The hematopoietic transcription factor RUNX1:

A structural view

by

Stefan Bäckström

Homo curiosus

From Umeå Centre for Molecular Pathogenesis
Umeå University
Umeå Sweden 2003
The hematopoietic transcription factor RUNX1:

A structural view

AKADEMISK AVHANDLING

Som med vederbörligt tillstånd av rektorsämbetet vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt förvar i Major Groove, Institutionen för Molekylärbiologi, byggnad 6L, fredagen den 5 september, kl.10.00. Avhandlingen kommer att förvaras på engelska.

av

Stefan Bäckström

Umeå Centrum för Molekylär Patogenes
Umeå Universitet

Fakultetsopponent: Prof. Keith Wilson, Department of Chemistry, University of York, UK.
The hematopoietic transcription factor RUNX1: A structural view

Abstract
The malfunction of the transcriptional regulator RUNX1 is the major cause of several variants of acute human leukemias and its normal function is to regulate the development of the blood system in concert with other transcriptional co-regulators. RUNX1 belongs to a conserved family of heterodimeric transcription factors that share a conserved DNA binding domain, the Runt domain (RD), named after the first member of this group – Runt - found in Drosophila melanogaster. The binding partner CBFβ serves as a regulator of RUNX by enhancing its DNA binding affinity through an allosteric mechanism. The main focus of my thesis work has been the crystallization and structural analysis of the RUNX1 RD and involved also more technical methodological aspects that can be applied to X-ray crystallography in general.

The high resolution crystal structure of the free RD shows that this immunoglobulin-like molecule undergoes significant structural changes upon binding to both CBFβ and DNA. This involves a large flip of the L11 loop from a closed conformation in the free protein to an open conformation when CBFβ and/or DNA are bound. We refer to this transition as the “S-switch”. Smaller but significant conformational changes in other parts of the RD accompany the “S-switch”. We suggest that CBFβ triggers and stabilizes the “S-switch” which leads to the conversion of the RD into a conformation enhanced for DNA binding.

During the structural analysis of the RD we identified two chloride ions that are coordinated by residues otherwise involved in DNA binding. In electrophoretic mobility-shift analyses (EMSA) we demonstrated a chloride ion concentration dependent stimulation of the DNA binding affinity of RUNX1. We further showed by NMR line width broadening experiments that the chloride binding occurred within the physiological range. A comparable DNA binding stimulation of RUNX1 was seen in the presence of negative amino acids. This suggests a regulation of the DNA binding activity of RUNX1 proteins through acidic amino acid residues possibly provided by activation domains of transcriptional co-regulators that interact with RUNX1.

The use of the anomalous signal from halide ions has become a powerful technique for obtaining phase information. By replacing the sodium chloride with potassium bromide in the crystallisation conditions of the RD, we could demonstrate in a single wavelength anomalous diffraction (SAD) experiment that the anomalous signal from 2 bromide ions were sufficient to phase a 16 kDa protein. Due to lack of completeness in the low-resolution shells caused by overloaded intensities, density modification schemes failed and the resulting electron density maps were not interpretable. By combining the high-resolution synchrotron data with low-resolution data from a native data set collected on a home X-ray source, the density modified bromide phases gave easily traceable maps.

Keywords: RUNX1, Runt domain, CBFβ, transcription factor, leukaemia, protein crystallography, anomalous diffraction

Language: English ISBN: 91-7305-476-3 Number of pages: 30 + 4 papers

Signature: Date:
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Papers in this thesis</td>
<td>3</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>4</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2. Molecular biology of RUNX transcription factors</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Three members of the Runt family of transcription factors are present in mammals.</td>
<td>6</td>
</tr>
<tr>
<td>2.2 CBFβ, the heterodimerization partner of the RUNX proteins</td>
<td>8</td>
</tr>
<tr>
<td>3. Structural biology of the RUNX1 transcription factor RD</td>
<td>8</td>
</tr>
<tr>
<td>3.1 DNA binding</td>
<td>9</td>
</tr>
<tr>
<td>3.2 Heterodimerization with CBFβ</td>
<td>10</td>
</tr>
<tr>
<td>3.3 Allosteric control of DNA binding by CBFβ</td>
<td>11</td>
</tr>
<tr>
<td>3.4 Structural connection to disease models.</td>
<td>11</td>
</tr>
<tr>
<td>4. Determination of protein structures with X-ray crystallography</td>
<td>13</td>
</tr>
<tr>
<td>4.1 Why do we use X-rays</td>
<td>13</td>
</tr>
<tr>
<td>4.2 Why do we need crystals</td>
<td>13</td>
</tr>
<tr>
<td>4.3 The phase problem and ways to overcome it</td>
<td>15</td>
</tr>
<tr>
<td>4.4 Anomalous scattering</td>
<td>16</td>
</tr>
<tr>
<td>4.5. Single wavelength anomalous diffraction – SAD</td>
<td>16</td>
</tr>
<tr>
<td>4.5.1 The anomalous contribution to the total scattering</td>
<td>16</td>
</tr>
<tr>
<td>4.5.2 The breakdown of Friedel's Law</td>
<td>17</td>
</tr>
<tr>
<td>4.5.3 Determining the substructure of the anomalous scatterers</td>
<td>17</td>
</tr>
<tr>
<td>4.5.4 Breaking the phase ambiguity</td>
<td>18</td>
</tr>
<tr>
<td>5. The aim of this study – Results.</td>
<td>21</td>
</tr>
<tr>
<td>5.1 Preparation and crystallization of the RD (paper I)</td>
<td>21</td>
</tr>
<tr>
<td>5.2 Comparison of the free and the complexed RD structures</td>
<td>22</td>
</tr>
<tr>
<td>(paper II)</td>
<td></td>
</tr>
<tr>
<td>5.3 Chloride and acidic amino acid binding to the RD (paper II and III)</td>
<td>24</td>
</tr>
<tr>
<td>5.4 The use of halides for phase determination (paper IV)</td>
<td>25</td>
</tr>
<tr>
<td>6. Future</td>
<td>26</td>
</tr>
<tr>
<td>7. Acknowledgement</td>
<td>27</td>
</tr>
<tr>
<td>8. References</td>
<td>28</td>
</tr>
</tbody>
</table>
Abstract

The malfunction of the transcriptional regulator RUNX1 is the major cause of several variants of acute human leukemias and its normal function is to regulate the development of the blood system in concert with other transcriptional co-regulators. RUNX1 belongs to a conserved family of heterodimeric transcription factors that share a conserved DNA binding domain, the Runt domain (RD), named after the first member of this group – Runt - found in *Drosophila melanogaster*. The binding partner CBFβ serves as a regulator of RUNX by enhancing its DNA binding affinity through an allosteric mechanism. The main focus of my thesis work has been the crystallization and structural analysis of the RUNX1 RD and involved also more technical methodological aspects that can be applied to X-ray crystallography in general.

The high resolution crystal structure of the free RD shows that this immunoglobulin-like molecule undergoes significant structural changes upon binding to both CBFβ and DNA. This involves a large flip of the L11 loop from a closed conformation in the free protein to an open conformation when CBFβ and/or DNA are bound. We refer to this transition as the “S-switch”. Smaller but significant conformational changes in other parts of the RD accompany the “S-switch”.

We suggest that CBFβ triggers and stabilizes the “S-switch” which leads to the conversion of the RD into a conformation enhanced for DNA binding.

During the structural analysis of the RD we identified two chloride ions that are coordinated by residues otherwise involved in DNA binding. In electrophoretic mobility-shift analyses (EMSA) we demonstrated a chloride ion concentration dependent stimulation of the DNA binding affinity of RUNX1. We further showed by NMR line width broadening experiments that the chloride binding occurred within the physiological range. A comparable DNA binding stimulation of RUNX1 was seen in the presence of negative amino acids. This suggests a regulation of the DNA binding activity of RUNX1 proteins through acidic amino acid residues possibly provided by activation domains of transcriptional co-regulators that interact with RUNX1.

The use of the anomalous signal from halide ions has become a powerful technique for obtaining phase information. By replacing the sodium chloride with potassium bromide in the crystallisation conditions of the RD, we could demonstrate in a single wavelength anomalous diffraction (SAD) experiment that the anomalous signal from 2 bromide ions were sufficient to phase a 16 kDa protein. Due to lack of completeness in the low-resolution shells caused by overloaded intensities, density modification schemes failed and the resulting electron density maps were not interpretable. By combining the high-resolution synchrotron data with low-resolution data from a native data set collected on a home X-ray source, the density modified bromide phases gave easily traceable maps.

**Keywords:** RUNX1, Runt domain, CBFβ, transcription factor, leukaemia, protein crystallography, anomalous diffraction
Papers in this thesis

This thesis is based on the following articles and manuscript, which will be referred to in the text by their Roman numerals (I-IV).


IV Bäckström S, Sauer UH. The importance of low resolution data: A SAD bromide example. Manuscript
**Abbreviations**

**AML**  acute myeloid leukemia  
**CBFβ**  core binding factor beta, also denoted PEBP2β (polyoma virus-enhancer binding protein 2β)  
**CCD**  cleidocranial dysplasia  
**DNA**  deoxyribonucleic acid  
**RD**  runt domain  
**NLS**  nuclear localization signal  
**NMR**  nuclear magnetic resonance  
**FWHH**  full width at half height  
**EMSA**  electrophoretic mobility-shift assay  
**FPD/AML**  familial platelet disorder with propensity to acute myeloid leukaemia  
**MAD**  multiple wavelength anomalous dispersion  
**SAD**  single wavelength anomalous dispersion  
**RNA**  ribonucleic acid  

**RUNX1**  runt related transcription factor 1, also denoted AML1 (acute myeloid leukemia protein 1), CBFα2, (core binding factor α2) or PEBP2αB (polyoma virus-enhancer binding protein 2αB)  
**RUNX2**  runt related transcription factor 2, also denoted AML3/ CBFα1/ PEBP2αA  
**RUNX3**  runt related transcription factor 3, also denoted AML2/CBFα3/ PEBP2αC
1. Introduction

The genetic material (genome) of all organisms consists of deoxyribonucleic acid (DNA) that serves as a template for transcription into ribonucleic acid (RNA). Some RNA molecules have structural or enzymatic functions but most RNA’s, the so-called messenger RNA’s (mRNA), carry information that is interpreted by the ribosome. The ribosome is a large, complex factory that follows the instructions written in the genetic code as it assembles 20 different amino acids into polypeptide chains – proteins - that in some cases are additionally modified/processed. Protein molecules serve numerous tasks in an organism. They can have structural functions for example in skin, hair and muscle or they can be enzymes, regulators, receptors, hormones, ion channels et c.

A large protein/RNA complex that among other proteins include RNA polymerase carries out the transcription of genes. DNA binding regulatory proteins called transcription factors control this process and thus the RNA production. Certain factors activate or enhance transcription (co-activators) whereas others reduce or repress transcription (co-repressors). Malfunctioning in the systems controlling gene expression can give rise to various diseases including different forms of cancers. These malfunctions are often caused by point mutations and deletions in the transcription factors or in components that controls the activity of the transcription factors.

The focus of this thesis is the regulatory protein – RUNX1 – that is involved in several crucial processes during blood system development. RUNX1 is a transcription factor and binds to DNA in concert with many other proteins and controls the mRNA production of genes involved in the complex formation and maturation of various blood cells (hematopoiesis).

To understand the structure and function of various classes of proteins, it is important to know their molecular structures at very high resolution. In principle, one should be able to predict the three-dimensional structure of a protein from its amino-acid sequence. However this is not yet possible and the structure must be determined experimentally. The major methods of obtaining structural information of proteins today are X-ray crystallography, the related electron and neutron crystallography, nuclear magnetic resonance (NMR) and electron microscopy (EM). The methods complement each other but the technique that provides the most detailed information to date is X-ray crystallography, which is the method used in this study.
2. Molecular biology of RUNX transcription factors

The Runt domain (RD) is the evolutionarily conserved DNA-binding domain of a family of heterodimeric eukaryotic transcription factors found in a diverse range of species ranging from Caenorhabditis elegans (a small soil dwelling round worm) to Homo sapiens (Figure 1) [1]. In mammals, three RD containing proteins have been identified – RUNX1-3. The RUNX transcription factors are relatively weak activators on their own but become effective transcriptional enhancers or repressors when they cooperate with other transcription factors, co-activators and co-repressors. [2-6]

Figure 1. Sequence alignment of the conserved RD from 20 members of the Runt domain family. The secondary structure assignment is shown at the top. Amino acids that make base specific interactions are marked with a star. Abbreviations: Mm1, M. musculus Runx1, etc.; Tr1, T. rubripes Runx1, etc.; Ci, C. intestinalis; Sp, S. purpuratus; AgA, A. gambiae RunxA; DmA, D. melanogaster RunxA; DmB, D. melanogaster RunxB; AgL, A. gambiae Lozenge; DmL, D. melanogaster Lozenge; AgR, A. gambiae Runt; DmR, D. melanogaster Runt; Ce, C. elegans; Pl, Pacifastacus leniusculus; Cs1, Cupiennius salei Run-1, etc.; Mh, Meloidogyne hapla. Adapted from [1]

2.1 Three members of the Runt family of transcription factors are present in mammals.

RUNX1 is essential in hematopoiesis, the development of the blood system, and the homozygous disruption of the gene encoding RUNX1 leads to a total lack of definitive hematopoietic stem cells in the mouse embryo and no blood cells are developed. [2, 7-9] The RUNX1 gene is a target for chromosomal translocations and point mutations and is disrupted in approximately one-fourth of all de novo acute leukemias. Haploinsufficiency of RUNX1 leads to familial platelet disorder with propensity to acute myeloid leukemia (FPD/AML). [10]
Apart from the essential role in hematopoiesis, RUNX1 is also important in the development of blood vessels (angiogenesis). [11]

![Diagram of RUNX1](image)

**Figure 2.** Schematic overview of RUNX1 (based on [12]). RD – Runt domain, NLS – Nuclear localisation signal, TE – transactivating element, ID – inhibitory domain, NRDB – negative regulatory region for DNA binding, VWRPY – conserved TEL/Groucho binding motif.

RUNX1 consists of several functional modules (Figure 2). The RD is necessary and sufficient for DNA binding and for heterodimerization with core binding factor beta, CBFβ. At the C-terminal end of the RD there is a nuclear localisation signal (NLS), and regions that inhibit the DNA binding are found both N- and C-terminal to the RD. C-terminal to the RD are several elements that have effects on transactivation. This C-terminal part was shown to be associated with the nuclear matrix, and contains at the very end a conserved motif, VWRPY, which is known to be involved in the repression of transcription. [13-16]

RUNX2 is essential for osteoblast differentiation and bone formation. Haploinsufficiency of RUNX2 leads to the dominant disease cleidocranial dysplasia (CCD), which is characterized by multiple skeletal abnormalities. [17, 18]

RUNX3 is a tumor suppressor gene essential for antiproliferation and apoptosis of the gastric epithelium. It is needed for the proper development of the gastrointestinal and RUNX3 expression is lost in about 60% of stomach cancers. [19, 20] RUNX3 has also been shown to act in the projection of dorsal root ganglion neurons. [21]

All three RUNX proteins bind to the same DNA-consensus sequence. This competitive binding leads to a complex transcriptional regulation depending on which RUNX transcription factors and cofactors are present. Incorrect expression of RUNX members outside their normal tissue interfere with the RUNX member that is normally present and induce oncogenic effects through competitive interference. [5]
Table 1. Functions and oncogenic effects of the three RUNX family members. Adapted from [5]

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alternative gene names</th>
<th>Proposed essential function</th>
<th>Mouse(./.) phenotype</th>
<th>Tumorigenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>runx1</td>
<td>aml1/cbfa2/pebp2αB</td>
<td>Definitive hematopoiesis</td>
<td>Embryonic lethal. Absence of fetal liver hematopoiesis</td>
<td>Hemizygocity in humans predisposes to acute myeloid leukemia. Frequent translocation breakpoint. Common insertion site in retrovirus-induced mouse leukemia</td>
</tr>
<tr>
<td>runx2</td>
<td>aml3/cbfa1/pebp2αA</td>
<td>Bone ossification</td>
<td>Dies at birth from respiratory failure.</td>
<td>Transgenic Runx2 overexpression predisposes to T-cell lymphomas. Common insertion site in retrovirus-induced mouse leukemia</td>
</tr>
</tbody>
</table>

2.2 CBFβ, the heterodimerization partner of the RUNX proteins

CBFβ is the heterodimerization partner of all RUNX transcription factors and is an essential protein that is ubiquitously expressed. The homozygous disruption of the gene coding for CBFβ gives the same mouse phenotype as the knock-out of the runx1 and runx2 genes.[22-25]. CBFβ is localized in the cytoplasm and is transferred into the nucleus by its RUNX partner where it improves the DNA binding of the RUNX proteins without contacting the DNA itself. CBFβ does this in two ways: First, in a direct way by inducing a 6-fold increase in DNA binding affinity of the isolated RD and secondly in an indirect manner by neutralizing the negative regulatory regions for DNA binding (NRDB) that flank the RD. [2, 7-9, 26]

Furthermore, CBFβ has a stabilizing effect by protecting RUNX proteins from ubiquitin-proteasome mediated degradation. [27]

3. Structural biology of the RUNX1 transcription factor RD

Extensive structural studies of the RD by several independent groups show that the RD adopts the fold of an immunoglobulin(Ig)-like b-barrel. [28-32] The RD thus belongs to a superfamily of transcription factors that share a common Ig-type DNA-binding domain. Other members in this superfamily are STAT, NFAT, NFκB, T-domain and p53 families of transcription factors as classified in SCOP (Figure 3). [33] The structural features of the DNA binding domains of these proteins are very similar in spite of no apparent sequence
similarities and they all interact with DNA by using loops and a C-terminal linker located on the same end of the β-barrel.

Figure 3. Diversity tree constructed on the structural relationship between the Runt domain and transcription factors containing an Ig-like DNA-binding domain. The structurally conserved DNA binding Ig domains are shown in black.

3.1 DNA binding

Structural studies of RD-DNA complexes provide details of the direct contacts formed between the RD and DNA. By using three loop regions situated at one side of the β-barrel, the RD makes base-specific contacts to both the major and minor grooves of DNA. [29, 30] Two loop regions, β3-L3 and β12-L12, interact with the major groove whereas the β9-L9 region make minor groove interactions. Three arginine residues mediate the sequence specificity: Arg80 at the end of beta strand β3 and Arg174 and Arg177 both situated in loop L12. These arginines recognize three guanine residues in the consensus sequence of DNA (Figure 4). In addition, important contacts are formed between Asp171 and two cytidines situated on the complementary strand and Arg142 in loop L9 that makes direct contact with two bases in the minor groove although this latter recognition does not appear to be sequence specific.
3.2 Heterodimerization with CBFβ

The co-factor CBFβ binds to the RD at a face of the protein that is distinct from the DNA binding region. [28-30] The interaction contains two hydrophilic binding surfaces (Area I and Area II), of which the first one is close to the DNA binding end of the β-barrel (Figure 4). Area I encompasses β-strands β10 and β5-L5, which are closely linked to the DNA binding loop L9 through a hydrogen-bonding network. Area II consists of loops L1, L10 and β-strand β11. Binding to CBFβ creates a short intermolecular parallel β-sheet formed between β11 of the RD and β5 of CBFβ.
The heterodimer interacting surface between the two proteins increases from 1900 Å² to 2190 Å² upon DNA-binding and the number of stabilizing hydrogen bonds increase from 10 to 16.

3.3 Allosteric control of DNA binding by CBFβ

The allosteric regulation mechanism of RD DNA-binding by CBFβ is not entirely clear, in spite of the number of RD structures available. Since the structures of the RUNX1 RD bound to either CBFβ, DNA or both CBFβ and DNA are very similar, one drew the conclusion that CBFβ does not structurally affect the RD and does not induce any major structural changes. [28-30] Together with experiments showing a thermal stabilization of the RD upon CBFβ and DNA binding, this implicates that part of the effect that CBFβ exerts is due to the stabilisation of RD loop regions involved in the minor groove binding. This stabilization is mediated directly from CBFβ to DNA via the Area I interaction and by tightening the hydrogen-bonding network between loop L5 and L9.

3.4 Structural connection to disease models.

Point mutations in RUNX1 that lead to Acute myelogenous leukaemia and related diseases are exclusively positioned in the DNA binding region of the RD, impairing DNA binding but not the ability to form a heterodimer with CBFβ. [29] This explains the dominant-negative behaviour of these point mutations since the mutated RUNX1 protein competes with wild-type RUNX for the binding of CBFβ and other partners. In the case of cleidocranial dysplasia (CCD), the point mutations are more scattered over the RUNX2 RD and can be divided into mutations that effect the DNA binding directly and mutations that probably drastically effects the overall fold of the RD. This leads to the destruction of both DNA and CBFβ binding, explaining the haploinsufficiency of these mutations.
Figure 5 The point mutations found in AML patient affect residues that are involved in DNA binding whereas the point mutations found in CCD interfere with proper folding of the protein as well as DNA and CBFβ binding. Adapted from [29]. The figure was generated with MOLSCRIPT [34].
4. Determination of protein structures with X-ray crystallography

4.1 Why do we use X-rays?

For several centuries, scientists have been using microscopes to visualise small objects. The resolution limit at which two objects can be separated is about \( \frac{1}{2} \) the wavelength of the light used to visualise it. Therefore, in order to resolve individual atoms in a molecule it is necessary to use electromagnetic radiation at wavelengths comparable to the atomic bond distances (around \( 1 \times 10^{-10} \) metres or 1 Å). For this purpose X-rays have suitable wavelengths around 1 Å, whereas visual light having wavelengths between 4000 to 7000 Å cannot be used to study molecules at the atomic resolution level.

4.2 Why do we need crystals?

The scattering information from individual molecules is far too weak to measure. Crystals, which are three-dimensional arrays of molecules, are required for X-ray diffraction experiments and act as amplifiers by increasing the scattering.
signal due to the highly ordered 3D arrays of molecules it contains. A crystal with the size of 0.1x0.1x0.1 mm can contain as many as $10^{13}$ molecules and due to the regular arrangement of the molecules this leads to a tremendous amplification effect.

**Figure 7**

a) A crystal consists of many building blocks – unit cells – that are regularly packed in an ordered 3D array. Depending on the symmetry of the crystal, the unit cell can contain one or more asymmetric units, the smallest independent repeating unit within the crystal. The amplification of the scattering signal in the “mini-crystal” shown would be approximately 36 times = the number of unit cells in the crystal. (courtesy of Prof Read, University of York) b) The theoretical molecular transform from a single molecule of the hen egg-white lysozyme. (courtesy of Prof. Hajdu, Uppsala University) c) A protein crystal diffraction, which is a combination of the unit cell transform and the crystal diffraction.
When X-rays interact with core electrons of individual atoms, they are elastically scattered and all the scattered waves interfere with each other, either in a constructive fashion by adding up or in a destructive way by cancelling each other out. The total interference of waves from a crystal is the sum of the molecular transform that describes the atomic structure of the protein content in the smallest repeated unit (unit cell or asymmetric unit), and the crystal diffraction that reflects the internal symmetry and lattice properties of the crystal. The observed intensities of the diffraction spots are determined by the underlying “molecular diffraction” whereas the positions of the spots are determined by the properties of the crystal lattice. Whenever the conditions for constructive interference are satisfied (Bragg’s law), there will be a diffraction spot. An important effect of the additional properties of waves is that each spot in the diffraction pattern contains information that is sampled from the entire unit cell.

4.3 The phase problem and ways to overcome it

In a microscope, light hits an object and is diffracted in various directions, bringing information about the object with it. Lenses subsequently re-focus the diffracted waves and a magnified image is reconstructed. In this way the intensity information is preserved together with the information about the phase relationships of the scattered waves. Since diffracted X-rays cannot be re-focused, we are stuck with the diffraction pattern. The recreation of the molecular structure cannot be done directly from diffraction images since the phase information of the diffracted waves originating from the molecular structure is lost. Each spot of the diffraction pattern has an amplitude - in principle the square-root of the intensity of the spot - and a phase. The phase cannot be measured and has to be recovered indirectly. There are several methods to overcome this problem.

I. Direct methods
   If the number of atoms in the studied molecule is relatively low, the phases can be “guessed”, and then recombined with the experimental amplitudes.

II. Solving heavy atom sub-structure
   The protein phase information is recovered by determining the positions of electron-dense atoms by direct and Patterson methods.

III. Molecular Replacement (MR)
   The phase information can be extracted from another known, similar structure and used together with the experimental amplitudes to solve the structure.
4.4 Anomalous scattering

Anomalous scattering is a resonance effect that occurs when the wavelength (or energy) of the incident X-rays is close to the excitation energy of core electrons around the nucleus. “Normal”, elastic scattering occurs when the incoming X-rays set the electrons around the nucleus in vibration and radiation of the same energy (wavelength) is re-emitted in all directions. If the energy of the incoming radiation is high enough, some electrons are excited into higher orbitals or completely expelled, and when the electrons fall back, radiation of the same wavelength is re-emitted with a change in phase compared to the elastically scattered waves, and this effect is called anomalous scattering. Some of the energy is used to bring about the transition, thus reducing the intensity of the scattering. In a typical X-ray experiment, light atoms such as N, O, C have no anomalous contribution whereas atoms commonly found as protein ligands such as Cu, Co, Fe and Zn and ions that bind to protein in an unspecific manner such as halides show significant anomalous scattering. Normally this difference is concealed in the experimental errors, but if the anomalous signal is large and the data is collected with high accuracy, these small differences in intensities can be detected.

4.5. Single wavelength anomalous diffraction – SAD

The following section is based on the reference “One-and-a-half wavelength approach” by Dauter. [35]

If there are anomalous scattering atoms present in the crystal the simplest way of determining the protein phases of the diffraction spots is single wavelength anomalous diffraction (SAD), since it only requires a single dataset collected at one wavelength. This is achieved in two steps: I determining the substructure of the anomalous scatterers present in the molecule using the differences in intensities between Friedel mates (\(DF_{hkl} = |F_{hkl}| - |F_{-h-k-l}|\)). II Using this known anomalous contribution (partial structure) to obtain protein phase information.

4.5.1 The anomalous contribution to the total scattering

If there are no anomalous scatterers present in the molecule, the total structure factor for a reflection originating from “normal” atoms is \(F_N(h)\):

\[
\hat{F}_N(\vec{h}) = \sum_{i} f_i^0(\theta) \exp\left[2\pi i \vec{h} \cdot \vec{r}_i\right]
\]
Where $f^0_i$ is the scattering factor of atom $i$, $\mathbf{h}$ is a reciprocal lattice vector (composed of three integers – $h$, $k$, $l$) and $\mathbf{r}_i$ is the position vector of atom $i$, (composed of three coordinates $x,y,z$) and $N$ is the number of atoms in the unit cell.

The normal scattering magnitude decreases with the scattering angle $\theta$, but is not dependent on the wavelength of the incoming X-rays.

If there is a mixture of $N$ normal atoms and $A$ anomalous atoms, then the total scattering $\mathbf{F}_T$ is:

$$F_T(h) = \sum_i^N f_i^n(\theta)\exp[2\pi i (\mathbf{h} \cdot \mathbf{r}_i)] + \sum_j^A \left(f_j^n(\theta) + i f_j'(\lambda)\right)\exp[2\pi i (\mathbf{h} \cdot \mathbf{r}_j)]$$

Where $\mathbf{F}_N$ is the structure factor of the normal atoms. The anomalous atoms contribute with a normal scattering, $\mathbf{F}_A$, that depends on the scattering angle ($\theta$) in the same way as $\mathbf{F}_N$, and two anomalous contributions $\mathbf{F}'_A$ (real) and $\mathbf{iF}''_A$ (imaginary) that are dependent on the wavelength ($\lambda$) of the incident X-rays but not on the resolution. The real component $\mathbf{F}'_A$ is anti-parallel to the normal scattering $\mathbf{F}_A$ since there is some loss of energy when the electrons are excited.

### 4.5.2 The breakdown of Friedel’s Law

In the absence of anomalous scatterers, the intensities of Friedel pair reflections are identical. Since the anomalous dispersion component $\mathbf{iF}''_A$ is always rotated 90° counter clockwise of the normal scattering $\mathbf{F}_A$, the phase and amplitude of the Friedel mates $\mathbf{F}_T(+)$ and $\mathbf{F}_T(-)$ are no longer equal. This effect can be detected if the anomalous signal is large enough and the data is collected with high accuracy.

### 4.5.3 Determining the substructure of the anomalous scatterers

The magnitude of the difference between the Friedel mates is called the Bijvoet difference. It can be shown that the Bijvoet differences depends sinusoidally on the difference between the phases of the total scattering vector ($\mathbf{\phi}_T$) and the anomalous scattering vector ($\mathbf{\phi}_A$).
This means that the maximum Bijvoet difference occurs when the two phases differ in angle by ± 90°. When this occurs, the Bijvoet difference is directly proportional to $\mathbf{F}_A$, which we need to know to be able to determine the substructure of the anomalous scatterers:

$$\Delta F^\pm = 2F_A^\pm \sin(90°) = 2F_A^- = 2F_A (f^- / f^+)$$

The large Bijvoet differences can then be used in direct method approaches or Patterson searches or combinations of these methods (these will not described here) to locate the positions of anomalous scattering atoms within the asymmetric unit. When this is done, we know the length and angle of the anomalous scattering vectors $\mathbf{F}_A$ (values for $\mathbf{F}'_A$ and $\mathbf{F}''_A$ for a certain wavelength is taken from tables) and the length of the total scattering vector $\mathbf{F}_T$ of the vector equation. From this we can deduce the length and angle of the normal scattering vector $\mathbf{F}_N$ and consequently the missing phase angle of $\mathbf{F}_T$.

4.5.4 Breaking the phase ambiguity

Even if the measurements were error-free, there are always two solutions that satisfy the vector relations placed symmetrically around $\varphi A - 90°$. This means that for a typical dataset with > 30 000 reflections, each reflection can have either one of the two possible angles and it is not possible to try all the solutions for this problem. If experimental noise is considered you do not get two distinct solutions but a continuous probability curve of the phase angle which makes this an even more difficult task. Ramachandran & Raman [36] first postulated that the total phase $\mathbf{F}_T$ that lies closer to the anomalous phase should be chosen for map calculation. When we add the phase probability curve from the anomalous atoms, also termed the partial known structure, the total probability distribution shows that the SAD solution that is closer to the anomalous scatterer vector is more likely to be correct. This is only true if there is a large fraction of anomalous scatterers within the molecule. Another method of reaching the correct phase angle is to use a method called iterative single-wavelength anomalous scattering (ISAS). [37] By adding the two possible SAD solutions ($\mathbf{F}_T$’s) and calculating an electron density map from this, the total map contains the proper molecular structure placed on more or less homogeneous noise. By using noise-filtering methods (density modification) such as histogram matching and solvent flattening, it is then possible to distinguish protein from solvent and through an iterative process the total scattering vector is shifted towards the correct solution.
Figure 8. Argand diagrams showing the basic steps in a SAD phasing experiment, adapted from [35]. In this and all subsequent figures, the total scattering of all atoms are shown in blue, the normal scattering atoms are shown in black, the normal part of the anomalous scatterer is shown in orange, the real part of the anomalous scattering is shown in magenta and the imaginary part is shown in red. a) When there is a mixture of normal and anomalous scatterers, Friedel’s Law does not hold since the imaginary part of the anomalous contribution is always rotated 90° in front of the normal scattering for the anomalous scatterer. The Friedel mates of the total scattering is thus not equal in amplitude and phase. b) The negative Friedel mate \( F_{-} \) is usually reflected across the horizontal axis and represented as a complex conjugate \( ^*F_{-} \) as it more clearly shows the relations between Friedel mate vectors and phases. c) When the phase difference between the total average structure factor \( F_{\tau} \) and the normal structure factor \( F_{\Lambda} \) of the anomalous atoms differs by 90° the Bijvoet difference reaches a maximum and when the difference is 0° or 180°, no Bijvoet difference is observed. d) After the substructure of the anomalous scatterer have been solved by anomalous difference Patterson methods and/or direct methods, there are two solutions for the phase of the total average structure symmetrically placed around \( \phi_{A-90°} \) that for fill the vector equation. e) This phase ambiguity can be solved by various methods. The combination of the phase probability distribution of the heavy atom substructure (Ppart) also termed the partial structure with the phase probability distribution of the total scattering vector (Pano), shows that the total phase closer to the phase of the anomalous scatterer is more likely to be correct. However, this is only true when a there is a substantial amount of anomalous scatterers in the molecule. f) Wang’s ISAS – iterative single-wavelength anomalous scattering - method. A map is calculated from the vector sum of the two equal solutions and protein features are enhanced in an iterative process, thereby resolving the phase ambiguity.
5. The aim of this study – Results.

In spite of the extensive structural studies of the RUNX1 RD there was still missing pieces in the puzzle. The structure of the uncomplexed RD had not been determined and it was thus not possible to get a full understanding of the effect that DNA and CBFb had on the RD. In the light of this we managed to crystallize the RUNX1 RD in its unbound form and produce crystals diffracting to high resolution. The analysis of the free RD structure revealed two bound chloride ions that gave rise to new questions. We pursued then with NMR line width analysis to examine the binding affinities of the chloride ions that were coordinated by DNA binding residues. By EMSA we also examined whether chloride ions and acidic amino acids could affect DNA binding. During this process we replaced Cl\(^-\) with Br\(^-\), which brought us to more technical aspects of X-ray crystallography where we tested the power of SAD phasing using anomalous signal from bromide ions. The encouraging results from these experiments have initiated the use of SAD phasing as a standard method in our lab.

5.1 Preparation and crystallization of the RD (paper I)

Previous attempts to solve the structure of the RUNX1 RD with NMR have proved to be very difficult due to aggregation problems. To improve solubilisation, a construct comprising the RD (residues 46-185) of Runx1 with the two cysteins mutated to serines was prepared and the modified protein was more prone to crystallize than the wild-type construct. Initial crystals in space group C2 grew very quickly with heavy nucleation. These crystals were not well ordered as judged by the high mosaic spread of the X-ray diffraction pattern. The crystals were fragile and were often destroyed during heavy atom soaks. In spite of extensive screening for suitable heavy atoms using various Pt, Os, Hg salts and Xe, we were unable to obtain heavy atom derivatives. Glutaraldehyde treatment stabilized the crystals, but it was still not possible to incorporate heavy atoms.

By adding DNA that contained the high affinity RUNX1 consensus binding site to the crystallization solution, the crystal growth was slowed down at the same time as microcrystals and protein aggregates were eliminated. This lead to an improvement of the size and diffraction quality of the crystals and the resolution increased from 2.6 Å to 1.7 Å with a lower mosaic spread of the diffracting spots. The crystals were also less fragile and the glutaraldehyde fixation was no longer necessary. A salting-in procedure where the initial salt concentration in the drop was about 1.1 M lead also to the formation of well behaving crystals in space group R32 diffracting to 2.0 Å. These findings,
together with the production of Se-Met labelled crystals, were crucial for solving the RUNX1 RD structure at the high resolution discussed in paper II.

5.2 Comparison of the free and the complexed RD structures (paper II)

The structure of the uncomplexed RD was determined in space group C2 from crystals diffracting to 2.2 Å resolution using a three wavelength Se-Met MAD. The MAD data was collected from a single crystal in three passes at three different wavelengths. One additional Se-Met labelled crystal was used to collect a high-resolution data set to a d-spacing of 1.25 Å. The asymmetric unit in space group C2 contained two RD molecules. The R32 structure was determined by molecular replacement using a C2 monomer as search model and contained one molecule in the asymmetric unit.

In this manner we obtained three independent copies of the RD in two crystal forms that provided a good picture of the normal flexibility and variation within the molecule. The overall shape of the individual molecules is very similar but there are some significant differences. The most significant difference is found at the N-terminal a-helix (residues Ser50-Asp57), which is only visible in molecules of the C2 crystal form where it showed some positional variation. This helix is part of the conserved RD and has not been reported in previous RD structures. In addition, the tip of the DNA-binding loop L9 (residues Gly141-Ser145) displays large displacements between all three molecules and reflects the flexibility of this loop.

Comparison of the apo-RD structure with the previously published RD/CBFβ, RD/DNA and RD/CBFβ/DNA structures revealed significant structural changes in three regions of the RUNX1 RD. The major changes occur in loop L11 where the imidazole ring of His163 swings 12 Å towards CBFβ from a closed position in the free RD to a open conformation in the complex structures (the “S-switch”) (Figure 9).
Figure 9. The largest conformational changes occur in loop L11, the so-called “S-switch” region. The stereo figure show the free RD in dark grey and the RD in complex with DNA and CBFβ in light grey.

Following this movement, strands β1-β2 close in against the core of the RD and loops L5 and L9 shift towards the DNA-binding end of the RD, causing the removal of three buried water molecules. In contrast to the above mentioned changes, the major groove binding motif comprising loops L3, L12 and strands β3 and β12 remains unaffected by complex formation with either DNA or CBFβ. This motif, which in addition includes a buried water molecule that hydrogen bonds the loops and strands, also exist in the STAT family of transcription factors (Figure 10).
The conserved major groove binding motif shared between STAT and RUNX proteins includes a conserved water molecule that is coordinated by residues from loop L3 and strand β9 in the RD and the equivalent regions in STAT. The RD is shown in dark grey, and STAT1 is shown in light grey. The conserved, buried water molecules are shown as black spheres in both the RD and STAT.

The combined changes seem to be crucial for sustained DNA binding and the fact that CBFβ can induce the closed to open transition suggests that one role of CBFβ is to trigger the S-switch and to stabilize the Runt Domain in a conformation that is enhanced for DNA binding mainly by affecting the L5 and L9 loops.

5.3 Chloride and acidic amino acid binding to the RD (paper II and III)

From the solved apo structures, we identified two putative chloride ions specifically bound to the RD. We verified their positions by a bromide SAD experiment where NaCl was replaced with KBr in the crystallization buffer. Through their anomalous signal, bromide ions were identified at the identical positions as the chloride ions (paper II). Both chloride ions are situated in regions that make both major and minor groove interactions. The first chloride is coordinated by the Nε and Nη1 side chain atoms of the DNA binding residue Arg139 in loop L9 and the Val170 main-chain nitrogen in loop L12, binding to the major groove of DNA. The second chloride is situated between loop L5 and...
and L9, coordinated by the N\textsuperscript{82} atom of Asn112 and the backbone nitrogen of Glu116.
Chloride ion binding was verified with \(^{35}\text{Cl}\) FWHH NMR studies and the binding constant \(K_{d,\text{Cl}}\) could be determined to 34 mM for the free RD and 56 mM for the RD/CBF\(\beta\) complex. This study also showed that both chloride ions are lost upon DNA binding (paper III). Through EMSA we could in addition demonstrate that chloride ions and acidic amino acids have a positive effect on DNA binding in vitro (paper II). The physiological relevance of the chloride induced DNA-binding stimulation is unclear, but the stimulation observed for acidic amino acids suggests a regulatory mechanism through the cooperative interaction with other transcription factor activation domains.

5.4 The use of halides for phase determination (paper IV)

In the era of structural genomics there is a need for fast, simple and reliable phasing methods. Promising candidates for being “universal protein phasers” are negatively and positively charged ions displaying significant anomalous signals at the wavelengths normally used in crystallography. Such candidates include Halides (Br, I), monovalent cations (Cs\(^+\), Rb\(^+\)) and lanthanide ions (Gd\(^{3+}\), Eu\(^{3+}\), Sm\(^{3+}\) or Ho\(^{3+}\)) all shown to bind proteins in a very general manner by replacing water molecules.

We wanted to investigate the quality of the SAD phase information obtained from two bromide atoms bound to a 15 kDa protein molecule. The program SOLVE was used in a SAD procedure to analyse the data and the correct bromide ion sites were easily identified from the anomalous difference Patterson maps. The initial protein phases calculated from the bromide ion positions produced poor maps that were not traceable even after density modification. We noted that a large portion of the strong low resolution data were systematically absent due to overloaded reflections. We could overcome this problem by merging the low resolution data from a native crystal with the bromide SAD data, the completeness of the data increased and the density modification scheme worked successfully, providing clearly interpretable electron density maps.
6. Future

The work described in this thesis has provided new structural insights into the regulation of a conserved family of transcription factors and emphasises the importance of studying all individual components in a system in order to obtain a detailed view of a complex system.

A question that arises is if the S-switch and the structural changes that accompany it represent a general mechanism valid for all members of the Runt family of transcription factors that have a CBFβ partner. In *C. elegans* there is a Runt member but no CBFβ homologue has been identified. We have initiated structural studies of this Runt member to find out what conformation it has in the absence and presence of DNA.

We have shown that the N-terminal part of the conserved Runt motif is folded as a flexible α-helix but we have not been able to assign a function to this helix. It might be involved in the interaction with other components of the transcriptional machinery.

Recently our collaborators have shown a direct interaction between the Runt domain and the Calmodulin dependent phosphatase Calcineurin that indicates a regulatory function for the N-terminal α-helix and we are now following up this with structural studies.
7. Acknowledgement

Först vill jag tacka för min tid hos min flygande och hårt roende handledare Uwe, och för de många erfarenheter jag har samlat på mig. De kommer att vara till stor nytta för mig i framtiden. Jag vill speciellt tacka för den stora frihet jag har fått i mitt arbete.

Någon har sagt till mig att bond-döttrar har fötterna extra stabilt på jorden, och det måste gälla Liz. En positiv, effektiv och stödjande prefekt som UCMP inte klarar sig utan.

Utan min vapendragare Fredrik Ekström hade det inte varit lätt. Tack för allt ditt stöd och din uppmuntran, speciellt när det varit jobbigt…

Andreas, Erik, Fredrik och 黄胜华 (Shenghua Huang) (och alla andra på UCMP…) och vår före detta labbmedlem Therese: TACK för att jag har fått snoka och rota i era problem och allmänt lägga min näsa i blöt. Jag tror även att Jeanette, Ulrika, Karina och Anders Olofsson ligger i farozonen för framtida blötnäsor.

Christin – Wilket lyft för gruppen (på alla sätt) när du kom. Och vilket lyft för ordningen i mina labbanteckningar, och bland mina prover…

Anders Karlsson – Det är alltid lika roligt att hitta likasinnade pryl- och teknikbögar… Synd bara att jag inte har några pengar att slösa bort.

Tobias, vår labbguru. Du bist die kompetenteste Person, der ich begegnet bin.

Marek, vår datorsmiskare, tack för att jag fick 1 Gb internminne före de andra…


Alla andra på UCMP: Ledzen, ni blir inte av med mig… så vi hinner fika och äta lunch fler gånger… och jobba hårt såklart.

I synnerhet behöver nog Calle mig även i forstsättningen. Han är en så dålig förlorare att jag måste låta honom vinna då och då i pingis för att hans självkänsla inte ska brytas ner. För då vet man aldrig hur det går.


Slutligen vill jag tacka de människor (släkt och vänner…) utanför jobbet som ger mig stöd – de ingår i La Mia Familla – och speciellt mamma och pappa som kontinuerligt hjälper mig med allt möjligt.

Och Ritu och mina två underverk – Oskar å Sophie…

Utan gröngölingsboken hade jag nog aldrig varit i denna situation. En underbar bok…
8. References


