Understanding the interaction of antibodies and transcription factors with their ligands through structural biology

Využití strukturní biologie ke studiu interakce protilátek a transkripčních faktorů s jejich ligandy

Dissertation thesis

Supervisor: RNDr. Pavlína Maloy Řezáčová, Ph.D.
This dissertation describes my original work except where acknowledgement is made in the text. It is not substantially the same as any work that has been, or is being submitted to any other university for any degree, diploma or any other qualification.

Prague, 29.4.2015

Mgr. Jana Škerlová
I hereby declare that Mgr. Jana Škerlová contributed significantly to all scientific publications that are part of this dissertation. She performed most of the experiments by herself and also took part in their planning; she interpreted all results and wrote the publications. Her contribution to the individual publications is expressed in percentage below.

1. Crystallization of the effector-binding domain of repressor DeoR from *Bacillus subtilis*

2. Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis*

3. Optimization of the crystallizability of a single-chain antibody fragment

4. Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85

Prague, 29.4.2015

............................................................

RNDr. Pavlína Maloy Řezáčová, Ph.D.
This work was elaborated in the Laboratory of Structural Biology at the Institute of Molecular Genetics and the Institute of Organic Chemistry and Biochemistry, AS CR, v.v.i.

I acknowledge the funding sources that made my work possible. I was funded by the Grant Agency of the Charles University in Prague, project GA UK No. 567012. This work was also supported in part by projects RVO 61388963 and 68378050 awarded by the Academy of Sciences of the Czech Republic, by the Ministry of Education of the Czech Republic, projects LK11205 (program “NAVRAT”), LO1304 (program “NPU I”), and ME08016, and also by project GA203/09/0820 awarded by The Grant Agency of the Czech Republic. I also thank the team of the BESSY II electron storage ring, Helmholtz-Zentrum Berlin, for the opportunity to collect X-ray diffraction data.

First and foremost I want to thank my supervisor Pavlína Řezáčová for her patient guidance, encouragement, and for being an inspiration. I am also grateful to Václav Veverka for introducing me to NMR and for sharing his expertise.

My special thanks belong to my colleagues Milan Fábry and Vlastimil Král, for being my “lab mentors” in all aspects of scientific work and life, and also to Irena Sieglová who has always been ready to help and give me support. I wish to thank Petr Pachl and Jiří Brynda, who helped me with all crystallography hardware and software, Jitka Kredbová and Věra Mrkvičková for technical assistance with protein preparation, and Petr Těšina for his help with NMR assignments.

I would also like to express my appreciation to everyone from the Laboratory of Structural Biology for being such a great team. I am grateful to all colleagues and collaborators who contributed to the publications included in this thesis.

Last but not least, I thank my family for the unceasing encouragement, support and attention, most of all to my husband Jan for his care and patience.
# Table of contents

1 Abbreviations .......................................................................................................................... 6  
2 Abstract .................................................................................................................................. 8  
3 Abstrakt.................................................................................................................................. 10  
4 Preface ................................................................................................................................... 12  
5 Introduction ............................................................................................................................. 14  
  5.1 Regulation of metabolism in *Bacillus subtilis* ...................................................................... 14  
    5.1.1 Carbohydrate catabolism .............................................................................................. 14  
    5.1.2 Carbon catabolite repression of catabolic operons ........................................................ 15  
    5.1.3 Induction of catabolic operons ....................................................................................... 16  
    5.1.4 Metabolic repressors in *B. subtilis* ............................................................................. 18  
    5.1.5 Transcriptional regulator DeoR form *B. subtilis* .......................................................... 19  
  5.2 CD44 antigen .......................................................................................................................... 22  
    5.2.1 CD44 function ................................................................................................................. 22  
    5.2.2 CD44 structure ............................................................................................................... 23  
    5.2.3 Hyaluronate-binding domain of CD44 ........................................................................... 24  
    5.2.4 CD44 regulation ............................................................................................................. 27  
    5.2.5 CD44 in pathology ......................................................................................................... 28  
    5.2.6 CD44 as a target in cancer therapy .................................................................................. 29  
    5.2.7 Anti-CD44 antibodies in cancer therapy ....................................................................... 30  
    5.2.8 Monoclonal antibody MEM-85 against CD44 ............................................................... 32  
6 Aims of the study ...................................................................................................................... 33  
7 Publications ............................................................................................................................... 34  
  7.1 Crystallization of the effector-binding domain of repressor DeoR from *Bacillus subtilis* ........................................................... 35  
  7.2 Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis* ......................................................................................................... 42  
  7.3 Optimization of the crystallizability of a single-chain antibody fragment ............................ 57  
  7.4 Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85 .......................................................... 66  
8 Discussion .................................................................................................................................. 99  
  8.1 Structural studies of deoxyribonucleoside regulator form *Bacillus subtilis* ...... 99  
  8.2 Epitope mapping for the anti-CD44 monoclonal antibody MEM-85 ......................... 102  
9 Conclusions .............................................................................................................................. 106  
10 References ............................................................................................................................... 108
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AraR</td>
<td>Arabinose repressor</td>
</tr>
<tr>
<td>CBS</td>
<td>Cystathionine beta synthase domain</td>
</tr>
<tr>
<td>CcpA</td>
<td>Catabolite control protein A</td>
</tr>
<tr>
<td>CcpN</td>
<td>Catabolite control protein N</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>CD44</td>
<td>Hyaluronate receptor</td>
</tr>
<tr>
<td>CD44s</td>
<td>Standard isoform of CD44</td>
</tr>
<tr>
<td>CD44v</td>
<td>Variant isoform of CD44</td>
</tr>
<tr>
<td>C-DeoR</td>
<td>C-terminal effector-binding domain of deoxyribonucleoside</td>
</tr>
<tr>
<td>CggR</td>
<td>Central glycolytic genes regulator</td>
</tr>
<tr>
<td>c-myc</td>
<td>Transcription factor, proto-oncogene</td>
</tr>
<tr>
<td>CP</td>
<td>Cytoplasmic region</td>
</tr>
<tr>
<td>cre</td>
<td>Catabolite-responsive element</td>
</tr>
<tr>
<td>CrH</td>
<td>Catabolite repression Hpr-like protein</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>DBTBS</td>
<td>Database of transcriptional regulation in <em>B. subtilis</em></td>
</tr>
<tr>
<td>deo</td>
<td>Deoxyribonucleoside catabolism operon from <em>E. coli</em></td>
</tr>
<tr>
<td>DeoR</td>
<td>Deoxyribonucleoside regulator</td>
</tr>
<tr>
<td>Dra</td>
<td>Deoxyriboaldolase</td>
</tr>
<tr>
<td>Drm</td>
<td>Phosphodeoxyribomutase</td>
</tr>
<tr>
<td>EBD</td>
<td>Effector-binding domain</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>ERM</td>
<td>Ezrin, radixin, moesin</td>
</tr>
<tr>
<td>FadR</td>
<td>Fatty acid metabolism regulator</td>
</tr>
<tr>
<td>FrlR</td>
<td>Fructosamine repressor</td>
</tr>
<tr>
<td>FruR</td>
<td>Fructose repressor</td>
</tr>
<tr>
<td>GamR</td>
<td>Glucosamine repressor</td>
</tr>
<tr>
<td>glp</td>
<td>Glycerolphosphate operon</td>
</tr>
<tr>
<td>GmuR</td>
<td>Glucomannan operon repressor</td>
</tr>
<tr>
<td>GntR</td>
<td>Gluconate operon repressor</td>
</tr>
<tr>
<td>HABD</td>
<td>Hyaluronate-binding domain</td>
</tr>
<tr>
<td>HER2</td>
<td>Human epidermal growth factor receptor 2</td>
</tr>
<tr>
<td>HPr</td>
<td>Phosphocarrier protein</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>HutC</td>
<td>Histidine utilization repressor</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IolR</td>
<td>Myo-inositol operon repressor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LacI</td>
<td>Lactose operon repressor</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption/ionization, time of flight</td>
</tr>
<tr>
<td>MEM-57</td>
<td>Murine monoclonal antibody against human CD3</td>
</tr>
<tr>
<td>MEM-85</td>
<td>Murine monoclonal antibody against human CD44</td>
</tr>
<tr>
<td>Mfd</td>
<td>Mutation frequency decline (transcription repair coupling factor)</td>
</tr>
<tr>
<td>NagC</td>
<td>N-acetylglucosamine repressor</td>
</tr>
<tr>
<td>NagR</td>
<td>N-acetylglucosamine repressor</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NupC</td>
<td>Nucleoside uptake protein C</td>
</tr>
<tr>
<td>O</td>
<td>Ordered conformation</td>
</tr>
<tr>
<td>PD</td>
<td>Partially disordered conformation</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank (<a href="http://www.rcsb.org">http://www.rcsb.org</a>)</td>
</tr>
<tr>
<td>Pdp</td>
<td>Pyrimidine nucleoside phosphorylase</td>
</tr>
<tr>
<td>pelB</td>
<td>Pectate lyase B leader sequence</td>
</tr>
<tr>
<td>PRD</td>
<td>Phosphotransferase regulator domain</td>
</tr>
<tr>
<td>PTS</td>
<td>Phosphoenolpyruvate: sugar phosphotransferase system</td>
</tr>
<tr>
<td>RAT</td>
<td>Ribonucleic antiterminator sequence</td>
</tr>
<tr>
<td>RbsR</td>
<td>Ribose operon repressor</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small-angle X-ray scattering</td>
</tr>
<tr>
<td>scFv</td>
<td>Single-chain variable antibody fragment</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SorC</td>
<td>Sorbitol operon regulator</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>TCR</td>
<td>T-cell receptor</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane region</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>TreR</td>
<td>Trehalose repressor</td>
</tr>
<tr>
<td>V_H</td>
<td>Variable domain of the antibody heavy chain</td>
</tr>
<tr>
<td>V_L</td>
<td>Variable domain of the antibody light chain</td>
</tr>
<tr>
<td>XylR</td>
<td>Xylose operon repressor</td>
</tr>
<tr>
<td>YbgA</td>
<td>Alternative name for GamR</td>
</tr>
<tr>
<td>YdhQ</td>
<td>Alternative name for GmuR</td>
</tr>
<tr>
<td>YvvoA</td>
<td>Alternative name for NagR</td>
</tr>
<tr>
<td>σ^A</td>
<td>RNA polymerase σ-43 factor</td>
</tr>
<tr>
<td>σ^L</td>
<td>RNA polymerase σ-54 factor</td>
</tr>
</tbody>
</table>
2 Abstract

Understanding protein function highly benefits from the knowledge of its three-dimensional structure, especially in the case of protein-ligand complexes. Structural biology methods such as X-ray crystallography, SAXS and NMR are therefore widely used for structural studies of protein-ligand interaction. In this work, these methods were used to understand two biological processes involving protein interactions: X-ray structural analysis was used to study binding of effector molecule to a prokaryotic transcription factor. NMR and SAXS techniques were used to study interaction of a monoclonal antibody with its protein antigen.

Transcriptional regulator DeoR negatively regulates the expression of catabolic genes for the utilization of deoxyribonucleosides and deoxyribose in *Bacillus subtilis*. DeoR comprises an N-terminal DNA-binding domain and a C-terminal effector-binding domain (C-DeoR), and its function is regulated by binding of a small-molecular effector deoxyribose-5-phosphate. We determined crystal structures of C-DeoR both in the free form and in complex with deoxyribose-5-phosphate. Structural analysis revealed unique covalent binding of effector molecule through a reversible Schiff-base double bond with an effector-binding-site lysine residue. The physiological nature of this binding mode was confirmed by mutational analysis and mass spectrometry of the Schiff-base adducts formed in solution. Comparison of the free and effector-bound structures of C-DeoR explained on a molecular level the mechanism of DeoR function as a molecular switch.

CD44, the cell receptor for hyaluronate, is involved in cell adhesion, migration, and tumor metastasis. CD44 binds hyaluronate through the hyaluronate-binding domain (HABD), which triggers an extensive conformational rearrangement in HABD that induces CD44 shedding. We constructed a single-chain variable fragment (scFv) of antibody MEM-85. MEM-85 is of therapeutic interest, recognizes a poorly characterized epitope in HABD, blocks hyaluronate binding, and induces CD44 shedding. Antibody scFv fragments exhibit poor homogeneity caused by scFv oligomerization. We developed a novel optimization approach, which we used for the preparation of a homogeneous scFv sample for NMR-based mapping of the epitope of MEM-85 in HABD. This mapping together with mutational analysis identified epitope residues in the C-terminal tail of the HABD. Additionally, we built a rigid-body model of the HABD – scFv MEM-85 complex based on the SAXS data. These biophysical methods provided a detailed insight into the
mechanism of MEM-85 action: MEM-85 binds to an epitope located outside the hyaluronate-binding site and induces a conformational change in the HABD tail, similar to that induced by hyaluronate. The mechanism of MEM-85 cross-blocking of hyaluronate binding is not a direct physical competition, but rather an allosteric, relay-like effect.
3 Abstrakt

K pochopení funkce proteinu významně přispívá znalost jeho trojrozměrné struktury, obzvláště v případě komplexů s ligandy. Metody strukturní biologie, jako je rentgenová krystalografie, SAXS nebo NMR, jsou proto ve strukturní analyse interakcí proteinů s ligandy široce využívány. V této práci byly tyto metody využity k objasnění molekulární podstaty dvou biologických dějů využívajících proteinové interakce: rentgenostrukturologické analýzy byla použita ke studiu vazby efektorové molekuly k prokaryotnímu transkripčnímu faktoru, metody NMR a SAXS byly využity ke studiu interakce monoklonální protilátky s proteinovým antigenem.

Transkripční regulátor DeoR negativně reguluje expresi katabolických genů pro utilizaci deoxyribonukleosidů a deoxyribosy v bakterii Bacillus subtilis. DeoR se skládá z N-koncové DNA-vazebné domény a z C-koncové efektor-vazebné domény (C-DeoR). Jeho funkce je regulována vazbou nízkomolekulárního efektoru deoxyribosa-5-fosfáty. Vyřešili jsme krystalové struktury C-DeoR ve volné formě a v komplexu s deoxyribosa-5-fosfátem. Strukturní analýza odhalila unikátní vazbu mezi molekulou efektoru a lysylovým zbytkem efektor-vazebného místa prostřednictvím reversibilní dvojné vazby typu Schiffovy base. Fysiologická relevance této vazby byla potvrzena mutačními experimenty a také analyzou tvorby kovalentních aduktů v roztoku pomocí hmotnostní spektrometrie. Porovnání struktur volné C-DeoR a komplexu C-DeoR s efektorem vysvětlilo na molekulární úrovni funkci DeoR jako molekulárního spínače.

CD44, buněčný receptor pro hyaluronát, se účastní buněčné adhese, migrace a metastazování nádorů. CD44 váže hyaluronát prostřednictvím hyaluronát-vazebné domény (HABD), což indukuje v HABD rozsáhlou konformační změnu, vedoucí k proteolytickému uvolnění CD44 z povrchu buněk. Vytvořili jsme jednořetězový variabilní fragment (scFv) protilátky MEM-85. MEM-85 má terapeutický potenciál, rozpoznává nedostatečně popsaný epitop v HABD, blokuje vazbu hyaluronátu a indukuje odstěpení CD44. Fragmenty scFv jsou kvůli své oligomeraci obvykle nehomogenní. Vyvinuli jsme nový optimalizační přístup, pomocí něhož jsme připravili homogenní vzorek scFv vhodný pro NMR mapování antigen-vazebného epitopu protilátky MEM-85 v CD44 HABD. Toto mapování společně s mutační analýzou vedlo k identifikaci protilátkového epitopu v C-koncové části HABD. Na základě SAXS analýzy jsme také vytvořili model komplexu HABD s scFv MEM-85. Tyto biofyzikální metody přinesly
detailní vhled do mechanismu funkce MEM-85: MEM-85 váže epitop, který se nachází mimo hyaluronát-vazebné místo, a indukuje v C-konci HABD strukturní změny, jež jsou podobné změnám vyvolaným vazbou hyaluronátu. Blokace vazby hyaluronátu protilátkou MEM-85 proto nespočívá ve fyzické kompetici o vazebné místo, ale jde spíše o allosterický přenosový efekt.
4 Preface

This work includes a combination of two research projects: structural studies of prokaryotic transcriptional regulators and structural studies of antibody-antigen interaction. The common basis of both projects is the understanding of protein-protein and protein-ligand interactions and their effects on protein structure and function through structural biology. The first project is based on protein X-ray crystallography technique, whereas the second one utilizes NMR and SAXS techniques, illustrating thus the powerful synergy of these methods in structural biology.

X-ray crystallography is the principal method for determination of structures of biological macromolecules. It has made an immense impact in the field of biochemical and biological research; there have been 25 Nobel laureates connected to X-ray crystallography. Determination of protein three-dimensional structure is driven by the objective to gain the knowledge on the protein’s biological activity, which is very often associated with the characterization of ligand binding. In an ideal case, X-ray crystallography enables structure determination for protein-protein and protein-ligand complexes at a resolution high enough to directly observe atomic details of covalent and non-covalent interactions in three dimensions. This allows for further analysis of the protein’s mode of action on the molecular level, as well as for the design of small-molecular inhibitors of the protein’s function. Moreover, the knowledge of high-resolution three-dimensional structures of the protein in the free and ligand-bound states constitutes the basis for understanding the ligand-induced changes in the protein conformation and thus unravels the effects of the ligand binding on the function of the protein.

The use of X-ray crystallography is irreplaceable for the detailed studies of the function of prokaryotic transcriptional regulators on the molecular level. These transcription factors function as molecular switches and their action usually includes extensive ligand-induced conformational changes associated with the modulation of oligomeric state. The first part of this thesis comprises X-ray structural study of deoxyribonucleoside regulator DeoR from Bacillus subtilis with the aim to understand the regulation of its function by the small-molecular effector, deoxyribose-5-phosphate.
NMR methods play a prominent role in biochemistry and molecular biophysics for probing molecular structure and interactions. X-ray crystallography is still a simpler and more straightforward method of structure determination, especially in the case of larger molecules and complexes. However, solution NMR is the method of choice in structure determination of molecules that are resistive to forming well-ordered crystals. Moreover, the classical NMR applications, where the complete structure determination is not the goal, demonstrate the wide-range impact of NMR in biological and biophysical research. In particular, solution NMR has become the paramount method for studying molecular dynamics and for characterization of protein-ligand and protein-protein interactions. The method of choice for this application is the differential chemical shift perturbation method, which is based on the fact that the NMR-active nuclei that are located at the interaction interface undergo changes in their chemical shift, which are not observed for distant nuclei. Therefore, the location of the binding site on the protein can be mapped.

Recently, this widely applicable approach has been increasingly used to characterize antigen – antibody complexes, usually for the definition of the antigen-binding epitopes of therapeutic antibody fragments. The second part of this thesis is focused on the characterization of antibody binding to a cancer marker CD44. We have prepared a single-chain variable fragment of antibody MEM-85, which recognizes human CD44 antigen and which is of therapeutic potential and employed the NMR-based differential chemical shift perturbation method to map the antibody epitope.
5 Introduction

5.1 Regulation of metabolism in *Bacillus subtilis*

In order to be evolutionarily successful, bacteria need to be competitive both under limited and excessive supply of carbon and energy resources. They have to be able to efficiently select the preferred, most energetic carbon source out of a mixture of nutrients available in the environment at a given moment. The underlying regulatory mechanisms have been extensively studied in *Escherichia coli* as well as in *Bacillus subtilis*, the model Gram-positive bacterium [1, 2].

*Bacillus subtilis* is a soil bacterium, that can use at least 18 different mono- or disaccharides as sole sources of carbon and energy, as it produces extracellular hydrolases to digest polysaccharides of plant biomass [3]. The most preferred carbon and energy source is glucose and the expression of the genes involved in the catabolism of alternative nutrients is strictly controlled by two key mechanisms: carbon catabolite repression and induction. The carbon catabolite repression mechanism ensures that the genes encoding enzymes for the utilization of the non-preferred carbon sources are not expressed in the presence of the preferred nutrient. Additionally, the more specific induction mechanism ensures that genes for alternative catabolic pathways are not expressed, unless the respective alternative substrate is present in the environment. This regulation network includes a wide range of specific repressor proteins and functions at the level of transcription of metabolic operons [1, 4].

5.1.1 Carbohydrate catabolism

*B. subtilis* excretes various enzymes including glucanases, proteases and nuclease to decompose complex nutrients present in the soil. Amylases and glucanases are of major importance, as they degrade glucose polymers and produce maltodextrins, maltose, and glucose, the most preferred nutrient. The resulting mono-, di-, or oligosaccharides are imported into the cell by facilitated diffusion, secondary active transporters, ATP-binding cassette transporters, or the phosphoenolpyruvate: sugar phosphotransferase system (PTS), which phosphorylates the transported substrates. The PTS system not only transports nutrients into the cell, but also plays role in signal transduction. The system consists of two universal proteins: enzyme I and phosphocarrier protein (HPr), and of numerous
substrate-specific enzymes II. Enzymes II are protein complexes that perform the actual transport of the nutrients which are simultaneously phosphorylated by the cooperative action of enzyme I, HPr, and components of enzyme II complex [1, 3].

Once in the cell, the di- or oligosaccharides or their phosphorylated forms are cleaved by the respective glycosidases or phosphoglycosidases, respectively. The non-phosphorylated products are phosphorylated by kinases and enter the pentose phosphate pathway and glycolysis, yielding pyruvate which is converted by pyruvate dehydrogenase complex to acetyl-coenzyme A to be further metabolized in the tricarboxylic acid cycle. Apart from sugars, \textit{B. subtilis} can also import and utilize various carbohydrate derivatives, such as aminosugars, hexitols, hexuronic acids, deoxyribose, and glycerol [3].

### 5.1.2 Carbon catabolite repression of catabolic operons

Carbon catabolite repression mechanisms ensure that catabolic genes and operons for alternative nutrients are not transcribed as long as preferred carbon and energy sources are present in the environment. This regulation can be achieved by three mechanisms: (1) control of the catabolic genes transcription by global sensor proteins registering the presence of preferred nutrients, (2) regulation of the intracellular availability of the catabolic genes inducers (inducer exclusion), or (3) induction prevention by operon-specific regulators controlled by preferred nutrients [2, 5, 6]. \textit{B. subtilis} combines these mechanisms while using the PTS system and HPr kinase as the intracellular sensors of nutrient availability that trigger further carbon catabolite repression pathways [1, 7].

HPr can be phosphorylated by an ATP-dependent HPr kinase activated by specific metabolites, which leads to decreased PTS activity. HPr kinase functions both as kinase and phosphatase. The kinase activity is stimulated by glycolytic intermediates and inhibited by inorganic phosphate, which also stimulates the phosphatase activity of HPr kinase. Phosphorylated HPr protein can than act as a cofactor of the PTS system, or phosphorylate other regulatory proteins. Phosphorylated HPr binds to the C-terminal effector-binding domain of catabolite control protein A (CcpA), a transcriptional regulator that binds through its N-terminal DNA-binding domain to operator palindromic sequences \textit{cre} (catabolite-responsive element) in various operons and genes. The interaction of CcpA with the phosphorylated HPr downregulates the expression of catabolic genes by binding the \textit{cre} elements or by direct contact of the HPr-bound CcpA with RNA polymerase. CcpA
can also function as a transcriptional activator of glycolytic and overflow metabolism genes in the presence of preferred nutrients by interacting with the cre sites located upstream of the promoters of their operons, which enables very fast growth [1, 7].

The presence of preferred nutrients in the cell can prevent transport or formation of effector molecules inducing the expression of catabolic operons. An example of this inducer exclusion mechanism is B. subtilis glp operon, where the PTS is involved in formation of the inducer glycerol-3-phosphate. Monophosphorylated HPr can phosphorylate glycerol kinase and enhance its kinase activity [8]. The concentration of this monophosphorylated form of HPr is lower in the presence of preferred carbon sources, as it is depleted as a cofactor of the PTS by the nutrient-specific enzyme II or by phosphorylation of other HPr residues by HPr kinase. Therefore, glycerol kinase is less active and lower amount of the inducer glycerol-3-phosphate is formed under carbon catabolite repression conditions [1, 9].

Several catabolic operons for the non-preferred nutrients are positively regulated by activators or antiterminators in the presence of the inducer. Transcriptional factors of this type are inactivated in the absence of the inducer by the PTS-dependent phosphorylation at their phosphotransferase regulator domains (PRDs). Some of these proteins, which are subject to carbon catabolite repression, are only active (binding DNA) when phosphorylated at other sites in their PRDs by the monophosphorylated HPr protein. This additional HPr-dependent phosphorylation of the PRDs prevents induction of the catabolic operons for the non-preferred nutrients [1].

5.1.3 Induction of catabolic operons

The operons coding for the enzymes required for the catabolism of specific nutrients are not expressed, unless the respective carbon source is present in the environment. This process is controlled at the level of transcription by three different mechanisms: transcriptional repressors, activators, and antiterminators [3].

Transcriptional repressors in B. subtilis negatively control expression of operons for the catabolism of numerous nutrients including glucose, glucosamine, N-acetyl-glucosamine, trehalose, glucomannan, fructose, fructosamine, inositol, gluconate, arabinose, ribose, deoxyribose, xylose, and galactose. In the absence of the nutrient, the repressor binds through the N-terminal DNA-binding domain to the DNA operator site, usually upstream of the operon. In the presence of the effector (the nutrient or its catabolic intermediate), the
The repressor binds the effector through the C-terminal effector-binding domain and the resulting conformational change results in lower affinity towards the DNA operator or even promotes dissociation of the repressor from the operator (see Figure 1) [1].

![Figure 1: Regulation of catabolic genes expression by a transcriptional repressor](image)

The molecular-switch function of a transcriptional repressor is schematically illustrated.

Transcriptional activators regulate the expression of catabolic genes for nutrients such as oligo-β-glucoside, fructose, mannitol, mannose, and glucitol and can be divided into two classes: (1) activators functional only in the complex with their specific inducer and (2) activators modulated by phosphorylation dependent on the presence of the substrate. Vast majority of *B. subtilis* activators are phosphorylation-regulated and can be further subclassified as activators of the σ^A^ or σ^L^-dependent promoters. The activator proteins typically contain an N-terminal DNA-binding domain and several other domains with additional functions and conserved phosphorylation sites, which are phosphorylated by PTS components (HPr or enzyme II) in the absence of the respective nutrient, which results in the inactivation of the activator protein [1, 10].

Transcriptional antiterminators regulate the expression of catabolic operons coding for utilization of nutrients including glucose, sucrose, β-glucosides, and glycerol. Most of the transcriptional antiterminators consist of an N-terminal RNA-binding domain and two PRD domains. The antiterminator-regulated operons are transcribed constitutively, but the transcription stops at transcriptional terminator sequences located upstream of the open reading frames. The 5’-parts of the terminators can be incorporated into stem-loop structures formed by ribonucleic antiterminator inverted repeat sequences (RATs). The antiterminator regulatory proteins bind the RAT sequences through their RNA-binding
domains, and their binding promotes transcription. In the absence of the specific nutrient, the PRDs are phosphorylated by PTS components (HPr or enzyme II), resulting in inactivation of the respective antiterminator protein [1, 10].

5.1.4 Metabolic repressors in B. subtilis

Majority of the transcriptional repressors regulating catabolic genes in *Bacillus subtilis* that have been characterized so far are members of following families of prokaryotic transcriptional factors: DeoR, GntR, SorC or LacI (see Table 1). The classification of the families is based on the sequence similarity of the N-terminal helix-turn-helix DNA-binding domains of the repressors. These conserved DNA-binding domains are combined with C-terminal effector-binding domains belonging to various structural families. The C-terminal domains are responsible for effector molecule recognition and for repressor oligomerization (see Figure 1, Table 1). This chimeric domain organization allows for great structural and functional variety of metabolic repressors. Several *B. subtilis* repressors have been studied on the structural level by X-ray crystallography (Table 1), usually as individual domains with the exception of the recently determined crystal structure of the full-length repressor NagR, also in the complex with its DNA operator [11, 12].

**Table 1: Repressors in carbohydrate utilization in Bacillus subtilis**

<table>
<thead>
<tr>
<th>Sugar</th>
<th>Operon</th>
<th>Repressor</th>
<th>DBD family (PDB code)</th>
<th>EBD family (PDB code)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabinose</td>
<td>araABDLMNOPQ-abfA</td>
<td>AraR</td>
<td>GntR (4EGY [13])</td>
<td>LacI (3TB6 [14])</td>
</tr>
<tr>
<td>Deoxyribose</td>
<td>dra-nupC-pdp</td>
<td>DeoR</td>
<td>SorC</td>
<td>SorC</td>
</tr>
<tr>
<td>Fructose</td>
<td>yarPONML</td>
<td>FrlR</td>
<td>GntR</td>
<td>HutC (2IKK)</td>
</tr>
<tr>
<td>Fructose</td>
<td>fruRBA</td>
<td>FruR</td>
<td>DeoR</td>
<td>DeoR</td>
</tr>
<tr>
<td>Galactose</td>
<td>galKT</td>
<td>AraR</td>
<td>GntR (4EGY [13])</td>
<td>LacI (3TB6 [14])</td>
</tr>
<tr>
<td>Glucomannan</td>
<td>ydhMNOPQRST</td>
<td>GmuR (YdhQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucomate</td>
<td>gunRKPZ</td>
<td>GntR</td>
<td>GntR</td>
<td>FadR</td>
</tr>
<tr>
<td>Glucosamine</td>
<td>gamAP</td>
<td>GamR (YbgA)</td>
<td>GntR</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>gapB-speD</td>
<td>CcpN</td>
<td>DeoR</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>cggR-gapA-pgk-tpiA-pgm-eno</td>
<td>CggR</td>
<td>SorC (YbgA)</td>
<td>SorC (2OGK [15])</td>
</tr>
<tr>
<td>Inositol</td>
<td>iolABCDEFGHIJ</td>
<td>IolR</td>
<td>DeoR</td>
<td>DeoR</td>
</tr>
<tr>
<td>N-acetyl-glucosamine</td>
<td>nagP-nagAB-yvoA</td>
<td>NagR (YvoA)</td>
<td>GntR (2WV0 [11], 4WWC [12])</td>
<td>HutC (2WV0 [11], 4WWC [12])</td>
</tr>
<tr>
<td>Ribose</td>
<td>rbsRKDMCB</td>
<td>RbsR</td>
<td>LacI</td>
<td>LacI</td>
</tr>
<tr>
<td>Trehalose</td>
<td>trePAR</td>
<td>TreR</td>
<td>GntR</td>
<td>HutC (2OGG [16])</td>
</tr>
<tr>
<td>Xylose</td>
<td>xylAB</td>
<td>XylR</td>
<td></td>
<td>NagC/XylR</td>
</tr>
</tbody>
</table>

*acompiled from [1], Pfam database [17], database of transcriptional regulation in Bacillus subtilis (DBTBS [18]), Protein data bank (PDB, http://www.rcsb.org), and UniProt (www.uniprot.org); *domains were not unambiguously classified; DBD, DNA-binding domain; EBD, effector-binding domain.
5.1.5 Transcriptional regulator DeoR form *B. subtilis*

*Bacillus subtilis* can utilize deoxyribonucleosides and deoxyribose as a source of carbon and energy. The deoxyribonucleosides are imported into the cell by nucleoside uptake protein C (NupC) and phosphorolytically cleaved by pyrimidine nucleoside phosphorylase (Pdp), yielding a base and deoxyribose-1-phosphate, which is converted to deoxyribose-5-phosphate and further processed by deoxyriboaldolase (Dra) to acetaldehyde and glyceraldehyde-3-phosphate. The genes encoding these three key enzymes of this catabolic pathway are grouped into the *dra-nupC-pdp* operon, which is under negative transcriptional control of the deoxyribonucleoside regulator, DeoR [19]. Another enzyme involved in this catabolic pathway is phosphodeoxyribomutase (Drm) which converts deoxyribose-1-phosphate to deoxyribose-5-phosphate. This enzyme is encoded by the *drm* gene located far from the *dra-nupC-pdp* operon in *B. subtilis* and its expression is therefore probably modulated by some other regulatory protein [19].

The *deoR* gene is located immediately upstream of the *dra-nupC-pdp* operon [19] and codes for the 313 amino acids long DeoR repressor protein that recognizes a palindromic operator sequence located between the positions -60 and -43 relative to the transcription start site for the *dra-nupC-pdp* operon. Also, a direct repeat sequence identical to the 3′-half of the palindrome, that might be another possible DeoR binding site, is located downstream to the palindrome (see Figure 2, Table 2) [20].

The *dra-nupC-pdp* operon is subject to another negative regulation mechanism, carbon catabolite repression by glucose [20, 21]. The operon carries one *cre* element, located downstream of the *dra* transcription start site (Figure 2). The catabolite repression by HPr can be in this case substituted by the homologous CrH protein [21], which is also phosphorylated by HPr kinase and then induces CcpA binding to *cre*, similarly to phospho-HPr [22, 23]. Mfd protein is also involved [21], which is a strand-specific DNA repair protein that is able to displace RNA polymerase form downstream-located *cre* sites [24].
Figure 2: Schematic overview of the deoR-dra-nupC-pdp operon
Genome positions of the individual genes in the minus DNA strand are schematically illustrated: deoR is the deoxyribonucleoside regulator gene, dra is the deoxyriboaldolase gene, nupC is the nucleoside uptake protein C gene, and pdp is the pyrimidine nucleoside phosphorylase gene. Promoter and operator sites for RNA polymerase σ-43 factor (σA), deoxyribonucleoside regulator (DeoR), and catabolite control protein A (CcpA) are also indicated in colors. Compiled from [18, 19].

Table 2: Promoter and operator sites of the dra-nupC-pdp operon in Bacillus subtilis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Location</th>
<th>Binding sequence (cis-element)a</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcpA</td>
<td>+60 – +82</td>
<td>GCTTTGAAACCGCATACACAAAA</td>
<td>[19]</td>
</tr>
<tr>
<td>DeoR</td>
<td>-60 – -22</td>
<td>ATTGAAACAAAATTTCAATTACCAATTTACAT</td>
<td>[19, 20, 25]</td>
</tr>
<tr>
<td>σA</td>
<td>-60 – +10</td>
<td>ATGGACAAATTTCAATTACCAATTTACAT ATGTTCAAAAGTCGGTTATGCTAAAAATATCTTAACAA</td>
<td>[19]</td>
</tr>
</tbody>
</table>

The minimal operator binding sites for catabolite control protein A (CcpA) and deoxyribonucleoside regulator (DeoR) are highlighted in red. Compiled from [18].

Both deoxyribose-1-phosphate and deoxyribose-5-phosphate are suggested to function as inducers of the catabolic operon, but only deoxyribose-5-phosphate appears to be the true effector [20]. Despite the participation of Pdp and Drm in the catabolism of pyrimidine ribonucleosides, no induction of expression of the dra, pdp, or drm genes has been observed in B. subtilis in the presence of ribonucleosides [19]. The carbon catabolite repression of the dra-nupC-pdp operon only occurs when the operon is not repressed by DeoR [21]. Also, neither CcpA, nor DeoR binds strong enough to totally prevent expression of the dra-nupC-pdp. This might be a protective mechanism against the toxic effects of deoxyribose-5-phosphate [19], or it may also be a result of the action of phosphodeoxyribomutase, which constitutively produces very low amounts of deoxyribose-5-phosphate from deoxyribose-1-phosphate [26].
The DeoR regulator from *Bacillus subtilis* is not related to the DeoR from *Escherichia coli* [18, 25], which belongs to the DeoR family and regulates the *deo* operon and has been studied in great detail [27]. Although both DeoR regulators from *B. subtilis* and *E. coli* contain a similar helix-turn-helix motif typical for DNA-binding proteins [28, 29] there is very little sequence similarity between them, and there is no similarity between their DNA operator sites [20, 25, 30]. The N-terminal DNA-binding domain of the *B. subtilis* DeoR is homologous to the SorC family of regulator proteins from various bacteria species that play role in transcriptional regulation of sugar alcohol catabolism genes [25]. The C-terminal effector-binding domain of *B. subtilis* DeoR is homologous to the SorC-type regulators of glucose metabolism genes that probably recognize glyceraldehyde-3-phosphate as the effector molecule [25]. The domain organization of DeoR is shown in Figure 3.

![Figure 3: Domain organization of DeoR](image)

The modular domain organization of deoxyribonucleoside regulator (DeoR) is schematically illustrated with residue numbers indicating domain boundaries according to [17].

The only one SorC-type transcriptional regulator characterized so far in *B. subtilis* is the central glycolytic genes regulator, CggR [18]. The sequence identity of DeoR and CggR is 20% (according to the UniProt Align tool at www.uniprot.org/align). Crystal structure of the CggR effector-binding domain, which is also classified as the SorC-type effector-binding domain, has been determined and revealed a double-Rossmann fold core with three accessory peripheral subdomains involved in repressor oligomerization and effector binding. CggR binds four different ligands: fructose-1,6-bisphosphate, dihydroxyacetone phosphate, glucose-6-phosphate, and fructose-6-phosphate. Each of these ligands induces individual specific conformational changes that influence the dimer interface in the effector-binding domain and affect the distribution of the repressor oligomers [15].
5.2 CD44 antigen

The CD44 antigen, a type I transmembrane glycoprotein also known as lymphocyte homing/adhesion receptor, epican, or Hermes antigen, was first described in 1983 [31-33]. CD44 is the principal cell receptor for hyaluronate [34], which is a major component of extracellular matrix. CD44 can also mediate binding to other components of extracellular matrix, including fibronectin, collagen, laminin, and fibrin and to several cell-surface ligands, such as E-, P-, and L-selectin [35-39]. Intracellular portion of CD44 is involved in signaling pathways [39, 40]. It interacts through ankyrin adaptor protein with spectrin and is involved in hyaluronate-dependent cell adhesion and motility [41]. Other binding partners of the intracellular CD44 domain are ezrin, radixin and moesin (ERM proteins) involved in the regulation of cell migration, shape, and protein resorting in the plasma membrane [42]. The ERM proteins mediate the interaction of CD44 intracellular domain with filamentous actin [43] which results in CD44 clustering [39, 40].

CD44 is expressed in most mammalian tissues, mainly in the gastrointestinal tract, skin, spleen, and leukocytes. Lower expression level can be observed in erythrocytes, neurons, and brain. CD44 is not expressed in kidney, hepatocytes, and mature thymocytes [32, 44]. The expression of CD44 on blood cell progenitors is strictly regulated during differentiation and it can be induced in many cell types only after cell activation [45, 46].

5.2.1 CD44 function

The CD44 receptor is involved not only in cell adhesion, but also in cell survival and proliferation, growth factor presentation, apoptosis, and cell motility and migration [47]. CD44 is a key player in lymphocyte trafficking and recirculation, including lymphocyte extravasation and homing to the peripheral lymphoid organs or the site of inflammation. These processes are initiated by a CD44-mediated rolling adhesion of lymphocytes to endothelial cells presenting hyaluronate (see Figure 4) [48]. CD44 also mediates and regulates the mobilization and distribution of blood cell progenitors during hematopoiesis in bone marrow [39, 49]. CD44 is also important for hyaluronate metabolism as it executes hyaluronate internalization which allows subsequent hyaluronate degradation by acid hydrolases within low-pH cellular compartments such as lysosomes [50]. This process is regulated by a palmitoylation-dependent association of CD44 with lipid rafts [51].
Figure 4: CD44 in lymphocyte homing
Schematic representation of CD44-mediated rolling of lymphocytes on endothelium is shown. T-cell capture and rolling are mediated by CD44 (orange) recognition of hyaluronate (green) presented on the endothelial cells. Firm adhesion is induced by the binding of integrins (dark blue) activated by inflammatory chemokines (red) to their counter-receptors (light blue) on the endothelium, which results in T-cell diapedesis, extravasation, and migration to the site of inflammation. Amended from http://www.glycoforum.gr.jp/.

5.2.2 CD44 structure

The 363 amino acids long CD44 glycoprotein (residues 21-383; see Figure 5) belongs to the cartilage Link protein superfamily [52, 53]. The extracellular portion of CD44 comprises the N-terminal 158 amino acids long extended Link domain coded by exons S1-S4 followed by the 112 amino acids long stem region coded by exons S5, S6, and S7 [54]. The stalk-like structure can be enlarged by the insertion of variable exon products V1-V10 between residues 222 and 223 due to alternative splicing [45, 55]. The extracellular portion of CD44 is followed by the 21 amino acids long hydrophobic transmembrane region coded by exon S8. Exons S9 and S10 code for the 72 amino acids long intracellular C-terminal domain [56, 57].

CD44 is subject to posttranslational modification. The N-terminal part is heavily N-glycosylated (Asn25, 57, 100, 110, and 120 [58]), the membrane-proximal serine- and threonine-rich extracellular region is O-glycosylated, an intracellular serine residue can be phosphorylated, and two conserved cysteine residues in the transmembrane domain and in the membrane-proximal cytoplasmic region can be palmitoylated [45, 51, 57]. The stem
region contains putative proteolytic cleavage site that allows for CD44 shedding [59]. After the ectodomain shedding, the intracellular domain can be released by an intramembrane presenilin-dependent γ-secretase cleavage [60] and the released cytoplasmic CD44 domain can then function as a transcription factor in the nucleus, promoting expression of numerous genes [61].

![Figure 5: CD44 gene and protein structure](image)

**Figure 5: CD44 gene and protein structure**

CD44 gene composition is shown comprising 10 constant exons (S1-S10) and 10 alternatively spliced variable exons (V1-V10), coding for the extracellular, transmembrane (TM), and cytoplasmic (CP) portions of CD44. The CD44 protein products, the standard (CD44s) and variant (CD44v) isoforms are schematically illustrated with the color coding indicating the extracellular, transmembrane, Link, and stalk domains, the variable exons, and also the interaction sites for ankyrin and for the ezrin, radixin, and moesin (ERM) proteins. N-glycosylation sites are represented by grey circles, O-glycosylation sites by orange circles, and glycosaminoglycan-binding sites by yellow circles. S-S indicates disulfide bonds. Adapted from [57].

### 5.2.3 Hyaluronate-binding domain of CD44

The N-terminal extended Link domain is responsible for hyaluronate binding and therefore, it is often referred to as the CD44 hyaluronate-binding domain (CD44 HABD). CD44 HABD binds hyaluronate molecules in a wide range of molecular weights. The minimum length of hyaluronate chain to be recognized by HABD is a hexamer and optimal
binding was achieved for hyaluronate octamer [62, 63]. Three-dimensional structures of mouse and human HABD have been determined by X-ray crystallography and NMR, respectively [62-64]. These structures revealed that HABD has a novel type of C-lectin fold, which can be designated as: \( \beta_0 - \beta_0' - \beta_1 - \alpha_1 - \beta_2 - \beta_3 - \alpha_2 - \beta_4 - \beta_5 - \beta_6 - \beta_7 - \beta_8 - \beta_9 - \alpha_3 \) (see Figure 6). The 100 amino acids long hyaluronate-binding Link module (residues 32-132, in bold) contains two highly conserved disulfide bridges and its sequence is generally conserved within the family of hyaluronate-binding proteins (80 – 90% sequence similarity) [33, 40, 45, 65]. Link module is the most evolutionarily conserved part of the whole CD44 antigen [66, 67]. In contrast to other Link superfamily members, the Link module in CD44 HABD is equipped with an extension lobe formed by \( \beta_0 \) and \( \beta_0' \) at the N-terminus and by \( \beta_7, \beta_8, \beta_9, \) and \( \alpha_3 \) at the C-terminus of HABD [62]. This extension is crucial for structural integrity as it contains an additional disulfide bridge important for the correct folding of the Link module [68] and it also plays an important role in hyaluronate binding [65, 69]. The key interactions for the hyaluronate binding to human CD44 are hydrogen bonds with residues Arg41, Tyr42, Arg78 a Tyr79 from the Link module, but residues Arg29, Arg150, Arg154, Lys158 a Arg162 from the extension lobe also take part in hyaluronate binding, suggesting a more complex hyaluronate-binding mode than in the case of other hyaluronate-binding proteins of the Link family [62, 63, 65].

The solution structure of HABD in complex with hyaluronate determined by NMR [64] unraveled a major allosteric conformational change in HABD induced by the ligand binding (Figure 6): \( \beta_9 \) strand and \( \alpha_3 \) helix become an unstructured random coil, and \( \beta_8 \) and \( \beta_0 \) strands are repositioned relative to each other, which results in an enhanced flexibility of the C-terminus of the HABD [64]. The ligand-free state of HABD is referred to as “ordered” (O) and the ligand-bound state as “partially disordered” (PD). The O-state structure of the ligand-bound murine CD44 HABD has been also captured by X-ray crystallography, probably because the rigid ordered conformation is more likely to form crystal contacts [63]. There is a two-state equilibrium between the O- and PD-conformational states of CD44 HABD and both conformations are capable of hyaluronate binding [70]. However, the PD-conformation exhibits higher affinity towards hyaluronate and it is also more resistant to cell detachment from a hyaluronate-coated surface in comparison to the O-conformation [70]. This conformational equilibrium has been proposed to be a mechanism essential for the proper CD44-mediated cell rolling, specifically for the formation of new contacts of the CD44 receptor with hyaluronate at the cell rolling front and for their disruption at the rear edge [70]. The ligand-induced shift in
the equilibrium in the favor of the PD-conformation also results in the elongation of the stalk-like membrane-proximal region of CD44. This makes the stem region more easily accessible to matrix metalloproteases and thus enhances the proteolytic susceptibility of the cleavage sites [64], which results in hyaluronate-induced HABD shedding and regulates thus cell motility and migration [59, 71, 72] (Figure 7).

Figure 6: Structures of the CD44 hyaluronate-binding domain
A: CD44 hyaluronate-binding domain (HABD) topology is depicted in a cartoon diagram of the ligand-free CD44 HABD (left, PDB code 1POZ [62]) and hyaluronate-bound CD44 HABD (right, PDB code 2I83 [64]). Secondary structure elements are labeled; α-helices are in red, β-strands are in yellow, and random coil is in green. B: Hyaluronate-induced structural changes are depicted in a cartoon-surface diagram of the ligand-free CD44 HABD (left, PDB code 1POZ [62]) and hyaluronate-bound CD44 HABD (right, PDB code 2I83 [64]). The Link module is shown in blue and the extension lobe in gray. The hyaluronate octamer from the crystal structure of the hyaluronate-bound murine CD44 HABD (PDB code 2JCQ [63]) is superposed onto the structures to mark the hyaluronate-binding site.
Figure 7: Hyaluronate-induced CD44 shedding
Schematic representation of the hyaluronate-induced structural changes (see also Figure 6) in CD44 hyaluronate-binding domain (HABD) is shown, which promotes proteolytic cleavage in the stem region resulting in CD44 ectodomain shedding. Illustrated according to [64, 70].

5.2.4 CD44 regulation

Hyaluronate binding to CD44 is regulated by a number of factors including cell activation state, receptor clustering, glycosylation pattern, phosphorylation of intracellular Ser291 and receptor dimerization through intramembrane cysteine residues [39, 73-76]. Especially the O-glycosylation pattern together with the cytoplasmic tail of CD44 can affect membrane subdomain localization which can modulate the interaction of CD44 with hyaluronate [77]. CD44 glycosylation also directly controls its binding capacity to fibrin and immobilized fibrinogen [38, 78]. Moreover, CD44 has an inducible character. It is in an inactive (low-affinity) state on the surface of resting T-cells and monocytes and it only binds hyaluronate after the transition to a high-affinity state due to cytokine activation or T-cell activation through the TCR-CD3 [46]. The transition to the high-affinity state is induced by enzymatic removal of terminal sialic acids from the N-linked glycosylation chains on Asn25 and Asn120 [75, 79]. This mechanism targets selected populations of peripheral blood cells to the inflammation site, where they recognize hyaluronate presented on the surface of cytokine-activated endothelial cells and leave the blood stream [62, 63].

Some of the CD44 functions are determined by alternative splicing of the CD44 precursor mRNA [57, 80]. The CD44 gene, located on chromosome 11, spans 50 thousand base pairs of genomic DNA [40] and comprises 20 exons, out of which 10 exons are variable (V1-V10; see Figure 5) [57, 80]. Omitting all these exons gives rise to the
standard or hematopoietic form CD44s (80-85 kDa, 270 extracellular amino acids [45, 56]), which is expressed in most cell types [81], typically on activated leukocytes, where it plays role in their adhesion and trafficking to the inflammation site [82]. The variable exons code for CD44 membrane-proximal region and give rise to at least 20 variant mRNA molecules and at least six out of them were observed also in the form of variant isoform protein product CD44v [82]. Some of the variant isoforms exhibit constitutive hyaluronate-binding activity [73, 83]. Inclusion of exon V3 creates a site for chondroitin sulfate and heparan sulfate modification. This modification enables binding of several growth factors and chemokines, which amplify the immune response [84, 85]. Inclusion of exon V6 creates a binding site for hepatocyte growth factor and vascular endothelial growth factor [85, 86].

5.2.5 CD44 in pathology

CD44 also plays a role in various types of pathological processes, such as extensive inflammatory diseases [87] and cancer, especially in tumor-cell migration and metastasis [47, 72, 81, 88]. Autoimmune inflammatory disorders are characterized by elevated counts of lymphocytes bearing the activated high-affinity-state CD44 [89]. This phenomenon has been observed in patients suffering from lupus erythematosus and chronic arthropathies [90] and in rheumatoid arthritis patients [91]. CD44 has therefore been suggested to play a crucial role in the maintenance and propagation of autoimmune inflammatory diseases [45].

Expression of CD44 variant isoforms can induce metastatic phenotype in locally growing tumor cells [88, 92] and usually correlates with bad prognosis [88]. Inclusion of variant exon V6 has been shown to induce tumor metastasis [92] which is probably caused by a loss of contact with the neighboring cells due to altered adhesion properties of the cells. Proteolysis of the extracellular matrix induced by CD44 variant isoforms [71, 93] facilitates the detachment of tumor cells from the tumor site and their spread outside the original tumor.

Another explanation for the metastatic behavior might be an increased mobility of the cells along the hyaluronate-rich extracellular matrix due to the overexpression of CD44 standard variant CD44s [94]. A significant number of acute lymphoblastic leukemia cases exhibit very high or low CD44 expression on leukemic B-cells when compared to normal B-cell progenitors. This can be used for the detection of minimal residual disease [39, 95,
Low expression of CD44 has been correlated with the aggressive phenotype of some solid tumor types [97-99]. CD44 in cancer cells promotes angiogenesis and the tumor growth, as CD44 has proliferative and anti-apoptotic effects in general [67, 80]. Expression of CD44 is important in the progression of many tumor types [81], is often observed on cancer stem cells [100], and can be a marker of poor prognosis [101]. The cytostatic and pro-apoptotic function of p53 tumor suppressor protein [102] is connected with the inhibition of CD44 expression. The p53 factor binds to the promoter region of the CD44 gene and thus prevents excess mitotic activity in the cell which prevents tumor growth [103]. CD44 interacts with the HER2 tyrosine kinase from the EGF receptor family. Overexpression of HER2 by breast and ovarian cancer cells is an indicator of poor prognosis. CD44s isoform can form interchain disulfide bonds with HER2 and hyaluronate binding to CD44 then stimulates the tyrosine kinase activity of HER2 which triggers signaling pathways that promote cell growth and proliferation [49, 104]. Moreover, CD44 receptor clusters are capable of recruiting matrix metalloprotease 9 to the cell membrane and thus localizing its proteolytic collagenase activity to the tumor cells. This leads to extracellular matrix degradation which by itself facilitates tumor growth but it also releases growth factors and angiogenic factors sequestered in the extracellular matrix, which further stimulates tumor progression [105].

5.2.6 CD44 as a target in cancer therapy

Many approaches in cancer therapy are focused on CD44. CD44 has been identified as a marker of cancer stem cells which represent only a minor population of tumor cells, but are essential for tumor progression and tumor survival during treatment. Cancer stem cells often resist the treatment and this causes cancer relapse or recurrence at distal sites [106]. Cancer initiating cells might thus be targeted through CD44 [57].

One of promising therapeutic approaches is CD44 blocking, the abruption of CD44 interaction with hyaluronate [107]. Hyaluronate induces anti-apoptotic signaling pathways through interaction with the activated CD44 during inflammation and tissue healing or with the constitutively active CD44 on cancer cells [73, 107]. The CD44 inactive state structure on normal cells is different as it is not activated by the specific conditions of the environment [88]. The constitutive hyaluronate-binding by CD44 can be exploited in cancer therapy by four different approaches [107]: (1) small hyaluronate oligomers can be used to compete with the endogenous hyaluronate, (2) soluble CD44 or other
hyaluronate-binding proteins can be used to outcompete the endogenous cell-surface CD44 for hyaluronate binding, (3) the post-transcriptional expression of CD44 variant isoforms can be inhibited using CD44 siRNAs, and (4) hyaluronate binding can be blocked by an anti-CD44 antibody recognizing the hyaluronate-binding site [108-112]. Recently, another approach emerged: (5) first small-molecular inhibitors of the hyaluronate binding to CD44 have been identified by a combination of fragment screening and structure-based optimization of the lead compounds. These compounds are based on 1,2,3,4-tetrahydroisoquinoline and they might be potential therapeutic agents in cancer, but also in chronic inflammatory diseases [113].

Small hyaluronate oligomers or antibodies can be further equipped with cytostatic drugs, possibly encapsulated in liposomes, internalized through CD44 [50] and activated inside the tumor cell [114-116]. This approach also increases the drug solubility [116]. The next part is focused on the use of antibodies in CD44-targeted cancer therapy.

### 5.2.7 Anti-CD44 antibodies in cancer therapy

Known anti-CD44 antibodies recognize a wide variety of antigenic determinants in the CD44 molecule [117]. The standard variant CD44s contains three distinct groups of antigen epitopes. Antibodies recognizing epitope group 1 inhibit T-cell clustering, antibodies recognizing epitope group 2 block hyaluronate binding to CD44, and antibodies recognizing epitope group 3 block lymphocyte adhesion to the epithelium of blood vessels [39, 44]. Antibodies recognizing CD44 can be used to block hyaluronate binding to CD44 and disrupt its interactions with the extracellular matrix [117, 118] or even induce CD44 ectodomain shedding [119]. Activating antibodies can be used to induce CD44 pro-apoptotic signaling [120-122], and anti-CD44 antibodies in general can be used for targeting anti-cancer active substances to the tumor site [123-125].

Several anti-CD44 antibodies targeting cancer stem cells show a promising potential in the preclinical stage [126, 127]. Murine monoclonal antibody H90 activates CD44 signaling that triggers pro-apoptotic and proliferation-inhibitory pathways in human acute myeloid leukemia [121, 122, 128, 129]. Murine monoclonal antibody P245 reduces tumor growth in the mouse model of human basal-like breast cancer and was able to eliminate cancer stem cells and thus prevent breast cancer relapse [130]. Murine monoclonal antibody H4C4 inhibits human pancreatic tumor growth, metastasis, and recurrence in mouse xenografts by inhibiting specific CD44 signaling pathways [127]. Recombinant
human monoclonal antibody GV5 against CD44v8-v10 isoform completely inhibited tumorigenesis of human cervix and larynx carcinomas in mouse models, by inducing antibody-dependent cell cytotoxicity mechanisms [131]. Recombinant humanized antibody RO5429083/RG7356 recognizes a glycosylated part of CD44 present in all CD44 isoforms. The antibody interferes with hyaluronate binding, but does not induce CD44 shedding. It inhibits the growth of head and neck squamous cell carcinoma tumor cells, breast cancer cells, and chronic lymphocytic leukemia cells in vitro and in mouse xenografts [118, 132, 133]. The RO5429083/RG7356 antibody has recently reached phase I clinical trials in patients with CD44-positive malignant solid tumors and a study with the antibody alone or an immunotoxin conjugate of the antibody with cytarabine in patients suffering from acute myeloid leukemia [127].

The use of antibodies targeting the CD44v6 variant isoform appears to be the most successful approach in CD44-based cancer immunotherapy [134, 135]. For example, monoclonal antibody 1.1ASML against CD44v6 has been shown to block metastases formation in animal models of cancer [119]. These antibodies are especially potent in the form of radioimmunoconjugates or immunotoxin conjugates. Several clinical phase I trials were performed with murine (U36, BIWA 1), chimeric (chU36), and humanized (BIWA 4) monoclonal antibodies recognizing CD44v6 isoform [135]. Murine antibody U36 conjugated with technetium $^{99m}$Tc was successfully tested as a diagnostic tool for the detection of lymph node metastases [136], however, it induced human anti-mouse immune response in later radioimmunotherapy studies with the rhenium $^{186}$Re-conjugate [137] and had to be chimerized for further therapeutic applications [138]. The phase I study with the chimerized chU36 antibody has shown reduced adverse immune response while maintaining the effective antibody responses [139-142]. Murine monoclonal antibody BIWA 1 conjugated with $^{99m}$Tc has been proven an effective and relatively safe diagnostic tool in a radioimmunoscinigraphy study [143]. Due to the human anti-mouse adverse immune response in the patients [143], a humanized version, BIWA 4 was constructed for further therapeutic applications, also known as Bivatuzumab. Bivatuzumab reached the phase I clinical trials in the form of a radioimmunoconjugate with $^{186}$Re [123, 144]. Bivatuzumab was also tested in phase I clinical trials in the form of an immunotoxin conjugate with the cytotoxic drug mertansine. Although these trials seemed very promising, they were abruptly ended after the death of one patient. [124, 145-148].
5.2.8 Monoclonal antibody MEM-85 against CD44

Murine monoclonal antibody MEM-85 of IgG2 class recognizes the CD44 antigen [149]. It was raised against peripheral blood monocytes of a patient suffering from large granular lymphocyte leukemia [149]. The antibody reacts with all peripheral blood monocytes, with granulocytes, and with cell lines including CEM, MOLT-4, ARH-77, HEL, HL-60. MEM-85 reacts very weakly or it does not react at all with cell lines such as Jurkat, Nalm-6, Daudi, and KM3 [32].

MEM-85 has been shown to block hyaluronate binding to CD44 [149, 150] and the antibody binding also induces CD44 shedding, similar to that induced by hyaluronate binding [151]. The HABD epitope recognized by MEM-85 has not yet been unambiguously identified. The epitope is supposed to be localized in the N-terminal part of the HABD [32, 33, 44] and mutational analysis results suggest that residues Lys38, Arg41, and Tyr42 are important for MEM-85 binding [33]. This represents an epitope from the epitope group 2 [32, 44, 150] discussed above. Some consider the MEM-85 epitope to be structural [33], but others have observed MEM-85 binding to SDS-denatured CD44 under non-reducing conditions, which suggests a sequential epitope, stabilized by a disulfide bridge [32].

Antibody MEM-85 might play an important role in diagnostics of some types of malignancies, especially for the detection of minimal residual disease. It might also be of therapeutic interest, as it has been shown to inhibit lung cancer cells growth in murine models (Godar, S., personal communication). This antibody can also be used as a tool for better understanding hyaluronate binding to CD44 and of the subsequent cellular processes, which would be useful for structure-based design of inhibitors of the hyaluronate binding.
6 Aims of the study

This work is a part of two long-term projects: (1) X-ray crystallographic studies of bacterial transcription factors and (2) recombinant antibody fragments targeted against therapeutically relevant human antigens.

Structural studies of selected metabolic repressors from the model bacterium *Bacillus subtilis* aim to elucidate the mechanisms of metabolite repression on the molecular level. Recombinant fragments of monoclonal antibodies targeting selected human antigens are constructed to be used as potential diagnostic or immunotherapeutic agents, and as molecular tools for better structural and functional characterization of their antigens.

**Specific research aims:**

- Structural characterization of deoxyribonucleoside regulator from *Bacillus subtilis*, the binding of the ligand deoxyribose-5-phosphate, and the effect of the ligand-induced structural changes on the function of the regulator.

- Definition of the binding epitope of monoclonal antibody MEM-85 in the hyaluronate-binding domain of human CD44 receptor and structural characterization of the antibody-induced conformational changes in the antigen and their impact on CD44 function.
7 Publications

1. Crystallization of the effector-binding domain of repressor DeoR from *Bacillus subtilis*
   IF$_{2013}$ = 4.558

2. Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis*
   IF$_{2013/2014}$ = 3.986

3. Optimization of the crystallizability of a single-chain antibody fragment*
   IF$_{2013/2014}$ = 0.568

4. Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85

*Part of the results published in this article was included in the master’s thesis of Jana Škerlová (nee Písačková).
7.1 Crystallization of the effector-binding domain of repressor DeoR from *Bacillus subtilis*

**Background**

Carbon catabolite repression in *Bacillus subtilis*, the model Gram-positive bacterium, has been intensively studied on cellular and molecular level. However, little investigation has been done on structural characterization of the proteins involved in the regulation of the expression of metabolic genes. Structural studies of these repressors would provide the needed insight into the carbon catabolite repression in Gram-positive bacteria as well as the general mechanisms of gene regulation. In this work, we focused on the preparation of crystals of the C-terminal effector-binding domain of the deoxyribonucleoside regulator, DeoR, suitable for X-ray diffraction analysis in order to gain structural information on the effector binding and the effector-induced conformational changes.

**Summary**

We prepared a recombinant C-terminal effector-binding domain of DeoR (C-DeoR) from *Bacillus subtilis*. Initial crystallization trials yielded thin needle-shaped crystals which could only be optimized to poorly-diffracting rod-shaped crystal clusters. In order to obtain high quality monocrystals suitable for X-ray diffraction experiments, we employed Thermofluor assay to screen for a buffer composition optimal for protein stability, homogeneity, and therefore, crystallizability. In this assay, C-DeoR exhibited the highest melting temperature in citrate buffer, which was identified as the optimal protein storage buffer. The subsequent crystallization screening experiments gave rise to new crystal forms for both the free and effector-bound C-DeoR, that were further optimized to large three-dimensional monocrystals. We collected three complete diffraction data sets for the free and effector-bound C-DeoR. The processed data can be used to solve the structures by molecular replacement using some of the available three-dimensional structures of homologous bacterial repressor proteins. Some of the key results of the publication are summarized in Figure 8.
Figure 8: Optimization of the C-DeoR crystals using Thermofluor assay
A: Needle-shaped C-DeoR crystals from the initial crystallization screening in the original protein storage buffer are shown. B: Three-dimensional C-DeoR crystals obtained from the crystallization optimization experiments in the optimized protein storage buffer are shown. C: Screening for the optimal protein storage buffer composition by Thermofluor assay is illustrated by a plot of the C-DeoR melting temperatures ($T_m$) in various buffers. The original Tris-based protein storage buffer is shown in black and the optimized protein storage buffer in blue.

My contribution

I prepared recombinant C-DeoR protein, performed protein storage buffer optimization by the Thermofluor assay, conducted both the screening and optimization crystallization experiments, participated in X-ray diffraction data collection and processing, and wrote the manuscript.

Crystallization of the Effector-Binding Domain of Repressor DeoR from Bacillus subtilis

Published as part of the Crystal Growth & Design virtual special issue on the 14th International Conference on the Crystallization of Biological Macromolecules (ICCBM14)

Jana Písačková,†,* Katěrìna Procházková,† Milan Fábry,‡ and Pavlína Řezáèová∗†,*†

†Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, Prague 6, Czech Republic
‡Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videnska 1083, Prague 4, Czech Republic

ABSTRACT: The DeoR repressor of Bacillus subtilis negatively regulates the expression of the enzymes needed for the metabolism of deoxyribonucleosides and deoxyribose. To gain structural information on the regulation of DeoR function by a small molecular ligand, we prepared crystals of the C-terminal domain of DeoR in its free form and in complex with the effector molecule deoxyribose-5-phosphate. To develop the optimal procedure for crystallization, we have employed the thermostability assay as a tool for the optimization of protein crystallizability. The monocryals of three crystal forms were used to collect complete sets of diffraction data at a synchrotron radiation source and will be used to determine the DeoR crystal structure.

INTRODUCTION

The Gram-positive soil bacterium Bacillus subtilis can utilize deoxyribonucleosides and deoxyribose as a source of carbon and energy. After the deoxyribonucleoside uptake into the cell, the catabolic pathway begins with the phosphorolytic cleavage by the nucleoside phosphorylase, yielding the respective base and deoxyribose-1-phosphate, which is converted to deoxyribose-5-phosphate by phosphodeoxyribonuclease to be cleaved to the acetaldehyde and glyceraldehyde-3-phosphate by deoxyribululolase. The three genes encoding the enzymes required for the metabolism of deoxyribonucleosides and deoxyribose, deoxyriboululolase (dra), nucleoside uptake protein (mupC), and pyrimidine nucleoside phosphorylase (pdp) are coded by the dra-mupC-pdp operon. Their expression is negatively regulated at the level of transcription initiation by the DeoR repressor protein. The deoR gene, which codes for the 313 amino acid long repressor protein, is located immediately upstream of the dra-mupC-pdp operon.1 The DeoR-specific operator consists of a palindromic sequence, which is located between the positions –60 and –43 relative to the transcription start site. Also, a direct repeat sequence that is identical to the 3′ half of the palindrome, located downstream to the palindrome, might function as a part of the DeoR binding site.2 Both deoxyribose-1-phosphate and deoxyribose-5-phosphate induce the expression of the operon, but deoxyribose-5-phosphate seems to be the preferred effector.3

The DeoR regulator from Bacillus subtilis shows a significant sequence similarity to regulator proteins from the SorC family which occur in various bacteria species. The N-terminal DNA-binding domain is homologous to SorC-like proteins involved in the transcriptional regulation of sugar alcohol catabolism genes. The C-terminal effector-binding domain is homologous to SorC-like proteins involved in the transcriptional regulation of glucose metabolism genes and is probably binding with glyceraldehyde-3-phosphate as the effector molecule.4 Regulation of the deo operon by the DeoR regulator in Escherichia coli has been studied in great detail.5 Although both the DeoR regulators from B. subtilis and E. coli contain a helix-turn-helix motif that is typical for many DNA-binding proteins6,7 and they both form an octamer in the absence of DNA, there is very little sequence similarity between them, and there is no similarity between their DNA operator sites.5,7

To gain structural information on the effector binding by DeoR, we have prepared a recombinant effector-binding domain of the DeoR repressor protein from B. subtilis (referred to as C-DeoR) for X-ray crystallography analysis. We have developed the optimal protocols for the crystallization of C-DeoR free form and for C-DeoR complex with the effector molecule deoxyribose-5-phosphate. We have employed the thermostability assay as a tool for the optimization of protein crystallizability5,9 and obtained monocryals of three crystal forms which are suitable for diffraction data collection.

Received: October 22, 2012
Revised: December 12, 2012
Published: December 19, 2012
Crystal Growth & Design

EXPERIMENTAL SECTION

Protein Preparation. The coding sequence for the C-terminal effector-binding domain of DeoR, referred to here as C-DeoR (residues 56–313 of the NCBI reference sequence NP_391822.1), was cloned into the pET15b/T-TOPO vector (Invitrogen, USA). The vector coded for an N-terminal His6 tag, V5 epitope, and the tobacco etch virus (TEV) protease recognition site.

The C-DeoR protein was expressed and purified analogously to the protocol developed for other bacterial effectors. The protein was overexpressed in E. coli BL21 (DE3) at 310 K on an LB broth (Sigma Aldrich, USA) supplemented with 0.8% (v/v) glycerol and 20 μg/mL ampicillin. At an OD600max value of about 1, the protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside. The cultivation proceeded at 293 K for 5 h and the cells were harvested by centrifugation.

The bacterial cells were lysed by sonication in 5 vol of lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% (v/v) glycerol, and 1 mM phenylmethylsulfonyl fluoride). Clarified lysate was purified using nickel-chelation chromatography on a Hi-select Ni2+ Affinity Gel column (Sigma Aldrich, USA) equilibrated in lysis buffer. To elute His6-tagged protein, 250 mM imidazole in lysis buffer was used. The His6 tag was cleaved off by a recombinant Hi6x-cleaved TEV protease during an overnight dialysis against the lysis buffer with 286 K. The TEV protease and the cleaved His6 tag were removed using a second nickel-chelation chromatography step. The protein was dialyzed against the protein storage buffer (either 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.02% (v/v) β-mercaptoethanol or 20 mM trisodium citrate pH 7.0, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol), concentrated on Amicon Ultra concentrators (Millipore, USA), aliquoted, and stored at 203 K.

The protein was purified with a yield of 8 mg of protein from 1 L of bacterial culture. The protein purity was assessed by a silver-stained SDS-PAGE to be greater than 95% (see Figure 1, lane 2).

![Figure 1. SDS-PAGE analysis of the C-DeoR protein preparation and C-DeoR crystals. 15% gel visualized by silver staining, lane 1 - protein molecular weight marker (BioRad, USA), lane 2 - 1.3 μg of purified C-DeoR protein, lane 3 - dissolved C-DeoR co-crystal with deoxyribose-5-phosphate, lane 4 - dissolved C-DeoR crystal.](image)

In an attempt to optimize the crystal quality, the protein was further purified using ion-exchange chromatography on the MonoQ column (Amersham Biosciences, USA) in 10 mM Tris-HCl pH 7.5 containing 0.02% (v/v) β-mercaptoethanol with a segmented gradient of sodium chloride using the Äkta Basic FPLC system (Amersham Biosciences, USA). The protein eluted as a single peak at 320 mM NaCl.

Crystallization Screening. The C-DeoR protein in the concentration of 14.4–22 mg/mL in a protein storage buffer was used for the initial crystallization screening. In some experiments the protein solution was supplemented with 50 mM deoxyribose-5-phosphate. The screening was performed by the vapor diffusion method with the help of the Gryphon crystallization workstation (Art Robbins Instruments, USA) in 96-well sitting drop Inneltplates (Art Robbins Instruments, USA) using the JCSG+ Suite (QIAGEN, Netherlands). A total volume of 450 μL of the protein and precipitant solution mixture in the ratio 1:2 or 2:1 was equilibrated over a 30 μL reservoir solution. Crystallization plates were stored at 291 K and checked regularly.

Crystallization Optimization. The experiments were set up manually in the hanging drop format in VDX48 plates (Hampton Research, USA) by mixing the protein and precipitant solutions in the ratio 1:2, 2:1 (total volume of 1.5 or 3 μL, or 1:1 (total volume of 2 μL). The reservoir volume was 150 or 200 μL and the temperature for all experiments was 291 K.

During the optimization of the initial crystallization conditions, the concentration of the crystallization solution as well as the protein concentration were varied, the effect of various additives (agarose, n-acetyl-D-galactosamine, and n-dodecyl-D-maltoside, glycerol, ethylene glycol, and 2-methyl-2,4-pentanediol) was tested. In some experiments, the reservoir was overlaid by 100 μL of oil barrier (Alfa Oil, Hampton Research, USA) to slow down the evaporation. To lower the number of crystallization nuclei and precipitation observed immediately after drop mixing, a crystallization mixture containing 6 μL of C-DeoR in the concentration of 22 mg/mL supplemented with 50 mM deoxyribose-5-phosphate, and 3 μL of condition number 77 of the JCSG+ Screen was incubated at 291 K. The mixture was centrifuged (for 20 min at 18500g) at time intervals of 30, 90, and 180 min from mixing, and the crystallization drops were set up after centrifugation. This approach did not help to lower the number of crystallization nuclei or the precipitation.

Counter-diffusion crystallization experiments were set up in single capillaries using the three-layer configuration. Glass capillaries of 75 mm length and inner diameters of 0.2 mm (Hirschmann, Germany) were placed into the Granada Crystallization Boxes (TriaSciTech, Spain) for storage. Four microLiter of the protein solution at a concentration of 16 mg/mL were separated from the 40 μL of precipitant solution (200 mM trisodium citrate pH 5.5, 25% (v/v) PEG 3350) by 4 μL of 0.5% (v/v) low melting agarose (BioRad, USA).

The composition of protein crystals was analyzed by a Lammelli SDS-PAGE analysis of the crystals extensively washed by mother liquor and dissolved in water.

Thermofluor. Thermofluor assay was performed using real-time PCR LightCycler 480 II (Roche, Switzerland). The stabilizing effect of 16 buffers at a concentration of 100 mM was tested in the presence or absence of 200 mM NaCl following the procedure described in Ericsson et al., 2006. Phosphate buffers were excluded from the screening, because of possible interference with the ligand binding.

The protein at a concentration of 0.08 mg/mL was tested in the presence of 8X concentrated Sypro Orange (Invitrogen, USA) in the total reaction volume of 25 μL in the LightCycler 480 Multwell Plate 96 (Roche, Switzerland). The plate was sealed with the LightCycler 480 Sealing Foil (Roche, Switzerland), and a temperature gradient from 293 to 368 K with a rate of 0.83 K/min was applied. The fluorescence intensity was recorded at 580 nm, and the excitation wavelength was 465 nm. The melting temperature values, corresponding to the inflection points of the melting curves, were determined as the minima of the negative first derivative using the Roche LightCycler 480 SW 1.5 software.

Diffraction Data Collection and Processing. For the diffraction experiments, crystals were mounted into nylon loops (Hampton Research, USA), quick-soaked (for 3–5 s) in cryoprotectant solution containing mother liquor supplemented with 10% or 20% (v/v) glycerol, flash-cooled, and then stored in liquid nitrogen.

Three complete data sets were collected at 100 K with a 0.9184 Å wavelength and a MAR Mosaic (225 mm) detector at beamline MAX14.2 of the BESSY, Berlin, Germany. The crystal parameters and data collection statistics are listed in Table 1. All diffraction data sets were processed using the HKL 3000 package.

dx.doi.org/10.1021/ijog301511v Crys. Growth Des. 2013, 13, 844–848
Table 1. Data Collection Statistics for the Crystals of C-DeoR

<table>
<thead>
<tr>
<th>crystal form</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>deoxyribose-5-phosphate</td>
<td>0.91841</td>
<td>0.91841</td>
<td>0.91841</td>
</tr>
<tr>
<td>wavelength (Å)</td>
<td>50–165 (1.71–1.65)</td>
<td>50–160 (1.66–1.60)</td>
<td>50–130 (1.97–1.90)</td>
</tr>
<tr>
<td>space group</td>
<td>P222₁</td>
<td>C222₁</td>
<td>P3₁ or P3₂</td>
</tr>
<tr>
<td>unit-cell parameters (Å, °)</td>
<td>40.4, 78.4, 91.3; 90.0, 90.0, 90.0</td>
<td>78.9, 78.9, 91.4; 90.0, 90.0, 90.0</td>
<td>63.3, 63.3, 150.0, 90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>total no. of measured intensities</td>
<td>210465</td>
<td>194099</td>
<td>251655</td>
</tr>
<tr>
<td>number of unique reflections</td>
<td>3263 (2799)</td>
<td>34609 (2284)</td>
<td>52424 (5253)</td>
</tr>
<tr>
<td>multiplicity</td>
<td>6.6 (5.0)</td>
<td>5.6 (3.3)</td>
<td>4.8 (4.5)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>59.9 (2.0)</td>
<td>42.8 (2.8)</td>
<td>44.7 (2.1)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>35.6 (33.9)</td>
<td>40.0 (25.5)</td>
<td>4.4 (75.1)</td>
</tr>
<tr>
<td>overall R factor from Wilson plot (Å³)</td>
<td>28.4</td>
<td>18.6</td>
<td>33.6</td>
</tr>
<tr>
<td>Matthews coefficient (Å³/Da)</td>
<td>2.56</td>
<td>2.42</td>
<td>3.02</td>
</tr>
<tr>
<td>no. of molecules in asymmetric unit</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>solvent content (%)</td>
<td>51</td>
<td>49</td>
<td>59</td>
</tr>
</tbody>
</table>

The numbers in parentheses refer to the highest resolution shell. The criterion for the observed reflection I/σ(I) > 0. The Rmerge = 100 \sum_i \sum_k |I(ikl)| \sum_k |l(ikl)| where I(ikl) is the individual intensity of the i-th observation of reflection khl and \langle l(ikl) \rangle is the average intensity of reflection khl with summation over all data.

RESULTS AND DISCUSSION

The C-terminal effector-binding domain of DeoR from *B. subtilis* (residues 56–313, referred to as C-DeoR) was expressed in *E. coli* and purified to homogeneity (Figure 1, lane 2). The recombined C-DeoR contained 258 amino acid residues of its native sequence and an N-terminal sequence GIDPFPT, which remained upon cleavage by the TEV protease.

The initial crystallization screening of C-DeoR in the presence of 50 mM deoxyribose-5-phosphate was performed with a protein stored in a buffer used in protein purification and with a JCSG+ Suite (Qagen, Netherlands). Needle-shaped crystals (typical crystals shown in Figure 2A) were obtained after several days in 8 out of 96 conditions (see Figure 3A). Condition number 2 from the JCSG+ Screen (Figure 2A) was chosen for further optimization.

In the optimization experiments, a wide range of approaches was used to lower the number of crystals and increase their size.

Figure 2. C-DeoR crystals obtained from the initial screening and optimization in the protein storage buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, and 0.02% (v/v) β-mercaptoethanol: panel A: crystals from the initial screening (14.6 mg/mL C-DeoR, 50 mM deoxyribose-5-phosphate, 450 nL droplet, protein to precipitant ratio 1:2, condition number 2 from the JCSG+ Screen (61 M trisodium citrate pH 5.5, 20% (v/v) PEG 3000); panel B: optimized crystals (14.6 mg/mL C-DeoR, 50 mM deoxyribose-5-phosphate, 2 μL droplet, protein to precipitant ratio 1:1, 20% (v/v) ethylene glycol, 80% (v/v) condition number 2 from the JCSG+ Screen: 0.08 M trisodium citrate pH 5.5, 16% (v/v) PEG 3000). The crystals grew to their final size in 1 day.

Figure 3. The crystallization screening and protein storage buffer optimization by the thermofluor assay: panel A: the crystallization screening plate scheme: the conditions of JCSG+ Suite (Qagen, Netherlands) yielding crystals or microcrystals are highlighted in the colors for the original and the optimized protein storage buffers in the presence (+) or absence (−) of deoxyribose-5-phosphate (βRSP); panel B: the melting temperatures (Tm) obtained by thermofluor assay are shown for 16 screened buffers and water (black columns), the original protein storage buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, and 0.02% (v/v) β-mercaptoethanol (black column) and the most stabilizing buffer containing 100 mM trisodium citrate pH 7.0 (blue). Tm values in Kelvins are indicated for the selected buffers.

and volume: (1) protein concentration was varied (8–18.5 mg/mL); (2) precipitant concentration was varied (12–20% (v/v) PEG 3000); (3) various additives known to affect protein solubility and nucleation were added (0.10-0.101% (w/v)
agarose; 0.5% (w/v) of n-octyl-β-D-glucoside or n-dodecyl-β-D-maltoside; 10–40% (v/v) of glycerol or ethylene glycol or 2-methyl-2,4-pentanediol); (4) the capillary crystallization method was performed to screen a wide range of precipitant concentrations; and (5) the purity of the protein was improved by purification on ion-exchange chromatography. This extensive effort yielded plate-like crystals with maximal dimensions of 300 × 80 × 20 μm (Figure 2B) and a poor diffraction quality.

As optimization of the crystallization conditions was not successful we decided to change the protein storage buffer to achieve maximal protein solubility and stability. The optimization of the protein storage buffer composition was performed using thermolysin-based assay following the procedure described by Ericsson et al., 2006.6 The effect of 16 buffers on protein thermal stability was tested in the presence or absence of 200 mM NaCl. The results of the analysis (Figure 3B) revealed the specific stabilizing effect of trisodium citrate pH 4.7 and pH 5.5; the optimal pH value for protein stability was around pH 7.5. The buffer combining a higher pH with trisodium citrate (100 mM trisodium citrate pH 7.0) showed the highest stabilizing effect with the melting temperature (Tm) being about 6 K higher than the original protein storage buffer. Phosphate buffers were not considered in this analysis, because of the possible interference with the ligand binding. The binding of the phosphate ion by bacterial repressors was shown before.14,15 The addition of 200 mM NaCl did not increase protein stability in most buffers, and the difference between the Tm values for 100 mM trisodium citrate pH 7.0 in the presence and absence of 200 mM NaCl was not significant (data not shown). In the optimized buffer composition, 150 mM NaCl was included, because the precipitation of protein at high concentration was observed in the buffer without any salt.

On the basis of the results of thermolysin assay, the composition of the protein storage buffer was thus changed to 20 mM trisodium citrate pH 7.0, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol.

C-DeoR protein was dialyzed into the protein storage buffer with an optimized composition and the crystallization screening procedure using the JCSG+ Suite was repeated. C-DeoR supplemented with 50 mM deoxyribose-5-phosphate yielded crystals in 8 out of 96 conditions which were different from those obtained in the screening in the original protein storage buffer (Figure 3A). Most of the crystals were needle-shaped or plate-like, and often clustered, but condition number 77 (0.1 M HEPES pH 7.5, 12% (w/v) PEG 3350, 5 mM CoCl₂, 5 mM MnCl₂, 5 mM MgCl₂, and 5 mM NaCl) yielded three-dimensional crystals which grew to their final size of 70 × 60 × 40 μm in 5 days. This condition was therefore selected for further optimization.

An analogous crystallization screening experiment was conducted with C-DeoR in the optimized storage buffer in the absence of deoxyribose-5-phosphate. The initial crystals were obtained in 31 out of 96 conditions (see Figure 3A) after several days. Most of the crystals were needle-shaped or plate-like, and often clustered, but condition number 15 (0.1 M Bis-Tris pH 8.5, and 20% (w/v) PEG 6000) yielded three-dimensional crystals with a size of 130 × 120 × 120 μm in one day and was therefore selected for further optimization. Three-dimensional crystals were also obtained in condition number 28 (0.1 M HEPES pH 6.5, and 10% (w/v) PEG 6000) after 7 days, but these were not optimized, because the optimization of the crystals from condition number 15 was already successful.

The optimization of the protein and precipitant concentration led to large three-dimensional crystals with a good diffraction quality allowing the collection of three complete data sets from three crystal forms (Figure 4). The presence of

Figure 4. The optimized C-DeoR crystals and the corresponding X-ray diffraction images: panel A: crystal form I, panel B: crystal form II, panel C: crystal form III. For crystallization conditions see text; for diffraction data statistics see Table 1.

The C-DeoR protein in the crystals was confirmed by an SDS-PAGE analysis of the crystals washed in a cryoprotectant solution and dissolved in water as shown in Figure 1. The crystals belonging to the P2221 space group (crystal form I, Figure 4A, Table 1) were obtained for C-DeoR at 22 mg/mL in the presence of 50 mM deoxyribose-5-phosphate in conditions containing 0.08 M HEPES pH 7.5, 9.6% (w/v) PEG 3350, 4 mM CoCl₂, 4 mM MnCl₂, 4 mM MgCl₂, and 4 mM NaCl. A complete data set at a resolution of 1.65 Å was collected from a crystal soaked in mother liquor supplemented with 20% (w/v) glycerol (see Table 1 for data statistics). The crystals belonging to the C2221 space group (crystal form II, Figure 4B, Table 1) were obtained for C-DeoR at 16.5 mg/mL in the presence of 50 mM deoxyribose-5-phosphate in conditions containing 0.08 M HEPES pH 7.5, 9.6% (w/v) PEG 3350, 4 mM CoCl₂, 4 mM MnCl₂, 4 mM MgCl₂, and 4 mM NaCl. A complete data set at a resolution of 1.60 Å was collected from a crystal soaked in mother liquor supplemented with 20% (w/v) glycerol (see Table 1 for data statistics).
The crystals exhibiting the symmetry of space group P3 with systematic absences indicating the presence of a 3, or 3_2 screw axis (crystal form III, Figure 4C, Table 1) were obtained for C-DeoR at 12.5 mg/mL in the absence of deoxyribosyl-5-phosphate. The crystallization conditions contained 0.1 M Bicine pH 8.5, and 20% (w/v) PEG 6000. A complete data set at a resolution of 1.90 Å was collected from a crystal soaked in mother liquor supplemented with 10% (w/v) glycerol (see Table 1 for data statistics).

The calculated Matthews coefficient values (Table 1) suggested the presence of one molecule of C-DeoR in the asymmetric unit for the ligand-bound C-DeoR and two molecules for the ligand-free form. This correlates with the results of the oligomerization analysis done by size-exclusion chromatography which showed that a dimer is the biological unit of C-DeoR (data not shown).

The diffraction data will be used to determine the structure of C-DeoR by the molecular replacement using the available structures of homologous proteins: the putative sugar-binding transcriptional regulator from *Arthrobacter aurantiacus* TC1 (PDB code 3NZ6; 52% sequence similarity, 30% sequence identity for 259 residues), probable sor operon regulator from *Sphingobacter flexneri* (PDB code 3FEB; 49% sequence similarity, 31% sequence identity for 265 residues) and sorbitol operon regulator SorC from *Klebsiella pneumoniae* (PDB code 2W48, 50% sequence similarity, 30% sequence identity for 246 residues).

![AUTHOR INFORMATION](image1)

**Corresponding Author**
*Tel.: +420210831144; e-mail: rezacova@uochb.cas.cz.*

**Notes**
The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**
This work was supported by the Ministry of Education of the Czech Republic (research projects ME08016 and LK11205), and research projects AV0Z40550506 and AV0Z40520514 awarded by the Academy of Sciences of the Czech Republic and research project GA203/09/0820 awarded by Grant Agency of the Czech Republic. Diffraction data have been collected on MX14.2 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron storage ring (Berlin-Adlershof, Germany). Authors wish to thank Devon Maloy for critical proofreading of the manuscript.

**ABBREVIATIONS**
DeoR, deoxyribonucleoside regulator; C-DeoR, C-terminal effector-binding domain of DeoR; dra, deoxyribonucleoside degrading gene; mmpC, nucleoside uptake protein coding gene; pdp, pyrimidine nucleoside phosphorylase coding gene; dso, E. coli operon coding for enzymes involved in deoxyribonucleo- sides metabolism; SorC, sorbitol operon regulator; d8SP, deoxyribose-5-phosphate; T_m, melting temperature; sop, operon encoding enzymes for the metabolism of d-glucitol, sucrose and l-sorbose.

**REFERENCES**


7.2 Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis*

**Background**

In this work, which followed the results of the previous publication “Crystallization of the effector-binding domain of repressor DeoR from *Bacillus subtilis*”, we aimed at a detailed structural characterization of the C-terminal effector-binding domain of deoxyribonucleoside regulator, DeoR, a repressor protein taking part in the carbon catabolite repression in *Bacillus subtilis*. In particular, we were interested in the structural aspects of effector binding by DeoR. From the known three-dimensional structures of other metabolic regulators from *Bacillus subtilis*, we can say that although the effector-binding domains of the individual repressors represent common conserved folds, the molecular mechanisms of the regulatory function of each repressor are quite unique. Therefore, a detailed study of the DeoR molecular mechanics would contribute to complex understanding of prokaryotic transcriptional regulation.

**Summary**

We determined crystal structures of the effector-binding domain of the deoxyribonucleoside regulator (C-DeoR) from *Bacillus subtilis* both in the free form and in the complex with the effector deoxyribose-5-phosphate. The effector-bound C-DeoR structure revealed a unique mode of effector binding through a reversible covalent Schiff-base double bond to a lysine residue in the effector-binding site. This is the first case of a transcriptional factor that binds the effector molecule covalently. The physiological nature of this covalent effector binding mode was confirmed by mutational analysis combined with differential scanning fluorimetry (Thermofluor assay) and by mass spectrometry analysis of the formation of the Schiff-base complex in solution. Detailed structural analysis of the free and effector-bound C-DeoR structures, supported by the size-exclusion chromatography data, allowed us to propose a model for the DeoR function as a molecular switch. Some of the key results of the publication are summarized in Figure 9.
Figure 9: C-DeoR crystal structure and a model of DeoR function

A: Overall structure cartoon diagram with the topology description is shown for C-DeoR in complex with the effector deoxyribose-5-phosphate (shown in sticks). Double-Rossman-fold core (green) and peripheral subdomains (P1, P2, P3) are highlighted in colors. 

B: A detail of the effector molecule covalently bound to the effector-binding-site lysine residue is shown with the difference electron density map for the ligand contoured at 2.5 $\sigma$. 

C: Schematic representation of the tetrameric assembly of the full-length DeoR on the DNA operator site is shown in the high-affinity free state (left) and in the low-affinity effector-bound induced state (right).

My contribution

I participated in determination of crystal structure of the effector-bound C-DeoR, I determined the structure of free C-DeoR, refined both structure models and performed complete structure analysis. I purified the mutant and full-length C-DeoR proteins, performed the differential scanning fluorimetry (Thermofluor) assay and size-exclusion chromatography, prepared C-DeoR – effector adduct samples for mass spectrometry, and wrote the manuscript.

Structure of the effector-binding domain of deoxyribonucleoside regulator DeoR from Bacillus subtilis

Jana Škerlová1,2, Milan Fáby2, Martin Hubálek1, Zbyszek Otwinowski3 and Pavlína Řezáčová1,2

1 Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic
2 Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
3 UT Southwestern Medical Center, Dallas, TX, USA

Keywords
dimeric interface; effector binding; Schiff base; transcription repressor; X-ray crystallography

Correspondence
P. Řezáčová, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, Prague 6, Czech Republic
Fax: +420 220 163 144
Tel: +420 220 183 144
E-mail: rezacova@uochb.cas.cz

(Received 27 March 2014, revised 15 May 2014, accepted 16 May 2014)
doi:10.1111/febs.12856

Deoxyribonucleoside regulator (DeoR) from Bacillus subtilis negatively regulates expression of enzymes involved in the catabolism of deoxyribonucleosides and deoxyribose. The DeoR protein is homologous to the sorbitol operon regulator family of metabolic regulators and comprises an N-terminal DNA-binding domain and a C-terminal effector-binding domain. We have determined the crystal structure of the effector-binding domain of DeoR (C-DeoR) in free form and in covalent complex with its effector deoxyribose-5-phosphate (dR5P). This is the first case of a covalently attached effector molecule captured in the structure of a bacterial transcriptional regulator. The dR5P molecule is attached through a Schiff base linkage to residue Lys141. The crucial role of Lys141 in effector binding was confirmed by mutational analysis and mass spectrometry of Schiff base adducts formed in solution. Structural analyses of the free and effector-bound C-DeoR structures provided a structural explanation for the mechanism of DeoR function as a molecular switch.

Databases
Atomic coordinates and structure factors for crystal structures of free C-DeoR and the covalent Schiff base complex of C-DeoR with dR5P have been deposited in the Protein Data Bank with accession codes 4QQQ and 4QQQ, respectively.

Structured digital abstract
- C-DeoR and C-DeoR bin by X-ray crystallography (View interaction)
- DeoR and DeoR bind by molecular sieving (Q2)

Introduction

The Gram-positive soil bacterium Bacillus subtilis can utilize deoxyribonucleosides and deoxyribose as sources of carbon and energy. The deoxyribonucleoside catabolic pathway involves three steps. First, the deoxyribonucleosides enter the cell, a process mediated by nucleoside uptake protein C. In the second step, the deoxyribonucleosides are cleaved by pyrimidine nucleoside phosphorylase (Pdp), yielding

Abbreviations
AU, asymmetric unit; C-DeoR, C-terminal domain of deoxyribonucleoside regulator; CcgR, central glycolytic gene regulator; DeoR, deoxyribonucleoside regulator; dR1P, deoxyribose-1-phosphate; dR5P, deoxyribose-5-phosphate; Dna, deoxyribodiolase; DSF, differential scanning fluorimetry; GalR, putative α-aminobutyrate aminotransferase; Pdp, pyrimidine nucleoside phosphorylase; SorC, sorbitol operon regulator; TLS, translation, liberation, screw.
deoxyribose-1-phosphate (dR1P), which is then converted to deoxyribose-5-phosphate (dR5P) by phosphodeoxyribomutase. In the third key step, dR5P is cleaved by deoxyriboladolase (Dra) to acetaldehyde and glyceraldehyde-3-phosphate, which can enter further metabolic pathways. The genes that encode the three key enzymes of deoxyribonucleoside catabolism are grouped in one operon: *dru-nucP-c-pdp*. Expression of this operon is repressed at the level of transcription initiation by deoxyribonucleoside regulator (DeoR) protein, which is encoded by the *deoR* gene located immediately upstream of the *dru-nucP-c-pdp* operon [1].

The DeoR transcriptional regulator is homologous to the sorbitol operon regulator (SorC) family of regulatory proteins present in a wide range of bacterial species. These proteins typically contain a small N-terminal DNA-binding domain, which is responsible for the recognition and binding of the operator DNA, and a large C-terminal effector-binding domain, which is responsible for oligomerization of the regulator and transmission of the effector signal to the DNA-binding domain. The classical helix–turn–helix DNA-binding domain of DeoR (residues 1–55) shows significant sequence similarity to that of SorC-like proteins that regulate the catabolism of sugar alcohols [2,3]. DeoR recognizes a palindromic operator sequence located between positions 60 and 43 relative to the transcription start site, possibly together with an additional direct repeat sequence identical to the 3′ half of the palindromic, which is located downstream of the palindromic: 5′-ATTGAACAAAAATTCATACACAAATCTACTCAGTGGCGAA-3′ (the palindromic sequence and the direct repeat are underlined) [2].

The effector-binding domain of DeoR (residues 56–313) is homologous to that of SorC-like proteins that regulate glucose metabolism [5]. Both dR1P and dR5P have been shown to induce expression of the *dru-nucP-c-pdp* operon, but dR5P is considered the preferred effector [2]. To date, no crystal structures of DeoR from *B. subtilis* or its individual domains are available.

To gain structural information about effector recognition by DeoR, we prepared a recombinant effector-binding domain of DeoR from *B. subtilis* (referred to as C-DeoR) and determined crystal structures of C-DeoR in its free form and in complex with the effector dR5P. This structural information revealed that the effector molecule forms a Schiff base linkage with a lysine side chain in the C-DeoR effector-binding site. The physiological relevance of this interaction was supported by mutational and mass spectrometry analyses.

Results and discussion
Overall structure description and quality
The structure of C-DeoR (residues 56–313) in complex with the effector dR5P was solved by molecular replacement using phases derived from a homologous putative sugar-binding regulator (PDB code 3NZE). The structure was refined to a resolution of 1.6 Å (Table 1). The centred orthorhombic crystal contained 49% of solvent and one protein molecule per asymmetric unit (AU), from which residues 58–311 could be modelled into the electron density map. Four terminal residues (56, 57, 312, 313) were omitted from the final crystallographic model. Non-protein electron density in some regions of the structure was modelled as 264 water molecules and six cations (Co2+, Cd2+, Mg2+ and three Ni2+). The cations originated from crystallization solutions (the crystallization droplets contained 1.3 mM CoCl2, 1.3 mM CdCl2, 1.3 mM MgCl2 and 1.3 mM NiCl2) and are not known to be important for protein function. Three Ni2+ ions bind to three histidine residues (His103, His288 and His304). The remaining three ions bind at interfaces with symmetrically related molecules distant from the effector-binding site and their binding most probably represents a crystallization artefact: Co2+ is coordinated by Asp226 and by residues His78 and Tyr97 from the symmetrically related molecule, Mg2+ is coordinated by Glu93 and symmetrically related His92, and Cd2+ is coordinated by Glu189 and His191 and symmetrically related His146.

The structure of C-DeoR in free form was solved by molecular replacement using the previously determined structure of the C-DeoR-effector complex and refined to 1.8 Å resolution (Table 1). The primitive trigonal crystal contained 58% of solvent and two protein molecules per AU. In both molecules, only residues 59–312 could be modelled into the electron density map, and four terminal residues (56, 57, 58, 313) were omitted from the final crystallographic model. The two molecules in the AU represent the bioologically active dimer. The RMSD for the superposition of the 254 Ca atoms of the two protein chains in the AU is 0.090 Å, a typical value for different crystal structures of identical proteins [4]. Non-protein electron density within the hydrogen bonding radius was modelled as 225 water molecules and two bicine molecules that originated from crystallization solutions.

The C-DeoR structure comprises a central NAD-binding double Rossmann fold core [5] composed of six parallel β-sheets (β2, β3, β4, β5, β8, β9) flanked by a total of five helices (α3, α4, α6, α10, α11) on both
Table 1. Crystal data and diffraction data collection and refinement statistics. Values in parentheses refer to the highest resolution shell. ADP, atomic displacement parameter. 

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>Effector-bound</th>
<th>Free</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.918</td>
<td>0.918</td>
</tr>
<tr>
<td>Space group</td>
<td>C2221</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit-cell parameters (Å, °)</td>
<td>76.9, 78.9, 91.4</td>
<td>63.1, 63.1, 150.1</td>
</tr>
<tr>
<td>No. of molecules in AU</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.1-1.80</td>
<td>50.1-1.60</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>34,660 (2,264)</td>
<td>57,687 (3,293)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.6 (3.9)</td>
<td>3.8 (2.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>93.5 (62.3)</td>
<td>93.1 (53.3)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>4.0 (25.5)</td>
<td>4.8 (17.9)</td>
</tr>
<tr>
<td>Average (l/σl)</td>
<td>42.8 (2.8)</td>
<td>72.6 (2.9)</td>
</tr>
<tr>
<td>Wilson B (Å²)</td>
<td>18.6</td>
<td>26.6</td>
</tr>
</tbody>
</table>

Refinement statistics

| Resolution range (Å)       | 19.77-1.60     | 27.31-1.80 |
| No. of reflections in working set | 32,880         | 54,842 |
| R (%)                      | 15.44 (23.7)   | 19.28 (31.9) |
| Rmerge (%)                 | 19.06 (28.6)   | 22.76 (34.4) |
| RMSD bond lengths (Å)      | 0.012          | 0.012 |
| RMSD angles (°)            | 1.488          | 1.439 |
| No. of atoms in AU         | 2398           | 4259 |
| No. of water molecules in AU | 264            | 225 |
| Mean ADP value, prot/solv/gnd (Å) | 23.4/33.5/17.0 | 52.9/45.9/—  |
| Residues in Ramachandran plot (favourable regions) (%) | 98.9 | 98.8 |
| Residues in Ramachandran plot (allowed regions) (%) | 100 | 100 |
| PDB code                   | 4D3P          | 4D3Q  |

*As determined by MolProbity [7].

46
characterized in terms of effector binding and oligomerization [10]. Superposition of the CggR structure with the C-DeoR structures provides RMSD values of 2.6 and 2.7 Å for effector-bound and free C-DeoR, respectively. The two proteins are similar in the area of the double Rossmann fold core but differ in their peripheral subdomains. Nevertheless, the position of the effector-binding site and binding orientation of the ligand in this site are very similar (Fig. 2A).

In addition to transcriptional regulators, we identified as structural homologues of C-DeoR many bacterial enzymes involved in sugar metabolism, such as 6-phosphogluconolactonase, glucose-6-phosphate-1-dehydrogenase and glucosamine-6-phosphate deaminase. This is an example of one common protein fold performing distinct enzymatic and regulatory functions. Interestingly, ribose-5-phosphate isomerase A from Burkholderia thailandensis (PDB code 3UW1) binds the linear form of ribose-5-phosphate in a position very similar to that of dR5P in C-DeoR (Fig. 2B) [11]. The sugar, however, does not form a covalent adduct in this B. thailandensis enzyme because the amino acid residue corresponding to Lys141 of C-DeoR is serine. Superposition of the ribose-5-phosphate isomerase structure with the effector-bound C-DeoR structure provides an RMSD value of 3.4 Å, although the sequence identity is only 12%.

The DALI structure similarity search [9] also identified several enzymes that carry pyridoxal phosphate as a prosthetic group bound via a Schiff base linkage with a lysine residue. These enzymes include 5-amino-levulinate synthase (PDB code 2BWN) [12], methionine gamma-lyase (PDB code 3ACZ), carbon-sulfur lyase involved in aluminium resistance (PDB code 3GWP) and L-allo-threonine aldolase (PDB code 3FM1). Superposition of any of these enzymes with the C-DeoR structures resulted in RMSD values of over 3.5 Å.

**Fig. 2.** Superposition of effector-bound C-DeoR with homologous structures. (A) Superposition of C-DeoR (green) with the central glycolytic gene regulator CggR (light blue, PDB code 2BXP) [11] in stereo view. (B) Detail of the ligand binding sites rotated by 50° along the x-axis relative to the view in (A). (C) Superposition of C-DeoR (green) with ribose-5-phosphate isomerase A from B. thailandensis (gold, PDB code 3UW1) [11] in stereo view. (D) Detail of the ligand binding sites rotated by 50° along the x-axis relative to the view in (C). The ligands are shown as sticks with carbon atoms in black (dR5P in C-DeoR), light blue (dihydroxyacetone phosphate in CggR) or gold (ribose-5-phosphate in ribose-5-phosphate isomerase A). Key amino acid residues in the ligand binding sites are highlighted in sticks. Nitrogen, oxygen and phosphorus atoms are coloured blue, red and orange, respectively.
Structural details of effector binding

The effector-binding site of DeoR is formed by residues Trp116, Thr118, Thr119 and Lys283 from the core subdomain, residue Tyr87 from peripheral subdomain P1, and residues Lys141, Thr208, Val209, Gly210, Asp239 and Cys241 from peripheral subdomain P2. The effector molecule dR5P forms a covalent Schiff base linkage with the amino group of Lys141 and is deeply buried in the effector-binding site. A 225.7 Å² surface area, which represents 65% of the dR5P total surface area, is involved in the binding interface. All effector hydroxyl groups are engaged in direct hydrogen bond interactions with the protein (Fig. 3A). Specifically, the O3 and O4 atoms of dR5P interact with the side chain of Asp239; the oxygen atoms from the phosphate moiety interact with Thr118, Thr119 and Lys283. The interaction of the phosphate moiety with the N-terminal tip of α3 of the double Rossmann fold (Thr118 and Thr119) is typical for dinucleotide binding Rossmann folds [5,13,14]. In addition, a water molecule (W 501) in the effector-binding site mediates hydrogen bond linkages of O5 and one oxygen atom of the phosphate moiety to amino acid residues Thr116, Thr119 and Met120 (Fig. 3A).

The Schiff base linkage of R5P to Lys141 is schematically shown in Fig. 3B. The imine bond of a Schiff base is generally formed by reaction of an aldehyde with an amine. In the case of C-DeoR, the aldehyde of dR5P reacts with the Lys141 side chain amine. In the C-DeoR crystal structure, the distance between Nε of Lys141 and dR5P C1 is 1.29 Å, which clearly indicates the presence of a double bond. Schiff bases occur in enzymes carrying pyridoxal phosphate as a prosthetic group [15] and in rhodopsins, in which the retinal group is covalently bound via a Schiff base to a lysine residue [16]. One of these pyridoxal-phosphate-containing enzymes, the putative γ-aminobutyrate aminotransferase GabR from B. subtilis, has lost its enzymatic activity and has adapted to function as a transcriptional regulator [17]. The covalent attachment of pyridoxal phosphate to this transcriptional regulator of γ-aminobutyric acid metabolism is crucial for binding of the effector molecule γ-aminobutyric acid [17].

Furthermore, formation of Schiff base intermediates between the substrate and the catalytic lysine residue is a critical step in the catalytic mechanism of some aldolases [18–22] and other enzymes, such as porphobilinogen synthase [23] and acetate decarboxylase [24]. Among these enzymes, dR5P aldolase contains dR5P Schiff base adduct with the active site lysine, which represents an intermediate in its catalytic mechanism [25]. The structure of Escherichia coli dR5P aldolase, which captures the dR5P Schiff base complex with Lys201, has been solved at atomic resolution (PDB code 1UC2) [25] and represents the only protein-dR5P adduct in the PDB to date. The Nε-C bond length of 1.29 Å and Nε-C-C angle of 125.3° are almost identical to the values found in our C-DeoR-dR5P structure (Nε-C bond length of 1.29 Å and Nε-C-C angle of 125.2°),

To our knowledge, the crystal structure of the C-DeoR adduct is the first example of a bacterial transcriptional regulator with a covalently bound effector molecule. To exclude the possibility that the covalent

Fig. 3. Binding of dR5P to C-DeoR. (A) The dR5P Schiff base adduct is shown as sticks with carbon atoms in black. The C-DeoR residues in direct hydrogen bonds or a covalent bond (Lys141) with the effector are shown as sticks with green carbon atoms. Other residues within the van der Waals distance (4.2 Å) from the effector are shown as sticks with grey carbon atoms (Thr208, Val209 and Gly210 are not shown for clarity). Direct hydrogen bonds are indicated by black dotted lines; water-mediated hydrogen bonds are indicated by blue dotted lines. (B) Scheme of the formation of the covalent adduct of dR5P on Lys141.
attachment of the effector molecule to Lys141 was a crystallization artefact, we performed mutational analysis and explored Schiff base adduct formation in solution.

**Mutational analysis**

To evaluate the role of Lys141 in dR5P binding, we prepared four C-DeoR variants mutated at position 141 and tested them for effector binding. Specifically, we assayed dR5P binding to K141N, K141Q, K141R and K141A C-DeoR variants using differential scanning fluorimetry (DSF), also known as ThermoFluor [26-28]. Ligand binding restrains the overall protein structure, resulting in higher stability. Thermal stabilization of a protein upon addition of ligand can therefore be used as a qualitative measure of effector binding which can be used as a hint for effector affinity. The thermal unfolding transitions of wild-type and mutant C-DeoR variants were monitored by the fluorescence of SYPRO® Orange dye, and the melting temperature values ($T_m$) were estimated for increasing concentrations of dR5P (Fig. 4). Addition of the dR5P effector molecule significantly increased the $T_m$ of wild-type C-DeoR (increase of ~9 °C), while the addition of dR1P did not stabilize the protein and even caused destabilization at high concentrations. We thus did not confirm dR1P binding to C-DeoR, although this sugar has been shown to induce expression of the dru-rup-pdp operon [2]. Thermal stabilization of the C-DeoR mutants upon addition of dR5P was lower than that of wild-type which suggested that mutation of Lys141 significantly lowers or even disrupts the protein’s affinity toward the effector molecule. We can therefore conclude that Lys141, which provides the free amino group for formation of a Schiff base with the carbonyl group of dR5P, is crucial for effector binding.

**Formation of C-DeoR-dR5P Schiff base adduct in solution**

We followed the formation of a covalent Schiff base adduct between dR5P and C-DeoR in solution at low protein and effector concentrations using mass spectrometry (Fig. 5A). The Schiff base product was identified by MALDI-TOF in C-DeoR incubated with 1.8 mM dR5P, which is close to the upper concentration limit of this metabolite in B. subtilis [29]. Free C-DeoR was detected as a peak corresponding to m/z 29 335, while the reduced Schiff base C-DeoR adduct was detected at m/z 29 533. Time-dependent increase in the abundance of the higher molecular mass peak was observed (Fig. 5A). The difference in m/z values of 198 corresponds to the molecular mass of the dR5P residue upon Schiff base formation and subsequent reduction to 3,4-dihydroxy-5- (phosphonoxy) pentyl (C$_7$H$_{13}$O$_5$P). Incubation of K141N C-DeoR with 1.8 mM dR5P did not result in adduct formation, even after 16 h (Fig. 5B).

This mass spectrometry analysis confirmed that C-DeoR-dR5P Schiff base adducts form in solution and were not merely crystallization artefacts. Furthermore, we confirmed that the presence of Lys141 is crucial for covalent binding of the effector to C-DeoR.

**Effect of ligand binding on the structure and oligomerization of C-DeoR**

To understand on a structural level how the effector molecule modulates DeoR function, we followed changes in C-DeoR structure upon effector binding. Superposition of the 237 Cz atoms of the effector-bound and free C-DeoR structures provides an RMSD value of 1.15 Å. The binding of dR5P induces major structural changes in the vicinity of the effector-binding site, mostly in peripheral subdomains P2 and P3. In contrast, the core double Rossmann fold and peripheral subdomain P1 structures remain almost unaffected (Fig. 6A). Major changes (with RMSD for main chain atoms over 1.8 Å) occur mainly in peripheral subdomain P2, specifically in helices 7 and 8 and residues 141–151, which form a loop connecting β3 and α4 (Fig. 6A). Structural changes in this loop region have been shown to be important for effector binding to the repressor CggR from B. subtilis [10]. While the loop region in CggR (residues 174–184)
undergoes a major structural change and acts as a flap that closes over the effector-binding site, the changes in C-DeoR are more subtle.

Upon attachment of dRSP to Lys141, which is located at the base of the loop, the downstream residues rearrange so that the tip of the loop extends toward the helical region of the P2 subdomain. The structural changes in the P2 subdomain can be qualitatively described as the grasp of an index finger (helical region) and thumb (loop region) around the effector-binding site (Fig. 6B). Helices 27 and 28 move toward the centre of the effector-binding site, and residues 220–223 in the region connecting 27 and 28 come into contact with the tip of the loop region (residues 146–149). This primary structural change induced by effector binding further propagates to peripheral subdomain P3 through residues 173–179, which are adjacent to the 141–151 loop region and to 25 (Fig. 6A).
Helix z5 and the 141–151 loop are involved in the dimeric interface, and we therefore investigated the effects of structural changes induced by effector binding on the protein's oligomeric state.

We analysed the oligomeric state of both C-DeoR and full-length DeoR in solution by analytical size-exclusion chromatography. The apparent molecular weight of C-DeoR as well as that of the full-length DeoR protein were estimated in the presence or absence of the effector molecule dR5P. These experiments revealed that C-DeoR forms dimers in solution which persist in the presence of dR5P. Full-length DeoR forms tetramers in solution and this oligomeric assembly is also observed in the presence of dR5P (Fig. 7). These results were corroborated by dynamic light scattering experiments which showed no change in particle size distribution upon addition of 50 μM dR5P to 750 μM C-DeoR and 120 μM full-length DeoR proteins, respectively (data not shown). Our results suggest that two distinct interfaces involved in oligomerization exist in the DeoR: the first one is formed by the C-terminal domains and the second one by the DNA-binding domains.

In an attempt to identify and characterize the interfaces mediating interaction of two C-DeoR, we performed a detailed analysis of the intermolecular contacts observed in both the free and effector-bound C-DeoR crystal structures. The crystal interfaces and the stability of potential assemblies were evaluated using methods available through the Protein Interfaces, Surfaces and Assemblies (risa) server [30]. In the free C-DeoR structure, we identified a major interface of 1367.3 Å², which results in formation of a stable dimeric assembly of the two molecules present in the AU. The interface area represents approximately 12% of the total solvent-accessible surface area of each monomer, comprising 40 interacting residues from each chain (Fig. 8, Table 2). In the effector-bound C-DeoR structure, the interface area decreased to 841.3 Å², which represents approximately 7% of the total solvent-accessible surface area of each monomer and comprises 25 interacting residues from each chain (Fig. 8, Table 2). Dimeric assemblies of both the free and effector-bound C-DeoR were classified as stable based on the value of the free energy of dissociation of a potential assembly calculated by the rasa server [30]. The dimer of free C-DeoR is tightly packed and gets looser upon effector binding as documented by the distance between the two N-termini changing from 78 to 84 Å (Fig. 8A, C).

The interfacing residues are located in peripheral subdomain P3 in z5 and its connecting loop (comprising residues 170–179) and in peripheral subdomain P2, specifically in the 141–151 loop and in the region connecting α7 and α8 (residues 221–224). Another part of the interface area is located in the region of the central double Rossmann fold core adjacent to the P2 subdomain, specifically β3, α4, β4, α6, and the region connecting α4 and β4. The structural changes upon effector binding result in an overall loss of contact, which is most pronounced in the 141–151 loop and in the central double Rossmann fold core α4 (Table 2).

The dimeric assembly in C-DeoR crystal structures is similar to dimers observed for other effector-binding domains of the SorC/DeoR family, such as CggR from B. subtilis [10], putative sugar-binding transcriptional regulator from A. aurescens TC1 (PDB code 3NZF) and sorbitol operon regulator SorC from Klebsiella pneumoniae (PDB code 2W4B) [31]. The crystal structure of SorC also revealed a tetrameric assembly.
formed by the full-length protein (Fig. 9A) which helped us to deduce the positions of N-terminal DNA-binding domains in DeoR, to understand DeoR oligomerization and to propose a model for modulation of DeoR function (Fig. 9B,C). DeoR was shown to bind DNA operator, which contains a palindromic binding sequence, as a tetramer or possibly cooperatively as two dimers [2,3]. In our model, the tetramer of DeoR with high affinity toward DNA is assembled through two distinct interfaces and can be regarded as a dimer of dimers (Fig. 9B). The first interface is mediated by a domain swap of the DNA-binding domains (similarly to the structure of SorC, Fig. 9A) while the second interface is formed by tightly packed effector-binding domains (interfaces observed in our crystal structures, Fig. 8). Upon effector binding, the interface between the two effector-binding domains loosens and results in a change in relative orientation of the N-terminal DNA-binding domains within the dimer (Fig. 9C). The tetrameric assembly is not disrupted but has a lower affinity toward the DNA and might be more easily displaced from the promoter region by RNA polymerase during transcription initiation.

This model is in agreement with our observation that the DeoR oligomERIC state is not affected by dR5P (Fig. 7). The change in relative distance between the two N-termini in C-DeoR observed for the dimer in free and effector-bound form (Fig. 8A,C) gives us a hint of understanding about how the structural change upon effector binding is transmitted into the DNA-binding domains. The exact structural mechanism, however, has yet to be unravelled. Determining a crystal structure of full-length DeoR or DeoR-DNA operator complex may provide further insight.

Materials and methods

Protein expression and purification

Wild-type and mutant C-DeoR and full-length DeoR proteins were expressed in E. coli BL21(DE3) and purified as previously described [8,32]. Briefly, His-tagged proteins were subjected to nickel-chelation chromatography followed by His-tag cleavage with tobacco etch virus protease, a second nickel-chelation chromatography step, and dialysis against protein storage buffer optimized for crystallization
Table 2. Residues involved in the C-DecR dimer interface. Amino acid residues with buried surface area (BSA) > 10% are listed; residues with main chain RMSD over 1.2 Å for the superposition of free C-DecR with effector-bound C-DecR are highlighted in bold. BSA was determined by the msa server [30].

<table>
<thead>
<tr>
<th>Residue</th>
<th>Location</th>
<th>C-DecR BSA (%)</th>
<th>C-DecR-dr5P BSA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLU 136</td>
<td>123</td>
<td>19.6</td>
<td>0.9</td>
</tr>
<tr>
<td>GLY 143</td>
<td>141-151</td>
<td>20.7</td>
<td>0</td>
</tr>
<tr>
<td>ILE 144</td>
<td>141-151</td>
<td>72.0</td>
<td>59.6</td>
</tr>
<tr>
<td>SER 145</td>
<td>141-151</td>
<td>81.2</td>
<td>20.4</td>
</tr>
<tr>
<td>HIS 146</td>
<td>141-151</td>
<td>79.5</td>
<td>75.1</td>
</tr>
<tr>
<td>SER 147</td>
<td>141-151</td>
<td>45.7</td>
<td>0</td>
</tr>
<tr>
<td>ARG 148</td>
<td>141-151</td>
<td>19.9</td>
<td>52.6</td>
</tr>
<tr>
<td>ASN 150</td>
<td>141-151</td>
<td>66.3</td>
<td>0</td>
</tr>
<tr>
<td>THR 151</td>
<td>141-151</td>
<td>15.0</td>
<td>0</td>
</tr>
<tr>
<td>ALA 154</td>
<td>α4</td>
<td>57.2</td>
<td>0</td>
</tr>
<tr>
<td>GLU 155</td>
<td>α4</td>
<td>45.1</td>
<td>0</td>
</tr>
<tr>
<td>GLN 158</td>
<td>α4</td>
<td>66.6</td>
<td>0</td>
</tr>
<tr>
<td>THR 160</td>
<td>α4-β4</td>
<td>51.0</td>
<td>4.4</td>
</tr>
<tr>
<td>MET 167</td>
<td>β4</td>
<td>72.9</td>
<td>31.1</td>
</tr>
<tr>
<td>PRO 168</td>
<td>β4</td>
<td>76.3</td>
<td>27.0</td>
</tr>
<tr>
<td>ARG 169</td>
<td>β4</td>
<td>45.4</td>
<td>34.3</td>
</tr>
<tr>
<td>TYR 170</td>
<td>β4</td>
<td>46.8</td>
<td>0</td>
</tr>
<tr>
<td>PRO 172</td>
<td>β4-α5</td>
<td>84.0</td>
<td>69.9</td>
</tr>
<tr>
<td>LEU 173</td>
<td>β4-α5</td>
<td>95.7</td>
<td>26.9</td>
</tr>
<tr>
<td>PRO 174</td>
<td>β4-α5</td>
<td>92.1</td>
<td>28.7</td>
</tr>
<tr>
<td>VAL 176</td>
<td>β4-α5</td>
<td>77.2</td>
<td>0</td>
</tr>
<tr>
<td>PHE 177</td>
<td>β4-α5</td>
<td>100.0</td>
<td>5.6</td>
</tr>
<tr>
<td>ASP 178</td>
<td>β4-α5</td>
<td>58.9</td>
<td>39.2</td>
</tr>
<tr>
<td>ASN 179</td>
<td>β4-α5</td>
<td>1.4</td>
<td>22.1</td>
</tr>
<tr>
<td>ASP 181</td>
<td>α5</td>
<td>1.7</td>
<td>18.2</td>
</tr>
<tr>
<td>VAL 182</td>
<td>α5</td>
<td>88.1</td>
<td>88.4</td>
</tr>
<tr>
<td>MET 185</td>
<td>α5</td>
<td>65.1</td>
<td>72.3</td>
</tr>
<tr>
<td>VAL 186</td>
<td>α5</td>
<td>100.0</td>
<td>65.6</td>
</tr>
<tr>
<td>ASP 189</td>
<td>α5-α6</td>
<td>87.8</td>
<td>70.2</td>
</tr>
<tr>
<td>ARG 190</td>
<td>α6</td>
<td>30.1</td>
<td>2.1</td>
</tr>
<tr>
<td>HIS 191</td>
<td>α6</td>
<td>82.0</td>
<td>61.4</td>
</tr>
<tr>
<td>ILE 192</td>
<td>α6</td>
<td>91.2</td>
<td>83.6</td>
</tr>
<tr>
<td>ILE 196</td>
<td>α6</td>
<td>59.2</td>
<td>0</td>
</tr>
<tr>
<td>GLY 222</td>
<td>αβ-αβ</td>
<td>35.7</td>
<td>42.8</td>
</tr>
<tr>
<td>TYR 223</td>
<td>αβ-αβ</td>
<td>76.8</td>
<td>0</td>
</tr>
</tbody>
</table>

Protein crystallization and diffraction data collection

Crystallization of C-DecR in complex with dr5P and collection of the diffraction data have been published elsewhere [8]. Crystals of free C-DecR were obtained at 18 °C using the hanging-drop vapour diffusion technique in VDX48 plates (Hampton Research, Aliso Viejo, CA, USA) with a reservoir solution (200 μL) of 0.08 M HEPES, pH 8.5, 16% (v/v) PEG 6000. Droplets were prepared by mixing 0.5 μL of 22 mg mL⁻¹ C-DecR stored in buffer containing 20 mM tri-sodium citrate, pH 7.0, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol with 1 μL of reservoir solution. Crystals were quick-soaked (45 s) in mother liquor supplemented with 10% (v/v) glycerol and flash-cooled in liquid nitrogen using nylon mounting loops (Hampton Research). A complete data set at 1.80 Å resolution was collected at −173 °C at beamline MX14.2 of the Berlin Electron Storage Ring Society for Synchrotron Radiation (BESSY), Germany [33]. The data set was processed using the HKL-3000 package [34]. The crystal parameters and data collection statistics are listed in Table 1. The low value of the completeness in the highest resolution shell is due to anisotropic diffraction of the crystals.

Structure determination and analysis

The structure of C-DecR was determined by molecular replacement with the program MOLREP [35] using the structure of the effector-binding domain of the putative sugar-binding transcriptional regulator from A. naescentis TC1 (PDB code 1ANL, residues 5-263) as a search model. Model refinement was performed using the program REFMAC 5.7.0032 [36] from the CCP4 suite [37] in combination with manual adjustments in COOT software [38]. In the final stages, translation, libration, screw (TLS) refinement was included [39] using 20 TLS groups for effector-bound C-DecR and 10 TLS groups for each chain in the free C-DecR dimer. The ligand was modelled after complete refinement of the C-DecR and solvent model. MOLPROBE server [7] was used for evaluation of the final model quality. The final refinement statistics are summarized in Table 1. The structures were analysed using the servers PISA [30] and DALI [9] and the programs ISOKAR (supersuperpose) [40], BANNERAGE and CONTACT from the CCP4 suite [37]. All the
Fig. 9. Schematic model of DeoR function. (A) Tetrameric assembly of SarC from K. pneumoniae (PDB code 2WAC) [31] is shown as a cartoon with differently coloured chains. Effector-binding domains (residues 60-214) are also representated as transparent surfaces. (B) Free DeoR tight tetramer bound to the DNA operator sequence; the palindromic is highlighted in red in the operator sequence. (C) Effector-bound less-tight DeoR tetramer loosely bound to the DNA operator sequence. dRSP is shown as yellow triangles. The colour coding in (B) and (C) is the same as in (A).

figures representing structures were created using rVnode. [41]. Atomic coordinates and structure factors for the crystal structures of free C-DeoR and the covalent Schiff base complex of C-DeoR with dRSP were deposited in the PDB with accession codes 4QQQ and 4QQP, respectively.

Differential scanning fluorimetry
Ligand titration was performed in 20 mM tri-sodium citrate, pH 7.5, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol as previously described [32]. A temperature gradient from 20 to 95 °C was applied at a rate of 0.83 °C·min⁻¹ [8]. The concentration of wild-type C-DeoR and C-DeoR mutants (K141A, K141R, K141N and K141Q) was 1 μM; dRSP or dRIP concentrations ranged from 0.1 to 50 mM. Three parallel measurements were performed for each ligand concentration, and the Tm values were averaged.

Mass spectrometry
The formation of the covalent Schiff base adduct was analysed by UltrafleXtreme MALDI-TOF/TOF (Bruker Daltonics, Billerica, MA, USA). Wild-type or K141N C-DeoR (1 μM) was incubated with 1.8 mM dRSP in 20 mM tri-sodium citrate, pH 7.5, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol for 1, 3, 6 and 16 h at 25 °C in the presence of 50 mM NaBH₄CN, which reduced – and thus stabilized – the Schiff base product, a method widely used to study protein glycation [42]. Schiff base formation was terminated by addition of 50 mM hydroxylamine. The samples were desalted prior to mass spectrometry analysis on centrifugal Amicon Ultracel 10K filters (Merck Millipore, Billerica, MA, USA) by washing with deionized water. Protein sample (1 μL) was mixed with 1 μL of a saturated solution of sinapinic acid (Sigma-Aldrich, Saint Louis, MO, USA) in 30% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid and spotted on a MALDI target. The spectra were acquired in linear mode in the 5-35 kDa range. Calibration was performed with a protein standard mixture.

Size-exclusion chromatography
Analytical size-exclusion chromatography of both C-DeoR and full-length DeoR was performed in 20 mM tri-sodium citrate, pH 7.5, 150 mM NaCl, 0.02% (v/v) β-mercaptoethanol using an Äkta Basic FPLC system with a Superdex 200 10/300 GL Tricorn column (Amersham Biosciences, Little Chalfont, UK) at a flow rate of 0.5 mL·min⁻¹. The molecular weight calibration standards (Sigma-Aldrich) contained blue dextran (2000 kDa), bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), cytochrome c (12.4 kDa) and aprotinin (6.5 kDa). In the case of full-length DeoR, standards of higher molecular weights were used (Sigma-Aldrich): blue dextran (2000 kDa), bovine thyroglobulin (669 kDa), horse spleen apoferritin (443 kDa), sweet potato β-amylase (200 kDa), yeast alcohol dehydrogenase (150 kDa), bovine serum albumin (66 kDa) and carbonic anhydrase (29 kDa).

Acknowledgements
The authors wish to thank the following colleagues from the Institute of Organic Chemistry and Biochemistry AS CR in Prague: Jiří Brynda, Petr Pachl and Evzen Bouřa for their help with diffraction data collection, and Jiřka Krejčová for her technical assistance with protein preparation. This work was supported by the Ministry of Education of the Czech Republic (project ME08016) and in part by research projects RVO 68378050 and 61388963 awarded by the Academy of Sciences of the Czech Republic. The use of MX14.2
operated by the Helmholtz-Zentrum Berlin at the BESSEY II electron storage ring (Berlin-Adlershof, Germany) to collect diffraction data is acknowledged. JS is a student supported in part by the Faculty of Science, Charles University in Prague.

**Author contributions**

JS planned experiments, performed experiments, analysed data and wrote the paper; MF performed experiments; MH performed experiments and analysed data; ZO analysed data; PR planned experiments, analysed data and wrote the paper.

**References**


7.3 Optimization of the crystallizability of a single-chain antibody fragment

Background

Single-chain variable antibody fragments (scFvs) are engineered proteins of major therapeutic and diagnostic potential. Understanding the antigen-binding properties of the antibodies as well as employing various protein-engineering approaches including antibody humanization relies on the knowledge of antibody three-dimensional structure. Majority of scFv fragments tend to spontaneously oligomerize, which impairs their homogeneity, and therefore presents a major obstacle in crystallization. This work presents a new approach to scFv crystallizability optimization, demonstrated on a representative scFv fragment of anti-CD3 antibody MEM-57. The approach employs a Thermofluor-based assay to screen for optimal conditions both for stability and for oligomeric homogeneity of the scFv fragment. In the optimized conditions, scFv MEM-57 exhibited significantly improved ability to yield crystals.

Summary

We constructed a single-chain variable fragment (scFv) of anti-CD3 antibody MEM-57. Pre-crystallization biochemical and biophysical analysis of the purified protein using a combination of size-exclusion chromatography and dynamic light scattering revealed the presence of scFv dimer and higher oligomers and general heterogeneity of the sample, unfavorable for crystallization. Thermofluor assay was used to screen for the conditions optimal not only for protein stability, but also for protein sample oligomeric homogeneity, preferentially stabilizing scFv monomer. The beneficial effect of the optimized protein storage buffer was confirmed by the size-exclusion chromatography and dynamic light scattering data and finally by the results of crystallization screening trials. This approach may resolve the issue of unfavourable oligomerization of scFv fragments in general and result in scFv samples with improved crystallizability. The key results of the publication are summarized in Figure 10.
Figure 10: Optimization of crystallizability of scFv MEM-57
A: A schematic of the scFv MEM-57 gene construct and protein product is shown. \( V_H \) and \( V_L \) are the variable domains of the heavy and light IgG chains, respectively; pelB is a leader sequence targeting recombinant protein to the periplasm, c-myc and His\(_5\) are immunodetection and purification tags, respectively. B: Thermofluor-based protein-unfolding curves for the original (black) and optimized (green) storage buffers are shown along with representative multiple-inflection curves (blue). C: A schematic of the crystallization screening plate containing the PEGs Suite (Qiagen) is shown. Conditions yielding crystals are highlighted in green. D: An example of crystals obtained using 100 mM HEPES pH 7.5, 25% (w/v) PEG 3000 is shown.

My contribution
I expressed and purified recombinant scFv MEM-57, performed all the biochemical and biophysical experiments: size-exclusion chromatography, Thermofluor assay, and dynamic light scattering. I performed the crystallization screening experiments, and wrote the manuscript.

Optimization of the crystallizability of a single-chain antibody fragment

Jana Škerlová, Vlastimil Král, Milan Fábry, Juraj Sedláček, Václav Veverka and Pavlína Řezáčová


Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see https://journals.iucr.org/services/authorrights.html

Acta Crystallographica Section F: Structural Biology Communications is a rapid all-electronic journal, which provides a home for short communications on the crystallization and structure of biological macromolecules. Structures determined through structural genomics initiatives or from iterative studies such as those used in the pharmaceutical industry are particularly welcomed. Articles are available online when ready, making publication as fast as possible, and include unlimited free colour illustrations, movies and other enhancements. The editorial process is completely electronic with respect to deposition, submission, refereeing and publication.

Crystallography Journals Online is available from journals.iucr.org

Optimization of the crystallizability of a single-chain antibody fragment

Single-chain variable antibody fragments (scFvs) are molecules with immense therapeutic and diagnostic potential. Knowledge of their three-dimensional structure is important for understanding their antigen-binding mode as well as for protein-engineering approaches such as antibody humanization. A major obstacle to the crystallization of single-chain variable antibody fragments is their relatively poor homogeneity caused by spontaneous oligomerization. A new approach to optimization of the crystallizability of single-chain variable antibody fragments is demonstrated using a representative single-chain variable fragment derived from the anti-CD3 antibody MEM-57. A Thermofluor-based assay was utilized to screen for optimal conditions for antibody-fragment stability and homogeneity. Such an optimization of the protein storage buffer led to a significantly improved ability of the scFv MEM-57 to yield crystals.

1. Introduction

Knowledge of protein three-dimensional structure is important for understanding protein function. The three-dimensional structure can, amongst other things, also provide a basis for protein-engineering approaches especially in the case of antibody single-chain variable fragments (scFvs). X-ray crystallography is the principal method used in protein structure determination; close to 90% of all of the protein structures deposited in the Protein Data Bank to date have been determined using this method. The crystallization of the protein represents a common bottleneck in the process of structure determination, as a crystal is an essential requirement for diffraction experiments. The crystallization process is influenced by numerous factors, with the properties of the protein sample being the most important variable (Dule et al., 2003). Therefore, pre-crystallization biochemical and biophysical characterization of the protein sample can help in successful crystallization (Zuhluf & D'Arcy, 1992; Ericsson et al., 2006). The use of pre-crystallization analysis is described here in detail for a single-chain variable fragment of the antibody MEM-57.

The monoclonal antibody MEM-57 (murine IgG2) recognizes the human CD3 antigen, which is expressed on the surface of T lymphocytes in complex with the T-cell receptor (Hilgert et al., 1999). CD3 plays a role in the transduction of the activation signal after antigen recognition by the T-cell receptor (Weiss, 1990). The antibody MEM-57 shows similar effects (Hilgert et al., 1999; Fan et al., 2006) to the therapeutic anti-CD3 antibody OKT3 (Kung et al., 1979; muromonab CD3, Orthocline OKT3), which is used as an immunosuppressant in organ transplantation (Goldstein, 1987; Sgro, 1995; Chatenoud, 2005). The antibody MEM-57 could be used in diagnostics of autoimmune diseases, for T-cell lymphoma classification, as an immunosuppressant in transplantation (Hilgert et al., 1999) or in cancer therapy in the therapeutic antibody format of a bispecific T-cell engager (Wolf et al., 2005). Structural information on scFv MEM-57 would facilitate humanization of the antibody for therapeutic applications.

ScFv fragments are constructed as a fusion of the variable domains of the heavy (VH) and light (VL) immunoglobulin chains connected by a flexible linker, which stabilizes the noncovalent interaction between the VH and VL (Huston et al., 1988; Bird et al., 1988). The linker sequence Gly4Ser3 (Huston et al., 1988) is universally used, as
it is flexible, hydrophilic, protease-resistant and does not interact with either domain (Paparasoundis et al., 1999). A length of the linker of between 12 and 25 amino-acid residues (Jüger & Plückthun, 1999; Pantoliano et al., 1991) allows the correct assembly of the two variable domains, resulting preferentially in the monomeric scFv form (Zdanov et al., 1994; Whitlow et al., 1994; Korti et al., 1994). However, a fraction of domain-swapped scFv dimers and higher oligomers can be also formed, which prevail when shorter linker sequences are used (Power et al., 2003; Korti et al., 1997; Holliger et al., 1993; Atwill et al., 1999; Carmichael et al., 2003). The preferred oligomeric form of the scFv depends not only on the length of the linker but, amongst other variables, also on the protein concentration and on the pH and the ionic strength of the storage buffer (Arnold et al., 1998), which can be the subject of further optimization. Here, we use the scFv MEM-57 fragment as a model to demonstrate the general approach of optimization of the storage-buffer composition for scFv fragments in order to enhance their crystallizability.

2. Experimental

2.1. Protein expression and purification

The coding sequences for IgG heavy-chain and light-chain variable domains were obtained from hybridoma total RNA using RTP-PCR. The N-terminal primers corresponded either to the amino-acid sequence experimentally determined by N-terminal sequencing or, in the case of a blocked N-terminus, a mixture of degenerate primers (Orlandi et al., 1989) was used. The C-terminal primers corresponded to the conserved sequence of the constant regions adjacent to the variable domains (Kabat et al., 1991). The sequenced PCR products were re-amplified using modified primers to allow assembly of the (VH-pGlu-ScFv-(VL)scFv format, followed by a c-Myc tag sequence for anti-Myc antibody detection. Expression vectors were constructed using a modified pET22b(+)-vector, in which the scFv coding sequence is preceded by the pl vector signal sequence, targeting the recombinant product into the periplasmic space, and is followed by a His tag for purification.

The expression and isolation procedure was based on previously published protocols (Grivtsov et al., 1999; Kipriyanov et al., 1997; Kaperádek et al., 1991). E. coli B(L)41 (DE3) cells transformed by an expression plasmid were inoculated to a final OD$_{600nm}$ of 0.1 in LB medium supplemented with 0.1 mg mL$^{-1}$ ampicillin and 0.45% (v/v) glycerol and were cultured at 30°C. After reaching an OD$_{600nm}$ of 1.0, the cell cultures were supplemented with IPTG to a final concentration of 0.25 mM to induce recombinant protein expression and were cultured at 293 K for an additional 4 h. The cells were harvested by centrifugation and the soluble recombinant protein was isolated from the periplasmic space by hypotonic shock. The cells were then subjected to a 10% (w/v) buffer $300	ext{mM NaCl}$, 10 mM Tris pH 8.0, 1 mM EDTA, 1 mM PMSF; 10 mM of buffer per gram of cell paste] and centrifuged. The periplasmic space was then lysed by 45 min incubation at 273 K and 200 rev min$^{-1}$ in cold hypotonic buffer (5 mM MgSO$_4$, 10 mM Tris pH 8.0, 1 mM PMSF); a volume of 20 mL buffer was used per gram of cell paste. The periplasmic protein fraction was cleared by centrifugation, dialyzed against 50 mM Na$_2$HPO$_4$ pH 8.0, 300 mM NaCl at 277 K and loaded onto an Ni-NTA His-Tag affinity chromatography matrix. The His$_6$-tagged protein was eluted using a step gradient of imidazole (5, 20, 50 and 200 mM). The eluted protein was dialyzed against 20 mM dithiothreitol (pH 8.4) and further purified by ion-exchange chromatography using an ÄktaBasic FPLC system on a 1 mL MonoQ 5/50 GL Tricorn column (Amersham Biosciences, Uppsala) at a flow rate of 1 mL min$^{-1}$. The His$_6$-tagged protein was eluted using a segmented gradient of NaCl (0–200 mM NaCl) in 20 mL, 200 mM 1 M NaCl in 10 mL, 1 M NaCl in 5 mL and 1–0 M NaCl in 5 mL. The purified protein was concentrated by ultrafiltration using Amicon Ultra concentrators (Merck Millipore, USA) and stored in 20 mM dithiothreitol (pH 8.4), 100 mM NaCl or 100 mM sodium phosphate pH 7.5, 200 mM NaCl at 203 K. Protein purity was monitored by silver-stained SDS–PAGE without reducing conditions.

2.2. Size-exclusion chromatography

Analytical size-exclusion chromatography (SEC) was performed using an ÄktaBasic FPLC system on a 2.5 mL Superdex 200 10/300 GL Tricorn column (Amersham Biosciences, Uppsala) at a flow rate of 1 mL min$^{-1}$. The protein was eluted using a segmented gradient of NaCl (0–200 mM NaCl) in 20 mL, 200 mM 1 M NaCl in 10 mL, 1 M NaCl in 5 mL and 0 M NaCl in 5 mL. The purified protein was concentrated by ultrafiltration using Amicon Ultra concentrators (Merck Millipore, USA) and stored in 20 mM dithiothreitol (pH 8.4), 100 mM NaCl or 100 mM sodium phosphate pH 7.5, 200 mM NaCl.

2.3. Thermofluor assay

The Thermofluor assay was performed according to a previously published protocol (Püschl et al., 2013; Procházková et al., 2012) based on the procedure described previously (Eriksen et al., 2006; Reinhard et al., 2013). The stabilizing effect of 5 mM Na$_2$HPO$_4$ at a concentration of 100 mM was tested in the presence or absence of 200 mM NaCl; the protein concentration was 0.2 mg mL$^{-1}$. The melting temperatures ($T_m$) were determined as the minima of the negative first derivatives of the melting curves using the LightCycler 480 Software (Roche, Switzerland).

2.4. Dynamic light scattering

Dynamic light-scattering (DLS) measurements were performed prior to setting up the crystallization trials using 15 μL concentrated (15 mg mL$^{-1}$) centrifuged protein samples at 352 nm, 90° angle and 293 K on a Laser-Spektroskatter 201 (GmbH Network RNA-Technologien, Germany).

2.5. Protein crystallization

The scFv MEM-57 protein at a concentration of 15 mg mL$^{-1}$ was used for initial crystallization screening. Screening was performed at 291 K by the sitting-drop vapour-diffusion method with the help of a Gryphon crystallization workstation (Art Robbins Instruments, USA) in 96-well sitting-drop Intelli-Plates (Art Robbins Instruments, USA) using the commercial screening kit The PEGs Suite (Qiagen, Netherlands). A total volume of 450 μL of a mixture of the protein and precipitant solution in a 2:1 ratio, respectively, was equilibrated against 50 μL reservoir solution.
3. Results and discussion

The single-chain variable fragment (scFv) of the antibody MEM-57 was constructed: variable domains of the heavy and light chain were joined by a flexible linker and further equipped with an N-terminal pellB leader sequence and a C-terminal c-Myc tag and His tag. A longer version of the classical (Gly4Ser3) linker (Huston et al., 1998) with the sequence (Gly4Ser4) was used to favour the formation of scFv monomers (Table 1, Fig. 1). Fig. 1(b) illustrates the formation of inter-chain domain-swapped dimers, a typical feature of scFv fragments, which occurs in solution. To our knowledge, no other type of dimerization has been observed for scFv fragments.

The N-terminal pellB leader sequence targets the recombinant product into the periplasmic space of E. coli and is cleaved off during transport. In this compartment, disulfide bridges are formed and the recombinant antibody fragment usually accumulates in a soluble form. The recombinant protein was isolated through selective opening of the periplasmic compartment by osmotic shock. A two-step purification protocol employing nickel-chelation chromatography and ion-exchange chromatography produced a moderate yield of pure protein, as analyzed by nonreducing denaturing SDS-PAGE (Fig. 1c). The yield of 3 mg scFv MEM-57 per litre of bacterial culture is comparable to the typical yields obtained for other recombinant antibody fragments prepared from E. coli periplasm (Kipriyanov et al., 1997; Bayly et al., 2002). The antigen-binding activity of the scFv MEM-57 fragment was confirmed by flow cytometry on CD3-positive Jurkat cells (data not shown).

The ratio between the monomeric and the multimeric forms of the purified scFv fragment was determined by analytical size-exclusion chromatography (SEC) under native conditions. An equilibrium was established between monomer, dimer and higher oligomers of scFv MEM-57 in the original storage buffer, which was 20 mM diethanolamine pH 8.4, 100 mM NaCl (Fig. 2a). The monomeric form represented about 65% of the total protein. Dynamic light scattering (DLS) was also used to evaluate the dispersity of the protein preparation at the high concentration used for crystallization experiments. The DLS analysis revealed protein sample polydispersity (Fig. 2b), which indicated a low probability of successful crystallization of the protein (Zulauf & D'Arcy, 1992).

The equilibrium between oligomeric scFv forms was established even after the monomeric and dimeric forms were separated using ion-exchange chromatography (MonoQ column, 20 mM diethanolamine pH 8.4, NaCl gradient; for details, see §2.1). The scFv MEM-57 eluted in two peaks that probably corresponded to the scFv monomer and dimer (data not shown). When each fraction was separately analyzed by SEC analysis, an equilibrium between the monomer, dimer and higher oligomers similar to that shown in Fig. 2(a) was observed.

Because of the inhomogeneity of the protein originating from the equilibrium between scFv monomer and oligomers, it is not favorable for protein crystallization, we decided to screen for a buffer
composition which would stabilize the protein in one of the existing forms. The Thermofluor assay (also known as differential scanning fluorimetry) is currently routinely used to screen for optimal protein storage buffer compositions for protein stability and conformational homogeneity (Ericson et al., 2006; Reinhard et al., 2013). In our case, we not only used Thermofluor to evaluate protein stability, but we also followed the protein oligomeric state, which was reflected by a specific shape of the melting curve. A screen consisting of 22 buffers at pH 4.5–8.5 was used in the presence or absence of 200 mM NaCl (Ericson et al., 2006). Conditions in which the protein exhibited a multiple-inference melting curve, indicating the presence of multiple oligomeric forms, were excluded from further analysis of the melting temperature (T_m). In the T_m analysis, the buffer in which the protein exhibited a single-inference curve with the highest T_m (331 K) was selected as the optimal protein storage buffer (100 mM sodium phosphate pH 7.5, 200 mM NaCl; see Fig. 3). The protein-unfolding curve for the original protein storage buffer also included a second inflection, which was more apparent in the negative first derivative of the curve as a second minimum.

SEC and DLS analyses were performed on the scFv MEM-57 protein in the optimized buffer. SEC confirmed an improved oligomeric homogeneity of the protein sample, as illustrated in Fig. 4(a). The equilibrium between the monomer and dimer has shifted significantly in the favor of the monomeric form; the monomer represented 88% of the total protein (compared with 65% of monomer in the original storage buffer). DLS analysis revealed a monomodal particle-size distribution (Fig. 4b).

The positive effect of the optimized storage buffer composition on protein crystallizability was confirmed by the results of the initial crystallization trials: protein crystals were obtained in 41 out of 96 screened conditions (Fig. 5).

Selected crystals obtained in the screening diffracted to a maximal resolution of 6 Å (data not shown) when tested for X-ray diffraction at 120 K using a MAR345 image-plate system and a Nonius FR/1 rotating-anode generator (Bruker Nonius B.V., Netherlands). Optimization of the diffraction quality of the crystals is currently in progress. This includes optimization of the crystallization conditions, cryoprotection optimization and optimization of the protein construct (specifically introducing cleavable tags at either the C-termius or the N-termius). The use of an epitope-derived peptide to stabilize the oligomeric state and/or antigen-binding site, a strategy successfully used previously for other scFv fragments (Razvova et al., 2001, 2005), cannot be used for our antibody fragment as the epitope of MEM-57 is not linear but structural.

4. Conclusions

Pre-crystallization analysis of protein samples using a combination of biochemical and biophysical methods such as size-exclusion chromatography, dynamic light scattering and the Thermofluor assay can be used to help protein crystallization. We have employed these methods to optimize the stability and also the oligomeric homogeneity of scFv MEM-57 for crystallization. This approach resolved

Figure 3
Thermofluor assay for scFv MEM-57. (a) Thermofluor-based protein-unfolding curves for the original (black) and optimized (green) storage buffers are shown along with representative multiple-inference curves which were excluded from further melting-temperatures (T_m) analysis (blue). (b) Melting temperatures for storage buffers with the melting temperatures of other single-inference buffers. The original and optimized storage buffers are highlighted in black and green, respectively.

Figure 4
Homogeneity of scFv MEM-57 in the optimized storage buffer (100 mM sodium phosphate pH 7.5, 200 mM NaCl). (a) Size-exclusion chromatographic analysis of scFv MEM-57 at 7.5 mg ml⁻¹. Apparent molecular weights are indicated for each peak together with the percentage of the individual oligomeric forms in the total protein content. The discrepancy in the apparent molecular weights for the scFv dimer in the original and optimized buffer (compare with Fig. 3b) might be caused by the degradation and renaturation of the SEC column between these two runs. (b) The DLS particle-size distribution reveals monodispersity of scFv MEM-57 at 15 mg ml⁻¹.
the issue of unfavourable oligomerization of the scFv fragment by stabilization of the monomeric form and thus resulted in a protein sample with improved crystallizability.

This work was supported by the Charles University in Prague, project GA UK No. 567012. This work was also supported in part by projects RV06 309863 and 68376009 awarded by the Academy of Sciences of the Czech Republic and by the Ministry of Education of the Czech Republic – LK110125 project ‘NAVRAIT’.

References


laboratory communications


7.4 Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85

Background

Human hyaluronate receptor CD44 plays an important role in various physiological processes including cell adhesion and migration, such as hematopoiesis, lymphocyte homing, tissue healing, but also tumor metastasis. The transmembrane glycoprotein CD44 binds hyaluronate through the extracellular hyaluronate-binding domain (HABD), which induces an allosteric conformational change, resulting in CD44 HABD shedding. Murine monoclonal antibody MEM-85 recognizes CD44 HABD and might be of therapeutic interest as it has been shown to inhibit growth of lung cancer cells in murine models. MEM-85 cross-blocks hyaluronate binding to CD44 and also induces CD44 shedding, similar to hyaluronate-induced CD44 shedding, by an unknown mechanism. The epitope recognized by MEM-85 has not been unambiguously identified. Therefore, a detailed structural study of the MEM-85 – CD44 HABD interaction would shed light on the molecular mechanism of MEM-85 function.

Summary

We constructed a single-chain variable fragment (scFv) of murine antibody MEM-85 against the hyaluronate-binding domain (HABD) of human CD44. Flow cytometry and fluorescence microscopy confirmed that scFv MEM-85 binds native CD44 on cell surface. ThermoFluor assay revealed high thermal stability of the antibody fragment. We prepared recombinant $^{15}$N/$^{13}$C-labeled CD44 HABD, which was able to bind hyaluronate, as shown by NMR. We analyzed the antigen binding of both scFv and IgG MEM-85 using surface plasmon resonance (SPR). Preparation of a stable homogeneous complex of scFv MEM-85 and CD44 HABD allowed us to employ NMR and small-angle X-ray scattering (SAXS) analyses. We acquired a set of three-dimensional NMR spectra for the sequence-specific backbone resonance assignments for CD44 HABD. We used the minimal shift approach to assess the extent of the chemical shift perturbation upon scFv MEM-85 binding. By this approach, we identified the epitope, located in the C-terminal portion of CD44 HABD, which was corroborated with the results of a mutational analysis of the identified epitope residues combined with SPR measurements. Finally, we obtained a low-resolution ab initio SAXS structure of the complex together with a rigid body model of the complex using the
available high-resolution structures. The key results of the publication are summarized in Figure 11.

**Figure 11: Identification of MEM-85 binding epitope in CD44 HABD**

A: Two-dimensional $^{15}$N/$^1$H HSQC spectra are shown for free $^{15}$N/$^1$C-CD44 HABD (black) and $^{15}$N/$^1$C-CD44 HABD in complex with scFv MEM-85 (red). B: Histogram of minimal combined chemical shift versus the protein sequence is shown for $^{15}$N/$^1$C-CD44 HABD. C: Significantly perturbed residues are mapped on the surface of free CD44 HABD. D: Significantly perturbed residues are mapped on the surface of hyaluronate-bound CD44 HABD. E: SAXS *ab initio* envelope (grey) is shown in superposition with the rigid body model of scFv MEM-85 (red) and CD44 HABD (green).

**My contribution**

I participated in the construction of the CD44 HABD expression vectors, prepared protein samples for all experiments, participated in the flow cytometry and fluorescence microscopy experiments, performed size-exclusion chromatography, dynamic light scattering, and Thermofluor assay. I performed the sequence-specific backbone NMR assignments and the chemical shift perturbation analysis by the minimal shift method. I prepared samples for SAXS measurements, participated in SAXS data analysis, and I wrote the manuscript.

Molecular mechanism for the action of the anti-CD44 monoclonal antibody MEM-85

Jana Škerlová\textsuperscript{a,b,c}, Vlastimil Král\textsuperscript{b}, Michael Kachala\textsuperscript{d}, Milan Fábry\textsuperscript{b}, Ladislav Bumba\textsuperscript{e}, Dmitri I. Svergun\textsuperscript{d}, Zdeněk Tošner\textsuperscript{f}, Václav Veverka\textsuperscript{a,\ast}, and Pavlína Řezáčová\textsuperscript{a,b,\ast}

\textsuperscript{a}Institute of Organic Chemistry and Biochemistry, AS CR, v.v.i., Flemingovo nam. 2, Prague 6, 166 10, Czech Republic; \textsuperscript{b}Institute of Molecular Genetics, AS CR, v.v.i., Videnska 1083, Prague 4, 142 20, Czech Republic; \textsuperscript{c}Department of Biochemistry, Faculty of Science, Charles University in Prague, Albertov 6, 128 40 Prague 2, Czech Republic; \textsuperscript{d}European Molecular Biology Laboratory, Hamburg Outstation, c/o Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, D-22603 Hamburg, Germany; \textsuperscript{e}Institute of Microbiology, AS CR, v.v.i., Videnska 1083, Prague 4, 142 20, Czech Republic; \textsuperscript{f}Faculty of Science, Charles University in Prague, Albertov 6, 128 40 Prague 2, Czech Republic

\textsuperscript{\ast}To whom correspondence should be addressed: Pavlína Řezáčová, Institute of Organic Chemistry and Biochemistry of the ASCR, v.v.i., Flemingovo nam. 2, Prague 6, 166 10, Czech Republic; tel: +420 220 183 144; fax: +420 220 183 144; e-mail: rezacova@uochb.cas.cz

Correspondence may also be addressed to Václav Veverka, Institute of Organic Chemistry and Biochemistry of the ASCR, v.v.i., Flemingovo nam. 2, Prague 6, 166 10, Czech Republic; tel: +420 220 183 135; fax: +420 220 183 144; e-mail: veverka@uochb.cas.cz

1
Abstract
The hyaluronate receptor CD44 plays role in cell adhesion and migration and is involved in tumor metastasis. The extracellular domain of CD44 comprises the hyaluronate-binding domain (HABD) and the membrane-proximal stem region; the short intracellular portion interacts with adaptor proteins and triggers signaling pathways. Binding of hyaluronate to CD44 HABD induces an allosteric conformational change, which results in CD44 shedding. A poorly characterized epitope in human CD44 HABD is recognized by the murine monoclonal antibody MEM-85, which cross-blocks hyaluronate binding to CD44 and also induces CD44 shedding. MEM-85 is of therapeutic interest, as it inhibits growth of lung cancer cells in murine models. In this work, we employed a combination of biophysical methods to determine the MEM-85 binding epitope in CD44 HABD and to provide detailed insight into the mechanism of MEM-85 action. In particular, we constructed a single-chain variable fragment (scFv) of MEM-85 as a tool for detailed characterization of the CD44 HABD – antibody complex and identified residues within CD44 HABD involved in the interaction with scFv MEM-85 by NMR spectroscopy and mutational analysis. In addition, we built a rigid body model of the CD44 HABD – scFv MEM-85 complex using a low-resolution structure obtained by small-angle X-ray scattering. The MEM-85 epitope is situated in the C-terminal part of CD44 HABD, rather than the hyaluronate-binding groove, and the binding of MEM-85 induces a structural reorganization similar to that induced by hyaluronate. Therefore, the mechanism of MEM-85 cross-blocking of hyaluronate binding is likely of an allosteric, relay-like nature.

Keywords
CD44, epitope mapping, monoclonal antibody, MEM-85, NMR, SAXS

Abbreviations

**INTRODUCTION**

CD44, a type I transmembrane glycoprotein expressed in many mammalian cells, is the principal receptor for hyaluronate (Aruffo et al., 1990), a major component of the extracellular matrix, and plays an important role in cell adhesion and migration. These cellular events are involved in numerous biological processes, including hematopoiesis and lymphocyte homing (Naor et al., 1997), which is initiated by a CD44 mediated rolling adhesion of lymphocytes to endothelial cells presenting hyaluronate (DeGrendele et al., 1996). CD44 also plays a role in various pathological processes, such as tumor-cell migration and metastasis (Nagano and Saya, 2004; Naor et al., 1997).

The extracellular portion of CD44 comprises an N-terminal hyaluronate-binding domain (HABD), followed by a glycosylated stem region. Alternative splicing of the CD44 gene results in several stem region variants. The cytoplasmic tail of CD44 interacts with the cytoskeleton and is involved in intracellular signaling (Ponta et al., 2003). Three-dimensional structures of mouse and human HABD have been solved both by NMR and X-ray crystallography (Banerji et al., 2007; Takeda et al., 2006; Teriete et al., 2004). HABD adopts a novel form of C-lectin-type fold, which can be denoted as: $\beta_0$ - $\beta_0'$ - $\beta_1$ - $\alpha_1$ - $\alpha_2$ - $\beta_3$ - $\beta_4$ - $\beta_5$ - $\beta_6$ - $\beta_7$ - $\beta_8$ - $\beta_9$ - $\alpha_3$ (Fig. 6A). The hyaluronate-binding Link module (in bold), conserved among the family of hyaluronate-binding proteins (Day and Prestwich, 2002), is extended in HABD by an extension lobe formed by $\beta_0$ and $\beta_0'$ at the N-terminus and a C-terminal extension comprising $\beta_7$, $\beta_8$, $\beta_9$, and $\alpha_3$ (Teriete et al., 2004). This additional extension lobe is essential for structural integrity and hyaluronate binding (Banerji et al., 1998; Peach et al., 1993).

The NMR structure of HABD in complex with hyaluronate (Takeda et al., 2006) revealed that the protein undergoes a significant allosteric conformational change upon ligand binding. In this rearrangement, the $\beta_9$ strand and $\alpha_3$ helix become unstructured, and the $\beta_8$ strand changes its position relative to the $\beta_0$ strand, which provides enhanced flexibility of the HABD C-terminus (Takeda et al., 2006). Therefore, the ligand-free state is referred to as “ordered” (O) and the ligand-bound state as “partially disordered” (PD). The murine CD44 HABD structure in the O-conformation has also been captured by X-ray crystallography in the ligand-bound state, since the ordered state is favorable for crystallization (Banerji et al.,
Both conformational states of CD44 HABD exhibit affinity toward hyaluronate and exist in a two-state conformational equilibrium (Ogino et al., 2010). In comparison to the O-conformation, the PD-conformation exhibits higher affinity toward hyaluronate and higher resistance to cell detachment from a hyaluronate-coated surface (Ogino et al., 2010).

The conformational equilibrium between the PD and O states has been proposed to be essential for formation of new receptor-ligand interactions at the rolling front and their disruption at the rear edge in CD44-mediated cell rolling (Ogino et al., 2010). The ligand-induced shift in equilibrium toward the PD-conformation also “extends” the membrane-proximal stem region of CD44 and thus enhances the proteolytic susceptibility of cleavage sites in the stem region (Takeda et al., 2006). This mechanism of ligand-induced CD44 shedding is involved in the regulation of cell motility and migration (Cichy and Pure, 2003; Nagano and Saya, 2004; Okamoto et al., 1999).

The murine IgG2 antibody MEM-85 raised against human CD44 HABD (Bazil et al., 1989) has been shown to block hyaluronate binding to CD44 (Bajorath et al., 1998; Sandmaier et al., 1998). MEM-85 antibody binding induces CD44 shedding, similar to hyaluronate-induced CD44 shedding (Bazil and Horejsi, 1992). MEM-85 might be of therapeutic interest, as it has recently been shown to inhibit growth of lung cancer cells in murine models (Godar, S., personal communication). The HABD epitope recognized by MEM-85 has not yet been unambiguously identified. Results from a mutational analysis suggested that the epitope lies at the N-terminal part of HABD (Anstee et al., 1991; Bajorath et al., 1998; Stefanova et al., 1989), and residues Lys38, Arg41, and Tyr42 have been shown to be important for MEM-85 binding (Bajorath et al., 1998). Some consider the MEM-85 epitope to be structural (Bajorath et al., 1998), while others have shown that MEM-85 binds CD44 under denaturing non-reducing conditions (Stefanova et al., 1989), which implicates a sequential epitope additionally stabilized by a disulfide bridge.

Here, we present a detailed characterization of the interaction of MEM-85, a monoclonal antibody with therapeutic potential, with hyaluronate receptor CD44 implicated in tumor-cell migration. With the aim of exploring rather enigmatic epitope recognized by MEM-85, we constructed a single-chain variable fragment (scFv) of MEM-85 and prepared its stable complex with CD44 HABD. We carried out epitope mapping by NMR chemical shift perturbation analysis, which allowed us to identify the epitope of MEM-85 antibody in CD44 HABD and uncover the structural changes induced by antibody binding. Using small-angle X-ray scattering (SAXS), we modeled the complex of CD44 HABD and scFv MEM-85, which provided further insight into the mechanism of MEM-85 action.
MATERIAL AND METHODS

Construction of CD44 HABD expression plasmid

Construction of pCD44HABD was based on the work of Banerji et al. (1998) and Teriete et al. (2004). HABD, which encompasses residues 21-178 of transcript variant 4 of human CD44, was amplified from cloned cDNA (Cat. No. SC128160, OriGene, Rockville, MD, USA) using the following primers: forward, 5'-GCCGGCTAGGCGAGATCGATTAGAATAACC; reverse, 5'-CCGAATTCAACGTCATCACTAGTAAGGGTG. To obtain a construct of CD44 HABD with a His6 tag, the vector pMCSG7 (T7 promoter-driven, originally designed for ligation-independent cloning) (Stols et al., 2002) was modified so that the N-terminal His6 tag was followed by a TEV (tobacco etch virus) protease cleavage site and amino acid residues SNAAS, where AS corresponds to an NheI site. The CD44 HABD coding fragment was, after sequence verification, cloned as an NheI - EcoRI insert (recognition sites are underlined in the primer sequences) into this modified vector. Cleavage with TEV protease thus yields a product in which CD44 HABD (residues 21-178) is preceded by SNAAS at the N-terminus.

CD44 HABD mutations D140 + T144A, Y161A, and E160A + T163A were introduced into the coding sequence of CD44 HABD in pUC18 according to the QuikChange Site-Directed Mutagenesis Protocol (Stratagene, La Jolla, CA, USA) using the following pairs of mutagenic primers: D140 + T144 (5'-CCAATGCCTTTGCGCAGGCATCTAATTTGG, 5'-CAATAGTTATGCGAATTGGGCACCGACAAAGGCAATGG), Y161A (5'-GTCCAGAAAGAAGCGGCACGAAATCTCTAGAGAC, 5'-GTCTTCAGGATTCGTCGCTGGCCGCTTCTCTTCTGAC), E160 + T163 (5'-GTCCAGAAAGAAGCATAACCGGCGGGAAT CCTGAAGAC, 5'-GTCTTCAGGATTCGTCGCTGGGTATGCTCTCTCTGGAC). After sequence verification, the NheI-EcoRI fragments carrying the respective mutations were used to replace the wild-type fragment in the pCD44HABD expression plasmid.

Expression and purification of CD44 HABD

The expression and purification of CD44 HABD was based on a published protocol (Banerji et al., 1998). Recombinant CD44 HABD was expressed in E. coli BL21 (DE3) in the LB medium. Transformed cells were grown at 37 °C to an optical density OD550 nm of 1.5 and induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. At 4 h post-induction, cells were harvested, suspended in wash buffer (50 mM

5
Tris, pH 8.0, 50 mM NaCl, 5 mM EDTA), taken through one cycle of freezing and thawing,
partially homogenized using a Dounce homogenizer and taken through four cycles of
homogenization on EmulsiFlex-C3 (Avestin, Ottawa, ON, Canada). The inclusion body pellet
was sonicated (three 60-s pulses) on ice in 10% (v/v) Triton X-100 in wash buffer and
centrifuged. The Triton-washed pellet was sonicated (three 60-s pulses) on ice in wash buffer
and centrifuged. This procedure yielded purified inclusion bodies, which were denatured in 50
mM MES, pH 6.5, 8 M urea, 0.1 mM EDTA, 0.1 mM DTT (120 mg/ml inclusion bodies) by
rotating overnight at 4 °C. After removal of insoluble material by centrifugation, the protein
was refolded by rapid 250-fold dilution into 100 mM Tris, pH 8.5, 400 mM L-arginine, 2.5
mM reduced glutathione, 2.5 mM oxidized glutathione (480 μM CD44 HABD) and slowly
stirred at 4 °C for 48 h before dialysis into 50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl for
nickel-chelation chromatography. Refolded protein was concentrated by nickel-chelation
chromatography on 9 ml HIS-Select Nickel Affinity Gel column (Cat. No. P6611, Sigma-
Aldrich, St. Louis, MO, USA) in 50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl with a flow rate of
1 ml/min and eluted using 200 mM imidazole. Eluted protein was dialyzed into phosphate-
buffered saline (PBS) and further purified by size-exclusion chromatography using an Äkta
Basic FPLC system with a 124 ml HiLoad 16/600 Superdex 200 pg column (GE Healthcare,
Little Chalfont, UK) in PBS at a flow rate of 1 ml/min. The N-terminal His₆ tag was removed
by incubation with His₆-encoded TRV protease in PBS at 4 °C for 48 h. The cleaved product
was separated from His₆-encoded TEV protease and uncleaved CD44 HABD by nickel-
chelation chromatography on 2 ml HIS-Select Nickel Affinity Gel column (Cat. No. P6611,
Sigma-Aldrich, St. Louis, MO, USA) in PBS with a flow rate of 1 ml/min. Recombinant
product was concentrated in an Amicon Ultra concentrator (Merck Millipore, Billerica, MA,
USA), and protein purity was analyzed by 15% SDS-PAGE according to Laemmli (1970).
For the preparation of 15N/13C-labeled protein, cells were grown in minimal medium
supplemented with 1 g/l (15NH₄)₂SO₄ (Cat. No. NLM-713-10, Cambridge Isotope
Laboratories Inc., Tewksbury, MA, USA) and 4 g/l 13C-D-glucose (Cat. No. CLM-1396-5,
Cambridge Isotope Laboratories Inc., Tewksbury, MA, USA) (Renshaw et al., 2004; Veverka
et al., 2006).

Construction of scFv MEM-85 expression plasmid
The coding sequences for the variable domains of the monoclonal antibody MEM-85 heavy
and light chains (VH and VL, respectively) were obtained by RT-PCR using total RNA from
MEM-85 hybridoma cells (kindly provided by Václav Hořejší, Institute of Molecular
Genetics, AS CR, v.v.i.). The N-terminal primers reflected the experimentally determined amino acid sequences, and the C-terminal primers corresponded to constant regions adjacent to the variable domains (Kabat et al., 1991). PCR products were sequenced and used as templates for reamplification with modified primers, allowing assembly of the \( V_L \) and \( V_H \) coding DNA fragments into the scFv molecule in the format (\( V_H \)-(Gly\(_2\)Ser\(_2\)-)(\( V_L \)). The scFv molecule contains 115 N-terminal residues from the heavy chain linked to 114 N-terminal residues from the light chain, followed by a c-myc tag sequence (EQKLISEEDL) to allow detection of scFv MEM-85 with anti-myc antibody and by a His\(_3\) tag to allow product isolation and purification by nickel-chelation chromatography. For expression in \( E. coli \) BL21(DE3) cells, we used the pET22b(+) derived vector, \( pscFvMEM-85 \), in which the scFv coding sequence is preceded by the pelB signal sequence, directing translocation of the product into the periplasmic space.

**Expression and purification of scFv MEM-85**

Recombinant scFv MEM-85 was expressed in \( E. coli \) BL21 (DE3) in the LB medium. Transformed cells were grown at 37 °C to the optical density OD\(_{550}\) of 1.0, induced by addition of IPTG to a final concentration of 0.25 mM, and cultured at 20 °C. At 4 h post-induction, cells were harvested. The product, which accumulated in a soluble form in the periplasmic space of the host cells, was isolated by hypotonic shock and purified by nickel-chelation chromatography on a 6 ml HIS-Select Nickel Affinity Gel column (Cat. No. P6611, Sigma-Aldrich, St. Louis, MO, USA) in 50 mM NaH\(_2\)PO\(_4\), pH 8.4, 300 mM NaCl with a flow rate of 1 ml/min. Bound protein was eluted using a segmented gradient of imidazole. The protein was further purified by ion-exchange chromatography on a 1 ml MonoQ 5/50 GL Tricorn column mounted on an Äkta Basic FPLC system (GE Healthcare, Little Chalfont, UK) in 20 mM diethanolamine, pH 8.4, using a segmented gradient of NaCl. Finally, the recombinant product was concentrated in Amicon Ultra concentrators (Merck Millipore, Billerica, MA, USA), and protein purity was analyzed by 15% SDS-PAGE according to Laemmli (1970).

**Biochemical and biophysical characterization of proteins**

*Size-exclusion chromatography*

Formation of the complex of CD44 HABD and scFv MEM-85 was analyzed by size-exclusion chromatography using an Äkta Basic FPLC system on a 23 ml Superdex 200
10/300 GL Tricorn column (GE Healthcare, Little Chalfont, UK) in phosphate-buffered saline (PBS) at a flow rate of 0.5 ml/min.

**Thermofluor assay**

Thermofluor assay was performed according to published protocols (Pisackova et al., 2013; Prochazkova et al., 2012; Skerlova et al., 2014), based on the procedure described by Ericsson et al. (2006). The stabilizing effects of 22 buffers (100 mM concentrations) were tested in the presence or absence of 200 mM NaCl; the protein concentration was 0.2 mg/ml.

**Fluorescence microscopy**

Fluorescence microscopy was performed to assay the antigen binding activity of recombinant scFv MEM-85. CD44^+ MOLT-4 cells were cultured overnight at 37 °C on poly-L-lysine-coated microscopic cover slides. After formaldehyde fixation and blocking with 10% fetal bovine serum (FBS) in PBS, the cells were incubated in a humidity chamber for 1 hour with scFv MEM-85 or IgG MEM-85 (Cat. No. 11-221-C100, Exbio, Vestec, Czech Republic) at concentrations ranging from 1 μM to 100 nM in 10% FBS in PBS, washed in PBS, and incubated for 1 hour with FITC-conjugated secondary antibodies: anti-c-myc/FITC clone 9E10 (Cat. No. 1F-433-C100, Exbio, Vestec, Czech Republic) for scFv MEM-85 and GAM/FITC (goat anti-mouse antibody conjugated with fluorescein isothiocyanate, Cat. No. 115-095-164, Jackson ImmunoResearch Laboratories, West Grove, PA, USA) for IgG MEM-85 (30 nM in 10% FBS in PBS). The slides were washed in PBS, treated with 78 mg/ml mowiol, 1.5 μg/ml N-propylgallate and 6.3 mg/ml DAPI, and analyzed using an Olympus IX81 fluorescence microscope (Olympus, Shinjuku, Tokio, Japan).

**Flow cytometry**

Flow cytometry was performed on CD44^+ MOLT-4 cells and CD44^+ Jurkat cells to assay the antigen binding activity of recombinant scFv MEM-85. Cells (2×10^5 cells per well) were incubated for 20 minutes on ice in a microtiter U-bottom plate (Nalge Nunc International, Penfield, NY, USA) with scFv or IgG MEM-85 (Cat. No. 11-221-C100, Exbio, Vestec, Czech Republic) at concentrations ranging from 10 μM to 10 nM in 1% BSA in PBS. After washing in PBS, cells were incubated for 20 minutes on ice with the following secondary FITC-conjugated antibodies: anti-c-myc/FITC clone 9E10 (Cat. No. 1F-433-C100, Exbio, Vestec, Czech Republic) for scFv MEM-85 and GAM/FITC (Cat. No. 115-095-164, Jackson ImmunoResearch Laboratories, West Grove, PA, USA) for IgG MEM-85 (30 nM in 1% BSA in PBS). After washing in PBS, cells were suspended in 200 μl PBS supplemented with 1 μg/ml propidium iodide and measured on FACS Aria (Becton, Dickinson and
Company, Franklin Lakes, NJ, USA). The data were analyzed using FlowJo 7.2.5 software (Tree Star, Ashland, OR, USA).

*Surface plasmon resonance*

All surface plasmon resonance (SPR) measurements were performed at 25 °C using an HTG sensor chip mounted on a ProteOn XPR36 Protein Interaction Array System (Bio-rad, Hercules, CA, USA). The His-tagged proteins were diluted to final concentrations of 20 µg/ml (scFv MEM-85) or 5 µg/ml (CD44 HABD wild-type, wt, or mutants) in running buffer (PBS with 0.05% Tween 20) and captured on a Ni²⁺-NTA-functionalized HTG chip (Bio-rad, Hercules, CA, USA) at a flow rate of 30 µl/min using a ProteOn HTG Capturing Kit (Bio-rad, Hercules, CA, USA). The analyte proteins were serially diluted in the running buffer (0.312, 0.625, 1.25, 2.5, and 5 µM CD44 HABD; 31, 62, 125, 250, and 500 nM IgG MEM-85 – Cat. No. 11-221-C100, Exbio, Vestec, Czech Republic) and injected in parallel over the immobilized proteins at a flow rate of 30 µl/min. The sensorgrams were corrected for the sensor background by interspot referencing (the sites within the 6×6 array which were not exposed to ligand immobilization but were exposed to analyte flow), and double-referenced by subtraction of analyte (channels 1–5) using a “blank” injection (channel 6). The data were analyzed globally by fitting both the association and dissociation phases simultaneously for five different analyte concentrations using the 1:1 Langmuir-type binding model and the bivalent analyte model to determine the kinetic rate constants. The Langmuir-type model assumes the interaction between proteins A and B resulting in a direct formation of the final complex: A + B ⇔ AB, where $k_a$ and $k_d$ are the association and dissociation rate constants, respectively. The equilibrium dissociation constant, $K_D$, was determined as $K_D = k_d/k_a$. The bivalent analyte model assumes a two-step process: A + B ⇔ AB ⇔ AB₂, where formation of the AB complex is followed by subsequent interaction of the AB complex with another B molecule. Parameters $k_{a1}$ and $k_{d1}$ are the rate constants for the formation of the AB complex, while $k_{a2}$ and $k_{d2}$ are the rate constants for the formation of the AB₂ complex. An apparent equilibrium dissociation constant, $K_{D1}$, was determined as $K_{D1} = k_{d1}/k_{a1}$.

*Epitope mapping*

*NMR spectroscopy*

Sets of double- and triple-resonance NMR spectra were acquired at 308 K on a 600 MHz Bruker Avance spectrometer with triple-resonance ($^{15}$N/$^{13}$C/$^1$H) cryoprobes (Bruker, Billerica, MA, USA) for a 300 µl sample containing 580 µM $^{15}$N,$^{13}$C-labeled CD44 HABD and for a 250 µl sample containing 150 µM $^{15}$N/$^{13}$C-labeled CD44 HABD and 300 µM unlabeled scFv
MEM-85. Two-dimensional $^{15}$N/$^2$H HSQC (heteronuclear single quantum coherence) spectra were acquired for a 350 μl sample containing 100 μM $^{15}$N-labeled CD44 HABD, and for a 350 μl sample containing 200 μM $^{15}$N-labeled CD44 HABD and 210 to 620 μM hyaluronate hexamer (Contipro Group, Dolni Dobrouc, Czech Republic). The sample buffer was PBS supplemented with 5% D$_2$O. Essentially complete sequence-specific resonance assignment was obtained for free CD44 HABD, as previously described (Renshaw et al., 2004; Veverka et al., 2006). To determine the epitope in CD44 HABD, antibody-induced perturbations of $^{15}$N/$^{13}$C-labeled CD44 HABD signals in HNCO spectra were monitored employing the minimal backbone chemical shift method ($^{15}$N, $^{13}$C(N), and $^1$H) (Cheng et al., 2013). The minimum changes in the positions of HNCO peaks for the free and antibody-bound CD44 HABD were calculated and plotted against the protein sequence. This histogram identified CD44 HABD residues significantly affected by antibody binding, which were mapped onto the structures of free CD44 HABD (Protein Data Bank, PDB, code 1POZ; Teriete et al., 2004) and CD44 HABD in complex with hyaluronate hexamer (PDB code 2183; Takeda et al., 2006) using the PYMOL Molecular Graphics System (Version 1.3r1 Schrodinger, LLC. 2010).

Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) data were obtained on the European Molecular Biology Laboratory P12 beamline on the PETRA III storage ring (DESY, Hamburg, Germany) with a PILATUS 2M pixel detector (DECTRIS, Baden, Switzerland). The measurements were performed with a sample-detector distance of 3.1 m and wavelength of 1.24 Å. A range of momentum transfer of 0.0028 < $s$ < 0.45 Å$^{-1}$ (scattering vector $s = 4\pi \sin(\theta)/\lambda$, where $\theta$ is the scattering angle) was covered with this set-up. The CD44 HABD - scFv MEM-85 sample was measured at 3 different concentrations ranging from 1.3 to 5.0 mg/ml. We observed little concentration dependence; therefore, we extrapolated to zero concentration. To test for radiation damage, a 1-s exposure time was divided into 20 50-ms frames. No radiation damage was detected when the frames were compared. The collected data were processed with PRIMUS (Konarev et al., 2003) and the forward scattering $I(0)$ and radius of gyration ($R_g$) were calculated using the Guinier approximation (Guinier, 1939). The same parameters as well as the pair-distance distribution function $P(r)$ and the maximum particle dimension ($D_{max}$) were computed with the program GNOM (Svergun, 1992). The excluded volume of the hydrated particle (the Porod volume) was also calculated with GNOM. The molecular mass of the complex was estimated by comparison of the forward scattering of the sample to
that of a reference solution of bovine serum albumin (72 kDa). The low resolution \textit{ab initio}
models of the CD44 HABD – scFv MEM-85 complex were built using the program
DAMMIN (Svergun, 1999). The algorithm employs simulated annealing to determine the
shape of the macromolecule represented as densely packed beads within a sphere with
diameter $D_{\text{max}}$ by minimizing the discrepancy ($\chi^2$) between experimental and modeled
scattering curves. For the CD44 HABD – scFv MEM-85 complex, 10 initial models were
calculated and then averaged with the DAMAVER (Volkov and Svergun, 2003) program. The
rigid body modeling of the CD44 HABD – scFv MEM-85 complex was performed with
SASREF (Petoukhov and Svergun, 2005) with specified contact conditions using the solution
structure of CD44 HABD without the flexible tail residues 164-178 (PDB code 2I83; Takeda
et al., 2006) and a model scFv MEM-85 created using the RosettaAntibody modeling server
(Sircar et al., 2009). The final model was fitted into the \textit{ab initio} envelope using the
SUPCOMB program (Kozin and Svergun, 2001). For modeling of the full-length structure
including the flexible tail (residues 164-178), CORAL software (Petoukhov and Svergun,
2005; Petoukhov et al., 2012) was used. The last 15 residues of CD44 were marked as
missing for CORAL input and were reconstructed by the program as a C-alpha trace. The
program combines rigid body modeling with the modeling of missing fragments (loops or
termini) as a C-alpha trace and optimizes the structure by simulated annealing.

RESULTS

Protein preparation

CD44 HABD was heterologously expressed in \textit{E. coli}, refolded from inclusion bodies
(Banerji et al., 1998), and purified to homogeneity using a combination of nickel-chelation
and size-exclusion chromatography with a final yield of 3 mg native monomeric CD44
HABD per 1 l cell culture. The recombinant HABD consisted of 158 amino acid residues
from the native CD44 sequence (residues 21-178) and an additional short N-terminal
sequence (SNAAS), which remained following cleavage of the polyhistidine affinity tag by
TEV protease. The correct folding of recombinant CD44 HABD was confirmed by
monitoring its hyaluronate-binding activity by NMR. In particular, $^{15}$N-labeled CD44 HABD
was titrated with hyaluronate, and changes in the positions of CD44 HABD backbone amide
signals in two-dimensional $^{15}$N/$^1$H HSQC spectra were followed. The hyaluronate-induced
changes are clearly visualized in an overlay of the $^{15}$N/$^1$H HSQC spectra for free $^{15}$N-CD44

11
HABD and $^{15}$N-CD44 HABD in complex with hyaluronate hexamer (Fig. A.1 in Appendix A).

The MEM-85 antibody fragment (scFv MEM-85) contained 115 N-terminal residues of the heavy chain connected to 114 N-terminal residues of the light chain by a flexible linker (GGGGS)$_4$ and was equipped with a C-terminal c-myc epitope and His$_6$ tag. The scFv MEM-85 protein was expressed as soluble protein in the periplasm of *E. coli* and purified to homogeneity using a combination of nickel-chelation and ion-exchange chromatography with a yield of 0.3 mg per 1 l cell culture.

We performed flow cytometry to test the binding of scFv MEM-85 to native CD44 expressed on MOLT-4 cells. The shift in the FITC fluorescence intensity of the live MOLT-4 cell population upon incubation of cells with scFv MEM-85 (Fig. 1A) confirmed the antigen-binding activity of the antibody fragment. These results were corroborated by results from fluorescence microscopy experiments. The green fluorescence of the FITC-conjugated secondary antibody, which recognizes the c-myc of scFv MEM-85 bound to CD44 on the MOLT-4 cell membrane, confirmed the antibody binding (Fig. 1B).

The kinetics of the interaction between CD44 HABD and scFv MEM-85 was assessed by surface plasmon resonance. His-tagged scFv MEM-85 was immobilized to the sensor chip and probed in parallel with serially diluted CD44 HABD. Fig. 1C shows a typical interaction kinetics characterized by the association and dissociation phases of the sensogram. Kinetic parameters of the interaction were calculated from global fitting of the concentration-dependent binding curves. The data fitted well to a Langmuir-type binding model, indicating a simple 1:1 interaction between CD44 HABD and scFv MEM-85. The ratio between the calculated dissociation ($k_d = 1.0(\pm 0.2) \times 10^{-2}$ s$^{-1}$) and association ($k_a = 2.5(\pm 0.4) \times 10^4$ M$^{-1}$ s$^{-1}$) rate constants allowed us to determine the equilibrium dissociation constant ($K_D$) with a value of 430 ± 60 nM.
Figure 1: Binding of scFv MEM-85 to CD44

A: Flow cytometry on MOLT-4 cells incubated with the secondary anti-myc/FITC antibody only (gray) and with scFv MEM-85-myc followed by the secondary anti-myc/FITC (black).

B: Fluorescence microscopy on MOLT-4 cells with DAPI stained nuclei (blue) incubated with scFv MEM-85-myc followed by the secondary anti-myc/FITC (green fluorescence on the cell membrane). C: One-shot kinetic analysis of the interaction between CD44 HABD and scFv MEM-85 is shown. His-tagged scFv MEM-85 was immobilized to the sensor chip and probed in parallel with serially diluted CD44 HABD (5.0, 2.5, 1.25, 0.625, and 0.312 μM). The experimental binding curves (gray lines) were globally fitted with a 1:1 Langmuir model and the fitted curves are superimposed as thin black lines on top of the sensograms.

Thermofluor assay was used to evaluate the thermal stability of scFv MEM-85 in various buffers. The antibody fragment was stable in all tested buffers, with melting temperatures ranging from 67 to 69 °C.

We prepared a stable equimolar complex of CD44 HABD and scFv MEM-85, as confirmed by analytical size-exclusion chromatography of the complex and individual binding partners (Fig. 2). The protein complex sample showed the high homogeneity and monodispersity
needed for SAXS experiments, as analyzed by dynamic light scattering at a concentration of 8.5 mg/ml (Fig. A.2 in Appendix A).

**Figure 2:** Recombinant CD44 HABD and scFv MEM-85 form a stable homogeneous complex.

A: Size-exclusion chromatograms for the CD44 HABD – scFv MEM-85 complex (dashed gray line), free scFv MEM-85 (black), and free CD44 HABD (gray) are shown. Apparent molecular weights are indicated for each peak with the ratios between the apparent molecular weight and the theoretical molecular weight in parentheses. B: SDS-PAGE analysis of size-exclusion chromatography fractions is shown. Lane 1: molecular weight standard; lane 2: empty; lane 3: CD44 HABD – scFv MEM-85 elution peak. Molecular weights in kDa are indicated.

**NMR mapping of the MEM-85 epitope in CD44 HABD**

We used NMR spectroscopy to identify the CD44 HABD residues involved in MEM-85 binding. Comparison of the two-dimensional $^{15}$N/$^1$H HSQC spectra for free $^{15}$N/$^{13}$C-labeled CD44 HABD and a complex of $^{15}$N/$^{13}$C-labeled CD44 HABD with unlabeled scFv MEM-85 (Fig. 3A) revealed shifts in the positions of CD44 HABD backbone amide signals. To facilitate a detailed analysis of the antibody-induced changes in the CD44 HABD NMR spectra, we obtained essentially complete sequence-specific backbone resonance assignments for free CD44 HABD in conditions suitable for both free proteins as well as the complex. Specifically, we assigned 95% of backbone $^1$H, $^{15}$N, $^{13}$C, and $^{15}$Ca/β CD44 HABD atoms based on a set of standard triple-resonance NMR experiments. Backbone amide signals were assigned for 92% of the residues, but could not be assigned for residues Tyr42, Ala98, Thr108, Asn110, Thr111, Ser112, and Cys129 due to significant line broadening within several regions, most likely caused by conformational heterogeneity on a timescale unfavorable for the used NMR experiments. We used a minimal shift approach to assess the
extent of changes in the positions of CD44 HABD backbone resonance signals ($^{15}$N, $^{13}$C', and $^1$H') resulting from the binding of scFv MEM-85. For this analysis, residues Cys77, Asn94, Ile96, Asn100, and Ile106 were not taken into account, as they exhibited very weak backbone amide signals. The histogram of normalized combined chemical shifts against the CD44 HABD sequence in Fig. 3B indicates that the major antibody-induced changes primarily lie within the C-terminal parts of CD44 HABD; only some of the minor changes occur closer to the N-terminus. The specific C-terminal antibody-affected residues are Asp140, Thr144, Glu160, Tyr161, Thr163, and Ser171. Significantly higher normalized combined minimal shifts were also identified for residue Tyr117. Binding of scFv MEM-85 to CD44 HABD also resulted in a substantial sharpening of the signals of C-terminal CD44 HABD residues, as inferred from a comparison of the relative signal heights for residues 172-177 in the free and antibody-bound states of CD44 HABD, suggesting increased flexibility within this region upon antibody binding.

**Figure 3:** NMR chemical shift perturbations of CD44 HABD upon scFv MEM-85 binding

A: Two-dimensional $^{15}$N/$^1$H HSQC spectra are shown for free $^{15}$N/$^{13}$C-CD44 HABD (black) and $^{15}$N/$^{13}$C-CD44 HABD in complex with scFv MEM-85 (red). B: Histogram of minimal combined shift versus the protein sequence is shown for $^{15}$N/$^{13}$C-CD44 HABD. Numbers of
residues with normalized combined chemical shift values higher than 2 are indicated above the corresponding bars.

**Mutational analysis**

To validate the hypothesis that the MEM-85 epitope is located in the C-terminal extension lobe of CD44 HABD, we prepared CD44 HABD mutants and evaluated their affinities for MEM-85. Five CD44 HABD residues that were significantly perturbed upon MEM-85 binding (Asp140, Thr144, Glu160, Tyr161, and Thr163; Fig. 3B) were mutated to alanines. In particular, a single mutant (Y161A) and two double mutants (D140A + T144A and E160A + T163A) were constructed, and their interactions with MEM-85 were investigated using SPR. The CD44 HABD mutants were immobilized through the N-terminal His tag to the sensor chip and probed in parallel with serially diluted MEM-85 antibody. Real-time interactions of MEM-85 with CD44 HABD wt and mutant proteins are illustrated in Fig. 4. Since antibodies possess two identical epitope-binding regions on a single molecule, the kinetic data were fitted to both 1:1 Langmuir-type and bivalent analyte models. As expected, the interaction between MEM-85 and CD44 HABD proteins was significantly better described (in terms of reduced $\chi^2$ and residual statistics) by the bivalent analyte model. The calculated association and dissociation rate constants ($k_{a1}$, $k_{a2}$, $k_{d1}$, and $k_{d2}$) as well as the apparent equilibrium dissociation constants ($K_{D1}$) for the bivalent binding model are listed in Table 1. The data analysis revealed that MEM-85 binds CD44 HABD with $K_{D1}$ in the nanomolar range ($K_{D1}$ of 4.0 ± 1.7 nM). Y161A and E160A + T163A mutants exhibited significant decrease in the binding affinity for IgG MEM-85 ($K_{D1}$ of 32.6 ± 13.4 nM and 34.3 ± 12.8 nM, respectively). This decrease was caused by a combination of slower association and faster dissociation rates of the complex, as reflected in $k_{a1}$ and $k_{d1}$ rate constant values, respectively (Table 1). The binding affinity of the D140A + T144A mutant ($K_{D1}$ of 5.4 ± 2.1 nM) was similar to that of the wt CD44 HABD.
Figure 4: SPR analysis of the interaction of IgG MEM-85 with immobilized CD44 HABD mutants

One-shot kinetic analyses are shown for interactions of IgG MEM-85 with wild-type CD44 HABD (A), Y161A CD44 HABD (B), D140A + T144A CD44 HABD (C), and E160A + T163A CD44 HABD (D). His-tagged CD44 HABD was immobilized to the sensor chip and probed in parallel with serially diluted IgG MEM-85 (500, 250, 125, 62, and 31 nM). The experimental binding curves (gray lines) were globally fitted with a bivalent analyte model and the fitted curves are superimposed as thin black lines on top of the sensorgrams.

Table 1: Kinetic and binding affinity constants for the interactions between MEM-85 and CD44 HABD mutants

<table>
<thead>
<tr>
<th>CD44 HABD</th>
<th>$k_{a1}$ (×10^4) [M⁻¹s⁻¹]</th>
<th>$k_{d1}$ (×10⁴) [s⁻¹]</th>
<th>$K_{D1}$ [nM]</th>
<th>$k_{a2}$ (×10⁴) [M⁻¹s⁻¹]</th>
<th>$k_{d2}$ (×10⁴) [s⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type</td>
<td>9.1 ± 3.2</td>
<td>3.8 ± 1.5</td>
<td>4.0 ± 1.7</td>
<td>5.5 ± 1.9</td>
<td>2.1 ± 0.8</td>
</tr>
<tr>
<td>Y161A</td>
<td>3.0 ± 1.2</td>
<td>9.8 ± 3.1</td>
<td>32.6 ± 13.4</td>
<td>4.1 ± 1.4</td>
<td>5.5 ± 2.1</td>
</tr>
<tr>
<td>D140A + T144A</td>
<td>8.1 ± 2.6</td>
<td>4.4 ± 1.6</td>
<td>5.4 ± 2.1</td>
<td>1.5 ± 0.6</td>
<td>7.3 ± 2.5</td>
</tr>
<tr>
<td>E160A + T163A</td>
<td>2.8 ± 0.9</td>
<td>9.7 ± 2.9</td>
<td>34.3 ± 12.8</td>
<td>3.6 ± 1.6</td>
<td>4.4 ± 1.9</td>
</tr>
</tbody>
</table>

The apparent equilibrium dissociation constant of the complex, $K_{D1}$, was determined as $k_{d1}/k_{a1}$.

Small-angle X-ray scattering (SAXS)

A low-resolution structure of the CD44 HABD – scFv MEM-85 complex in solution was obtained using SAXS. The experimental scattering curve is shown in Fig. 5A, and overall
parameters of the complex are presented in Table 2. The molecular weight of the CD44 HABD – scFv MEM-85 complex was estimated by comparing the forward scattering intensity of the sample with that of a reference solution of bovine serum albumin. The experimentally determined molecular weight of the sample was approximately 44 kDa, which corresponds to a monomeric complex with 1:1 stoichiometry, which is in agreement with our results from analytical size-exclusion chromatography (39 kDa, Fig. 2A) and the actual molecular weight calculated from the protein sequences (46 kDa). The excluded volume of the hydrated particle (Porod volume) was $5.7 \times 10^4 \, \text{Å}^3$, and the calculated dry volume (Peptide Property Calculator web tool available at http://www.basic.northwestern.edu/biotools/proteincalc.html; Harpaz et al., 1994) was approximately the same ($5.6 \times 10^4 \, \text{Å}^3$).

We calculated the radius of gyration ($R_g$) directly from the scattering curve using Guinier approximation; the $R_g$ was 26.9 Å for the CD44 HABD – scFv MEM-85 complex. The maximum dimension of the particle ($D_{\text{max}}$) derived from the pair-distance distribution function was 94 Å (Fig. 5B).

A low-resolution *ab initio* model of the complex was obtained by generation of 10 dummy atom structures using DAMMIN (Svergun, 1999) and subsequent generation of an average model (using DAMAVER; Volkov and Svergun, 2003) with a fixed core for another DAMMIN run. The final model obtained from this run (Fig. 5C) has an elongated shape.

### Table 2: Overall parameters of the SAXS experiment

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instrument</strong></td>
<td>EMBL P12 beamline storage ring</td>
</tr>
<tr>
<td><strong>Beam geometry [mm²]</strong></td>
<td>0.2 x 0.12</td>
</tr>
<tr>
<td><strong>Wavelength [Å]</strong></td>
<td>1.24</td>
</tr>
<tr>
<td><strong>q-Range [Å⁻¹]</strong></td>
<td>0.003 – 0.448</td>
</tr>
<tr>
<td><strong>Exposure time [s]</strong></td>
<td>1 (20 x 50 ms)</td>
</tr>
<tr>
<td><strong>Concentration range [mg/ml]</strong></td>
<td>1.3 – 5.0</td>
</tr>
<tr>
<td><strong>Temperature [K]</strong></td>
<td>298</td>
</tr>
<tr>
<td><strong>Structural parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$I(0)$ (relative) from $P(r)$</td>
<td>8.363 ± 135</td>
</tr>
<tr>
<td>$R_g$ [Å] from $P(r)$</td>
<td>28.1 ± 0.5</td>
</tr>
<tr>
<td>$I(0)$ [cm⁻¹] from Guinier</td>
<td>8.235 ± 23</td>
</tr>
<tr>
<td>$R_g$ [Å] from Guinier</td>
<td>26.9 ± 1.4</td>
</tr>
<tr>
<td>$D_{\text{max}}$ [Å]</td>
<td>94 ± 10</td>
</tr>
<tr>
<td>Porod volume estimate [Å³]</td>
<td>57,000 ± 200</td>
</tr>
<tr>
<td>Dry volume calculated from sequence [Å³]</td>
<td>56,177</td>
</tr>
</tbody>
</table>
Molecular mass determination

Molecular mass $MM$ [Da] from Porod volume ($V_p$-0.6) | 34,200
Molecular mass $MM$ [Da] from forward scattering | 43,750
Calculated monomeric $MM$ from sequence | 46,427

Software employed

Data processing | PRIMUS
Ab initio analysis | DAMMIN, DAMAVER
Rigid-body modelling | SASREF
3D graphic representations | PYMOL

$D_{max}$: maximal particle dimension; $MM$: molecular mass; $R_g$: radius of gyration; $V_p$: Porod volume. The dry volume was calculated using the Peptide Property Calculator web tool available at http://www.basic.northwestern.edu/biotools/proteincalc.html (Harzaz et al., 1994).

To further illustrate the binding of scFv MEM-85 to CD44 HABD, rigid body modeling of the CD44 complex was performed with the program SASREF (Petoukhov and Svergun, 2005) using the NMR solution structure of CD44 HABD in the PD-conformation without the flexible tail residues 164-178 (PDB code 2IR3; Takeda et al., 2006) and a model of scFv MEM-85 created using the RosettaAntibody modeling server (Sircar et al., 2009). In the first modeling series, six options for paratope (complementarity determining region loops L1, L2, L3, H1, H2, and H3) were specified as possible sites of interaction with the epitope (residues 160-163 of CD44 HABD). In ten modeling runs, only two heavy chain CDR loops (H2 and H3) were successfully selected for interaction with epitope. The next step of the modeling was performed with these paratopes specified in the contact conditions. Models obtained in five runs were almost identical to each other. The final model of CD44 HABD and scFv MEM-85 was superimposed with the ab initio envelope using the SUPCOMB program (Kozin and Svergun, 2001) (Fig. 5A and 5C). The SAXS data and the rigid body model of the CD44 HABD – scFv MEM-85 complex can be accessed in the Small Angle Scattering Biological Databank: http://www.sasbdb.org/data/SASDAG7/.
Figure 5: SAXS scattering data and rigid body model for the CD44 HABD – scFv MEM-85 complex

A: The solution scattering pattern for the CD44 HABD – scFv MEM-85 complex (black) is shown with the fit of the theoretical scattering of the rigid body model of the complex (shown in panel C) to the SAXS experimental data (gray), where $\chi^2 = 1.24$. B: The plot of the pair-distance distribution function $P(r)$ is shown for the CD44 HABD – scFv MEM-85 complex with a maximum particle distance ($D_{\text{max}}$) of 94 Å. C: The rigid body model of the complex of CD44 HABD (green; PDB code 2I83; Takeda et al., 2006) and scFv MEM-85 model (red; Sircar et al., 2009) is shown fitted into the SAXS ab initio envelope. The disordered C-terminal portion of CD44 HABD (residues 164-178) is excluded from the model. The epitope is indicated with an arrow, and residues Glu160, Tyr161, and Thr163 are shown as sticks.

The theoretical scattering curve calculated for the rigid body model of the complex in which the epitope was defined as residues 160-163 of CD44 HABD exhibited a very good fit when superimposed with the ab initio SAXS envelope ($\chi^2 = 1.24$; see Fig. 5 and Fig. A.3 in Appendix A). However, when this modeling was performed with the epitope defined as residues 38-42 of CD44 HABD, which had been previously identified as important for MEM-85 binding (Bajorath et al., 1998), the modeling program SASREF (Petoukhov and Svergun, 2005) could not fit the experimental scattering data with any model with a reasonably good fit ($\chi^2 = 2.01$ for the best fit; see Fig. A.3).
Modeling of the entire structure with the flexible C-terminus of CD44 HABD (residues 164-178) was performed in CORAL (Petoukhov and Svergun, 2005; Petoukhov et al., 2012) with the same contact conditions as for the SASREF (Petoukhov and Svergun, 2005) rigid body modeling. The program was run five times to check the consistency of modeling. The resulting models only differed in the position of the flexible tail part. The rigid part of CD44 HABD in the CORAL models is only slightly rotated compared to the models created without the tail part, but the mutual orientation of the molecules remains the same. This indicates that the position of the flexible tail has no major influence on the overall structure of the complex. A representative model of full-length CD44 HABD in complex with scFv MEM-85 is shown in Fig. A.4A in Appendix A with the fit of the theoretical scattering curve of the model into the experimental SAXS data (Fig. A.4B), where $\chi^2 = 1.29$.

**DISCUSSION**

We constructed a recombinant scFv fragment of the monoclonal antibody MEM-85, which specifically recognizes CD44 HABD and might have therapeutic potential (Godar, S., personal communication). We successfully confirmed scFv MEM-85 binding to native, cell-surface CD44 using flow cytometry and fluorescence microscopy (Fig. 1A and 1B) and to recombinant CD44 HABD using SPR (Fig. 1C). Comparison of the affinities of IgG and scFv MEM-85 for CD44 HABD (Figs. 1C and 4, Table 1) showed that the apparent affinity of IgG MEM-85 for CD44 HABD ($K_D = 4.0 \pm 1.7$ nM) is approximately two orders of magnitude higher than that of scFv MEM-85 ($K_D = 430 \pm 60$ nM). This finding is in agreement with the general trend of antibody scFv fragments having lower antigen affinity than their parental immunoglobulins (Adams and Schier, 1999). The lower binding affinity of scFv MEM-85 to CD44 HABD is predominantly due to the quick dissociation kinetics of scFv MEM-85 from its complex with CD44 HABD. This suggests that avidity of MEM-85 plays a role in binding of bivalent immunoglobulin to CD44 HABD.

The scFv fragment of MEM-85 has relatively high thermal stability, as revealed by ThermoFluor assay. Preparation of a highly stable, homogeneous, and monodisperse complex of scFv MEM-85 with CD44 HABD (Fig. 2 and Fig. A.2) allowed us to employ NMR and SAXS methods to characterize the CD44 HABD epitope recognized by MEM-85 in detail and analyze conformational changes in CD44 HABD induced by antibody binding.

We identified the CD44 HABD residues involved in scFv MEM-85 binding using chemical shift perturbation analysis of the backbone NMR signals ($^5$N, $^1$C', and $^1$H') of $^1$N/$^1$C-
labeled CD44 HABD (Fig. 3). To interpret these perturbations and outline the antibody epitope, we used models of the three-dimensional structures of CD44 HABD and located the affected residues on their surfaces. Both conformational states of CD44 HABD, the O-conformation (Teriete et al., 2004) and the hyaluronate-bound PD-conformation (Takeda et al., 2006), were used for analysis (Fig. 6). We found that the most affected residues are situated in the extension lobe and clustered into a patch formed by the far N-terminus (β0) and, more importantly, by the C-terminal region including β8 (significantly perturbed residues Asp140 and Thr144), β9 (Glu160), the loop between β9 and α3 (Tyr161 and Thr163), and also extending into the unstructured C-terminal tail (Ser171). Another highly perturbed residue is Tyr17, which is located in β6, deeply buried in the core of CD44 HABD. Some of the minor changes occur in the vicinity of the hyaluronate-binding groove (Tyr79, Val87, Arg90, Leu107, and Asp115). For CD44 HABD topology, see Fig. 6A.

**Figure 6:** Mapping of the CD44 HABD epitope recognized by MEM-85.

A: Cartoon representations of CD44 HABD in the O- and PD-conformations are shown with numbered secondary structure elements. The Link module is shown in gray and the extension lobe in black. In all panels, residues with normalized combined chemical shift values higher than 1 (yellow), 2 (salmon), and 3 (green) are highlighted. B: Residues perturbed upon MEM-
85 binding are mapped on the surface of the free CD44 HABD structure (PDB code 1POZ; Teriete et al., 2004). C: Residues perturbed upon MEM-85 binding are mapped on the surface of hyaluronate-bound CD44 HABD (PDB code 2I83; Takeda et al., 2006). Tyr161 is highlighted in sticks, and residues selected for mutational analysis are indicated in panels B and C. The hyaluronate octamer bound to the crystal structure of the murine CD44 HABD – hyaluronan complex (PDB code 2JCQ; Banerji et al., 2007) is superposed onto the structures in panels B and C.

The location of the MEM-85 epitope around residues 160-163 was corroborated by our SAXS model of the CD44 HABD – scFv MEM-85 complex. The rigid body model with this epitope definition provided better fit of the theoretical scattering curve to the SAXS experimental data ($\chi^2 = 1.24$) than the model with the epitope defined as residues 38-42 ($\chi^2 = 2.01$; see Fig. A.3). This is also supported by the results of mutational analysis, in which mutations of Glu160, Tyr161, and Thr163 (but not Asp140 and Thr144) to alanine significantly lowered the affinity of IgG MEM-85 towards CD44 HABD (Fig. 5, Table 1).

The C-terminus of CD44 HABD undergoes major allosteric rearrangements upon hyaluronate binding (compare Fig. 6B and 6C) that mainly involve $\beta_0$, $\beta_8$, $\beta_9$, and $\alpha_3$ from the extension lobe and lead to enhanced flexibility of the CD44 HABD C-terminus (Ogino et al., 2010; Takeda et al., 2006). The most important residue for transition between the O-conformation and the PD-conformation is Tyr161 (Ogino et al., 2010). We showed that Tyr161 has the highest normalized combined chemical shift upon MEM-85 binding, and we therefore suggest that similar conformational changes occur upon MEM-85 binding. Sharpening of the backbone NMR signals of CD44 HABD C-terminal residues upon antibody binding also points to a change toward a more flexible conformation in this region.

The hyaluronate-induced conformational transition into the O-state exposes the CD44 stem region, which is rendered more susceptible to proteolytic cleavage (Takeda et al., 2006), resulting in CD44 shedding from the cell surface. Interestingly, MEM-85 is also able to induce CD44 shedding (Bazil and Horejsi, 1992), and thus we might expect transition to the O-state upon MEM-85 binding. Based on our structural data, we suggest that this conformational change is triggered by direct contact of MEM-85 with the epitope surrounding Tyr161 in the extension lobe.

To explain the observation that MEM-85 interferes with hyaluronate binding (Bajorath et al., 1998; Sandmaier et al., 1998), we analyzed residues with moderate antibody-induced
perturbations. We hypothesized that primary rearrangement of the CD44 HABD C-terminus upon antibody binding is likely accompanied by an allosteric conformational change transmitted into the hyaluronate-binding groove. We observed moderate antibody-induced perturbations within the Link module in the vicinity of the hyaluronate-binding groove (residues 79, 87, 90, 107, and 115) and in the adjacent helix α1 (residues 45, 47, 48, and 53). These perturbations, however, are relatively small to be caused by direct contact with the antigen-binding loops of the antibody. Direct contacts of CD44 HABD residues with the complementarity determining regions of MEM-85, which contain a total of 11 aromatic residues (Kabat et al., 1991), would result in larger changes in the CD44 HABD spectra (Addis et al., 2014; Wilkinson et al., 2009). Additionally, one of the most greatly perturbed residues, Tyr117, lies deep in the core of the HABD domain in β6, which connects the hyaluronate-binding “subdomain” and the C-terminal “subdomain” and might be one of the key residues for transmission of the conformational change into the hyaluronate-binding groove area. Therefore, the mechanism of MEM-85 cross-blocking of hyaluronate binding (Bajorath et al., 1998; Sandmaier et al., 1998) might be of an allosteric, relay-like nature, rather than physical competition for the hyaluronate-binding site.

Residues Lys38, Arg41, and Tyr42, which had been previously identified by mutational analysis to be important for MEM-85 binding (Bajorath et al., 1998), exhibited low to moderate normalized combined chemical shifts (Tyr42 was not assigned). They are located in a patch stretching from the MEM-85 epitope toward the hyaluronate-binding groove, and they are very likely to play a role in the antibody-induced conformational change.

CONCLUSIONS

The work presented here led to an unambiguous determination of the MEM-85 binding epitope in CD44 HABD using a combination of NMR and SAXS techniques and suggested the mechanism of antibody interference with the biological function of CD44. Strikingly, MEM-85 does not directly compete with the binding of hyaluronate, but acts through an allosteric, relay-like mechanism. In particular, it binds to the C-terminal extension lobe region of CD44 HABD and induces a conformational rearrangement which is further propagated into the hyaluronate-binding groove. Antibody MEM-85 triggers a structural transition of CD44 from "ordered" into "partially disordered" state, similar to the one observed upon hyaluronate binding, leading to CD44 shedding from the cell surface.
Acknowledgments

J.S. thanks Petr Těšina for his help with the NMR sequence-specific resonance assignment. This work was supported by Charles University in Prague, project GA UK No. 567012. This work was also supported in part by projects RVO 61388963 and 68378050 awarded by the Academy of Sciences of the Czech Republic and by the Ministry of Education of the Czech Republic, projects LK11205 (program ‘NAVRAT’) and LO1304 (program ‘NPU I’), and project GACR 15-11851S awarded by The Grant Agency of the Czech Republic. M.K. acknowledges support from the European Commission (the 7th Framework Programme) Marie Curie grant IDPhyNMR (contract No 264257).

Conflict of interest disclosure statement and role of the funding source

We wish to confirm that there are no conflicts of interest associated with this publication. The funding sources had no role in the study design, data collection, analysis, and interpretation, in the writing of the report, or in the decision to submit the manuscript for publication.

References


26


Figure A.1: Binding of hyaluronate hexamer to CD44 HABD
Two-dimensional $^{15}\text{N}/^1\text{H}$ HSQC spectra are shown for free $^{15}\text{N}$-CD44 HABD (black) and $^{15}\text{N}$-CD44 HABD in complex with hyaluronate hexamer (red).

Figure A.2: The CD44 HABD – scFv MEM-85 complex is monodisperse.
Mean particle size distribution determined by dynamic light scattering indicates monodispersity of the complex at a concentration of 8.5 mg/ml.
Figure A.3: Rigid body model fit to SAXS data
A: Fit of the theoretical scattering of the rigid body model of the CD44 HABD – scFv MEM-85 complex to SAXS experimental data (black) is shown for epitope residues 160-163 (red), where $\chi^2 = 1.24$, and for epitope residues 38-42 (green), where $\chi^2 = 2.01$. B: The enlarged section of the curves in panel A is shown at low scattering vector values ($s$).

Figure A.4: CD44 HABD and scFv MEM-85 SAXS model with flexible tail
A: The CORAL model of the complex of CD44 HABD (green; PDB code 2I83) (5) and scFv MEM-85 model (6) (red) is shown fitted into the SAXS ab initio envelope. The disordered C-terminal portion of CD44 HABD (residues 164-178) is modeled as a C-alpha trace (shown in bold). B: Fit of the theoretical scattering of the CORAL model of the CD44 HABD – scFv MEM-85 complex to the SAXS experimental data (black), where $\chi^2 = 1.29$ is shown in gray.
8 Discussion

8.1 Structural studies of deoxyribonucleoside regulator form *Bacillus subtilis*

Structural data on several key transcriptional regulators from *B. subtilis* obtained so far indicate that although the effector-binding domains of the repressors represent conserved folds typical for individual protein families, each of them exhibits a unique mode of action determined by specific effector-induced structural changes and oligomerization pattern. Therefore, structural studies of several members of the *B. subtilis* families of transcriptional regulators are necessary for the full understanding of the distinct molecular mechanisms of carbon catabolite repression in Gram-positive bacteria and also gene regulation in general. In this part of the work, we focused on X-ray crystallographic analysis of the C-terminal effector-binding domain of deoxyribonucleoside regulator DeoR from *B. subtilis*.

Crystals suitable for X-ray diffraction analysis were obtained by elaborate optimization of crystallization conditions. First crystallization trials with recombinant C-DeoR expressed in *E.coli* yielded thin needle-shaped crystals in 8 out of 96 screened crystallization conditions. Despite the extensive effort to optimize the initial crystallization conditions, only rod-shaped crystals of poor diffraction quality were obtained. Therefore, we employed the Thermofluor-based optimization of the protein storage buffer to achieve maximal protein stability. This assay revealed protein thermal stabilization by phosphate ions, but these buffers were not taken into account for the crystallization, as binding of phosphate was shown before for other bacterial transcription factors [15, 16]. The Thermofluor assay also identified a combined stabilizing effect of citrate buffer and the pH value of around 7.5. A combination of these two conditions was used in the optimized protein storage buffer (based on tri-sodium citrate pH 7.0) for new crystallization screening trials both in the presence and absence of the effector (dxyribose-5-phosphate). These trials yielded crystals in 8 out of 96 conditions for the effector-bound protein, which were different from those for the original protein storage buffer. The free C-DeoR protein in the optimized storage buffer yielded crystals in 31 out of 96 screened conditions. Most of the crystals were again represented by needle or plate clusters, but new crystal forms were
obtained for both the free and effector-bound C-DeoR, which were three-dimensional monocrystals. These results proved that the Thermofluor-based buffer optimization technique generally used in order to improve protein stability, and therefore its crystallizability can be used to obtain new, better diffracting crystal forms of otherwise well crystallizing protein.

The new crystal forms of C-DeoR were easily optimized to large well-diffracting monocrystals suitable for X-ray diffraction experiments. Three complete data sets were acquired for the free C-DeoR (space group P3$_2$ or P3$_1$, resolution 1.90 Å) and for the effector-bound C-DeoR (space group C222$_1$, resolution 1.60 Å, and space group P222$_1$, resolution 1.65 Å). The sequence similarity of C-DeoR with the homologous transcriptional regulators with known three-dimensional structures suggested that C-DeoR crystal structures can be solved using the molecular replacement method. The putative sugar-binding transcriptional regulator from *Arthrobacter aurescens* (PDB code 3NZE) shows 52% sequence similarity and 30% sequence identity with 259 DeoR residues, the probable sorbitol operon regulator from *Shigella flexneri* (PDB code 3EFB) shows 49% sequence similarity and 31% sequence identity with 265 DeoR residues, and the sorbitol operon regulator SorC from *Klebsiella pneumoniae* (PDB code 2W48) shows 50% sequence similarity and 30% sequence identity for 246 DeoR residues [152].

The next step in this project was the determination of the C-DeoR structures. The structure of effector-bound C-DeoR was determined at resolution of 1.60 Å by molecular replacement using the homologous putative sugar-binding transcriptional regulator from *Arthrobacter aurescens* (PDB code 3NZE) as the search model. The refined effector-bound C-DeoR structure was then used as a molecular replacement search model for the determination of the unliganded C-DeoR structure using a dataset acquired at 1.80 Å resolution.

The C-DeoR structure is composed of a central double-Rossmann-fold core, typical for dinucleotide-binding proteins [153], which is supplemented on both sides with a total of 5 α-helices, resulting in an overall α/β fold of the core subdomain. Additionally, the C-DeoR is equipped with three, mostly helical peripheral subdomains, which take part in effector binding and repressor oligomerization. The effector-binding site, formed by the interface of the core subdomain and two peripheral subdomains, binds the effector molecule deoxyribonuceloside-5-phosphate in a linear form, through a reversible Schiff-base double bond with residue Lys141 located in one of the peripheral subdomains.
The phosphate moiety of the effector molecule interacts with the N-terminal end of one of the Rossmann fold α-helices, an interaction motif typical for dinucleotide-binding Rossmann-fold proteins [153-155].

C-DeoR shows structural homology [156] to numerous prokaryotic transcriptional regulators from the SorC family. C-DeoR is also homologous to central glycolytic genes regulator CggR from *B. subtilis* (PDB code 3BXF [15]), which exhibits a very similar organization of the effector-binding site and also binds its ligands in a very similar pose to that of C-DeoR. Moreover, C-DeoR structure is homologous to many enzymes involved in prokaryotic sugar metabolism, such as 6-phosphogluconolactonase, glucose-6-phosphate-1-dehydrogenase and glucosamine-6-phosphate deaminase. This indicates that one common protein fold can be adapted to performing various enzymatic and regulatory functions. One such enzyme, ribose-5-phosphate isomerase A from *Burkholderia thailandensis* (PDB code 3UW1 [157]) also binds its substrate ribose-5-phosphate in a linear form in an orientation very similar to that of deoxyribose-5-phosphate in C-DeoR. The Schiff base, however, cannot be formed here, as ribose-5-phosphate isomerase A contains a serine residue instead of the lysine in the active site. The Schiff-base adduct of deoxyribose-5-phosphate has been observed also in the structure of deoxyribose-5-phosphate aldolase form *E. coli*, where it represents a reaction intermediate (PDB code 1JCJ [158]). The length of the imine bond in the aldolase structure is 1.29 Å and the Nζ-C-C angle is 125°, which are identical values to those in the C-DeoR structure.

Our effector-bound C-DeoR structure is the only crystal structure of a transcriptional factor capturing a covalently bound effector molecule. The physiological nature of the covalent bond was confirmed by the mass spectrometry analysis of the Schiff-base products formed in solution under physiological concentrations of deoxyribose-5-phosphate [159]. The mass to charge ratio difference between the two detected MALDI-TOF peaks corresponding to free C-DeoR and the C-DeoR Schiff-base adduct was equal to the mass of the cyanoborohydride-reduced deoxyribose-5-phosphate adduct, 3,4-dihydroxy-5-(phosphonooxy)pentyl residue. Moreover, the abundance of the higher molecular mass peak increased with the incubation time. Mutational analysis combined with the Thermofluor-based differential scanning fluorimetry assay [160-162] confirmed the key role of Lys141 in effector binding. The titration of Lys141-mutant proteins K141N, K141Q, K141R, or K141A, respectively, by deoxyribose-5-phosphate showed only small increase in the thermal stability of the mutant proteins when compared...
to the thermal stabilization of the wild-type C-DeoR. This suggests that Lys141 is essential for proper effector binding.

The superposition of the Cα atoms of the free and effector-bound C-DeoR structures identified the location and the extent of the effector-induced structural changes. Major changes occur around the effector-binding site and in the adjacent peripheral subdomains, and are in some extent similar to those observed in CggR [15]. The structural changes in the parts of peripheral subdomains that are involved in the dimeric interface might affect the protein’s oligomeric state as they disrupt the dimer interface and reduce the interface area from 1367.3 Å² formed by 40 interacting residues in the free C-DeoR dimer to 841.3 Å² formed by 25 interacting residues in the effector-bound C-DeoR dimer. The interface is formed by a peripheral subdomain of C-DeoR and also in part by a portion of the double-Rossmann-fold core, in a manner similar to dimers of other members of the SorC family, such as CggR [15], putative sugar-binding transcriptional regulator from A. aurescens (PDB code 3NZE), and SorC [152]. The size-exclusion and dynamic light scattering experiments with the C-DeoR and full-length DeoR both in the presence and absence of deoxyribose-5-phosphate have shown that C-DeoR and full-length DeoR persist in the dimeric and tetrameric forms, respectively, even upon effector-binding. This suggests the presence of two dimerization interfaces in the full-length DeoR: one in the effector-binding domain, and the second in the DNA-binding domain. We suggested a model for the regulation of DeoR function: the tight tetramer of DeoR binds to a palindromic operator sequence [20, 25], but in the presence of deoxyribose-5-phosphate, the effector-induced structural change in the effector-binding domains disrupts the dimerization interfaces in both the effector- and allosterically also in the DNA-binding domains and the less tight tetramer can be easily displaced from the operator by RNA polymerase. The exact structural mechanism, however, still remains to be unraveled by the determination of the full-length DeoR structures, ideally in complex with the operator DNA.

8.2 Epitope mapping for the anti-CD44 monoclonal antibody MEM-85

Monoclonal antibodies are molecules with immense therapeutic and diagnostic potential. The same applies to their single-chain variable fragments (scFv), the smallest antibody fragments that retain the antigen-binding properties while having a relatively small size which makes them excellent molecular biology tools for the characterization of
antigen binding. The only drawback that might compromise their use for some applications is their spontaneous oligomerization. This issue might be, however, resolved by some optimization approaches, as described in the first part of this section on a model case of the scFv fragment of the anti-CD3 antibody MEM-57. The second part of this project is focused on antibody MEM-85, which is one of the therapeutically applicable antibodies targeted against the CD44 antigen. CD44 has been recently identified to be involved in many aspects of cancer and in the last few years, novel CD44-targeted immunotherapeutic approaches keep emerging and seem to be very promising. Antibody MEM-85 has been widely used as a molecular biology tool, but its epitope has not been unambiguously described so far, which was the goal of this project.

We constructed an scFv fragment of antibody MEM-57 for recombinant expression in *E. coli*. The variable domains of the heavy and light chains were connected via a (Gly<sub>4</sub>Ser)<sub>4</sub> linker sequence. A longer version of the conventional (Gly<sub>4</sub>Ser)<sub>3</sub> linker [163] was used here, which should reduce the spontaneous oligomerization of scFv fragments, observed typically for shorter linker sequences [164-168]. Size-exclusion analysis of the purified scFv revealed equilibrium between scFv monomeric and oligomeric forms, where the monomer represented only 65% of the total scFv. This equilibrium re-establish even after the separation of scFv monomers and dimers using ion-exchange chromatography. This oligomerization reduced the effective protein yield, but, more importantly, presented a disadvantage in the use of the scFv fragments for experiments that require high protein homogeneity.

The oligomeric equilibrium is not only a function of the linker sequence length, but it also depends on other aspects such as protein concentration and the pH and ionic strength of the protein storage buffer [169]. In order to resolve the oligomerization issue, we employed the Thermofluor assay to screen for conditions that would preferentially stabilize scFv monomers, and thus reduce the oligomeric heterogeneity of the protein sample. The Thermofluor-based thermal denaturation curve for a homogeneous protein has a typical shape containing only one inflection point which indicates the melting temperature point ($T_{m}$). Therefore, multi-inflection curves were excluded from the $T_{m}$ analysis and the buffer which provided scFv denaturation curve with one inflection point and the highest $T_{m}$ was selected as the optimal protein storage buffer for further experiments. Size-exclusion analysis and dynamic light scattering confirmed scFv homogeneity in the optimized protein storage buffer (the monomer represented 88% of the total protein). Our results
suggested that Thermofluor assay nowadays routinely used for protein stability optimization [162, 170] can be of help also in the efforts to resolve the undesirable oligomerization issue of scFv fragments in general, including for example the anti-CD44 scFv MEM-85 studied in the following part of this project.

For the anti-CD44 antibody MEM-85, we have constructed a recombinant scFv fragment in which the variable domains of the heavy and light chains were connected via a (Gly₄Ser)₄ linker sequence, similarly to MEM-57. The recombinant scFv MEM-85 was able to bind native CD44 on the surface of CD44-positive cells as confirmed by flow cytometry and fluorescence microscopy, and it also bound recombinant hyaluronate-binding domain (HABD) of CD44 as confirmed by surface plasmon resonance (SPR). SPR analysis of the binding affinities of scFv MEM-85 and IgG MEM-85 revealed lower binding affinity of scFv MEM-85 ($K_D = 430 \pm 60$ nM) when compared to IgG MEM-85 ($K_D = 4.0 \pm 1.7$ nM), which is in concordance with a generally observed trend [171]. Preparation of a highly stable and homogeneous scFv sample (employing the Thermofluor assay optimization) and a sample of the CD44 HABD – scFv MEM-85 complex (employing size-exclusion chromatography and dynamic light scattering) was necessary for the further NMR and SAXS experiments aimed at the characterization of MEM-85 epitope and the antibody-induced conformational changes in CD44 HABD. The MEM-85 epitope residues in CD44 HABD were identified by using the chemical shift perturbation method on the backbone NMR signals, acquired in the sets of three-dimensional experiments performed on the free $^{15}$N/$^{13}$C-labeled CD44 HABD and on the complex of the $^{15}$N/$^{13}$C-labeled CD44 HABD with scFv MEM-85. Locating the significantly perturbed residues on the surfaces of the available three-dimensional structures of CD44 HABD (unliganded O-conformation [62] and the hyaluronate-bound PD-conformation [64]) revealed that the epitope is located in the C-terminal part of the HABD extension lobe, clustered around the most perturbed residues 160-163. The role of these residues in antibody binding was confirmed by mutational analysis combined with SPR: mutations of residues 160, 161, and 163 to alamines significantly reduced MEM-85 affinity towards CD44 HABD. These data were also corroborated with the rigid body model of the CD44 HABD – scFv MEM-85 complex, that was created based on the low-resolution structure of the complex obtained by small-angle X-ray scattering (SAXS). The rigid body model with the epitope defined as residues 160-163 provided a perfect fit to
the experimental data, unlike the epitope defined as residues 38-42, based on the previous mutational experiments performed by Bajorath et al. [33].

The major conformational changes, that occur in the C-terminus of CD44 HABD upon hyaluronate binding [64, 70] provide enhanced flexibility to the unstructured HABD tail and induce CD44 proteolytic shedding [64]. The key residue for this rearrangement is Tyr161 [70], which is also the residue exhibiting the highest normalized combined chemical shift in our NMR experiments. This suggests that the binding of MEM-85 triggers a similar conformational change in HABD from the O- to the PD-conformation, which is also in agreement with our observation of sharpening of the NMR signals for the very C-terminal HABD residues in the antibody-bound state. Moreover, MEM-85 has also been shown to induce CD44 shedding similar to that induced by hyaluronate [151].

Taken together, our results propose that MEM-85 recognizes an epitope located in the C-terminal part of CD44 HABD, including residues 160-163, and that the antibody binding induces conformational changes in the C-terminal region by a direct contact with residues around Tyr161. MEM-85 competes with hyaluronate for binding to CD44 HABD [33, 150]. This could be explained by our finding that some of the residues that exhibit moderate antibody-induced chemical shift perturbations (including residues 38-42) are located in the vicinity of the hyaluronate-binding groove. These perturbations are too small to be caused by a direct binding of the antibody [172, 173] but rather suggest a secondary allosteric rearrangement triggered by antibody-induced conformational change in the C-terminal portion of HABD. This conformational change may be transmitted through residues 38-42 that are clustered in a linear patch reaching from residue Tyr161 towards the hyaluronate-binding groove. Therefore, MEM-85 does not directly compete with hyaluronate binding, but acts through an allosteric, relay-like mechanism.
9 Conclusions

This work combines principal techniques that are used to study protein-protein and protein-ligand interactions and the changes in protein conformation. X-ray crystallographic analysis was used to study the structure of deoxyribonucleoside regulator DeoR from *Bacillus subtilis* to understand the binding of an effector molecule. NMR-based mapping of the interaction interface was used to identify the location of the antigen-binding epitope of monoclonal antibody MEM-85 in the hyaluronate-binding domain of human CD44.

The first part of this work describes an important contribution to the field of bacterial transcription regulation. We have crystallized the C-terminal effector-binding domain of deoxyribonucleoside regulator DeoR from *Bacillus subtilis* (C-DeoR). The new crystal form of C-DeoR, optimal for X-ray diffraction experiments, was obtained based on an extensive biophysical characterization of the C-DeoR protein together with protein buffer optimization using Thermofluor assay. The crystal structure of DeoR in complex with the effector molecule deoxyribose-5-phosphate revealed that DeoR binds the effector molecule deoxyribose-5-phosphate covalently via a Schiff-base double bond with a lysine residue in the effector-binding site. The formation of the Schiff-base adduct was also observed in solution at low effector concentrations and the key role of the effector-binding-site lysine residue was confirmed by mutational analysis. These findings make DeoR the first case of transcriptional regulator that binds its small-molecular ligand through a reversible covalent bond. Structural analysis of both C-DeoR in the free form and in complex with the effector revealed effector-induced structural changes disturbing the dimeric interface in C-DeoR. These structural data allowed us to create a model describing the molecular mechanism of the DeoR function as a molecular switch.

In the second part of this work, we have constructed a recombinant single-chain variable fragment (scFv) of monoclonal antibody MEM-85 targeted against CD44, the human cell receptor for hyaluronate, which is involved in cancer. Antibody MEM-85 is of therapeutic interest and needed to be characterized in terms of the antigen-binding epitope location and the effects on CD44 structure and function. The scFv fragment is a suitable small-size molecular tool retaining the antigen-binding function and simplifying the whole antibody molecule. The highly homogeneous and stable sample of recombinant scFv
MEM-85 necessary for further experiments was prepared based on a general strategy of optimization of the biophysical properties of scFv fragments, developed also in the framework of this project. NMR-based epitope mapping was employed to unambiguously locate the epitope of MEM-85 in the C-terminal region of the hyaluronate-binding domain (HABD) of CD44. Moreover, the antibody binding induced allosteric conformational changes which lead to an increased disorder in the C-terminal tail of HABD. The character of this conformational rearrangement was similar to that induced by hyaluronate-binding. These important data shed light on the previous observations that MEM-85 cross-blocks hyaluronate binding to CD44 and also induces CD44 shedding. MEM-85 does not do so by a direct physical competition for the hyaluronate-binding site, but rather allosterically, by binding to a distant epitope located close to the membrane-proximal C-terminus of HABD. Small-angle X-ray scattering and mutational experiments combined with surface plasmon resonance allowed us to create a rigid body model for the complex of scFv MEM-85 with CD44 HABD that further corroborates the proposed MEM-85 mode of action.
10 References


