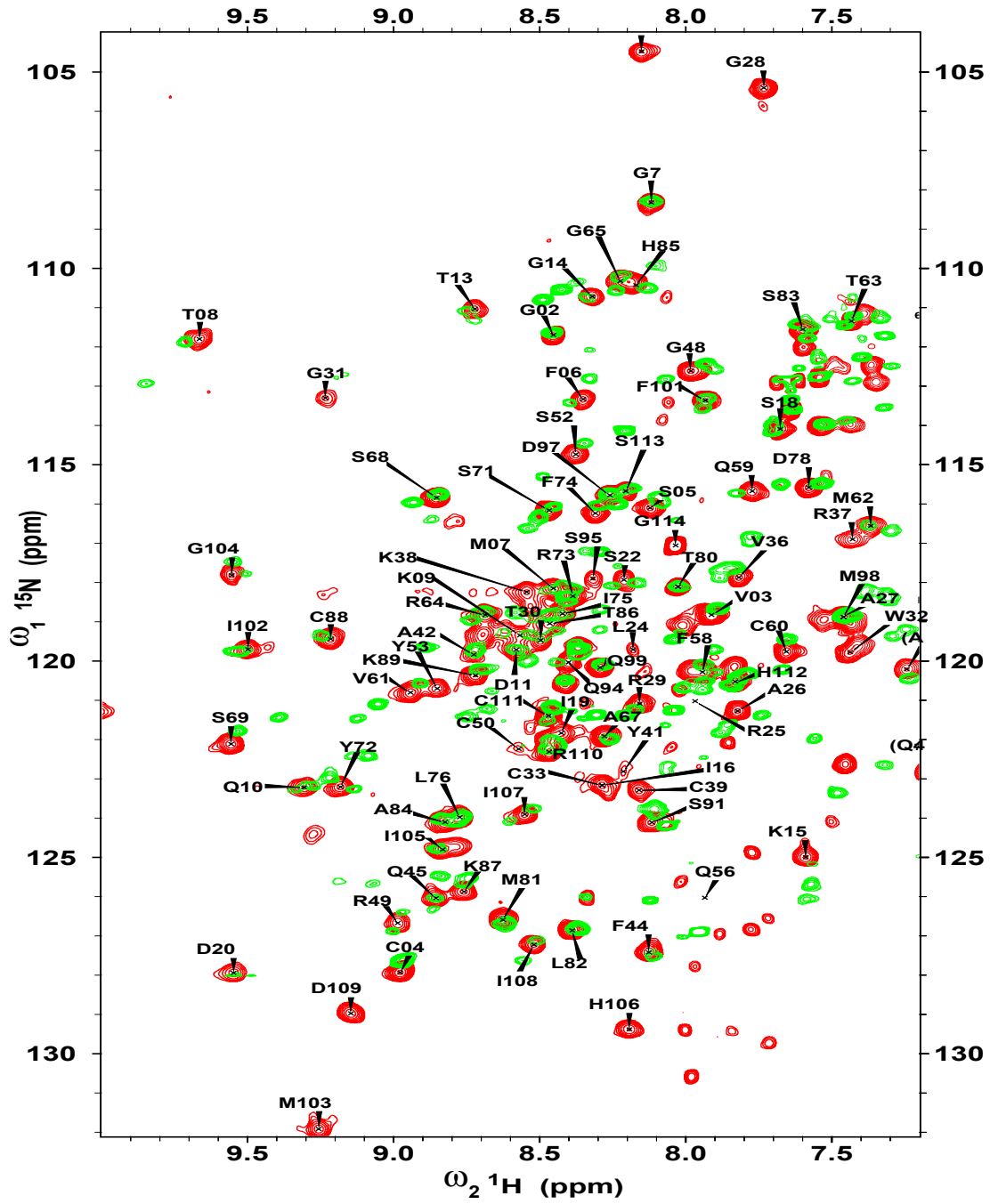

Figure 3.12: An overlay of the HSQC of NCoA-1(257-370) over the HSQC spectrum of NCoA-1 $_{(257-385)}$. The peaks in red with the assignment labels are of the 370 fragment and those in green are from the 385 fragment. The sharp peaks in the spectrum of the 385 fragment are from the 15 amino acids that are to the extreme C- terminal.

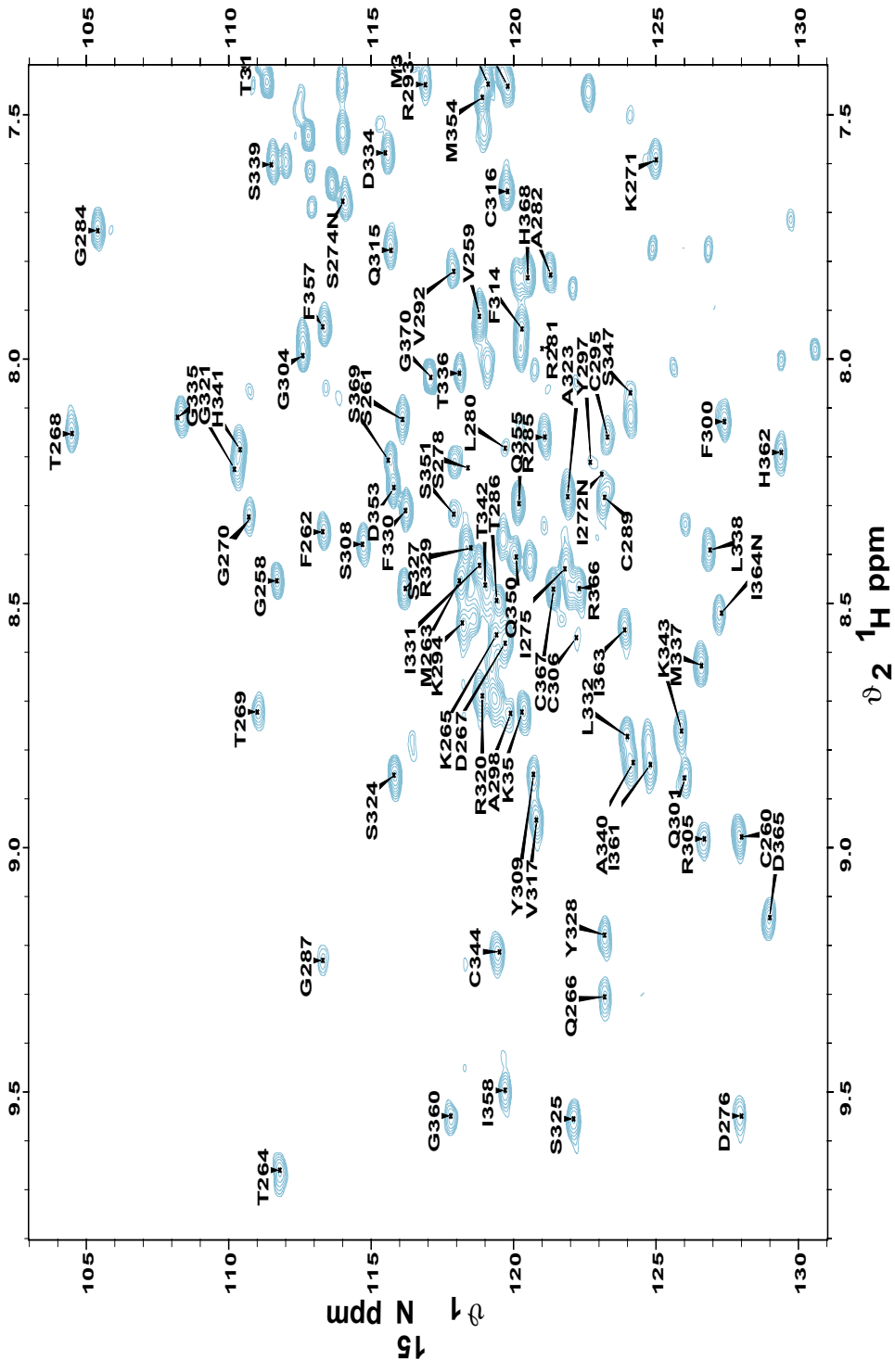

Figure 3.13: The HSQC spectrum of NCoA-1 (257-370) free form showing resonance assignment. The HSQC was measured with 16 scans.

Chemical shift index (CSI) as a measure of secondary structure

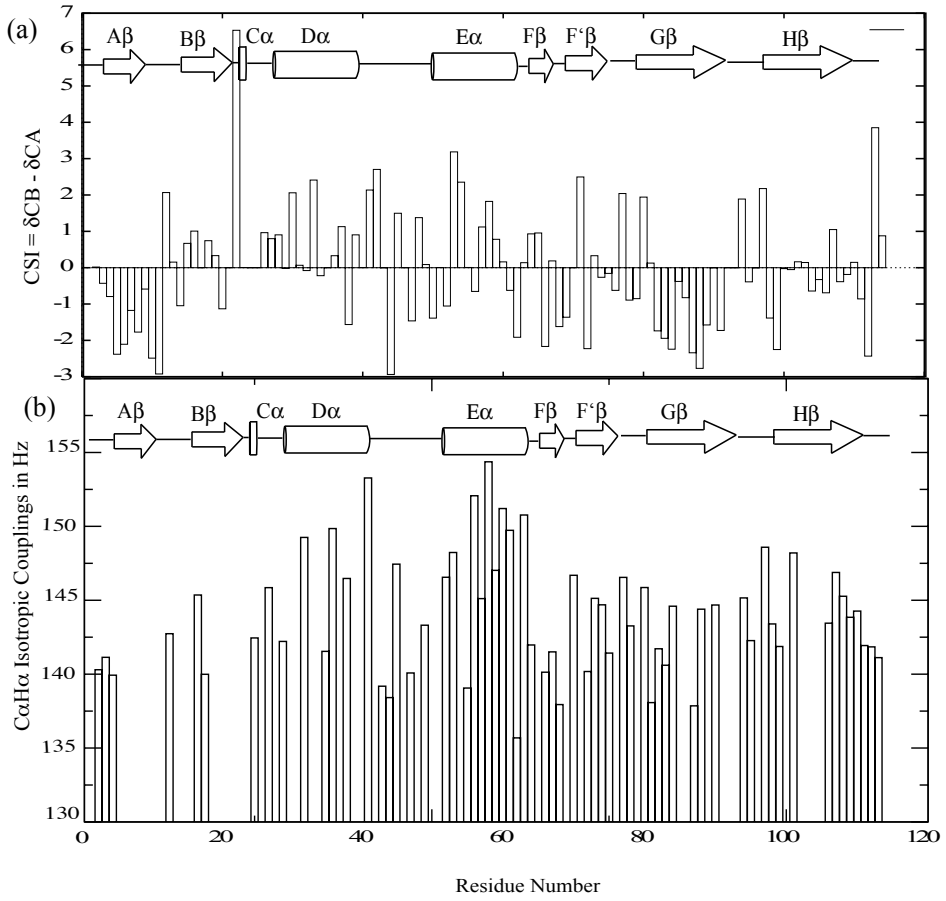

Figure 3.14: (a) Chemical shift index of the NCoA-$1_{257-370}$ free form compared with a secondary structure representation constructed out the PDB structure of the complex. CSI values are plotted against the residue numbers as per the primary sequence. Positive values indicate α helix while negative values indicate a β sheet. (b) The Isotropic ^{1}J couplings in Hz versus the residue number.

Figure 3.14(a) shows the CSI plot overlayed with the secondary structure cartoon of the NCoA-$1_{(257-370)}$ in complex. Clearly, the change in the conformation in the second β sheet (B_β) sheet from the N- terminus can be seen. This also supported by the fact that the residues involved in the interaction with STAT$6_{(794-814)}$ peptide like Ile272, Ile275 and Ile277 are located in the strand B_β. Significant change is also observed in the H_β strand in the C- terminus. Since the CSI values do not fall into a pattern

suggestive of either α helix or β sheet, the nature of conformational change in the H_β strand cannot be explained. Both the longer constructs, the NCoA-$1_{(257-385)}$ and NCoA-$1_{(257-420)}$ fragments could not be used for CSI studies because of the lack of sufficient signals.

$^1J_{C\alpha\, H\alpha}$ Isotropic couplings

There exists a relation between isotropic $^1J_{C\alpha\, H\alpha}$ couplings and the backbone ϕ and ψ angles. These couplings are particularly useful for identifying residues with positive ϕ angles. Residues with a positive ϕ angle have a $^1J_{C\alpha\, H\alpha}$ coupling $\leqslant$ 137 Hz. Figure 3.14(b) shows a plot of isotropic ^{1}J Cα $H\alpha$ couplings versus the residue number, compared with the secondary structure pattern of NCoA-$1_{(257-370)}$ fragment. The residues in helical regions have larger $^1J_{C\alpha\, H\alpha}$ values than others in extended regions. The two main helices present in the crystal structure of the complex are clearly represented in the plot of the free protein.

3.2.1 Titration of the peptide STAT6 $_{(794-814)}$ into NCoA-$1_{257-385}$

A HSQC spectrum of the NCoA-$1_{(257-385)}$ fragment showed peaks which were very weak in intensity. An overlay of the HSQC spectrum of the NCoA-$1_{(257-385)}$ on that of the NCoA-$1_{(257-370)}$ fragment showed that those peaks which were very sharp did not match with the peaks from the NCoA-$1_{(257-370)}$ (See Figure 3.12). It could be concluded that the peaks which were sharp arose out of the amino acids 271- 385 in the longer fragment, and they were disordered. The peptide was added in increasing molar ratios with respect to the protein as 0.2, 0.5, 1, and 1.5 times to the NCoA-$1_{(257-385)}$ concentrations. Figure 3.15(a) shows the overlay of the region from the different titration steps, of the peaks from the C- terminal region of the NCoA-$1_{(257-385)}$ fragment. The sharp peaks that were possibly from the 15 amino acids towards the C- terminus did not show any shifts from their position in the HSQC of the free form. In Figure 3.15(b), the peaks within the circles are the peaks that did not move upon titration. This suggests the possibility that the peaks are not affected by the addition of the STAT6 peptide.

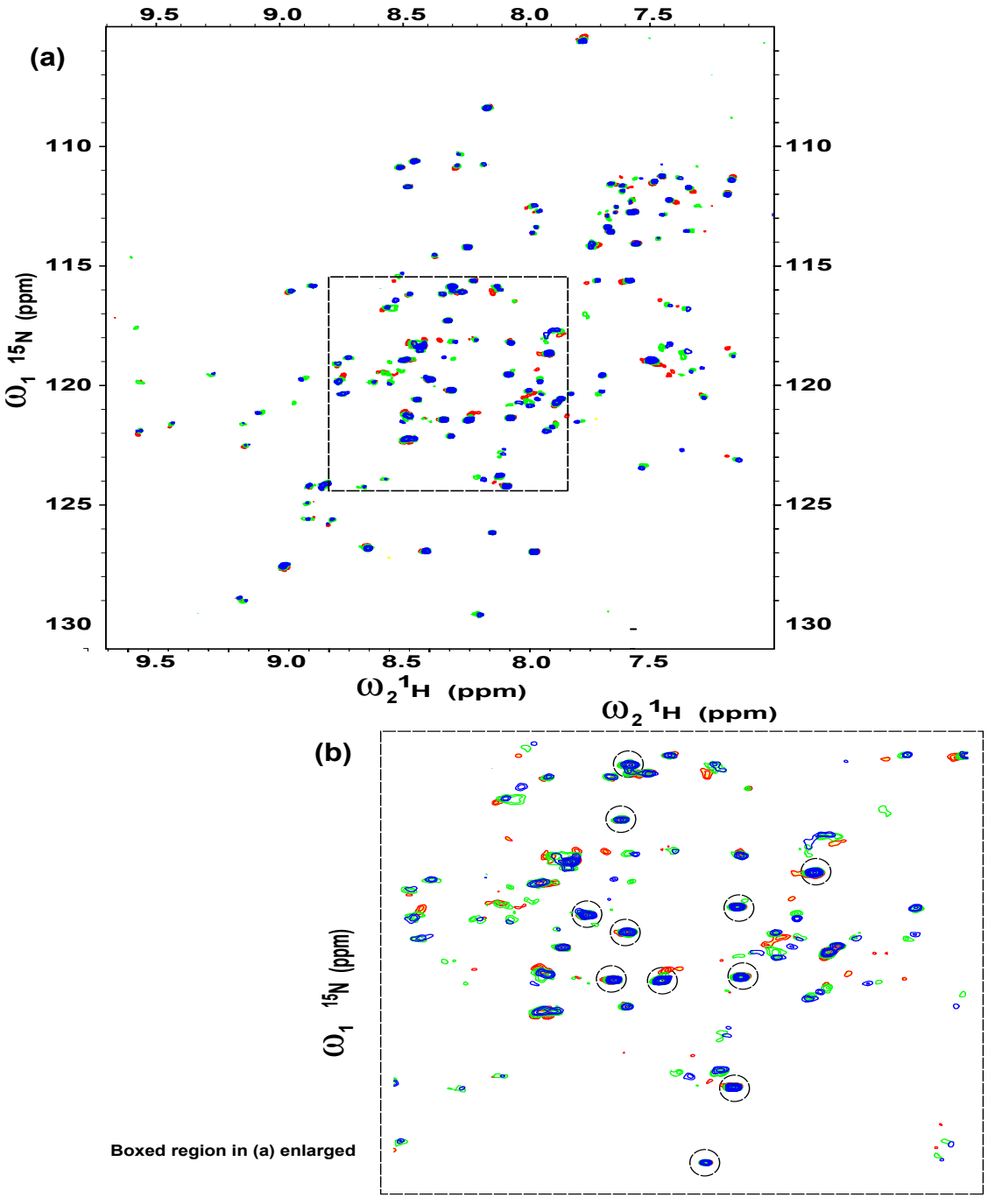

Figure 3.15: An overlay of HSQC spectra from different titration steps of the STAT6 peptide into the NCoA-1 (257-385) fragment. The peaks which have very sharp lineshapes (encircled peaks in (b)) do not move upon titration with the peptide.

3.2.2 Titration with the peptide STAT6 $_{(794-814)}$ into NCoA-1$_{257-370}$

The peptide ligand STAT6 (794-814) was titrated into to the NCoA-1 (257-370) in increasing molar ratios of 1:1, 2:1, 3:1, and 4:1 and the sample was taken in a Shigemi tube and a HSQC spectrum was measured using Bruker spectrometers of 600MHz and 700 MHz field strengths (Figure 3.16 (a)). The shifting of the peaks was not uniform Figure 3.16 (b). While the position of many of the peaks could be traced because their location was not very far from that of the peaks from the free form, some other peaks did not have any matching peak in the HSQC spectrum of the free form. Figure 3.17 shows the overlay of the HSQC spectra of the free form of NCoA-1(257-370) and the complex of NCoA-1(257-370) with STAT6(794-814) where the ratio of NCoA-1 protein to the STAT6 peptide is 1:4 Since the resonance assignment of the bound form was not complete, chemical shift mapping of the entire NCoA-1$_{(257-370)}$ could not be represented in a plot.

3.2.3 Studies of the free NCoA-1

Based on biological studies, the region between the amino acids 257-420 of the NCoA-1/SRC protein was identified as the interaction partner of the STAT6 transactivation domain. Limited proteolysis and gel Filtration experiments suggested that the region of the NCoA/SRC protein comprising of the amino acids 257-370 was sufficient for the interaction, although this construct did not crystallize on its own or in complex with its STAT6 interaction peptide. Rather a slightly longer construct, the NCoA-1$_{(257-385)}$ in complex with the STAT6 transactivation domain, crystallized. Crystallization trials with the free NCoA-1$_{(257-420)}$ or the complex with STAT6 transactivation domain have not been successful, so far. Circular dichroism spectra of the free and the bound form of NCoA-1$_{(257-370)}$ and the NCoA-1$_{(257-385)}$ showed that there is no significant difference in the overall secondary structure content between the two forms. But these spectra are not indicative of the intrinsic conformational details. Since the crystallization of the the free NCoA-1 LXXLL interaction domain was not possible, NMR was used to obtain structural information.

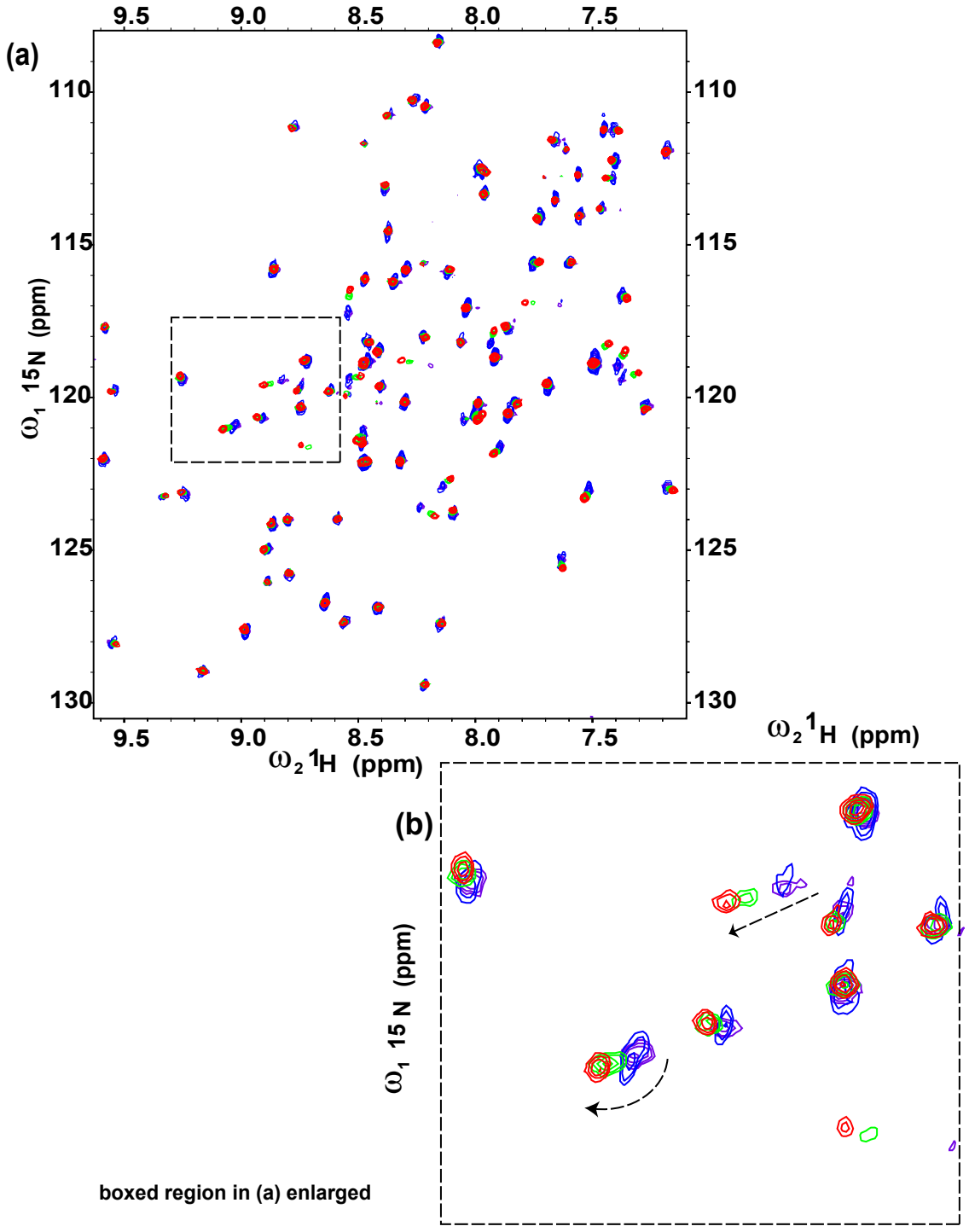

Figure 3.16: Overlay of titration steps of the NCoA-1 (257-370) at 700 MHz Color codes for titration steps : 1. blue - free form ; 2. purple - 1 ; 3. green - 2; 4. red - 3 times the molar ratio of the STAT6 peptide to the NCoA-1 (257-370) protein.

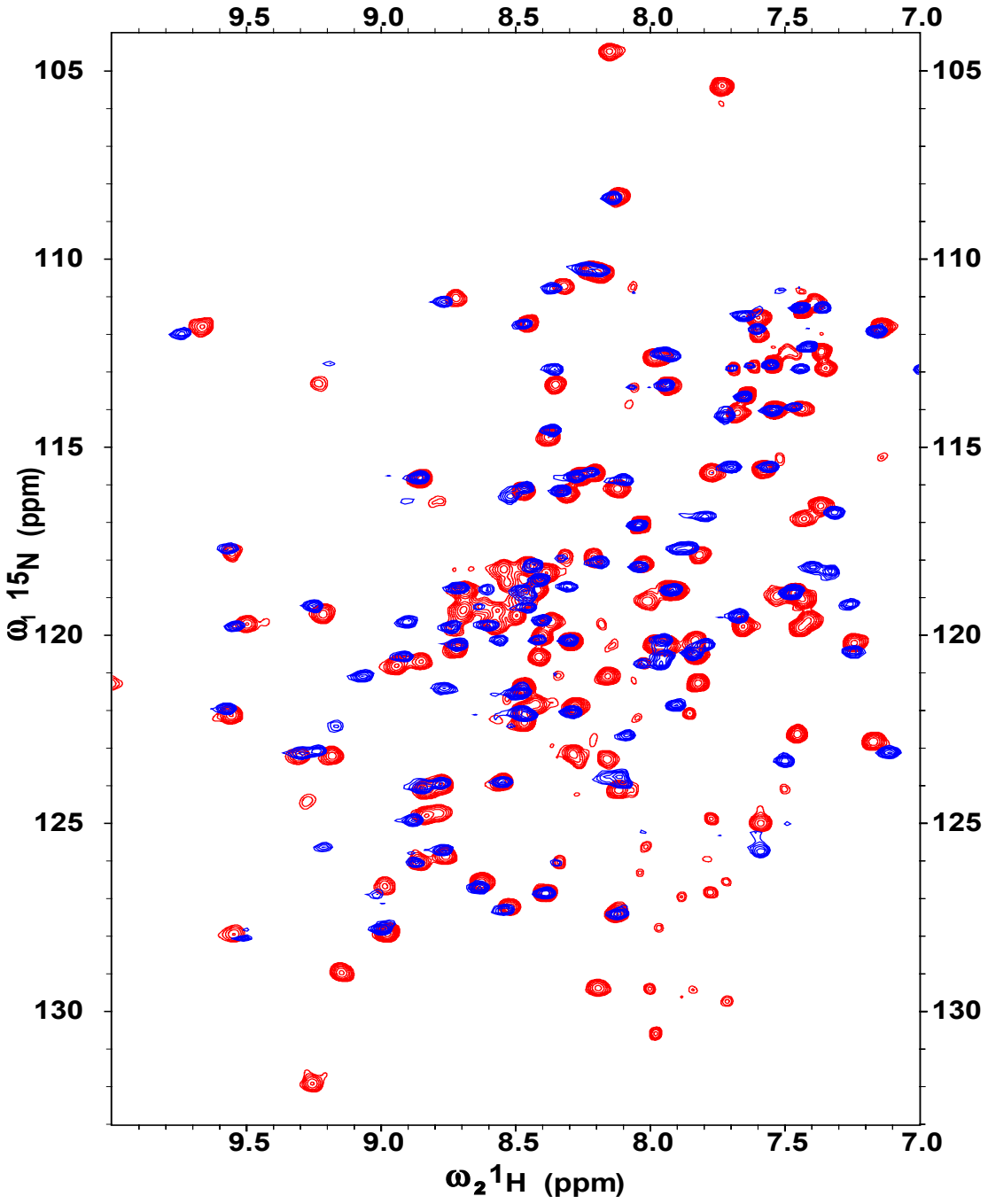

Figure 3.17: An overly of the HSQC spectra of the free form of NCoA-1(257-370) and the complex of NCoA-1(257-370) with STAT6(794-814) where the ratio of NCoA-1 protein to the STAT6 peptide is 1:4. Color code : red - free NCoA-1. blue - complex.

3.2.4 NMR studies of the free NCoA-1

The studies on ^{15}N and ^{13}C labeled NCoA-1$_{(257-420)}$ fragment showed that the amino acids 370-420 of the NCoA-1 constituted a highly flexible region. The studies on the NCoA-1$_{(257-385)}$ fragment also showed that the amino acids 370-385 were disordered. This is also the case in the crystal structure of the complex between NCoA-1$_{(257-385)}$ and STAT6 peptide. Amino acids from Leu366 were not visible as they were unstructured. The minimal domain identified through limited proteolysis, the NCoA-1$_{(257-370)}$ seems to be ordered well as indicated by the NMR spectra. Ninety percent of the backbone resonances amino acids could be assigned without ambiguity. Titration experiments indicate that there could be a possible conformational change as the chemical shifts of the residues with and without the peptide change in a non-uniform fashion.

3.3 Measurement of Residual Dipolar Couplings (RDCs)

3.3.1 Alignment of NCoA-1$_{(257-370)}$ Using Pf1 Phages

Different salt concentrations were used to test the alignment, and the protein concentration was kept uniform. Precipitation occurred below salt concentrations of 150 mM. The phage was pelleted out from the NMR sample by centrifuging at 45,000 r.p.m. for 1 hour, using a Beckman Coulter L80 ultracentrifuge. The pelleted phage had a whitish appearance. The supernatant was checked for protein at 280 nm UV absorption and the protein content was reduced to half the initial concentration. The spectrum was measured in a Bruker Avance 800 MHz spectrometer and the D_2O splitting of 8 Hz was found at a salt concentration of 600 mM using a ^{13}C,^{15}N labeled sample. The D_2O splitting gave an estimate of the concentration of the phage particles in the NMR sample. ^{15}N /^{13}C labeled NCoA-1$_{(257-370)}$ with a concentration 0.3 mM in 50 mM PO_4 buffer, pH 6.5 and 600 mM NaCl was aligned with 10 mg/ml phages (*Profos Biolabs, Munich*), and an IPAP HSQC type experiment was recorded to obtain N-H J+D couplings. The data were processed with NMRPipe[177] software from NIH and accessory scripts were used to calculate the couplings. Figure 3.18(a) shows a representative doublets from protein samples aligned using Pf1 phages. It is an overlay of the in-phase and anti-phase components from an IPAP-HSQC. Pf1 was added at a final concentration of 4mg/ml. The anti-phase component is not of the same intensity as the

in-phase component because of line-broadening. The reason for the line-broadening could be excessive alignment. Figure 3.18(b) shows the splitting of Thr366 in (1) 2mg/ml phage and (2) 5 mg/ml phage, which is 91 Hz and 93 Hz respectively. The increase in phage concentration did not increase the alignment further, but line-broadening was observed. One possible explanation for this behaviour of the protein is that a part of the protein interacts with the phages and aggregates, thus destroying the nematic phase, while the remaining protein aligns excessively with the phage particles that remained in solution.

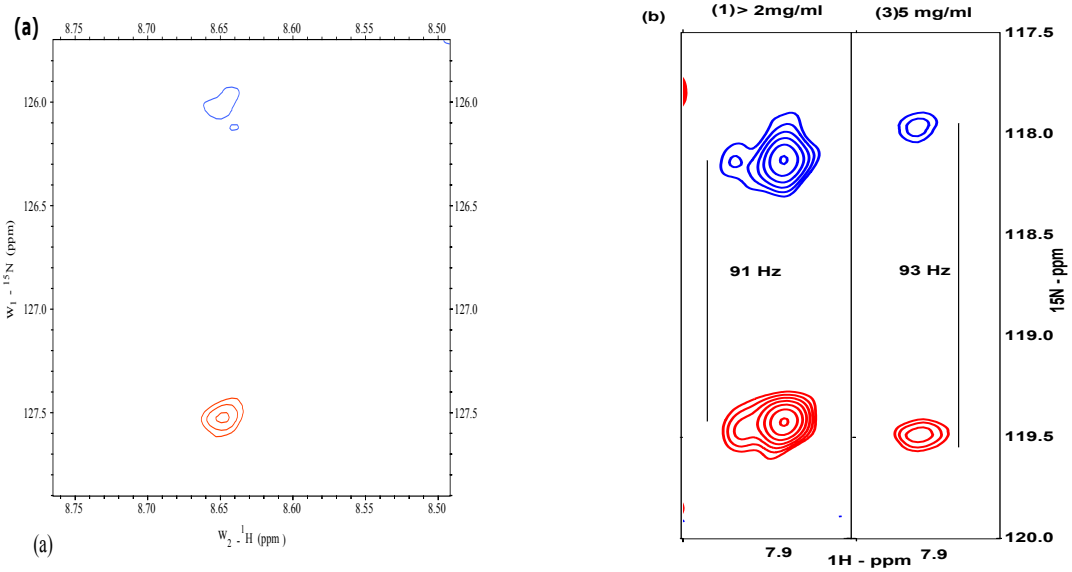

Figure 3.18: (a) Overlay of the inphase and antiphase components of an IPAP spectrum. (b) Splitting of Thr366 peak in Pf1 phage medium
(a)The peak in red shows the anti-phase component of the IPAP while the peak in blue is the in-phase component. Both spectra were measured in an interleaved manner and processed separately by extracting them. The above figure shows the overlay of the in-phase and anti-phase component of the signal from a representative residue. The anti-phase peak is not of the same intensity as of the in-phase peak because of the line-broadening. (b) J NH + D NH splitting of the amide residue Thr336 at (1). ¿ 2 (2) 5 mg/ml Pf1.

3.3.2 Measurement of Residual Dipolar Couplings

Since there was a huge line broadening effect that hindered the use of HNCO-TROSY type experiment for the measurement of dipolar couplings, perhaps due to a much

faster relaxation due to the interaction between the protein molecule and the bicelles. An IPAP type experiment was used to obtain NH dipolar couplings using a Bruker 600 MHz spectrometer, equipped with a cryoprobe. The quality of the spectrum was very poor with extremely broad peaks. It was not possible to get any couplings, from this trial. Addition of a positively charged lipid molecule like CTAB up to 5mM was expected to reduce the interaction between the protein and the biomolecule but did not improve the situation.

3.3.3 Polyethylene glycol based medium (Otting phase)

Precipitation was observed. This sample was used to check the quadrupolar splitting of deuterium using a Bruker 400 MHz spectrometer. Neither the D_2O splitting nor the protein signal was observable.

3.3.4 Polyacrylamide gels in alignment

A HSQC spectrum of NCoA-$1_{(257-370)}$ in 7 % was measured, which showed peaks only from the side chain region. Figure 3.19 shows the HSQC spectrum. The reason could be that during the diffusion of the protein sample inside the gel, most of the protein sample would have aggregated.

Paramagnetic Tags

The two cysteins present in the NCoA-$1_{(257-370)}$ fragment, Cys295 and Cys365 were suitable for the attachment of the tag. MALDI spectroscopy showed an increase of molecular weight but the increased weight did not correspond to an attachment of the tag whose molecular weight was 527 Da, either with one of the cysteins or with both. Table 3.3 shows the expected and observed molecular weights. The sample used for the trial was double labeled. The expected molecular weight is for a doubly labeled sample.

Different concentrations of phages and NaCl were used for alignment. When a phage concentration of 2 mg/ml was used 37 N-H RDCs could be obtained from an IPAP spectrum. Tables 3.4 and 3.5 show the details of the various trials done using Pf1 phages as alignment medium. Results from other alignment media are listed in Table 3.6.

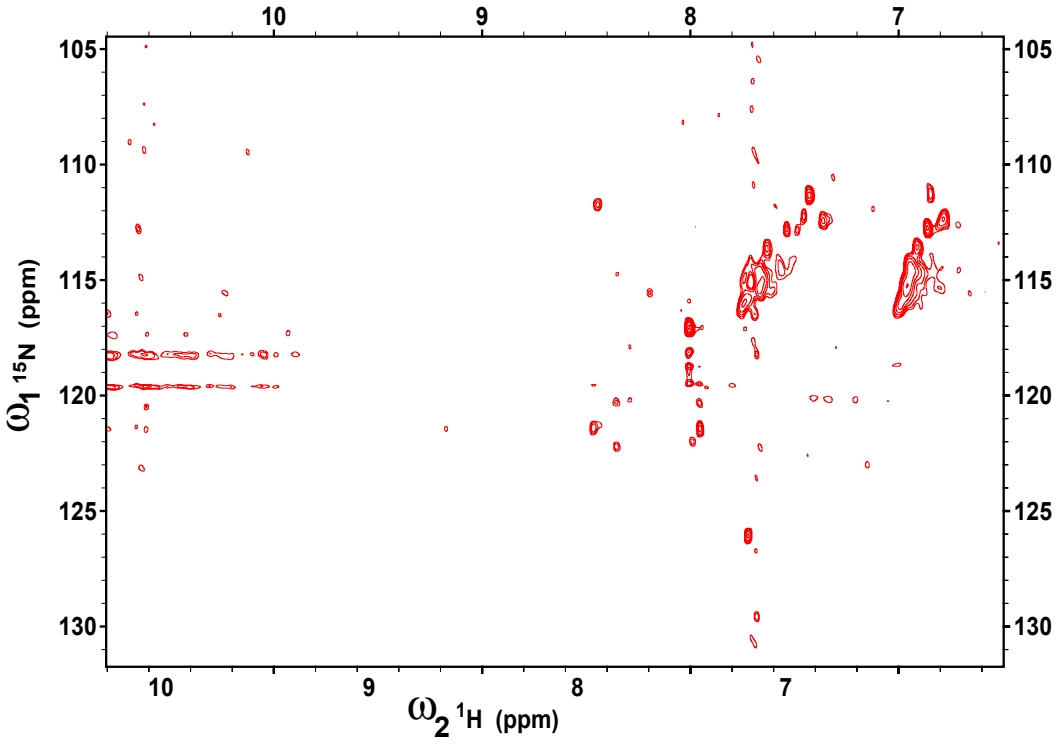

Figure 3.19: HSQC showing the NCoA-$1_{257-370}$ aligned in 7 %polyacrylamide gel. The concentration of the protein was 0.4 mM, with 15% D_2O. The spectrum was recorded using a Bruker 900 MHz spectrometer attached with a cryoprobe and spectral widths in the F1 and F2 dimensions were 2730 Hz and 4323.3 Hz, respectively. The number of t1 increments = 64.

Table 3.7 shows the parameters important in the analysis of the N-H dipolar couplings. Analysis of the couplings was done by using PALES[178] software. A correlation of the experimental dipolar couplings was done against the back-calculated dipolar couplings from the PDB file of the crystal structure of the complex of the NCoA-$1_{(257-370)}$ with STAT6 peptide.

The Q_{saupe} was calculated according to a previous work[146]. A correlation of 0.480 is not valid from a structural point of view.

RDCs were predicted from the crystal structure of the complex between STAT6$_{(794-814)}$ peptide (minimal transactivation domain of STAT6) and the NCoA-$1_{(257-385)}$ fragment. Measured RDCs were from the free form of the NCoA-$1_{(257-370)}$. The absolute value of the magnitude of the alignment tensor (D_a) obtained was 1.15×10^4 Hz. This value is too small that it falls within the root-mean-square

Molecule	*Expected molecular weight*	*Mol. Weight from MALDI in Daltons*
Free NCoA-1(257-370) (control)	13528.22	13825
NCoA-1(257-370) + 3 times molar excess paramagnetic tag	14562.22	15153

Table 3.3: Calculated and experimental molecular weights for the NCoA-$1_{(257-370)}$ unreacted and reacted with the paramagnetic tag

Expt.	2mg/ml	3.5 mg/ml	4 mg/ml	5 mg/ml
D_2O splitting	-	-	-	-
IPAP no. of couplings	~ 50	$\leqslant 10$	$\leqslant 10$	$\leqslant 10$
Salt titration	-	-	-	
HNCO-TROSY	-	-	37 couplings	-
Other 2D expts	-	-	-	-

Table 3.4: Different concentrations of phages and their effect on alignment.(1) The sample is unstable and aggregates in a few days, after the addition of phages.

uncertainty of the alignment tensor. This indicated that the average size of the dipolar couplings obtained is within the error range and therefore the correlation of 0.48 obtained using alignment with Pf1 phages is not structurally valid[178].

In another trial, the couplings were chosen from those regions where the secondary structure elements as indicated by the CSI plot do not show any change when compared to the predicted secondary structure from the crystal structure of the complex. In Figure 3.20, the correlation plot of predicted and experimental dipolar couplings is shown. There is no correlation between the measured and predicted RDCs as seen from the plot. This could mean that the total set of N-H RDCs obtained experimentally are not useful for comparison of the predicted RDCs from the crystal structure of the complex and the measured RDCs from the free NCoA-$1_{(257-370)}$ fragment. It is also possible that the structure of the free form differs drastically from the complex, but this not seen in the CSI plot, where marked changes are seen only in the Bβ and Hβ strands of the complex. While there is a helical tendency in the former strand whereas the changes are not explainable in the latter strand.

Expt.	8mg/ml	10 mg/ml	16 mg/ml	20 mg/ml
D_2O splitting	7.9Hz	10 Hz, aggregation	-	-
IPAP	-	-	-	-
Salt titration	5 steps	1 step	1 step	1 step
HNCO-TROSY	-	-	-	-
Other 2D expts	HNCO-ECOSY	-	-	-

Table 3.5: Different concentrations of phages and their effect on alignment. All the experiments were performed in the same buffer condition, that is, 50 mM PO_4, 150 mM NaCl, pH 6.5, 1 mM DTT. The sample is unstable and aggregates in a few days, after the addition of phages.1. Except for the 8mgs/ml experiment, all other conditions were tried using unlabeled sample.2. The D_2O splitting was not observable in higher concentrations, the reasons are unclear.

Alignment medium	*D_2O splitting*	*IPAP*
DHPC:DMPC bicelle	7 Hz	-
DHPC:DMPC bicelle + 5 mM CTAB	7 Hz	20 couplings
Polyacrylamide gel (7 %)	10 Hz	-
$C_{12}E_5$/hexanol	-	-
Paramagnetic tag	wrong mol. wt.	-
3% agarose	-	-

Table 3.6: Other alignment trials

Since the NCoA-1 is a highly basic protein, it is positively charged at a lower pH. Moreover, from the crystal structure it is evident that the peptide sits in a largely hydrophobic groove, which could possibly be exposed to water, in the absence of the peptide. Since the NMR measurement conditions require a pH of below 7.0, the protein is positively charged in the case of NMR buffer. As the Pf1 phages have a negative charge of -5 units, there is a very strong electrostatic interaction between the NCoA-1 and the phages at the pH of 6.5. Therefore there is too much alignment and consequently there are very broad NMR peaks. Therefore it was not possible to obtain dipolar couplings that are suitable for structural analysis. Three dimensional structure calculation using NOESY spectrum requires side chain resonance assignments. This is time consuming and prone to ambiguity. Given the availability of a crystal structure, measurement of dipolar couplings was considered as an easier alternative to get structural information. The nature of the free form of NCoA-1 did not allow it to be

Parameter	value
No. of couplings	39
RMS	3.028
Chi^2 index	357.468
Correlation coefficient	0.480
Q_{saupe}	0.951
D_a	-1.158391e-04
D_{a-NH}	-2.500

Table 3.7: Dipolar coupling analysis using PALES software

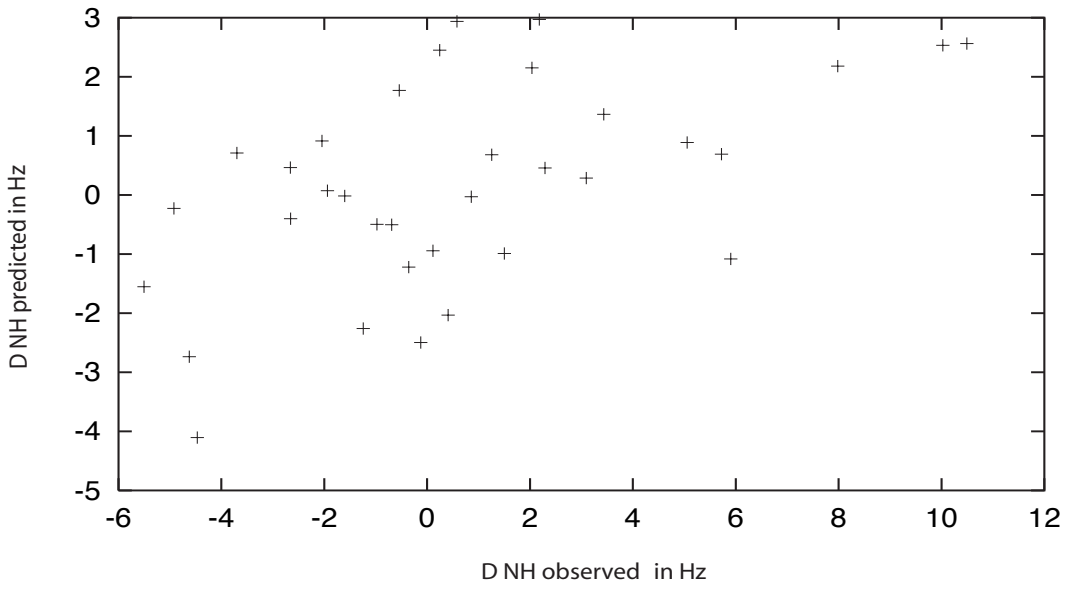

Figure 3.20: Correlation between experimental $^1D_{NH}$ values. Only those couplings from the residues that are in the regions where secondary structure elements predicted from the crystal structure and CSI values of the free NCoA-1$_{(257-370)}$ fragment are similar were used for correlation. A total of 33 dipolar couplings were used.

appropriately[1] aligned either sterically or electrostatically in existing alignment media commonly used for aligning proteins. Hence, it was not possible to obtain suitable[2] dipolar couplings that would give structural information in three dimensions.

[1]The alignment was either too strong, or there was no alignment because of protein precipitation.
[2]The number of couplings were not enough, and the magnitude of the couplings were small.

Chapter 4

Discussion

The NCoA-1$_{(257-385)}$ fragment and STAT6$_{(794-814)}$ eluted together in gel filtration trials indicating that the binding affinity of the peptide to the protein was very high. This was proved by the fact that MALDI (Matrix Assisted Laser Desorption and Ionization) experiments on the trypsin digested NCoA-1 fragments showed that the mass of the most stable fragment (14,795 Da.) corresponded to a fragment comprising of amino acids 257-385 of the NCoA-1 protein. N- terminal sequencing of the fragment by Edman degradation also confirmed the identities of the protein and the STAT6 transactivation domain. Secondary structure prediction showed that amino acids C- terminal to residue 366, in the NCoA-1$_{(257-420)}$, were disordered. Based on the limited proteolysis and secondary structure prediction two smaller constructs, the NCoA-1$_{(257-385)}$ and NCoA-1$_{(257-370)}$ were cloned. Crystallization screens were set up for these two shorter domains, in addition to theNCoA-1$_{(257-420)}$ fragment, by Ms. Karin Giller and Dr. Stefan Becker. All the three domains were used for the screening, each of the domains were in complex with STAT6$_{(794-814)}$ peptide, as well as in free form. One of the samples, the complex between STAT6$_{(794-814)}$ peptide and theNCoA-1$_{(257-385)}$ fragment crystallized. The structure was solved by Dr. Adelia Razeto and Dr. Stefan Becker, and the details of crystallization and structure elucidation have been published[179, 180].

The crystal structure revealed that the NCoA-1 $_{(259-367)}$ fragment in complex with STAT6$_{794-814}$ shows a PAS domain. PAS domains were first found in eukaryotes, and were named after homology to the *Drosophila* period protein (PER),the aryl hydrocarbon receptor nuclear translocator protein(ARNT) and the *Drosophila* single minded protein (SIM). PAS domains are located in the cytoplasm and are commonly

found in serine/threonine kinases [181], histidine kinases [182], photoreceptors and chemoreceptors [183], cyclic nucleotide phosphodiesterases [184], circadian clock proteins [185], voltage-activated ion channels [186], regulators of responses to hypoxia [187] and embryological development of the central nervous system [188].

The structure of the NCoA-1 PAS-B domain in complex with the STAT6 peptide now adds a new fold to this family of $\phi\chi\chi\phi\phi$ motif interacting domains. This nomenclature (PAS-B) was chosen as there is a another predicted PAS domain upstream, in the NCoA-1 molecule[90]. It extends the number of signal transduction modules, whose structures have been solved in complex with protein fragments containing unmodified signature motifs. Figures 4.6 and 4.1 shows the diagrammatic representations of the NCoA-$1_{(257-385)}$ and Photo active Yellow Protein (PYP) PAS folds respectively. It shows that NCoA-1 in complex with STAT6 transactivation domain adapts a canonical PAS domain.

The PAS domain of NCoA-1 consists of a five-stranded anti-parallel β-sheet from A_β to H_β and three α helices, C_α to E_α, which connect the second and third β strands, B_β and F_β (Figure 4.6). The STAT6 peptide binds between the helix D_α and strand B_β of the NCoA-1 PAS-B domain. In the STAT$6_{(794-814)}$ peptide, residues 799-807 form a two and a half turns α-helix, while residues 795-798 are in an extended conformation. The STAT6 helical axis is tilted both with respect to the axis of the D_α helix and to the B_β strand by an angle of approximately 50°(Figures 4.1 (a) and (b)).

Figures 4.2(a) and 4.2(b) show a comparison of the NCoA-1 PAS-B domain and the PAS domain of the Photoactive Yellow Protein(PYP) which has a similar basic topology. PYP is a bacterial light sensing protein, and together with the HERG (human ether-a-gogo-related gene) voltage-dependent K^+ channel, shares structural similarity with the NCoA-1 PAS-B domain according to the program DALI[189]. The whole structures of the PYP and HERG PAS domains when individually superimposed over the NCoA-1 PAS-B domain, have r.m.s.d. values of 3.0 $\mathring{A}$ and 2.7 $\mathring{A}$ respectively.

4.0.5 Interaction of the STAT6 peptide with the NCoA-1 PAS-B Domain

The electron density of the structured STAT6 peptide allows detailed analysis of interactions with the NCoA-1 PAS-B domain. The peptide is attached to binding cleft of the PAS-B domain, asymmetrically(Figure 4.1(a)). It is tightly packed against the

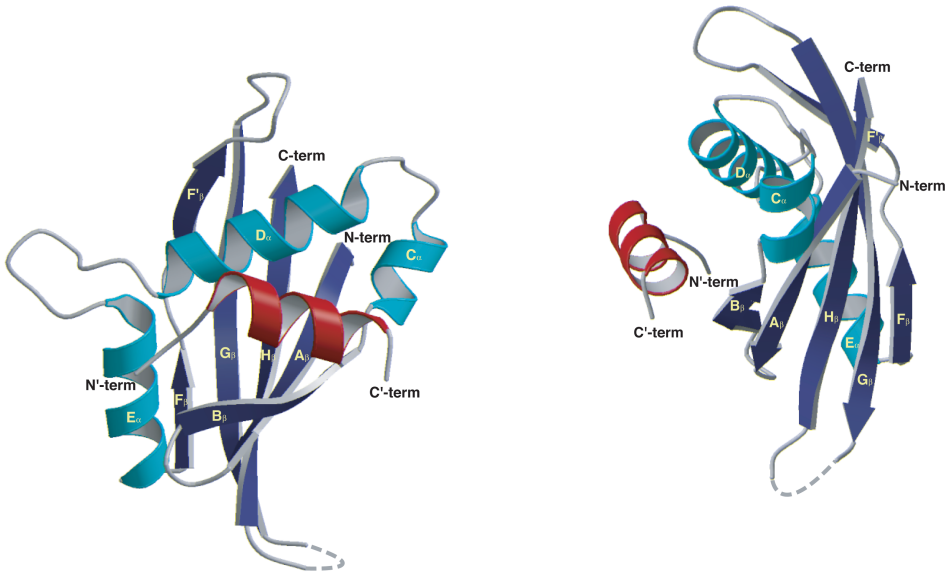

Figure 4.1: View of the STAT6/NCoA-1 PAS-B complex in mutually orthogonal orientations. The β sheet is depicted in dark blue and the α helix is shown in light blue. The STAT6 peptide is shown in red.

rim containing the *Arg293* anchor (Figure 4.3a). Due to this positioning, no ordered water molecules are found in this rim and the peptide. There is a relatively large space between the STAT6 peptide and the opposite rim. This space is occupied by a network of ordered water molecules, which mediate indirect interactions with the residues of the PAS-B domain. The STAT6 peptide binds into a shallow groove at the surface of the NCoA-1 PAS-B domain. The groove is about 20Å in length and 10 Å in width as estimated from C_α distances. Hydrophobic residues seal the floor of the groove (*Ile272*, *Ile273*, *Ile275* on the B_β strand, the methyl group of Thr277 on the C_α helix, *Trp288*, *Val292*, *Ile296* and *Phe300* on the D_α helix) forming a, "hydrophobic pocket", while the hydrophilic side-chains lining the walls of the groove, are directed outwards into the solvent region. The STAT6/NCoA-1 PAS-B domain interface buries 865 $Å^2$. The STAT6 peptide is amphipathic, presenting hydrophobic residues at the NCoA-1 interface and polar or hydrophilic residues on the solvent side(Figure 4.4). The Leu-sidechains of the STAT6 LXXLL signature motif are deeply embedded into the

hydrophobic floor of the PAS-B domain. The side chains of the *Leu'802* and*Leu'805* have the least accessible surface within the peptide, 6 $Å^2$ and 20 $Å^2$, respectively. Their substitution by alanine abolished binding to NCoA-1 *in vitro*, recruitment of NCoA-1 and consequently in its transactivation potential in cells[97]. Therefore not only the hydrophobicity of the motif, but also the size of the side-chains is essential for binding to the NCoA-1 PAS-B domain.This is clearly explicable by the crystal structure: the molecular surface of the peptide, at the signature motif, is complementary with the groove in the PAS domain(Figure 4.5).

Leu'802 fits perfectly into a deep pocket within the hydrophobic groove: it establishes Van der-Waals interactions with *Ile275* and *Ile296*. *Leu'805* fits into shallow pocket formed by *Ile272* and *Ile275* side chains and by C_α of *Ile273* and *Ser274*. *Leu'806* fits deep into deep pocket formed by the methyl group of *Thr277* and the side-chains of *Ile275* and *Trp288*. However *Leu'806* is more accessible (38$Å^2$) than *Leu'805* and *Leu'802*, because from one side it is not shielded from the solvent.

Pro'796 and *Pro'797* in the extended N-terminal part of the STAT6 peptide pack in a shallow depression formed by *Phe300* and *Ile296* at the C-terminal end of the PAS-B domain D_α helix. The STAT6 peptide is anchored by the *Arg293* side-chain through hydrogen bonds with the hydroxyl group of *Thr'803* and the backbone carbonyl of *Glu'799*. This anchor is strengthened by a weak salt bridge (3.17 Å)between the side-chains of *Arg293* and *Glu'799* (Figure4.4b). The hydrogen bond to the *Glu'799* carbonyl is the only direct contact of a PAS residue to the backbone of the STAT6 peptide. All other contacts between STAT6 and the PAS domain are side-chain to side-chain interactions (See Figures 4.3(a) and 4.3(b)). Therefore the peptide does not sit deep in the binding cleft, but is rather kept at a distance (Figures 4.3 and 4.3). The spacing between *Leu'802* and *Leu'805* is necessary to position these two residues and *Leu'806* in their respective pockets and only one orientation of the peptide in the groove is allowed.

A chargeclamp (hydrophilic clamp) is not found in the complex of the STAT6 peptide is not found in the complex of the STAT6 peptide with the NCoA-1 PAS-B domain. The side-chain of the Arg293 forms two hydrogen bonds and a salt bridge with the STAT6 peptide (Figure 4.4a and 4.4b). Due to the absence of such a chargeclamp, hydrophobic interactions play a central role in positioning the STAT6 fragment in the NCoA-1 PAS-B domain binding groove. Therefore the STAT6 fragment bound to the NCoA-1 PAS-B domain is the first activator/coactivator complex so far studied where

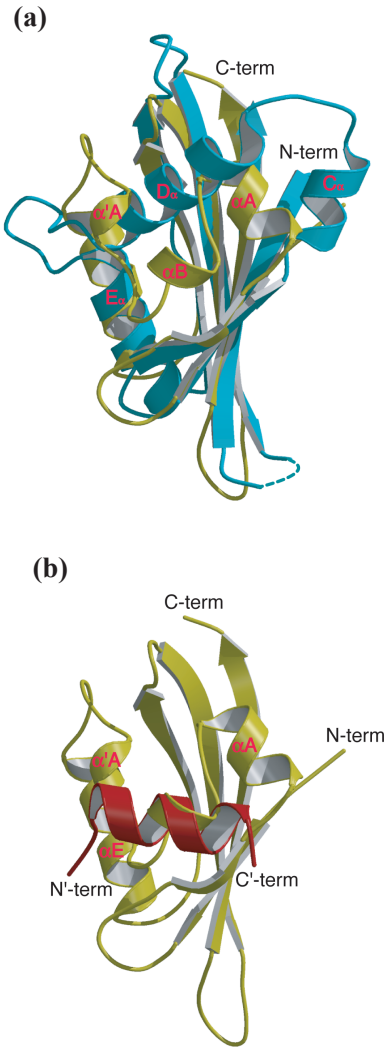

Figure 4.2: Superposition of the STAT6/NCoA-1 PAS B complex and the HERG PAS domain. The HERG domain is depicted in yellow, the STAT6 peptide is shown in red and the NCoA-1 PAS-B domain in light blue. For clarity the model of the STAT6 peptide in (a) and the model of the NCoA-1 PAS-B domain in (b) have been omitted.

the surface complementarity between the hydrophobic faces almost exclusively defines the orientation of the two proteins relative to each other.

Inspite of extensive research, the molecular basis of transactivation is poorly understood. One of the reasons is that there is little sequence similarity among transactivation domains. They generally have a poor intrinsic propensity to form secondary structure,

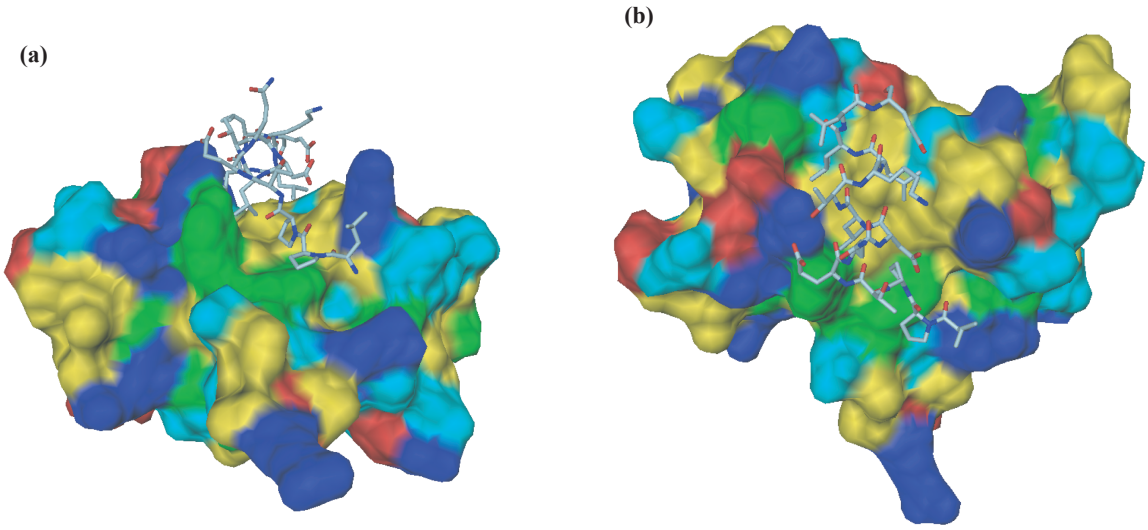

Figure 4.3: (a) Surface representation of the NCoA-1 PAS-B domain. Color code for the residues : acidic - red ; basic - blue; aliphatic - yellow; aromatic - green ; hydrophilic - cyan. (b) The complex is rotated by 90° about the horizontal axis of (b).

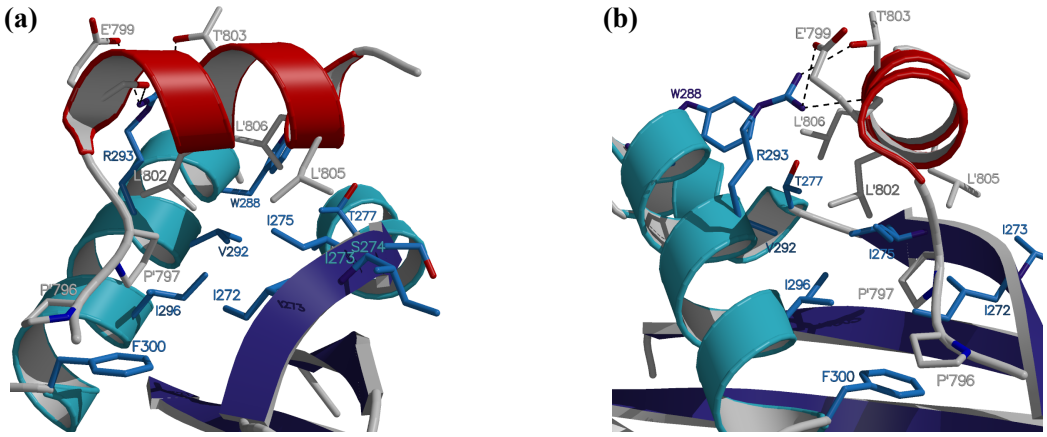

Figure 4.4: Close views of the interaction between the STAT6 peptide and the NCoA-1 PAS-B domain. (a) View along the helical axis of the STAT6 peptide. The residues of the STAT6 peptide involved in interactions are colored grey. Their partner residues in the NCoA-1 PAS-B domain are colored blue. Dotted lines indicate hydrophilic and electrostatic interactions. (b) Perpendicular view of the interface in an orientation rotated 90° about the vertical axis of (a).

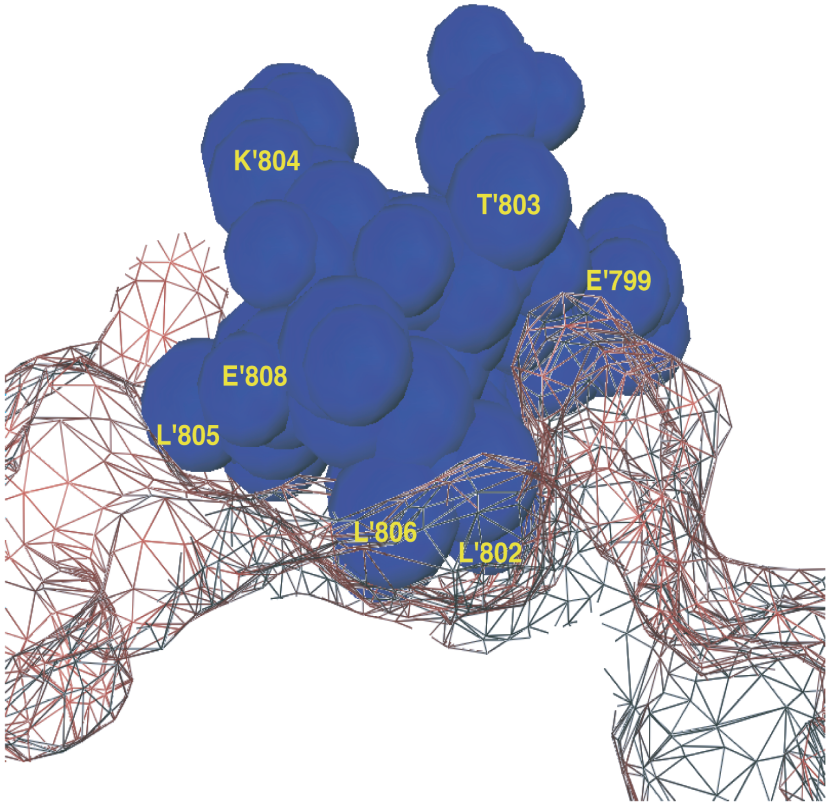

Figure 4.5: A cross section of the interface showing the complementarity of the STAT6 peptide with the PAS-B domain groove. The NCoA-1 PAS-B domain surface is represented as a wire mesh and the STAT6 peptide is in filling representation.

although they need to make specific interactions with several different target factors. This confirms the notion that minimal requirement for partner recognition is defined by very short stretches of amino acids, which become structured only when bound to target proteins. In the structure of the NCoA-1-PAS B domain complex with the STAT6(794-814) peptide the ordered part of the peptide (795-807) constitutes such a minimal structural entity. The LXXLL motif has been described before as a module mediating the interaction of the p160/SRC/NCoA cofactors with nuclear hormone receptors(NRs)[94]. In addition all the SRC/NCoA coactivators contain LXXLL motifs within a domain that mediates complex formation with the CBP, called the CBP Interacting Domain (CID).

There are a few cases where three dimensional structures of transactivation domains in complex with their coactivators have been solved. Those structures show a similar mechanism of transactivation as the STAT6 transactivation domain. The nuclear factor CREB activates transcription of target genes in part through direct interactions with the KIX domain of the coactivator CBP in a phosphorylation-dependent manner[190]. The solution structure of the complex formed by the phosphorylated kinase-inducible domain (pKID) of CREB with KIX reveals that pKID undergoes a coil to helix folding transition upon binding to KIX domain, forming two alpha helices. The amphipathic helix αB of pKID interacts with a hydrophobic groove defined by helices α1 and α3 of KIX. The other KID helix, αA, contacts a different face of the α3 helix. The phosphate group of the crucial phosphoserine residue of pKID is hydrogen bonded to the side chain of Tyr-658 of KIX. The structure provides a model for interactions between other transactivation domains and their targets. The structure of this complex shows that activating regions do fold and become ordered. Though KID is disordered in the free state, it folds into a helix upon binding to the KIX domain of the CBP. This adds to the previous notion that no specific structure is required for the recruitment process[191]. When the αA helix of the KID was destablilized, with mutations, the complex formation with the KIX domain of the CBP was disrupted, suggesting a role for the secondary structure in transcriptional activation[192]. NMR and biochemical evidence suggest that the minimal acidic domain of the herpes simplex virus VP16 undergoes an induced transition from a random coiled state to an α helix upon binding to its target protein human TFIID TATA box-binding protein associated factor[193]. This work attempted to tackle the question of the role of acidic residues in transactivation. The results from this work show that hydrophobic residues of the transactivation domains also play an important role. The Murine Double Minute 2 (MDM2) oncoprotein is a cellular inhibitor of the p53 tumor suppressor. It can bind the transactivation domain of p53 and down regulate its ability to activate transcription. The crystal structure of the 109-residue amino-terminal domain of MDM2 bound to a 15-residue transactivation domain peptide of p53 reveals a deep hydrophobic cleft in the MDM2 on which the p53 peptide binds as an amphipathic α helix[194]. The interface relies on the steric complementarity between the MDM2 cleft and the hydrophobic face of the p53 α helix and, in particular, on a triad of p53 amino acids–Phe19, Trp23, and Leu26–which insert deep into the MDM2 cleft[194]. These same residues of p53 are also involved in transactivation, supporting the hypothesis that MDM2 inactivates p53 by concealing its

transactivation domain. The structure also suggests that the amphipathic alpha helix may be a common structural motif in the binding of a diverse family of transactivation factors to the TATA-binding protein-associated factors.

4.0.6 The PAS domain is a new interaction module for LXXLL motif

While the LXXLL and similar motifs form structurally homologous amphipathic helices, their target proteins share no structural similarity. The binding grooves of the NR AF2 domain, the CBP KIX domain, the CBP glutamine rich domain and the NH2- terminal MDM2 domain are mostly made of different helical folds. The structure of the NCoA-1 PAS-B is a new fold in the existing family of $\phi\chi\chi\phi\phi$ motif interacting domains. It extends the number of signal transduction modules, whose structures have been solved in complex with protein fragments containing unmodified signature motifs. The importance of PAS domain in the NCoA/SRC activator function is not clear. A possible role in the formation of coactivator multiprotein complex has been suggested. Several three dimensional structures of eukaryotic proteins have already been solved. The crystal structure of NCoA-1/STAT6 complex is the first structure where detailed insight into the interaction of a PAS domain with its ligand is available. This forms the basis for studying this domain of NCoA-1 with other interaction partners and in general, the interaction of other PAS domains with their protein or peptide ligands.

4.0.7 The binding specificity between NCoA-1 and STAT6

NCoA-1 and not the proteins related to same family, NCoA-2 and NCoA-3, interact with STAT6[84]. To confirm that the binding specificity depends on the PAS-B domain, the interactions of the *in vitro* translated homologous PAS-B regions from the three NCoA isoforms, with the STAT6 transactivation domain were investigated in a GST pulldown experiment[97]. As expected, the NCoA-1 PAS-B domain strongly interacted with the GST fusion protein comprising of the residues 677-847 of the STAT6 transactivation domain.

4.0.8 Determinants of Binding Specificity

A sequence alignment of residues 259-367 of the NCoA-1-PAS B domain with the homologous sequences 267-376 of NCoA-2 and 262-370 of NCoA3 demonstrates, that those residues making specific-side chain contacts to the STAT6 peptide, are all well conserved or highly homologous between the three PAS domain sequences. Based on the available structure of the STAT6/NCoA-1 PAS-B complex we cannot identify individual residues in the NCoA-1 PAS-B domain, which cause specific binding to the STAT6 LXXLL motif. The inability of NCoA-2 and NCoA-3 to bind to STAT6, despite of the high sequence conservation in their PAS regions, strongly suggests that surface complementarity plays an essential role not only for the orientation of the STAT6 peptide in the NCoA-1 binding groove, but also for binding specificity. A dissociation constant in the order of 10^{-1} micromoles shows that the strong affinity for the STAT6 transactivation domain *in vitro*, under near physiological conditions shows that that PAS domains are able to function as interaction domains in intracellular signaling. The K_D value is also in accordance with the values found for intracellular protein-protein interactions. It was also clear from isothermal calorimetry that the interaction between NCoA-1 PAS-B domain is exothermic, and hence there is no additional energy needed for this particular step of signaling.

4.1 PAS domain

The first 3D structure of a PAS domain determined was the structure of the *Ectothiorhodospira halophila* blue-light photo receptor PYP (Photoactive Yellow Protein). PYP undergoes a self contained light cycle. Light-induced trans-to-cis isomerization of the 4-hydroxy-cinnamic acid chromophore and coupled protein rearrangements produce a new set of active hydrogen bonds. Resulting changes in shape, hydrogen bonding and electrostatic potential at the surface form a likely basis for signal transduction [195]. In the years after, a number of PAS domains have been determined. These include the 3D structure of the heme-binding domain of the rhizobial oxygen sensor FixL, from *Bradyrhizobium japonicum* [196] and from *Rhyzobium meliloti* [197]. The structural resemblance of the FixL heme domain to the PYP indicates the existence of a PAS structural motif, although both proteins are functionally different. The N-terminal domain of the human ether-a-go-go-related potassium channel, HERG which is the

first 3D model of a eukaryotic PAS domain [198], the FMN containing phototropin module of the chimeric fern Adiantum photoreceptor [199], and the solution structure of the N-terminal PAS domain of human PAS kinase [200], and the PAS like domains, periplasmic ligand binding domain of the sensor kinase, CitA [201] and the sensory domain of the two component fumarate sensor, DcuS [202] have seen solved. The beta-sheet of all the eight structures superimpose well [203].

Ponting and Aravind (1997) have proposed that the PAS sequence can be split into two separate regions, the PAS domain and the PAC motif. Although the PAC motif is proposed to contribute to the PAS domain structure, many PAS sequences in the SMART and the PFAM databases are not linked to a PAC motif, which raises a possibility of the existence of differences in the PAS domain superfamily. In the PFAM database there are sequences of almost of 1000 PAS domains representative of all kingdoms of life. Structural analysis of PAS domains in the PDB demonstrates that the PAS and PAC motifs split the five-stranded β-sheet into two sections. In the large scale modeling presented by [203], the PAS and PAC motifs are inseparable and together give rise to a structural fold.

The PAS fold represents an important sensory domain present in all kingdoms of life, and in the PFAM database many proteins appear to have more than one PAS domain. It is therefore possible that such proteins may utilize cofactors in multiple PAS domains to integrate different environmental signals. It has been shown that certain artificial chemicals can specifically bind into the well packed hydrophobic core of a PAS domain from PAS kinase [200], producing significant conformational changes therein. Studies of two PAS domains with natural cofactors, AsLOV and photoactive yellow protein, have shown that relatively small changes in these internally bound compounds are sufficient to displace α-helices bound onto the same β-sheet interface [204].

One class of PAS domains is the light-oxygen-volta-ge LOV domains which bind to flavin chromophores. Biophysical studies to define this pathway have produced conflicting data, because crystallographic studies of LOV domains show minimal light-induced conformational changes, whereas spectroscopic data suggest more pronounced structural rearrangements occur upon illumination. These conflicts were resolved by NMR spectroscopy, where it was shown that structural elements outside of a PAS domain core undergo light-induced conformational changes and thus provide a pathway for propagating structural changes to the kinase domain.

Among all the structurally characterized interactions of PAS domains with their

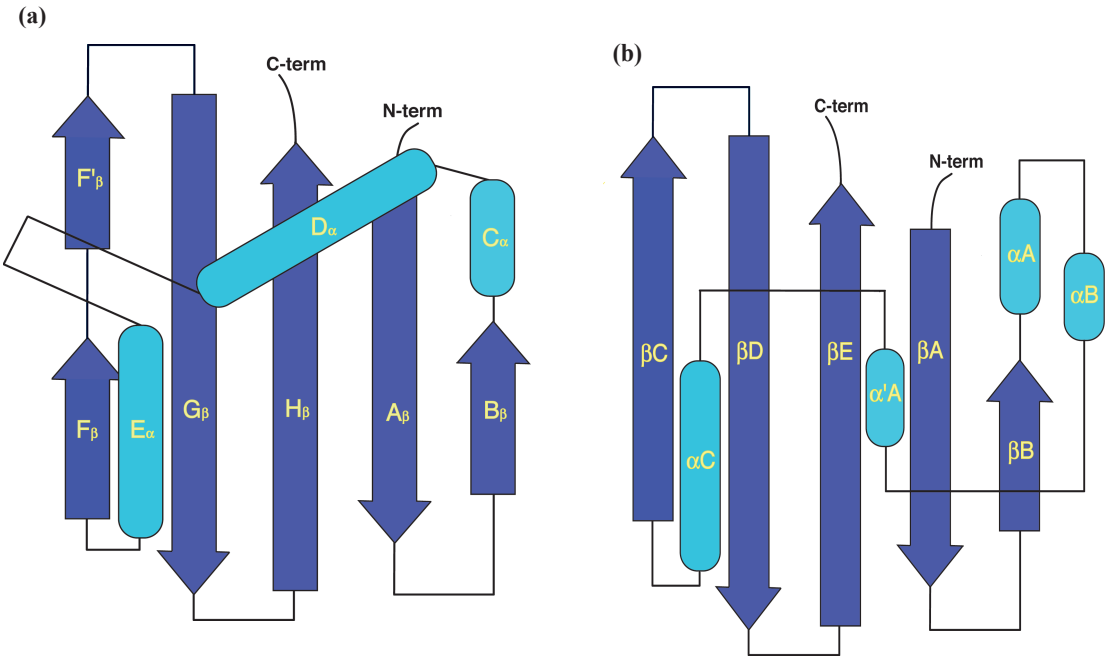

Figure 4.6: (a)The PAS fold the the NCoA-1. (b) Pictorial representation of the (Photoactive yellow protein domain (PYP)

ligands, the PAS domain complex of NCoA-1 and STAT6 transactivation domain is the first PAS domain to be solved along with a peptide ligand. Although the crystal structure of NCoA-1 PAS-B domain in complex with the STAT6 peptide, provides some detailed insight into the mechanism of transactivation by the recruitment of cofactors, it does not reveal the conformational changes in the cofactor that facilitate the specificity of this interaction. Therefore the results from the NMR studies of the free form are particularly important. Although PAS domains share a common fold, they have little sequence homology, except for a few residues which are conserved[205]. Most of these residues are located in the helix αA, which appears to act as a lid on the ligand binding pocket and undergoes conformational changes[195]. In the case of the free form of NCoA-1 however, as shown by the chemical shift index calculation, the conformational change is likely to be in the first β sheet. As shown by the titration studies, while majority of the residues do not undergo a major rearrangement, some of the residues do show significant changes in their location in two dimensional spectra. It is likely that the free form the amino acids from *Lys271* to *Lys280* constitute an α helix whereas in the complex they undergo a conformational change to form a part of the β sheet. This is in

confirmation with the previous studies on PAS domains where the αA/B segment moves in concert, and they all share not only a common structural motif but also a similar way of conformational flexibility. This evolution could have possibly have facilitated to accommodate the various input signals from different ligands and cofactors located at different positions in the domain and to transmit the signaling event to downstream transducer proteins.

4.2 Structural information from NMR

In order to calculate the three dimensional structure of proteins, structural constraints from Nuclear Overhauser Effect (NOE) alone or additionally Residual Dipolar Couplings (RDCs) should be obtained. NOEs are dependent on the distance and are short range interactions. Since fold determining constraints come from side chain contacts in the hydrophobic core of the proteins, side chain resonance assignments are a prerequisite for calculation of protein structures using NOEs alone. But the experiments that would facilitate the assignment of side chain resonances are relatively insensitive when compared to those used for backbone resonance assignments. Assignment of side chain resonances and NOE identification are time consuming. Measurement of RDCs was considered as an option for the assessment of the backbone structure of the free form of NCoA-1. However, use of RDCs alone for backbone structure elucidation is not without problems. The inherent ambiguity present in RDCs with respect to orientation, entails that additional alignment tensors are required to resolve the equivocacy of vectorial information. Since the alignment tensors are also dependent on the nature of the liquid crystal used for alignment, more such tensors can be obtained by using different alignment media. In the case of involvement of electrostatic attraction between the solute molecules and the medium as in the case of Pf1 phages, varying degrees of alignment and hence the corresponding tensors can be obtained simply by changing the ionic strength of the buffer used. Measurement of RDCs for NCoA-1 in various alignment media were marred by many factors like precipitation of the protein, unfavorable line broadening impeding with accurate estimation of dipolar couplings, and lack of signal. The reasons are manifold.

Some proteins exhibit extreme line broadening at Pf1 phage concentrations (20-50 ml/L) used to align acidic biomolecules, which would appear to limit the general applicability of such a technique[206]. Ojennus *et al.*,[207] reported that a *Src* homology domain, when aligned with an anomalously low concentration of the Pf1 phage, exhibited a low affinity binding mechanism that leads to an enhanced alignment. They showed that, at 2 mg/ml an optimum degree of alignment was achieved. Nitrogen linewidths are broadened approximately twofold at this concentration and higher concentrations of phage result in a prohibitively large line broadening. At this low concentration of phages required to obtain optimum alignment, modest line broadening does not interfere with the accurate measurement of large dipolar couplings. This effect was more pronounced in the case of the NCoA-1 domain, where large linewidths led to a spectrum of poor quality, making the evaluation of dipolar couplings difficult. Pf1 is negatively charged at pH 7.4, with an approximate linear charge density of 10 e/nm, and has a molecular weight 3.75×10^7 g/ml [208]. NCoA-1() is a highly basic protein with a pI of 9.6 and positively charged at pH 6.5. There could possibly be a high electrostatic interaction between the NCoA-1 and the Pf1 phages. The subunits of the Pf1 coat protein are connected to each other by extensive hydrophobic interactions. The exterior of the phages is composed of an elaborate array of N- terminal double hooks, made of six amino acids, Gly1 to Ser6[209]. It is probable that the hydrophobic groove of the NCoA-1 interacts with the exterior of the Pf1 phages which could possibly explain the irreversibility of the line broadening upon increasing ionic strength. Lowering the phage concentration to 1-2 mg/ml reduces the degree of induced order, but need not solve the line broadening problem [207]. Addition of salt increases the hydrophobicity of the protein, thereby facilitating other non-ionic interactions with the phage particles. Broadening also occurs when nucleic acids, apocalmodulin, and thioredoxin are aligned with phages [206]. But there is no significant change in the $T_{1\rho}$ and T_2 relaxation time constants[206]. The observed line broadening could be attributed to presence of additional ^{1}H -^{1}H dipolar couplings rather than a change in the correlation time. Filamentous phage are highly acidic[208] with a pI of 4.0. This could also possibly be the reason for a very high attraction exhibited by the NCoA-1 to the phage particles. Since the quality of the ^{15}N edited $T_{1\rho}$ spectra of the NCoA-$1_{(257-370)}$ aligned in Pf1 phages were also very poor with no well resolved and dispersed peaks it was not possible to estimate relaxation times for the protein aligned with the phages.

In the case of bicelles, the poor quality of the spectra could be attributed to the destructive effect of the salt concentration on the stability of bicelles. Since bicelles are unstable over a period of time usually a few days to a week, typical time length used for multidimensional NMR experiments, the combined effect of the high salt and commercial impurities lead to hydrolysis of the phospholipids making the whole environment hydrophobic[157]. The NCoA-1 protein is free to interact through its hydrophobic patch with the hydrolysed lipids, leading to instability of samples. It is not known, however, the reason for the inefficiency of CTAB to stabilize the sample. There is always a hydration term as well as the repulsive part of the Van der Waals interaction that keeps the bilayers from collapsing. But, a secondary minimum can arise at longer distances under certain combinations of electrostatic and Van der Waals terms. As the electrostatic term is modulated by ionic strength, these minima occur only above a certain salt concentration. It is these minima that are responsible for micelle and bilayer aggregation. For this to occur, the bilayer or micelle should be inherently charged, before this ionic strength dependent phenomenon happens. However, in the absence of weak charge there can be minima which result from the interplay of hydration and Van der Waals forces that also lead to aggregation. It is known that neutral lipid bilayers have an inherent tendency to aggregate in multilayer particles. The effects of ionic strength on what appears to be bicelle aggregation appears to be reminiscent of those on the aggregation of other bilayer and micelle systems [210]. There are two major forces that act between surfaces of lipids and influence interactions, such as the aggregation of bilayers. Van der Waal's forces occur between all solute molecules in water. At long distances, these forces are attractive [211]. If the interacting surfaces possess like partial charges, they too experience a repulsive electrostatic force. The total potential energy of interaction can be determined by the addition of repulsive and attractive potential energies. Commercially available DMPC and DHPC are only around 99% pure and rest is impurities which can be charged. Additional hydrolysis of phospholipids produces either a charged lipid or a partially ionized fatty acid. The released fatty acid could contribute to an amount of surface charge sufficient to stabilize, the neutral bicelles in water.

Other types of alignment media like polyacrylamide gels and paramagnetic tags were also used for alignment of the free NCoA-$1_{(257-370)}$ fragment. The high concentration of polyacrylamide needed to ensure mechanical stability is a serious limitation of

neutral polyacrylamide gels, for alignment of proteins. The high concentration of acrylamide will decrease the pore size of the gel and hence the rotational diffusion of the protein will be heavily affected even for weak alignment conditions leading to extensive line broadening[212]. When the free NCoA-$1_{(257-370)}$ fragment and the NCoA-$1_{(257-370)}$ fragment subjected to reaction with the paramagnetic tag were analyzed by mass spectroscopy, the expected and observed molecular weights did not match for both the free NCoA-$1_{(257-370)}$ fragment used as a control and NCoA-$1_{(257-370)}$ fragment subjected to reaction with the tag. The discrepancy between the experimental molecular weight and expected weight of the tagged NCoA-1 does not have an explanation.

4.2.1 Scope for further studies

Residual dipolar couplings obtained in the case of NCoA-1 were not suitable for structural analysis because of the fact they were too small to be outside the statistical error range. The reason for this was that the nature of the protein domain did not permit the right alignment conditions to obtain appropriately large dipolar couplings. A number of alignment media for proteins have been developed but none of them have universal application. Disk shaped phospholipid particles or bicelles have narrow temperature ranges and crystalline conditions are established upon a very delicate phase equilibrium, which can be disrupted by the soluble macromolecules. Since the filamentous phages are charged they tend to interact unfavorably with oppositely charged macromolecules which could destroy the nematic phase or induce excessive order resulting in unacceptably broad lines. Polyacrylamide gels should be highly concentrated for mechanical stability but this could result in difficulties in diffusing the protein samples into the gels. Paramagnetic tags cannot be used where there are too many cysteins in a polypeptide available for reaction with the tag or where none are available at all. Hence there is clearly a need to develop newer alignment media with wider applicability. In the case of NCoA-1 since the RDC based approach did not lead to any reasonable conclusion, the conventional methods of side chain resonance assignment along with NOE data could be used to calculate the structure.

It has been shown by Litterst *et al.*, that the amino terminal region of NCoA-1 comprising of amino acids 1 - 381 is required for its interaction with the transactivation

domain of another STAT protein, the STAT5. This region overlaps with the NCoA-1 domain used in the present study. Since the interaction between the NCoA-1 and STAT5 transactivation domain has not yet been structurally characterized, the backbone resonance assignments for a part of the amino terminal region of NCoA-1 from residues 257-370 could be used to investigate whether the same region of NCoA-1 that is essential for its interaction with the STAT6 transactivation domain is also involved in the interaction with STAT5 transactivation domain. It would be interesting to study the dynamics of such an interaction.

Bibliography

[1] M. Ptashne, Regulated recruitment and cooperativity in the design of biological regulatory systems, Phil. Trans. R. Soc. Lond. A 361 (2003) 1223–1234. 1

[2] D. Baltimore, Our genome unveiled., Nature 409 (2001) 814–816. 1

[3] T. Maniatis, S. Goodbourn, J. Fischer, Regulation of inducible and tissue-specific gene expression., Science 236 (1987) 1237–45. 1, 1

[4] B. Gravely, Alternative splicing : increasing diversity in the proteomic world, Trends in Genetics 17 (2001) 100–107. 1

[5] A. Kornblihtt, M. de la Mata, J. Fededa, M. Munoz, G. Nogues, Multiple links between transcription and splicing, RNA 10 (2004) 1489–98. 1

[6] F. Sauer, R. Tjian, Mechanisms of transcriptional activation : differences and similarities between yeast, drosophila and man., Current Opinion in Genetics and Development 7 (1997) 176–81. 1

[7] M. Levine, R. Tjian, Transcriptional regulation aand animal diversity, Nature 424 (2003) 147–151. 1

[8] G. Plowman, S. Sudarsanam, J. Bingham, D. Whyte, T. Hunter, The protein kinases of c. elegans : A model for signal tranduction in multicellular organisms., Proceedings of the National Academy of Sciences, USA 96 (1999) 13603–13610. 1

[9] E. Davidson, Genomic Regulatory Systems : Development and Evolution, Academic Press, 2001. 1

[10] J. Wyrick, R. Young, Deciphering gene expression regulatory networks, Current Opinion in Genetics and Development 12 (2002) 130–136. 1

[11] T. Pawson, Organization of cell-rgulatory systems through modular-protein-interaction domains., Philosophical Transactions of the Royal Society of London. A 361 (2003) 1251–1262. 1, 1, 1.0.2

[12] W. A. Lim, The modular logic of signaling proteins: building allosteric switches from simple binding domains, Current Opinion in Structural Biology 12 (2002) 61–68. 1, 1

[13] G. Cohen, R. Ren, D. Baltimore, Modular binding domains in signal transduction proteins., Cell 80 (1995) 237–248. 1

[14] T. Pawson, Protein modules and signaling networks., Nature 373 (1995) 573. 1

[15] R. Young, Rna polymerase ii., Annu Rev Biochem 60 (1991) 689–715. 1, 1

[16] P. Mitchell, R. Tjian, Transcriptional regulation in mammalina cells by sequence specific dna binding proteins, Science 245 (1989) 371–8. 1

[17] M. Crossley, S. Orkin, Regulation of beta-globin locus, Current Opinion in Developmental Genetics 3 (1993) 232–7. 1

[18] C. Pabo, R. Sauer, Transcription factors : structural families and principles of dna recognition., Annual Reviews in Biochemistry 61 (1992) 1053–95. 1, 1

[19] Z. Arany, W. Sellers, D. Livingston, R. Eckner, E1a-associated p300 and creb-associated cbp belong to a conserved family of coactivators., Cell 77 (1994) 799–800. 1

[20] E. Korzus, J. Torchia, D. Rose, L. Xu, R. Kurokawa, E. MacInerney, T. Mullen, C. Glass, M. Rosenfeld, Transcription factor-specific requirements for coactivators and their acetyltransferase functions., Science 279 (1998) 703–707. 1, 1.2.3, 1.2.8

[21] S. Smale, J. Kadonaga, The rna polymerase core promoter., Annual Reviews in Biochemistry. 1

[22] R. Roeder, Role of general and gene-specific cofactors in the regulation of eukaryotic transcription., in: CSH Symp. Quant. Biol., 1998. 1

[23] R. White, S. Jackson, The tata-binding protein: a central role in transcription by rna polymerases i, ii and iii., Trends in Genetics 8 (1992) 284–8. 1

[24] F. Sauer, S. Hansen, R. Tjian, Dna template and activator-coactivator requirements for transcriptional synergism by drosophila bicoid, Science 270 (1995) 1825–8. 1

[25] B. e. a. Burgess-Beusse, The insulation of genes from external enhancers and silencing chromatin., Proceedings of the National Academy of Sciences, USA 99 (2002) 16433–16437. 1

[26] A. Webber, R. Ingram, J. Levorse, S. Tilghman, Location of enhancers is essentials for the imprinting of h19 and igf2 genes., Nature 391 (1998) 711–715. 1

[27] A. Koleske, R. Young, An rna polymerase ii holoenzyme responsive to activators, Nature 3 (1999) 824–828. 1

[28] D. Bentley, Regulation of transcriptional elongation by rna polymerase ii, Current Opinion in Developmental Genetics 5 (1995) 210–16. 1

[29] W. Becker, L. Kleinsmith, J. Hardin, J. Raasch, The World of the Cell, 2000. 1

[30] R. Fassler, K. Martin, E. Forsberg, T. Litzenburger, A. Iglesias, Knockout mice: how to make them and why. the immunological approach., Int Arch Allergy Immunol. 106 (1995) 323–34. 1

[31] D. Martin, S. Tsai, S. Orkin, Increased gamma-globin expression in a nondeletion hpfh mediated by an erythroid-specific dna-binding factor., Nature 338 (1989) 435–8. 1

[32] V. Calhoun, A. Stathopoulos, M. Levine, Promoter-proximal tethering elements regulate enhancer-promoter specificity in the drosophila antennapedia complex., Proceedings of the National Academy of Sciences, USA 99 (2002) 9243–9247. 1

[33] T. Pawson, P. Nash, Protein-protein interactions define specificity in signal transduction., Genes and Development 14 (2000) 1027. 1

[34] M. Noble, J. Endicott, L. Johnson, Protein kinase inhibitors: insights into drug design from structure., Science 303 (2004) 1800–5. 1

[35] E. Lander, L. Linton, B. Birren, C. Nusbam, M. Zody, J. Baldwin, K. Devon, k. Dewar, M. Doyle, W. e. a. Fitzhugh, Initial sequencing and analysis of the human genome., Nature 409 (2001) 860–921. 1, 1.0.1

[36] C. Ponting, J. Schultz, R. Copley, M. Andrade, P. Bork, Evolution of domain families., Advances in Protein Chemistry 54 (2000) 185–244. 1

[37] S. Gosh, D. Baltimore, Activation in vitro of nfkb by phosphorylation of its inhibitor ikb, Nature 344 (1990) 678–682. 1.0.1

[38] R. Copley, T. Doerks, I. Letunic, P. Bork, Protein domain analysis in the era of complete genomes., FEBS Letters 513 (2002) 129. 1.0.1

[39] P. Nash, Multisite phosphorylation of a cdk inhibitor sets a threshold for the onset of dna replication, Nature 414 (2001) 514. 1.0.1

[40] J. Schlessinger, Sh2 and sh3 signaling proteins, Current Opinion in Genetics and Development 4 (1994) 25–30. 1.0.1

[41] B. Mayer, Sh3 domains: complexity in moderation., Journal of Cell Sciences 114 (2001) 1253. 1.0.1

[42] T. et al, Tsukazaki, Cell 95 (1998) 779. 1.0.1

[43] M. et al. Huse, Molecular Cell 8 (2001) 481. 1.0.1

[44] J. et al. Wu, Molecular Cell 8 (2001) 1277. 1.0.1

[45] J. et al. Massague, Nature Reviews in Molecular Cell Biology 1 (2000) 169. 1.0.1

[46] B. J. Mayer, Protein-protein interactions in signaling cascades, Mol Biotechnol 13 (3) (1999) 201–13. 1.0.1

[47] J. Kuriyan, D. Cowburn, Modular peptide recognition domains in eukaryotic signaling., Annual Reviews in Biophysics and Biomolecular Structure 26 (1997) 259–288. 1.0.1

[48] C. Heldin, Dimerization of cell surface receptors in signal transduction., Cell 80 (1995) 213–223. 1.0.2

[49] S. Triezenberg, Structure and function of transcriptional activation domains., Current Opinion in Genetics and Development 5 (1995) 190–196. 1.0.3

[50] G. Stark, I. Kerr, B. Williams, R. Silverman, R. Schreiber, How cells respond to interferons., Annual Reviews in Biochemistry 67 (1998) 227–264. 1.2

[51] J. Darnell, J. E., I. M. Kerr, G. R. Stark, Jak-stat pathways and transcriptional activation in response to ifns and other extracellular signaling proteins, Science 264 (5164) (1994) 1415–21. 1.2, 1.2

[52] D. Levy, J. Darnell, Stats: Transcriptional control and biological impact, Nature Reviews 3 (2002) 651–662. 1.2

[53] T. Hoey, U. Schindler, Stat structure and function in signaling, Curr Opin Genet Dev 8 (5) (1998) 582–7. 1.2, 1.3

[54] J. Hoeck, M. Woisetschlager, Stat6 mediates eotaxin-1 expression in il-4 or tnf-alpha-induced fibroblasts, J Immunol 166 (7) (2001) 4507–15. 1.2

[55] K. Takeda, T. Tanaka, W. Shi, M. Matsumoto, M. Minami, S. Kashiwamura, e. a. Nakanishi, K, Essential role of stat6 in il-4 signaling, Nature 380 (1996) 627–630. 1.2

[56] K. Shimoda, J. Duerson, M. Sangster, S. R. C. Sarawar, R. Tripp, Lack of il-4 induced th2 response and ige class switching in mice with disrupted stat6 gene., Nature 380 (1996) 630–633. 1.2

[57] M. H. Kaplan, A. L. Wurster, S. T. Smiley, M. J. Grusby, Stat6-dependent and -independent pathways for il-4 production, J Immunol 163 (12) (1999) 6536–40. 1.2

[58] J. OShea, Jaks, stats, cytokine signal recognition and immune recognition. are we there yet?, Immunity 7 (1997) 1–11. 1.2

[59] T. Kawata, A. Shevchenko, M. Fukuzawa, K. Jermyn, N. Totty, N. Zhukovskaya, A. Sterling, M. Mann, J. Williams, Sh2 signaling in a lower eukaryote : a stat protein that regulates stalk differentiation in dictyostelium., Cell 89 (1997) 909–916. 1.2, 1.2.2

[60] R. Yan, S. Small, C. Desplan, C. Dearolf, J. Darnell, Identification of a stat gene that functions in drosophila development, Cell 84 (1996) 421–430. 1.2

[61] Z. Zhang, S. Jones, J. Hagood, N. Fuentes, G. Fuller, Stat3 acts as a co-activator of glucocorticoid receptor signaling., Journal of Biological Chemistry 272 (1997) 30607–30610. 1.2.1

[62] D. Look, M. Pelletier, R. Tidwell, W. Roswit, M. Holtzman, Stat1 depends on the transcriptional synergy with sp1., Trends in Biochemical Sciences 270 (1995) 30264–30267. 1.2.1

[63] T. Mikita, C. Daniel, P. Wu, U. Schindler, Mutational analysis of stat6 sh2 domain, Journal of Biological Chemistry 273 (1998) 17634–17642. 1.2.1, 1.2.2

[64] U. Vinkemeier, S. L. Cohen, I. Moarefi, B. T. Chait, J. Kuriyan, J. E. Darnell, Dna binding of in vivo activated stat1alpha and stat1beta and truncated stat1 : Interaction between nh2-terminal domains stabilizes binding of two dimers to tandem dna sites, EMBO J. 15 (1996) 5616–5626. 1.2.2

[65] U. Vinkemeier, I. Moarefi, J. E. Darnell, J. Kuriyan, Structure of the amino-terminal protein interaction domain of stat-4, Science 279 (1998) 1048–1052. 1.2.2

[66] X. Xu, Y. Sun, T. Hoey, Cooperative dna binding and sequence selective recognition conferred by the stat amino terminal domain., Science 273 (1996) 794–797. 1.2.2, 1.2.2

[67] et al. Horvath, Trends in Biochem. Sciences 25 (2000) 496–502. 1.2.2

[68] T. Murphy, E. Giesal, J. Farrar, K. Murphy, Role of the stat4 n domain in receptor proximal tyrosine phosphorylation, Mol. Cell. Biol. 20 (2000) 7121. 1.2.2

[69] S. Becker, B. Groner, M. C.W., Three-dimensional structure of the stat3beta homodimer bound to dna, Nature 394 (1998) 145–51. 1.2.2, 1.2

[70] X. Chen, U. Vinkemeir, Y. Zhao, D. Jerusalami, J. Darnell, J. Kuriyan, Crystal structure of the stat-1 dimer bound to dna, Cell 93 (1998) 827–839. 1.2.2, 1.2.2

[71] A. Begitt, T. Meyer, M. van Rossum, U. Vinkemeier, Nucleocytoplasmic translocation of stat1 is regulated by a leucine-rich export in the coiled-coiled domain, PNAS 97 (2000) 10418–10423. 1.2.2

[72] T. Zhang, W. Kee, K. Seow, W. Fung, X. Cao, The coiled-coil domain of stat3 is essential for its sh2 domain-mediated receptor binding and subsequent activation induced by epidermal growth factor and interleukin-6, Mol. Cell. Biol. 20 (7132-7139). 1.2.2

[73] U. Schindler, P. Wu, M. Rothe, m. Brasseur, S. McKnight, Components of stat recognition code : evidence for two layers of molecular selectivity., Immunity 2 (1995) 689–697. 1.2.2, 1.2.3

[74] H. Seidel, L. Milocco, P. Lamb, J. J. Darnell, R. Stein, J. Rosen, Spacing of palindromic half sites as a determinant of selective stat (signal transducers and activators of transcription) dna binding and transcriptional activity., Proceedings of the National Academy of Sciences, USA 92 (1995) 3041–3045. 1.2.2

[75] E. Yang, Z. Wen, R. L. Haspel, J. J. Zhang, J. E. D. Jr., The linker domain of stat1 is required for gamma interferon-driven transcription, Mol. Cell. Biol. 19 (1999) 5106–5112. 1.2.2

[76] G. S, B. T, W. AJ, C. J, S. HK, D. B, D. RJ, Selective interaction of jnk protein kinase isoforms with transcription factors., EMBO J. 15 (1996) 1075–1084. 1.2.2

[77] K. Shuai, Modulation of stat signaling by stat-interacting proteins, Oncogene 19 (2000) 2638–44. 1.2.2

[78] et al. Schindler, Immunity 2 (1995) 689–697. 1.2.2

[79] et al. Schinder, Science 257 (1992) 809–813. 1.2.2

[80] U. Schindler, I. Strehlow, Cytokines and stat signaling, Adv. Pharmacol. 47 (2000) 113–174. 1.2.2

[81] et al. Decker, Oncogene 19 (2000) 2628–2637. 1.2.2

[82] A. Arimura, M. van Peer, A. J. Schroeder, P. B.Rothman, The transcriptional co-activator p/cip (ncoa-3) is up-regulated by stat6 and serves as a positive regulator of transcriptional activation by stat6, J. Biol. Chem 279 (2004) 31105–31112. 1.2.3

[83] S. Gingras, R. Moriggl, B. Groner, J. Simard, nduction of 3beta-hydroxysteroid dehydrogenase/delta5-delta4 isomerase type 1 gene transcription in human breast cancer cell lines and in normal mammary epithelial cells by interleukin-4 and interleukin-13., Molecular Endocrinology 13 (1999) 66–81. 1.2.3

[84] C. Litterst, E. Pfitzner, Transcriptional activation by stat6 requires the direct interaction with ncoa-1., The Journal of Biological Chemistry 276 (49) (2001) 45713–21. 1.2.3, 1.2.6, 1.2.7, 1.2.8, 4.0.7

[85] L. Xu, C. Glass, M. Rosenfeld, Coactivator and corepressor complexes in nuclear receptor function., Current Opinion in Genetics and Development 9 (1999) 140–147. 1.2.3

[86] T. Yao, G. Ku, N. Zhou, R. Scully, D. Livingston, The nuclear hormone receptor coactivator src-1 is a specific target of p300, Proceedings of the National Academy of Sciences, USA 93 (1996) 10626–10631. 1.2.3

[87] B. Lu, M. Reichel, D. A. Fisher, J. F. Smith, P. Rothman, Identification of a stat6 domain required for il-4-induced activation of transcription, J Immunol 159 (3) (1997) 1255–64. 1.2.4

[88] H. Chen, R. J. Lin, R. L. Schiltz, D. Chakravarti, A. Nash, L. Nagy, M. L. Privalsky, Y. Nakatani, R. M. Evans, Nuclear receptor coactivator actr is a novel histone acetyltransferase and forms a multimeric activation complex with p/caf and cbp/p300, Cell 90 (3) (1997) 569–80. 1.2.5, 1.2.8

[89] N. McKenna, R. Lanz, B. OMalley, Nuclear receptor co-regulators: cellular and molecular biology., Endocrinological Reviews. 20 (1999) 321–344. 1.2.5

[90] C. Leo, J. D. Chen, The src family of nuclear receptor coactivators, Gene 245 (1) (2000) 1–11. 1.2.5, 4

[91] M. et al, Cell 83 (1995) 835–839. 1.2.5

[92] B. D. Darimont, R. L. Wagner, J. W. Apriletti, M. R. Stallcup, P. J. Kushner, J. D. Baxter, R. J. Fletterick, K. R. Yamamoto, Structure and specificity of nuclear receptor-coactivator interactions, Genes Dev 12 (21) (1998) 3343–56. 1.2.5

[93] C. Murre, P. McGaw, D. Baltimore, A new dna binding and dimerization motif in immunoglobulin enhancer binding, daughterless, myod, and myc proteins., Cell 56 (1989) 777–883. 1.2.5

[94] D. M. Heery, E. Kalkhoven, S. Hoare, M. G. Parker, A signature motif in transcriptional co-activators mediates binding to nuclear receptors, Nature 387 (6634) (1997) 733–6. 1.2.5, 1.2.8, 4.0.5

[95] C. K. Glass, D. W. Rose, M. G. Rosenfeld, Nuclear receptor coactivators, Curr Opin Cell Biol 9 (2) (1997) 222–32. 1.2.5

[96] H. Kim, J. Kim, J. Lee, Steroid receptor coactivator-1 interacts with serum response factor and coactivates serum response element-mediated transactivations., Journal of Biological Chemistry 273 (1998) 28564–28567. 1.2.5

[97] C. Litterst, E. Pfitzner, An lxll motif in the transactivation domain of stat6 mediates recruitment of ncoa-1/src-1., The Journal of Biological Chemistry 277 (39) (2002) 36052–60. 1.2.5, 1.2.6, 1.2.8, 4.0.5, 4.0.7

[98] Y. Kamei, L. u, T. Heinzel, J. Torchia, R. Kurokawa, B. Gloss, S. C. Lin, R. A. Heyman, D. W. Rose, C. K. Glass, M. G. Rosenfeld, A cbp integrator complex mediates transcriptional activation and ap-1 inhibition by nuclear receptors, Cell 85 (3) (1996) 403–14. 1.2.8

[99] S. A. Oate, S. Y. Tsai, M.-J. Tsai, OMalley, B. W., Sequence and characterization of a coactivator for the steroid hormone receptor superfamily, Science 270 (5240) (1995) 1354–1357. 1.2.8

[100] J. Vogel, M. Heine, M. Tini, P. Chambon, H. Gronemeyer, The coactivator tif2 contains three nuclear receptor-binding motifs and mediates transactivation through cbp binding-dependent and -independent pathways. the coactivator tif2 contains three nuclear receptor-binding motifs and mediates transactivation through cbp binding-dependent and -independent pathways. the coactivator tif2 contains three nuclear receptor-binding motifs and mediates transactivation through cbp binding-dependent and -independent pathways. the coactivator tif2 contains three nuclear receptor-binding motifs and mediates transactivation through cbp binding-dependent and independent pathways, EMBO Journal. 1.2.8

[101] J. Torchia, D. Rose, J. Inostroza, Y. Kamei, S. Westin, M. Rosenfeld, The transcriptional co-activator p/cip binds cbp and mediates nuclear-receptor function., Nature 387 (1997) 677–684. 1.2.8

[102] S. L. Anzick, J. Kononen, R. L. Walker, D. O. Azorsa, M. M. Tanner, Q. Y. Guan, G. Sauter, O. P. Kallioniemi, J. M. Trent, P. S. Meltzer, Aib1, a steroid receptor coactivator amplified in breast and ovarian cancer, Science 277 (5328) (1997) 965–8. 1.2.8

[103] R. Kurokawa, D. Kalafus, M. H. Ogliastro, C. Kioussi, L. u, J. Torchia, M. G. Rosenfeld, C. K. Glass, Differential use of creb binding protein-coactivator complexes, Science 279 (5351) (1998) 700–3. 1.2.8

[104] S. B.M., H. R., Biological control through regulated transcriptional coactivators., Cell 2004 (119) 157–67. 1.3

[105] U. Laemmli, Cleavage of proteins during the assembly of the head of bacteriophage t4, Nature 15 (1970) 680–5. 2.2.6

[106] J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning : A laboratory manual., Vol. 3, Cold Spring Harbor Laboratory Press, 1987. 2.2.6

[107] A. Fontana, Limited Proteolysis of globular proteins occur at exposed and flexible loops., VSP Intl. Publ., 1989. 2.2.8

[108] A. Fontana, C. Vita, D. Dalzoppo, M. Zambonin, Limited Proteolysis as a tool to detect structure and dyanamic features of globular proteins: Studies on

thermolysin. In Methods in Protein Sequence Analysis, Springer-Verlag, Berlin., 1989. 2.2.8

[109] I. Schechter, A. Berger, On the size of active site in proteases. i. papain., Biochem. Biophys. Res. Commun. 27 (1967) 57–62. 2.2.8

[110] D. Herschlag, The role of induced fit and conformational changes of enzymes in specificity and catalysis, Bioorganic Chem. 16 (1988) 62–96. 2.2.8

[111] S. Hubbard, S. Campbell, J. Thornton, Molecular recognition : conformational analysis of limited proteolytic sites and serine proteinase inhibitors, Journal of Molecular Biology 220 (1991) 507–30. 2.2.8

[112] A. Fontana, Probing the conformational state of apomyoglobin by limited proteolysis, Journal of Molecular Biology 266 (1997) 223–30. 2.2.8

[113] J. Fernandez, M. DeMott, D. Atherton, S. Mische, nternal protein sequence analysis: enzymatic digestion for less than 10 micrograms of protein bound to polyvinylidene difluoride or nitrocellulose membranes., Analytical Biochemistry 201 (1992) 255–264. 2.2.9

[114] M. K. annd Franz Hillenkamp, Laser desorption inonization of proteins with molecular masses exceeding 10,000 daltons, Analytical Chemistry 60 (1988) 2299. 2.2.10

[115] O. Vorm, P. Roepstorff, M. Mann, Improved resolution and very high sensitivity in maldi-tof of matrix surfaces made by fast evaporation, Analytical chemistry 66 (1994) 3281–3287. 2.2.10

[116] O. Vorm, M. Mann, Improved mass accuracy in matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy of peptides, Journal of American Society of Mass Spectroscopy 5 (1994) 955–958. 2.2.10

[117] A. Smith, R. Young, S. Carr, D. Marshak, L. Williams, W. K.R., Techniques in Protein Chemistry, Academic Press, 1992. 2.2.12

[118] T. Wiseman, S. Williston, J. Brandts, L. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter., Analytical Biochemistry 179 (1989) 131–137. 2.2.13

[119] C. Griesinger, From molecular structure towards biology : Nmr spectroscopy as a tool for the determination of structure in biomolecules, Springer Verlag. 2.5

[120] D. Wishart, B. Sykes, F. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through nmr spectroscopy., Biochemistry 31 (1992) 1647–1651. 2.5.1, 2.6.5

[121] N. Clayden, R. Williams, J. Mag. Res. 49 (1982) 383–396. 2.5.1

[122] M. Karplus, Journal of Physical Chemistry 30 (1959) 11–15. 2.5.2

[123] J. Cavanagh, A. Palmer, N. Skelton, Protein nmr spectroscopy principles and practice, Academic Press. 2.6

[124] C. Griesinger, M. Hennig, J. Marino, B. Reif, H. Schwalbe, Biomolecular nmr eds. berliner l. and krishna, r., Plenum Press. 2.6.1

[125] G. Bodenhausen, J. Ruben, Chemical Physics Letters 69 (1980) 185–189. 2.6.2

[126] M. Ottiger, A. Bax, Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules, J. Biomol. NMR 12 (1998) 361–72. 2.6.3, 2.10

[127] K. Ding, A. Gronenborn, Simultaneous and accurate measurement ofone-bond (15)n-(13)c' and two-bond (1)h(n)–(13)c' dipolar couplings, JACS 125 (2003) 11504–5. 2.6.4

[128] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, J. Mag. Res. 89 (1990) 496–514. 2.6.5, 2.6.5

[129] S. Grzesiek, A. Bax, J. Mag. Res. 99 (1992) 201–207. 2.6.5, 2.6.5, 2.6.5

[130] R. Clubb, V. Thanabal, G. Wagner, J. Mag. Res. 97 (1994) 203–216. 2.6.5

[131] K. Gardner, L. Kay, Annu. rev. Biophys. Biomol. Struct. 27 (1998) 357–406. 2.6.5

[132] K. Pervushin, Q. Rev. Biophysics 33 (2000) 161–97. 2.6.5

[133] D. F., G. S., V. G.W., Z. G., P. J., B. A., J Biomol NMR 6 (1995) 277–93. 2.6.6

[134] J. H. Prestegard, H. M. al Hashimi, J. R. Tolman, Nmr structures of biomolecules using field oriented media and residual dipolar couplings, Q Rev Biophys 33 (4) (2000) 371–424. 2.7

[135] A. Bax, Weak alignment offers new nmr opportunities to study protein structure and dynamics, Protein Science 12 (2003) 1–16. 2.7

[136] J. H. Prestegard, C. M. Bougault, A. I. Kishore, Residual dipolar couplings in structure determination of biomolecules, Chem Rev 104 (8) (2004) 3519–40. 2.7

[137] H. M. Al-Hashimi, D. J. Patel, Residual dipolar couplings: synergy between nmr and structural genomics, J Biomol NMR 22 (1) (2002) 1–8. 2.7

[138] E. Bastiaan, C. Maclean, P. Van Zijl, A. Bothner-By, Annual Reports in NMR spectroscopy 19 (1987) 35. 2.7

[139] J. Emsley, J. Lindon, Nmr spectroscopy using liquid crystal solvents, Pergamon Press. 2.7

[140] C. Gayathri, A. BothnerBy, P. van Zijl, C. MacLean, Dipolar magnetic field effects in nmr spectra of liquids, Chemical Physics Letters. 2.7

[141] J. Tolman, J. Flanagan, M. Kennedy, J. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins-information for structure determination in solution., Proceedings of the National Academy of Sciences, USA 92 (1995) 9279–9283. 2.7, 2.14

[142] M. Mehring, High resolution nmr in solids, Springer. 2.7

[143] J. R. Tolman, H. M. Al-Hashimi, L. E. Kay, J. H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins, J Am Chem Soc 123 (7) (2001) 1416–24. 2.8

[144] J. A. Losonczi, M. Andrec, M. W. Fischer, J. H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition, J Magn Reson 138 (2) (1999) 334–42. 2.8, 2.9

[145] C. Schwieters, G. Clore, Internal coordinates for molecular dynamics and minimization in structure determination and refinement, J. Mag. Res. 152 (2000) 288–302. 2.8

[146] G. Cornilescu, J. Marquardt, M. Ottiger, A. Bax, Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase., JACS 120 (1998) 6836–6837. 2.9, 3.3.4

[147] M. Zweckstetter, A. Bax, Prediction of sterically induced alignment in a dilute liquid crystalline structure, JACS 122 (2000) 3791–3792. 2.9

[148] G. Kontaxis, F. Delaglio, A. Bax, Molecular fragment replacement approach to protein structure determination by chemical shift and dipolar homology database mining., Methods in Enzymology 394 (2005) 42–48. 2.9

[149] M. Rueckert, G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for nmr experiments, Journal of American Chemical Society 122 (2000) 7793–7797. 2.10, 2.12

[150] B. Ramirez, A. Bax, Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium, Journal of American Chemical Society 120 (1998) 9106–9107. 2.10

[151] H. M. Al-Hashimi, H. Valafar, M. Terrell, E. R. Zartler, M. K. Eidsness, J. H. Prestegard, Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings, J Magn Reson 143 (2) (2000) 402–6. 2.10

[152] C. Sanders, J. Schwonek, Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphophatidylcholine by solid-state nmr, Biochemistry 31 (1992) 8898–8905. 2.10

[153] G. M. Clore, A. M. Gronenborn, New methods of structure refinement for macromolecular structure determination by nmr, Proc Natl Acad Sci U S A 95 (11) (1998) 5891–8. 2.10

[154] M. Zweckstetter, A. Bax, Characterization of molecular alignment in aqueous suspensions of pf1 bacteriophage, J Biomol NMR 20 (4) (2001) 365–77. 2.10

[155] C. Sanders, B. Hare, K. Howard, J. Prestegard, Magnetically oriented micelles as a tool for membrane-associated molecules, Biophysics Journal 26 (1994) 421–444. 2.11, 2.12

[156] D. Yang, J. Tolman, N. Goto, L. Kay, An hnco based pulse scheme for the measurement of 13cagr 1hagr one bond dipolar couplings in 15n 13c labeled proteins, J. Biomol. NMR 12 (1998) 325–332. 2.11

[157] J. Lozonczi, J. Prestegard, Improved dilute bicelle solutions for high resolution nmr of biological macromolecules, J. Biomol. NMR 12 (1998) 447–451. 2.11, 4.2

[158] V. Vijayan, M. Zweckstetter, Simultaneous measurement of protein one-bond residual dipolar couplings without increased resonance overlap, Journal of Magnetic Resonance. 2.11.1

[159] M. Jonstroemer, R. Strey, Nonionic bilayers in dilute solutions: effect of additives, J. Phys. Chem. 96 (1992) 5993–6000. 2.12

[160] M. Penders, R. Strey, Lamellar and l3 phases in the simple h2o-c8e5-1-octanol system: Evidence of synergism, J. Phys. Chem. 99 (1995) 6091–6095. 2.12

[161] E. Freyssingeas, F. Nallet, D. Roux, Measurement of the membrane flexibility in lamellar and ”sponge” phases of the c12e5/hexanol/water system, Langmuir 12 (1996) 6028–6035. 2.12

[162] Y. Ishii, M. A. Markus, R. Tycko, Controlling residual dipolar couplings in high-resolution nmr of proteins by strain induced alignment in a gel, J Biomol NMR 21 (2) (2001) 141–51. 2.13

[163] H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, Solution nmr of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes, J Biomol NMR 18 (4) (2000) 303–9. 2.13

[164] J. J. Chou, S. Li, C. B. Klee, A. Bax, Solution structure of ca(2+)-calmodulin reveals flexible hand-like properties of its domains, Nat Struct Biol 8 (11) (2001) 990–7. 2.13

[165] D. Shortle, M. S. Ackerman, Persistence of native-like topology in a denatured protein in 8 m urea., Science 293 (2001) 487–489. 2.13

[166] T. Ikegami, L. Verdier, P. Sakhaii, S. Grimme, B. Pescatore, K. Saxena, K. M. Fiebig, C. Griesinger, Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions, J Biomol NMR 29 (3) (2004) 339–49. 2.14

[167] N. van Nuland, G. Kroon, K. Dijkstra, G. Wolters, R. Scheek, G. Robillard, The nmr determination of the iia(mtl) binding site on hpr of the escherichia coli phosphoenol pyruvate-dependent phosphotransferase system., FEBS Letters 315 (1993) 11–15. 2.15

[168] Y. Chen, J. Reizer, M. J. Saier, W. Fairbrother, P. Wright, Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, hpr and iiaglc., Biochemistry 32 (1993) 32–37. 2.15

[169] G. Otting, Experimental nmr techniques for studies of protein-ligand interactions, Current Opinion in Structural Biology 3 (1993) 760–768. 2.15

[170] D. Sitkoff, D. Case, Theories of chemical shift anisotropies in proteins and nucleic acids, Progress in Nuclear Magnetic Resonance Spectroscopy 32 (1998) 165–190. 2.15

[171] R. A. Williamson, M. D. Carr, T. A. Frenkiel, J. Feeney, R. Freedman, Mapping the binding site for matrix metalloproteinase on the n-terminal domain of the tissue inhibitor of metalloproteinases-2 by nmr chemical shift perturbation., Biochemistry 36 (1997) 13882–13889. 2.15

[172] F. W. Muskett, T. A. Frenkiel, J. Feeney, R. B. Freedman, M. Carr, High resolution structure of the n-terminal domain of tissue inhibitor of metalloproteinases-2 and characterization of its interaction site with matrix metalloproteinase-3, J. Biol. Chem. 273 (1998) 21736–21743. 2.15

[173] C. Geourjon, G. Deleage, Sopma, significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Computational and Applied Biosciences 11 (1995) 681–684. 3.1.3

[174] I. Wadso, Thermochimica Acta 267 (1995) 45–49. 3.1.6

[175] B. Bundle, D.R. Sigurskjold, Determination of accurate thermodynamics of binding by titration microcalorimetry, Methods in Enzymology 247 (1994) 288–304. 3.1.6

[176] M. Doyle, G. Louie, P. Dal Monte, T. Sokoloski, Tight binding affinities determined from thermodynamic linkage to protons by titration calorimetry., Methods in Enzymology 259 (1995) 183–194. 3.1.6

[177] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, Nmrpipe: a multidimensional spectral processing system based on unix pipes, Journal of Biomolecular NMR. 3.3.1

[178] M. Zweckstetter, A. Bax, Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by nmr, JACS 122 (2000) 3791–3792. 3.3.4, 3.3.4

[179] A. Razeto, V. Ramakrishnan, C. M. Litterst, karin Giller, C. Griesinger, T. Carlomango, N. Lakomek, T. Heimburg, M. Lodrini, E. Pfitzner, S. Becker, Structure of the ncoa-1/src-1 pas-b domain bound to the lxxll motif of the stat6 transactivation domain, J.Mol.Biol 336 (2004) 319–329. 4

[180] A. Razeto, E. Pfitzner, S. Becker, Crystallization and preliminary crystallographic studies of the ncoa-1/src-1 pas-b domain bound to the lxxll motif of the stat6 transactivation domain, Acta Crystallographica D 60 (2004) 550–552. 4

[181] C. Ponting, L. Aravind, Pas: a multifunctional domain family comes to light., Current Biology 7 (1997) 674–677. 4

[182] L. Alex, M. Simon, Trends in Genetics 10 (1994) 133–138. 4

[183] S. et al, J. Bacteriol. 175 (1993) 3096–3104. 4

[184] S. et al, PNAS 95 (1998) 8991–96. 4

[185] Schilber, Nature 393 (1998) 620–621. 4

[186] J. Warmke, B. Ganetzky, A family of potassium channel genes related to eag in drosophila and mammals, PNAS 91 (1994) 3438–42. 4

[187] B.-H. Jiang, E. Rue, G. L. Wang, R. Roe, G. L. Semenza, Dimerization, dna binding, and transactivation properties of hypoxia-inducible factor 1, JBC 271 (1996) 17771–78. 4

[188] N. JR, L. JO, W. K. Jr, C. ST, The drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of cns midline development., Cell 67 (1991) 1157–67. 4

[189] L. Holm, C. Sander, Protein structure comparison by alignment of distance matrices, J Mol Biol 233 (1) (1993) 123–38. 4

[190] I. Radhakrishnan, G. C. Perez-Alvarado, D. Parker, H. J. Dyson, M. R. Montimony, P. E. Wright, Structural analyses of creb-cbp transcriptional activator-coactivator complexes by nmr spectroscopy: Implications for mapping the boundaries of structural domains, Journal of Molecular Biology 287 (1999) 857–865. 4.0.5

[191] M. Ptashne, A. Gann, Transcriptional activation by recruitment, Nature 386 (1997) 569–577. 4.0.5

[192] D. Parker, U. Jhala, I. Radhakrishnan, M. Yaffe, C. Reyes, A. Shulman, L. Cantley, P. Wright, M. Montminy, Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation., Molecular Cell 2 (1998) 353–359. 4.0.5

[193] M. Uesugi, O. Nyanguile, H. Lu, A. J. Levine, G. L. Verdine, Induced alpha helix in the vp16 activation domain upon binding to a human taf, Science 277 (5330) (1997) 1310–3. 4.0.5

[194] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, N. P. Pavletich, Structure of the mdm2 oncoprotein bound to the p53 tumor suppressor transactivation domain, Science 274 (5289) (1996) 948–53. 4.0.5

[195] U. K. Genick, G. E. O. Borgstahl, K. Ng, Z. Ren, C. Pradervand, P. M. Burke, V. Srajer, T. Teng, W. Schildkamp, D. E. McRee, K. Moffat, E. D. Getzoff, Structure of a protein photocycle intermediate by millisecond time-resolved crystallography, Science 275 (1997) 1471–75. 4.1, 4.1

[196] W. Gong, B. Hao, S. S. Mansy, G. Gonzalez, M. A. G. Gonzalez, M. K. Chan, Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction, PNAS 95 (1998) 15177–82. 4.1

[197] H. Miyatake, M. Kanai, S. Adachi, H. Nakamura, K. Tamura, H. Tanida, T. Tsuchiya, T. Iizuka, Y. Shiro, Dynamic light-scattering and preliminary crystallographic studies of the sensor domain of the haem-based oxygen sensor fixl from rhizobium meliloti, Acta Cryst. D 55 (1999) 1215–18. 4.1

[198] J. H. Morais Cabral, A. Lee, S. L. Cohen, B. T. Chait, M. Li, R. Mackinnon, Crystal structure and functional analysis of the herg potassium channel n terminus: a eukaryotic pas domain, Cell 95 (5) (1998) 649–55. 4.1

[199] S. Crosson, K. Moffat, Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch, Plant Cell 14 (2001) 1067–75. 4.1

[200] C. A. Amezcua, S. M. Harper, J. Rutter, K. H. Gardner, Structure and interactions of pas kinase n-terminal pas domain:model for intramolecular kinase regulation, Structure 10 (2002) 1349–1361. 4.1

[201] S. Reinelt, The structure of the periplasmic ligand-binding domain of the sensor kinase cita reveals the first extracellular pas domain, J. Biol. Chem 278 (2003) 39189–39196. 4.1

[202] P. et al, The nmr structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) dcus of escherichia coli., JBC 278 (2003) 39185–188. 4.1

[203] M. et al, Eur. J. Biochem 271 (2004) 1198–1208. 4.1

[204] H. et al, Science 301 (2003) 1541–44. 4.1

[205] J. V. M. A. van der Horst, K. J. Hellingwerf, W. Crielaard, D. M. F. van Aalten, Pas domains. common structure and common flexibility, JBC 278 (2003) 18434–39. 4.1

[206] M. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous bacteriophage yields dipolar coupling interactions, Nat. Struct. Biol. 5 (1998) 1065–74. 4.2

[207] D. Ojennus, R. Mitton-Fry, D. Wuttke, Induced alignment and measurement of dipolar couplings of an sh2 doman through direct binding with filamentous phage, Journal of Biomolecular NMR 14 (1999) 175–179. 4.2

[208] K. Zimmermann, H. Hagedorn, C. Heuck, M. Hinrischen, H. Ludwig, The ionic properties of the filamentous bacteriophages pf1 and fd., J. Biol. Chem. 261 (1986) 1653–1655. 4.2

[209] T. DS, N. AA, Z. L, W. CH, O. SJ., Structural basis of the temperature transition of pf1 bacteriophage., J Mol Biol 341 (2004) 869–79. 4.2

[210] V. Degiorgio, M. Corti, Physics of amphiphiles, micelles, vesicles and microemulsions, North-Holland Physics publishing. 4.2

[211] J. Israelachvili, Q. Rev. Biophys. 6 (1973) 341–387. 4.2

[212] T. Cierpicki, J. H. Bushweller, Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins, J Am Chem Soc 126 (49) (2004) 16259–66. 4.2