UNIVERSITAT DE BARCELONA

Structure and function of multicomponent complexes

Arkadiusz Damian Szura

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L’ALIMENTACIÓ

Structure and function
of multicomponent complexes

ARKADIUSZ DAMIAN SZURA, 2023

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

PROGRAMA DE DOCTORAT
BIOTECNOLOGIA

Structure and function of multicomponent complexes

Memòria presentada per Arkadiusz Damian Szura per optar al títol de doctor per la universitat de Barcelona

Dra. Maria Sola Vilarrubias	Arkadiusz Damian Szura	Dra. Josefa Badia Palacin
Director de tesi	Doctorand	Tutora de tesi

ARKADIUSZ DAMIAN SZURA, 2023

Change is the only constant in life

-Heraclitus, a Greek philosopher

Contents

Abstract

The emergence and spread of antibiotic resistance among bacteria has become an undisputed global problem and one of the greatest threats to public health in the 21st century. The widespread, excessive and uncontrolled use of antibiotics, not only for therapeutic purposes but also in agriculture and animal husbandry, has resulted in a steady and rapid increase in the number of strains resistant to the drugs used.

In all living cells, ribonucleotide reductases (RNRs) are essential enzymes that constitute the only known *de novo* pathway of deoxyribonucleotide biosynthesis (dNTP biosynthesis) via the catalyzed reduction of ribonucleoside triphosphates (NTPs, such as ATP, CTP, GTP, and TTP) using radical chemistry, thereby forming the fundamental building blocks for DNA synthesis and repair. The transcriptional repressor NrdR controls the expression of RNR genes in most bacteria and in some archaea. Importantly, NrdR is missing in eukaryotes, and as it is found in antibiotic resistant pathogens such as *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*, it can be considered as a biomedical target. NrdR inhibits transcription of RNRs genes by binding its Zn-finger domain (ZFD) to a palindromic repeat of 16bp DNA that conform the so-called NrdR-boxes, which are upstream of RNRs promoter regions. An ATP-cone domain (ACD) present in NrdR, sensitive to the changes in concentration of (-deoxy) ribonucleoside triphosphates (dNTPs and NTPs), allosterically regulates the Zn-finger activity of NrdR. Several hypotheses on *RNR*s genes regulation by NrdR have been proposed, and in May 2022 a combination of biochemical and cryo-EM structural studies suggested mechanism of action for *Streptomyces coelicolor* (sc-) NrdR. Such a mechanism involves an ATP-loaded dodecamer, which cannot bind to DNA, a dATP/ATP-loaded octamer, and a dATP/ATP-loaded tetramer bound to the sc-*nrdRJ* promoter, which represses transcription of the RNR operon.

In this doctoral thesis, the crystal structure of *Escherichia coli* NrdR revealed key interactions whose mutations altered the multimerization. To test the functional roles played by the different residues, *in vitro* assays were carried out that showed in solution the WT NrdR dimer instability in the absence of nucleotides, and elution of different assemblies in the presence of AMP, ADP, ATP, and dATP. The same studies performed with fusion NrdR and designed mutants in which the multimerization interactions were altered, revealed the importance of residues Glu36, Glu42, Tyr131 at the ZFD, and of segment aa 132-149 from the ACD in NrdR oligomerization. The highest impact was noted for mutation Glu42Ala and the deletion of aa 132-149 segment. Thus, both the ZFD and the ACD are fundamental for oligomerization and essential for the protein function. *In vivo* studies performed with single-site mutants indicated that mutation E42A at the ZFD completely abolished NrdR ability to repress transcription of RNRs, while mutation to Ala of multimerization-sensitive residues Glu36 at the ZFD and Tyr131 at the ACD did not cause a decrease of the repression level. Therefore, amino acid Glu42 is pivotal for the repressive function of RNR. The abundance of the NrdR protein in bacteria and extrapolation of the results obtained for *E. coli* and *S. coelicolor* NrdR points to an ATP/dATP-orchestrated mechanism, in which the type of NrdR multimers change and coordinates the repression activity of the RNR operon.

Introduction

I1 Bacterial resistance to antibiotics

The emergence and spread of antibiotic resistance among bacteria has become an undisputed global problem and one of the greatest threats to public health in the 21st century [1]. The widespread, excessive and uncontrolled use of antibiotics, not only for therapeutic purposes but also in agriculture and animal husbandry, has resulted in a steady and rapid increase in the number of strains resistant to the drugs used [2].

There are two types of antibiotic and chemotherapeutic resistance: natural and acquired. Natural resistance, also known as innate resistance, is constant for a family, genus or species and is due to a lack of specific receptors or low affinity for the antibiotic, inability to penetrate and transport the drug through the cell wall and the production of enzymes that inactivate antibiotics [3]. Acquired resistance involves changes in bacterial genomes due to mutations and gene transfer between bacteria. Bacterial resistance to antibiotics stems from point mutations in genes targeted by antibiotics or in regulatory genes, such as resistance to rifamycin or fluoroquinolones [4, 5]. Many antibiotic resistance genes in *Proteobacteria*, including pathogenic bacteria, are derived from *Actinobacteria*, mainly *Streptomyces sp*., which are producers of antibiotics. The chloramphenicol resistance gene *cmx* identified in the genomes of the clinical strains *Pseudomonas aeruginosa*, *Klebsiella oxytoca* and *Enterobacter asburiae*, and the linezolid resistance gene *lmrA*, found in strains isolated from farm animals, have a high sequence similarity to *Streptomyces sp*. [6].

Most environments and organisms are assumed to play an important role in the flow of resistant bacteria and antibiotic resistance genes (ARGs). This flow is bidirectional, leading to the introduction of natural resistance genes occurring in environmental bacteria into bacteria pathogenic to humans and animals, and vice versa, important resistance genes are transferred to environmental microorganisms [7]. Transmission of antibiotic resistance genes among bacteria occurs via horizontal gene transfer (HGT). Evidence of the significant effect of lateral transfer of natural resistance genes from the environment to pathogens is the dynamic spread of the *blaCTX-M* gene, which determines resistance to a wide range of β-lactams, derived from the genome of environmental strains of *Kluyvera sp*. [8]. The *qnrA* gene associated with plasmid carrying resistance to fluoroquinolones is derived from environmental strains of *Shewanella algae* [9]. Vancomycin resistance genes, currently present in the strains

Enterococcus sp. and *Staphylococcus sp*., are found in the genomes of soil bacteria, including *C. Paenibacillus popilliae* **[10]**.

In 2017, the WHO published for the first time a list of resistant bacterial pathogens posing the greatest risk to human health and requiring prompt research into new drugs. "Critical" pathogens with particularly dangerous resistance mechanisms include carbapenem-resistant strains of *Acinetobacter baumannii*, *P. aeruginosa* and *Enterobacteriaceae*, which produce extended-spectrum of β-lactamases (ESBL). Vancomycin-resistant strains of *Enterococcus faecium* and *Staphylococcus aureus*, clarithromycin-resistant strains of *Helicobacter pylori*, fluoroquinolone-resistant strains of Campylobacter and Salmonella are also of high concern **[11]**. In order to combat the spread of resistance genes within bacterial species, pathways of bacterial migration with relevant mechanisms of resistance in the environment need to be identified. Such studies include not only the presence of resistant bacteria in human and animal ecosystems, but also antibiotic resistance and chemotherapeutic genes forming so-called resistomes.

I1.1 Bacterial resistome

The term resistome was introduced to understand the emergence and surprisingly rapid spread of bacterial resistance to antibiotics and chemotherapeutics in strains relevant for human health. Resistome refers to the collection of all antibiotic resistance genes found in the genomes of pathogenic bacteria, non-pathogenic antibiotic manufacturers, cryptic resistance genes and resistance gene precursors. The latter are genes coding for proteins with moderate antibiotic activity or with little affinity for antibiotics, which, if appropriate, can become effective resistance genes **[12]**. There are over 23,000 antibiotic resistance genes known, which have been classified into almost 400 genera. They confer resistance against more than 240 antibiotics, including aminoglycosides, β-lactams, macrolides, lincosamides, streptogramins, tetracyclines and glycopeptides, and transport pumps [13]. However, this is a fraction of the actual number of genes, not taking fully into account the soil bacterial resistome. The first studies of resistomes were conducted in soil microorganisms with evidence of precursors of the genes for resistance to aminoglycosides and vancomycin, currently present in pathogenic bacteria **[14]**. In the aquatic environment, as in the soil, there are microorganisms characterized by natural resistance to antibiotics. In addition, the impact of anthropogenic pollutants increases the incidence and spread of acquired resistance. Many aquatic environments such as surface water, river water, groundwater, marine water, aquaculture, wastewater and even water treatment systems contain bacteria with β-lactam resistance genes, aminoglycosides, glycopeptides, fluoroquinolones, tetracyclines, macrolides, lincosamides, streptogramins, rifamycin, sulphonamides and chloramphenicol. The presence of resistant bacteria and ARG in aquatic environments is a persistent organic pollutant and, like heavy metals, can endanger public health [15]. Selection pressure due to the use of antibiotics in agriculture and fish farming, increases the prevalence of resistance genes, also in pathogens such as *Escherichia coli* and *Salmonella sp*. in the aquatic environment and contributes to the circulation of resistance genes to and from the natural environment to the clinical environment [16].

I1.2 Bacterial resistance mechanisms to β-lactam antibiotics

β-lactam antibiotics are the most commonly used to treat many bacterial infections. The bactericidal mechanism of action of β-lactams is inhibition of penicillin-binding proteins (PBPs), which are involved in bacterial wall biosynthesis As a result, bacteria with damaged cell wall structure undergo lysis [17]. There are four main mechanisms of bacterial resistance to β-lactam antibiotics.

The first is found in both Gram-positive and Gram-negative bacteria and is associated with PBP proteins. The mechanism of resistance is the production of a new protein PBP2' or PBP2a, which has a reduced affinity for β-lactam antibiotics [18]. The second mechanism characteristic of Gram-negative bacteria is due to a reduction in cell membrane permeability of the antibiotic due to a reduction in the number of pore proteins present in the outer membrane. Consequently, the antibiotic is transported to the periplasmic space of the cell at lower concentrations. Another common, third mechanism of β-lactam resistance in gram-negative bacteria is the production of efflux pumps, which are involved in actively pumping out the antibiotic from the bacterial cell. This is a common mechanism of resistance in clinical strains and is also the least studied [19].

The main and most disturbing, fourth mechanism of resistance to β lactam antibiotics is their enzymatic hydrolysis by the production of β-lactamases by Gram-negative and Gram-positive bacteria. This mechanism has been known since 1940 when the first penicillin-inactivating β-lactam was discovered. The hydrolytic activity of β-lactamases depends primarily on the type of enzyme, the number of substitutions in the amino acid chain that increase the substrate spectrum, as well as the level of expression of genes coding for β-lactamase [19, 20].

I1.2.1 β-lactamase classification systems

The production of extended-spectrum β-lactamases (ESBL) is currently one of the most clinically and epidemiologically relevant mechanisms of resistance in *Enterobacteriacae* and, despite the emergence of other mechanisms, still remains the main source of resistance of *Enteromycetes* to β-lactam antibiotics [20, 21].

The β-lactamase classification uses two systems: a functional system developed by Bush and Jacoby, based on a comparison of the rate of hydrolysis of various types of β-lactamases and the sensitivity of β-lactamases to inhibitors, and a structural system, introduced by Ambler [17, 22, 23]. According to functional classification, the enzymes are divided into three groups: chromosomally coded cephalosporinases, serine β-lactamases divided into 12 subgroups, and metallic-β-lactamases. The structural classification is based on the amino acid sequence and divides β-lactamases into four classes: A, C and D which use serine for hydrolysis of β-lactams, whereas metalloenzymes of class B require the presence of a divalent zinc ion [17, 22, 23].

All β-lactamases described so far differ in their amino acid sequence and the range of hydrolysed substrates [24]. The genes encoding ESBL are often located within mobile genetic elements (MGE), most commonly within the structure of conjugative plasmids, transposons, and integron cartridges, allowing higher levels of the enzyme expression, as well as spread, between strains of one or different bacterial species [25, 26, 27]. ESBL-producing bacteria are found worldwide and are mainly associated with nosocomial infections [20]. ESBL enzymes are found in many strains of the *Enterobacteriaceae* family, mainly *E. coli* and *K. pneumoniae*, but also in *Salmonella* and *Shigella* and many non-fermenting rods, such as the *Pseudomonodaceae* family [21, 28]. All ESBL enzymes have a serine residue at their active site and are classified as Ambler Class A and D [17, 28].

I2 Ribonucleotide Reductases (RNRs) - the key enzymes in bacterial life

In all living cells, ribonucleotide reductases (RNRs) are essential enzymes that constitute the only known *de novo* pathway of deoxyribonucleotide biosynthesis (dNTP biosynthesis), thereby forming the fundamental building blocks for DNA synthesis and repair [29]. The ribonucleotide reductases are metalloproteins that catalyze the reduction of nucleoside triphosphates (NTPs, such as ATP, CTP, GTP, and TTP) using radical chemistry. As a result of this enzymatic reaction taking place in the active site of the protein, the hydroxyl group bound to the 2' ribose carbon of a ribonucleoside diphosphate or triphosphate (NDP or NTP, respectively) is reduced to hydrogen. In this manner, the corresponding deoxyribonucleoside diphosphate or triphosphate (dNDP or dNTP, respectively) is formed (Figure I1). RNRs not only contain the active site required to form and transfer the radical, but they additionally contain an 'activity site' (not all species contain it) and a 'specificity site' that are both involved in the allosteric control of the ribonucleotide reduction process, being highly conserved among RNRs [29, 30].

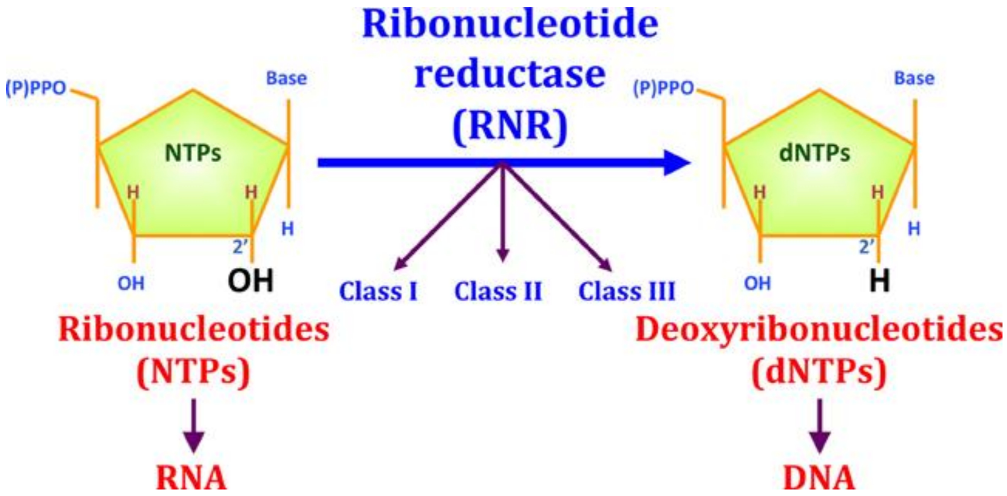


Figure I1. Schematic representation of the simplified reaction of ribonucleotides reduction to deoxyribonucleotides catalyzed by all three classes of ribonucleotide reductases (RNR). Adapted from (29).

I2.1 Classification of RNRs

RNRs are classified in three classes, namely classes I, II, and III, based on the mechanism employed to generate radicals, the type of metallo-cofactor required, the type of electron donor used, the differences in the structure of the protein complex, and the dependency on oxygen. Their main differences, explained in detail below, are summarized in Table I1. RNRs can be (but not always) constituted by two types of subunits, α and β. In class I, subunit α is the catalytic subunit whereas subunit β generates the required free radicals from Phe, Tyr or Cys; in class II, only subunit α exists and the radicals are formed from adenosylcobalamin (AdB); in class III likewise in class I, subunit α is the catalytic subunit, whereas subunit β is responsible for the radical generation. Despite the differences among the three classes of enzymes, their catalytic subunit shows high overall structural similarity and a highly conserved active site arrangement. Additionally, the two allosteric centers in RNRs that modulate specificity and activity are also highly conserved [29, 30, 31]. In bacteria, a single species can encode all of the known types of RNRs [29]. In contrast, in eukaryotes and eukaryotic viruses, only ribonucleotide reductases of Class Ia are found [28] (Table I1).

	Class Ia	**Class Ib**	**Class Ic**	**Class II**	**Class III**
Respiration	Aerobic	Aerobic	Aerobic	Aerobic/ Anaerobic	Anaerobic
Structural organization	$\alpha_2\beta_2/\alpha_6\beta_6$	$\alpha_2\beta_2$	$\alpha_2\beta_2$	$\alpha(\alpha_2)$	$\alpha_2 + \beta_2$
Encoding genes	*nrdAB*	*NrdHIEF*	*nrdAB*	*nrdJ*	*nrdDG*
Radical	Tyr… Cys	Tyr… Cys	Phe… Cys	AdB12… Cys	AdoMet…
Metallocofactor	Fe^{III}-O-Fe^{III}	Mn^{III}-O-Mn^{III} Fe^{III}-O-Fe^{III}	Mn^{IV}-O-Fe^{III}	Co	Fe^{II}-S^{II}
Substrate	NDPs	NDPs	NDPs	NDPs/NTPs	NTPs
Specificity site	Present	Present	Present	Present	Present

Activity site	Present	Absent	Present	Present/ Absent	Present
Inhibition of ATP	Yes	No	Yes	Yes/No	Yes
Distribution	Eukaryotes	Eubacteria		Eubacteria	
	Eubacteria			Archaea	
	Archaea			Bacteriophages	
	Bacteriophages				
	Virus				

Table I1. Overview of the characteristics of RNR classes. Adapted from (29). $\alpha_2\beta_2/\alpha_6\beta_6$ indicates dimers or hexamers of α (α_2 or α_6, respectively) and β (β_2 or β_6, respectively), $\alpha(\alpha_2)$ indicates monomer or dimer and $\alpha_2 + \beta_2$ indicates independent homodimers.

I2.1.1 Class I RNRs

RNRs class I shows three subclasses, Ia, Ib, and Ic, which can be distinguished by the presence or absence of the so-called ‘overall allosteric site’ in which the radical is produced, the specific metal cofactor they use, and the specific amino-acid residue where the radical is generated (Table I1). In eukaryotic organisms, only class Ia RNRs are present. In contrast, eubacteria and archaea utilize the three RNR Ia, Ib and Ic subclasses.

In bacteria, Class Ia ribonucleotide reductases are constituted by subunit α, which is encoded by *nrdA*, and subunit β encoded by *nrdB*. These catalytic subunits contain two allosteric regulatory sites, the activity site and the specificity site, which govern the ability to produce radicals, via a di-iron center (Fe^{III}-O-Fe^{III}). The typical quaternary structure of class Ia RNR is a dimer of homodimers ($\alpha_2\beta_2$), although higher oligomerization states are possible ($\alpha_6\beta_6$).

Class Ib RNRs are restricted to bacteria and bacteriophages [29]. Subunit α is encoded by *nrdE* and subunit β by *nrdF*. Subunit β of these enzymes may use (Fe^{III}-O-Fe^{III}) or the iron-free di-manganese (Mn^{III}-O-Mn^{III}) center. In *E. coli*, Mn^{III}-O-Mn^{III} is an alternative active ribonucleotide reductase that, in aerobic conditions, can be used under iron deprivation circumstances. The catalytic subunit α of RNRs in class Ib contain both the active site and the specificity allosteric site, however they do not have the overall activity site.

Class Ic RNRs are also limited only to bacteria [29]. The α and β subunits present in this subclass are codified by genes *nrdA* and *nrdB*, respectively, the same as in class Ia RNR. The RNRs in Class Ic contain a very specific manganese iron metal center (Mn^{IV}-O-Fe^{III}) inside *NrdB*, which is used to produce the radical required for the reduction reaction of nucleotides [32].

I2.1.2 Class II and III RNRs

Class II RNRs are found exclusively in bacteria, archaea, and some bacteriophages, but they were not observed in other organisms. The members of this class can take both ribonucleoside diphosphates and triphosphates as a substrate for reduction. These enzymes are formed by a single subunit α, considerably similar to the catalytic subunit of class I RNRs, and is expressed by the gene *nrdJ*. The generation of radicals in Class II RNRs is not mediated by a dedicated subunit of the enzyme, but rather by a direct interaction between 5'deoxyadenosylcobalamin, a modified form of vitamin B12 that interacts directly with the active site of the enzyme. Both vitamin B12 and 5’deoxyadenosylcobalamin that contain the cobalt atom need to be supplied. The cobalt atom is used as the metal center to generate a 5'-deoxyadenosyl radical. Allosteric sites are also present in the subunit α of class II RNRs, NrdJ protein. In spite of the fact that some particular class II RNRs contain the overall activity site such as the one in *Thermoplasma acidophilum*, the majority of NrdJ proteins do not contain this site and only have the specificity site [29, 30, 33].

The Class III RNRs enzymes can be found in both aerobic and anaerobic archaea, and in some bacteriophages that reduce ribonucleoside diphosphates or triphosphates. This class of RNRs is composed of two independent proteins, roughly equivalent to the two subunits seen in Class I RNRs: the catalytic protein α, encoded by the *nrdD* gene, and the activase protein β, encoded by *nrdG*. A pair of these proteins can be found in vivo as two independent homodimers ($\alpha_2 + \beta_2$). In this class of ribonucleotide reductases, the NrdG activase protein β is responsible for the radical generation. There are two allosteric sites in NrdD catalytic protein α, the activity and the specificity sites, as well as the active site that catalyzes the NTP’s reduction reaction through a mechanism that is analogous to those outlined above [29, 30, 34].

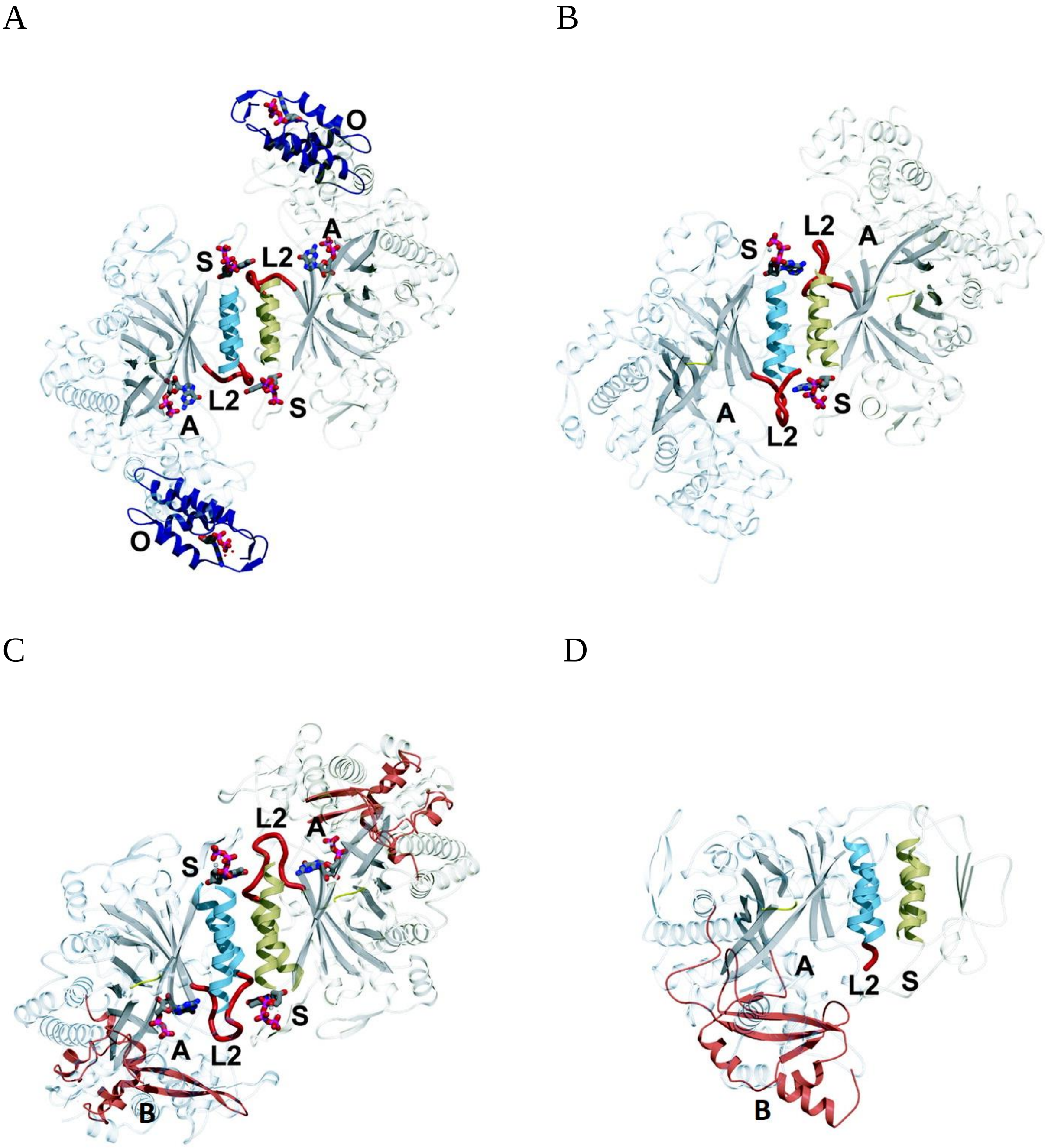
I2.2 Structures of RNRs

The crystallographic structures of all three RNR classes revealed a highly resembled α/β topology of the catalytic subunits, and structural similarity between class I and II RNRs mirrored by the high conservation of residues in the active-site, thus reflected by a similar catalytic mechanism. On the other hand, there is no significant sequence homology of the core α/β-barrel of the class I and II RNRs and of the class III RNR catalytic subunit [35, 36, 37].

The *E. coli* class I RNR contains an N-terminal ATP-cone domain formed by a small α-helical bundle carrying the binding site for effectors controlling the overall activity of RNR (Figure I2A). The C-terminal region is not visible in the structure, because of the flexibility of this region, required for electron transfer, from two cysteine residues of thioredoxin or glutaredoxin to the active site by reversible disulphide formation [38]. The *S. typhimurium* class Ib RNR consists of the active-site and a specificity site with a very close resemblance to *E. coli* class Ia RNR, but lacks the overall activity site in the N-terminal region [39] (Figure I2B).

The overall fold and the key active-site residues of the dimeric *T. maritima* and monomeric *L. leichmannii* class II RNRs are conserved with the class I RNR. Specificity site in *T. maritima* class II RNR is located in the C-terminal region, in which occurs dimerization essential for effector-binding, whereas in case of the monomeric *L. leichmannii* class II RNR specificity site of the missing subunit partner for effector-binding mimics a small structural extension in the C-terminal region of the protein (Figure I2C, D). The mock domain present in class II RNR monomer retains an allosteric binding site similar to the one of the class II RNR dimer. The structures of the *T. maritima* and *L. leichmannii* class II RNRs complexes with a coenzyme B12 analogues revealed a specific B12-binding region, with major fold of β-hairpin, different from one found in other B12-bound enzymes [36, 37].

The *bacteriophage T4* class III RNR catalytic α-subunit has a fold similar to the class I and II RNRs, with the active site occupying an identical position, although there is no significant sequence homology. The specificity sites for effector-binding composed of residues from both dimer subunits are localized on the dimer interface and are in a similar position on the catalytic α/β-barrel as in the other classes of RNRs [40] (Figure I2E).



E

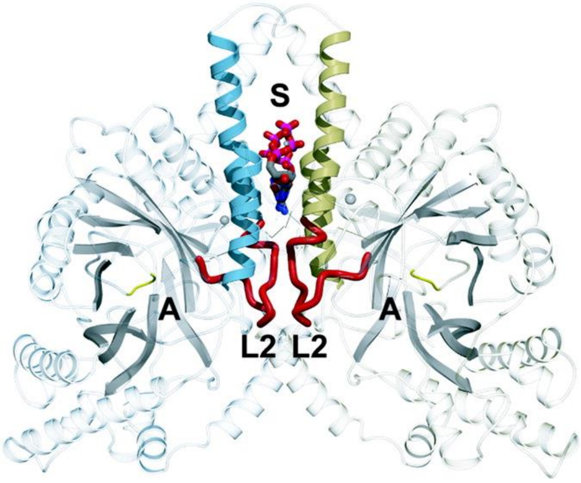


Figure I2. Overall structural organization of all three classes of RNRs. Three-dimensional structures of **(A)** *E. coli* RNR class Ia **(B)** *S. typhimurium* RNR class Ib **(C)** *T. maritima* RNR class II **(D)** *L. leichmannii* RNR class II **(E)** *bacteriophage T4* RNR class III are shown. Structural features essential for RNR regulation are distinguished and indicated as followed: O, overall activity allosteric site, the ATP-cone domain (blue) of class I; A, active site (dark grey); L2, loop 2 (red); S, specificity allosteric site (light blue, light green), B, B12-binding region of class II (wine red). Adapted from (35).

I2.3 Allosteric regulation mechanism of RNRs

The reduction of ribonucleotides is an essential activity, as an unbalanced dNTP pool may result in an increase of the DNA mutation rates and overall loss of DNA replication fidelity. Therefore, a tight regulation of the expression of different RNRs classes is required during DNA synthesis and repair [41].

In addition to a tight regulation of their gene expression (see I2.3), a second mechanism of RNR regulation is by binding of regulatory ligands to the allosteric sites of the protein. All four different NDPs and NTPs are reduced to their corresponding dNDPs and dNTPs by a single active site of RNRs. As a result of the binding of ATP or dATP to an allosteric site, CDP and UDP are enzymatically converted to their reduced forms. Alternatively, when the enzyme binds dGTP or dTTP, it reduces ADP and GDP, respectively [29, 35, 37] (Figure I3).

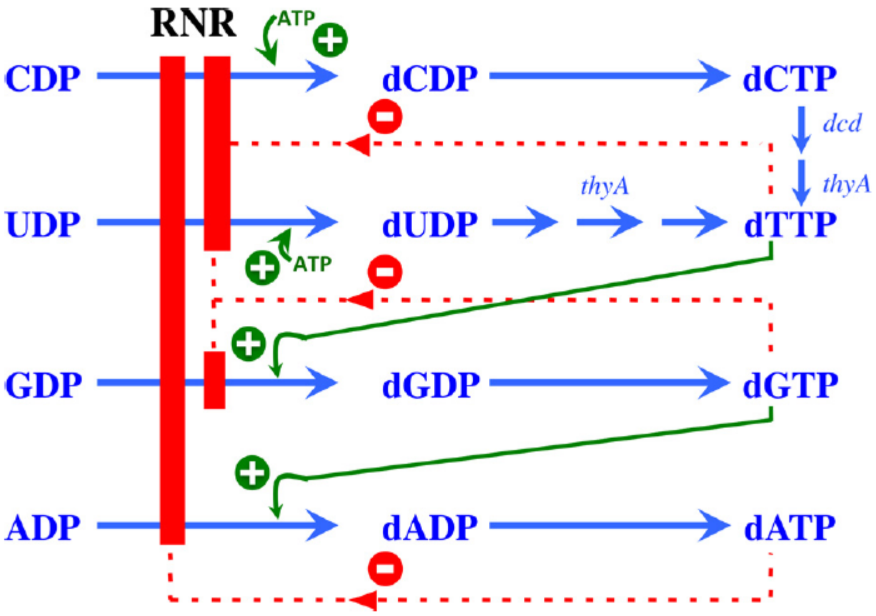


Figure I3. Allosteric regulation of class Ia RNR enzymatic activity. RNR activation induced by ATP binding at the substrate specificity site promotes the reduction of CDP to dCDP and UDP to dUDP. dTTP once formed from precursor dUDP induces the reduction of GDP to dGDP, which in turn promotes the reduction of ADP to dADP. A high dATP concentration inhibits the overall activity of RNR by binding to the allosteric activity site. Stimulation of RNR activity is indicated by green plus symbols, whereas inhibition is indicated by red bars. Adapted from (29).

The allosteric site of RNRs are at the cone domain, and when ATP or dATP bind to it, they trigger a change in the quaternary structure of the protein, thereby altering the enzyme overall activity but in opposing ways [42]. In the case of *Escherichia coli* class Ia RNR, this regulation is achieved by a change in the oligomerization state of the α and β subunits of the enzyme. As soon as ATP is bound to the allosteric site, a $\alpha_2\beta_2$ tetrameric complex is favored, thus a transient thiol radical is generated on Cys439 from subunit β, which promotes catalysis and regulates the production of dNTPs, which occurs in the active site of RNR. However, when the concentration of dATP surpasses a threshold, dATP binds to the allosteric site instead of ATP and, in consequence, an allosterically-mediated inactivation of the enzyme occurs, by inducing the formation of a dATP-bound $\alpha_4\beta_4$ RNR octameric ring, in which dNTP production is inhibited [43, 44]. Thus, ATP induces dATP synthesis, whereas dATP inhibits it. In *Pseudomonas aeruginosa* class Ia RNR, the N-terminal ATP-cone domain was described to play a key role in allosteric regulation. In addition, an internal second ATP-cone domain present in the enzyme is involved in the stabilization of the protein quaternary structure [45, 46].

I2.4 Transcriptional regulation mechanism of RNRs

As opposed to allosteric regulation that maintains the equilibrium between -oxy and deoxyribonucleotides, transcriptional regulation has the function of adjusting the expression of the ribonucleotide reductases according to amount of deoxyribonucleotides required at a specific moment, which for example are high when the DNA replication or repair occurs [29]. In bacteria, DNA replication is initiated primarily by the protein DnaA, a transcription factor involved in the regulation of the operon *nrdAB* of class Ia RNRs. In *E. coli*, DnaA-ATP binds to two DnaA DNA boxes at the *nrdAB* operon to activate the expression of the genes [47, 48]. Fis is a bacterial protein associated to the bacterial nucleoid, and is involved in coupling changes of DNA replication to the expression of *nrdAB* operon, thereby adjusting DNA replication initiation with the levels of deoxynucleotides. Fis, through binding up to five different binding sites, it activates the transcription of *nrdAB* genes [49]. Additionally, transcription of class Ia RNRs is positively regulated by IciA, an inhibitor that suppresses the initiation of the replication of DNA [50]. Finally, H-NS, a nucleoid-associated protein known as a global transcriptional repressor of environmentally-regulated genes, was found to negatively regulate the transcription of the class Ia RNR [51].

In those species in which class Ia and Ib RNRs coexist together, the transcriptional regulation of class Ib RNR is adjusted according to iron deprivation conditions. In *E.coli,* in the presence of iron, the class Ib RNR operon *nrdHIEF* is repressed by the Fur (ferric uptake regulator) protein **[52]**.

A recent study suggested that the transcriptional regulation of class II RNR from *P. aeruginosa* might be regulated by the AlgZR system involved in alginate production and inducer of mucoid biofilms. However, no specific transcription factors have been identified **[53]**. It was found that class Ia RNR is expressed during laboratory growth conditions, while class II and III RNR are expressed during infection **[29, 54]**.

The presence or absence of oxygen regulates the activity of oxygen-sensitive enzymes, which belong to class III RNRs. Under anaerobic conditions, a significant increase of the *E. coli nrdDG* transcription occurs until a point at which the regulator of the anaerobic metabolism, Fnr, starts to repress through the binding into two Fnr DNA boxes **[55]**. Described diverse transcription factors, responsible for regulating the RNRs expression during bacterial growth are listed in Table I2.

RNR class	Action	Transcription factor	Microorganism
Ia	Activator	DnaA / Fis / IciA	*E. coli*
Ia	Repressor	H-NS	*E. coli*
Ib	Activator	Fur	*E. coli*
II	Activator / Repressor	AlgZR system / non identified	*P. aeruginosa*
III	Repressor	Fnr	*E. coli*

Table I2. Transcriptional factors implicated in the regulation of RNRs. Adapted from (47-55).

The complexity of the transcriptional regulation increases in organisms that encodes more than one RNR class. This is because the expression of the different RNR genes in the organism requires a properly balanced amount of dNTPs **[29]**.

I3 NrdR – transcriptional repressor of RNRs

I3.1 Discovery of NrdR, functional predictions

The NrdR protein family is a highly conserved class of proteins that inhibit transcription for all prokaryotic RNR protein classes **[29]**. A novel gene named NrdR was discovered directly upstream of the *nrdJ* gene in *Streptomyces spp*. in 2002 by Ilya Borovok et al*.* **[56]**. A further study demonstrated that *nrdR* cotranscribes with RNR class II genes, forming an *nrdRJ* operon. In addition, it also represses the RNR class Ia operon *nrdABS* of *Streptomyces coelicolor* **[57]**. In parallel, by using bioinformatic methods, Rodionov et al. (2005) discovered that one orthologous protein, COG1327, was present only in species that contained repeated palindromic DNA sequences in their *nrd* operons. COG1327 corresponded to a family of proteins, and the protein was named NrdR and the correlated cis-elements NrdR-boxes. Hypothetically, NrdR was predicted as a transcriptional regulator of RNR expression. Studies performed by Rodionov et al. revealed that all currently known *Eubacteria* encode an *nrdR* gene except *Rickettsia*, *Helicobacter*, and *Desulfovibrio*, but also that it is completely absent in ε *Proteobacteria* **[58, 59]**. Interestingly, all species encoding NrdR contain NrdR-boxes in all RNR operons, regardless of the combination of encoded RNRs.

The NrdR-binding sites, also known as NrdR-boxes, are palindromic repeat sequences with 16 base pairs, roughly aligned with the consensus sequence acaCwAtATaTwGtgt. Taxonomic groups differ slightly in the consensus sequence of the NrdR-box, but NrdR-boxes are almost always encountered in tandem. The distance between the boxes is itself significant as the difference between their centers equals an integer number of DNA helix turns (21 bp for 2 turns, 31-32 bp for 3 turns or 41-42 bp for 3 turns). Distance between NrdR-boxes suggests protein-protein interactions between the NrdR molecules indicating the presence of oligomers bound to both boxes [29, 56, 58]. In the *nrd* promoters, the NrdR-boxes always overlap with the consensus sequences of the basal promoter, suggesting specific repression activity of NrdR [56, 57, 60]. In the *E. coli* genome, *nrd* promoters were bound *in vitro* by specific NrdR-boxes and NrdR was demonstrated to repress all three encoded reductases, class Ia, Ib and III RNRs [60].

I3.2 Structural characterization and functional analysis of NrdR

NrdR proteins consist of 140-220 aminoacids and contain two domains, an N-terminal DNA binding rubredoxin-like Zn-ribbon module formed by three β-strands that form a β-hairpin, and a C-terminal nucleotide binding ATP-cone structurally organized into four helices covered by three beta-strands. The Zn ribbon is characterized by the presence of the four cysteine residues Cys3, Cys6, Cys31 and Cys34 that coordinate the zinc ion, and four Arginines 26-29 (a highly conserved motif), participating in DNA binding, being highly conserved among all known NrdR sequences. ATP-cone domain contains a hydrophobic pocket featured for purine base binding and also a highly conserved sequence of residues participating in nucleotide binding ValXLysArgAspGly [56, 57, 63, 65].

Functional analysis aimed at the analysis of the transcriptional regulation by the zinc finger domain, allosteric regulation via nucleotide binding into ATP-cone domain, and their relationship with protein oligomerization of *S. coelicolor* NrdR were described by Grinberg I. and collaborators in 2006 and 2009, respectively [56, 61]. For the first time, the tridimensional structure of the *S. coelicolor* NrdR was determined by Grinberg I. and collaborators in 2022, revealing structure-function relationships of different *S. coelicolor* NrdR oligomers. Therefore, the transcriptional regulation mechanism of RNRs by *S. coelicolor* NrdR was explained [62]. Functional studies assumed NrdR to exist in different oligomeric forms [56, 60, 63]. The *S. coelicolor* NrdR oligomerization mechanism and its implications described by Grinberg I. in 2022 using a combination of biochemical and cryo-EM studies revealed three NrdR functional multimers at atomic resolution including a dodecamer, an octamer and a tetramer. Initially, in the presence of ATP NrdR forms the dodecamer, which has no ability to bind to DNA. Through the increase of the dATP concentration, one of the two ATP molecules present in each of the dodecamer ATP-cones is replaced by dATP, so that each ATP-cone is loaded with one molecule of ATP and dATP per NrdR monomer. This causes conformational changes that induce the octamer. A tetramer derived from this octamer then binds to the nrdRJ promoter and represses transcription of RNR (Figure I4) [62].

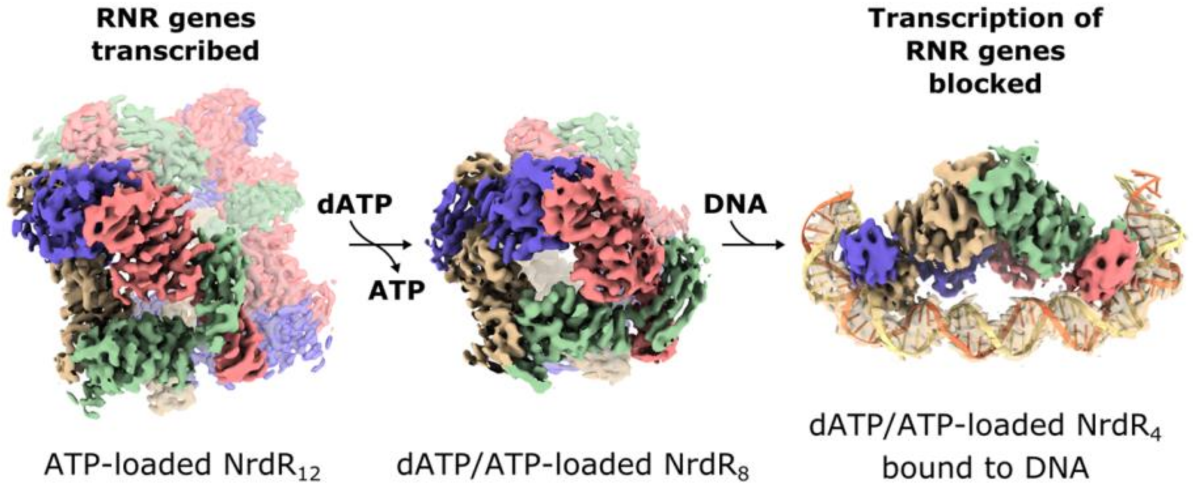


Figure I4. Surface representation of the cryo-EM maps for different oligomers reveals *S. coelicolor* NrdR mechanism of action. The ATP-loaded dodecamer **(left)** in the presence of dATP reoligomerizes into dATP/ATP-loaded octamer **(middle)**, which in the presence of *nrdRJ* promoter forms dATP/ATP-loaded DNA-bound tetrameric assemblies **(right)** inhibiting transcription of RNR genes. Adapted from (62).

The dodecameric NrdR assembly formed in the presence of ATP consists of three tetramers, each tetramer containing the A, B, C and D molecules. In each tetramer, the ATP-cones of molecules A and B, but also those of molecules C and D, interact with each other. Interestingly, interactions between the Zn ribbons occur within domains of molecules A and D, and B and C. These interactions generate an intertwined tetrameric structure. Interestingly, octameric dATP/ATP NrdR assembly consists of two intertwined tetramers, in which interactions between Zn ribbons and between ATP-cones appear in the same manner as in case of tetramers from ATP-loaded dodecamer. However, the relative orientations of the dimerized Zn ribbons and

ATP-cone domains differ between the tetramers. On the other hand, the tetrameric structure of dATP/ATP-loaded NrdR in complex with dsDNA reveals that binding of NrdR tetramer to the *nrdRJ* promoter is mediated by a Zn ribbons dimer interacting with each NrdR DNA box through seven arginines from each domain, including Arg12, Arg17, Arg26-29, Arg37, but also residue Asp15. Interestingly this NrdR tetramer shows two DNA-binding sites, therefore is able to establish the coordination of two NrdR DNA boxes distanced approx. three dsDNA helix turns from each other, which also plays a key role in the DNA-binding specificity (Figure I5) **[62]**.

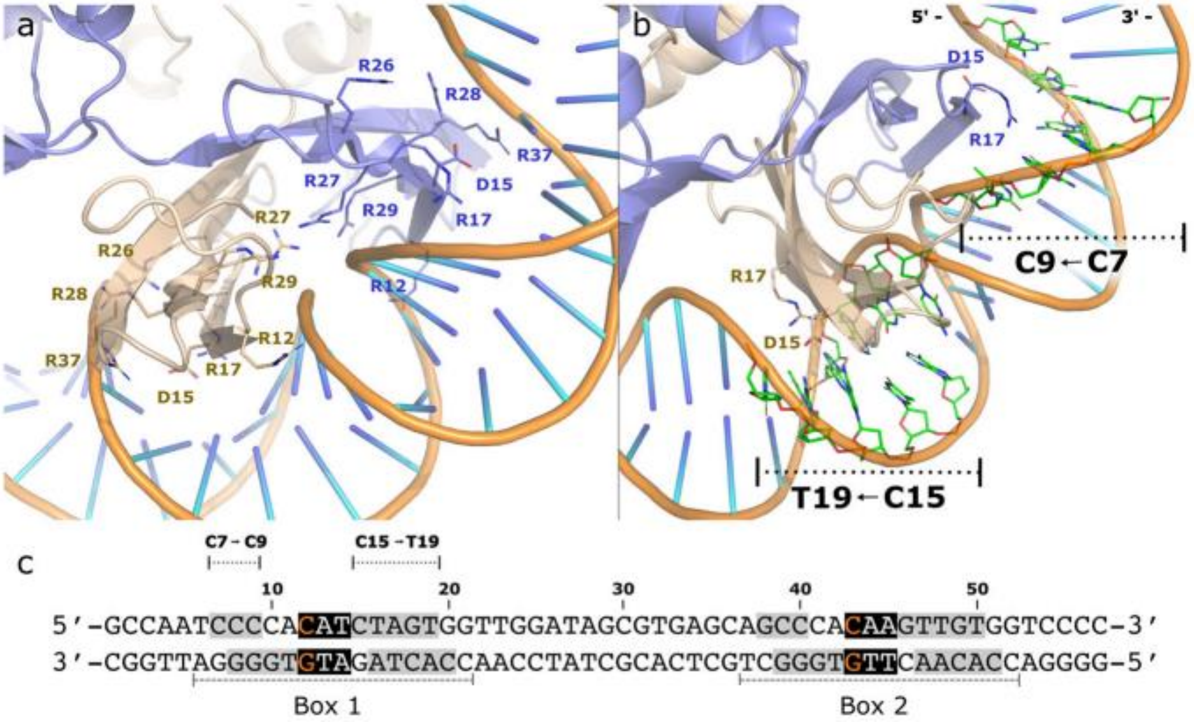


Figure I5. DNA-bound dATP/ATP-loaded *S. coelicolor* NrdR bound to *nrdRJ* promoter region. Interactions. (a) Interactions between arginine residues 12, 26–29, and 37 of chains A (beige) and D (blue) and the phosphate backbone of the dsDNA from NrdR box 1 are shown. **(b)** Close to the DNA bases in the major groove are located residues Asp15 and Arg17. **(c)** DNA sequence of the *S. coelicolor nrdRJ* promoter binding region (NrdR boxes) with distinguished nucleotides interacting with NrdR (grey); highly distorted interactions, with cross-base pair interactions (orange on black); middle of sharp bend (white on black) are shown. Adapted from (62).

Described by Grinberg and collaborators, the nucleotide-binding properties of the ATP-cone domain revealed the presence of two nucleotides bound at the same time, and defining two specific sites. The “inner” site, which is highly specific for ATP binding, similar to sites found in all known ATP-cones, and the “outer” site seen only in *S. coelicolor* NrdR ATP-cone, which binds two nucleotides. Therefore, residues participating in ATP coordination of *S. coelicolor* NrdR ATP-cone domain in the “inner” site are Val48, Lys50, Arg51, Glu56, Phe58, Val63, Thr102, Ile109, Tyr128, but also Lys69 and Glu72 from the neighboring molecule. On the other hand, binding of the second ATP in the “outer” site is conditioned by Lys62, Lys69, Gln72 and Tyr128. Interestingly, residue Tyr128 interacts with alpha phosphate of the inner ATP, but also with 2’-OH and 3’-OH groups of the ribose ring of the outer ATP. However, when in ATP-cone the outer ATP is replaced by dATP, the absence of the 2′-OH in dATP allows for a close interaction of the ribose ring with Phe124, Val127, and the backbone of Tyr128, which faces away from the nucleotide binding sites through driven conformational changes (Figure I6) **62**.

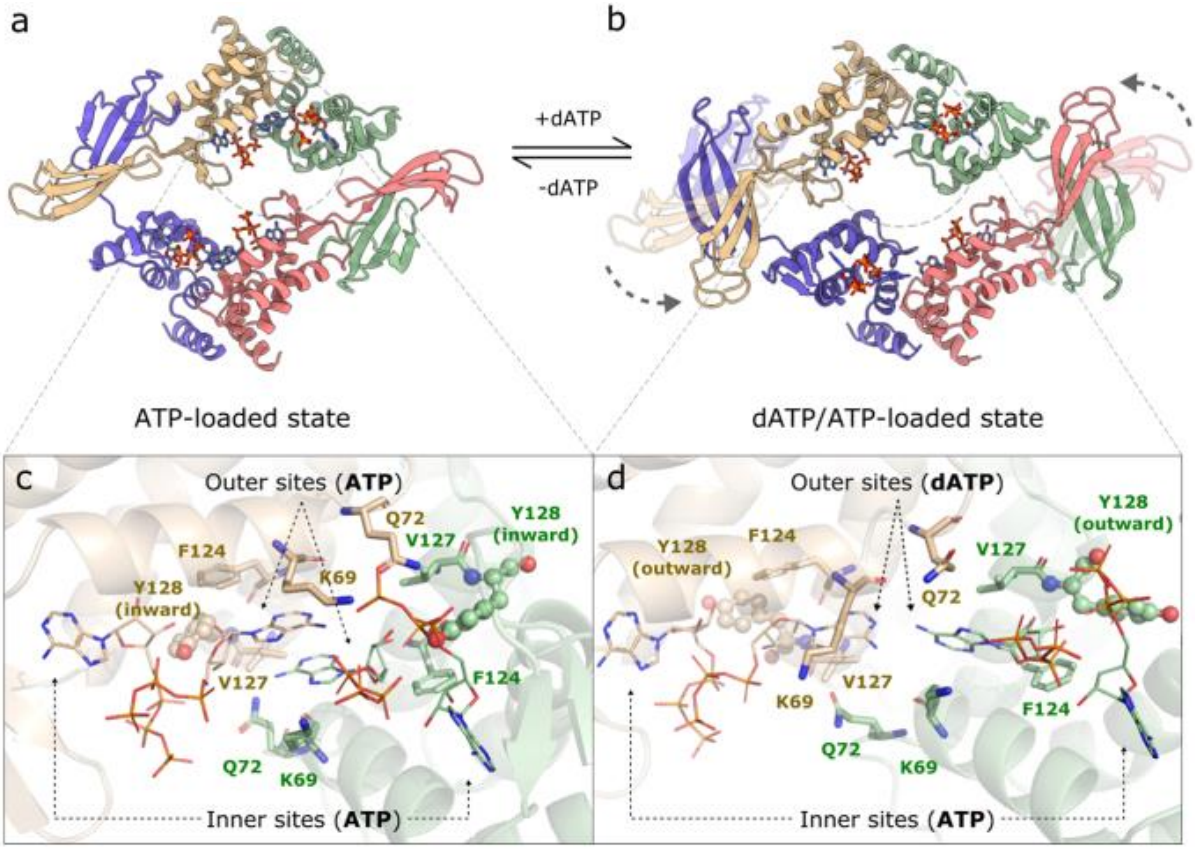


Figure I6. Mechanism of NrdR function involving transition from dodecameric to octameric, and tetrameric assemblies. Cartoon representation of **(a)** the ATP-loaded NrdR tetramer (from dodecamer) and **(b)** the dATP/ ATP-loaded tetramer (from octamer). Chains A, B, C, and D are represented in beige, green, pink, and blue, respectively. **(c)** Interfaces between the ATP-cones in chain A (beige) and chain B (green) in the ATP-loaded tetramer (from dodecamer) and **(d)** in the dATP/ATP-loaded tetramer (from octamer) are shown from the same perspective. Adapted from (62).

In earlier studies performed in 2009 by Grinberg and collaborators, single-site mutations of Val48, Lys50, Glu56, or Lys62 in *S. coelicolor* NrdR impaired nucleotide binding, changed the multimerization state of the repressor, but also interestingly inhibited the binding to DNA [61]. Likewise, studies performed in 2013 by McKethan and Spiro revealed that mutation of Lys53 in *E. coli* NrdR, the equivalent of *S. coelicolor* Lys50, also abolished DNA binding [63].

Structural studies performed by Grinberg and collaborators in 2022 were supported by multimerization analysis in solution, which revealed *S. coelicolor* NrdR *in vitro* dimerization in the absence of nucleotides, tetramerization in the presence of dATP, dADP, ADP, or a combination of dATP and ADP, but also a dodecamerization (major fraction) in the presence of ATP, and an equilibrium between octamers and tetramers when dATP and ATP were present simultaneously. In the presence of dAMP and cyclic di-AMP, NrdR appeared as a dimer/tetramer equilibrium [62]. Since the ATP-cone domain alone can only form dimers as it was demonstrated by Grinberg et al. in 2009 for *S. coelicolor* NrdR mutant lacking the Zn-ribbon domain [61], there is a strong indication that the higher oligomeric forms, which includes tetra-, octa-, and dodecamers, seen both in the crystal structures and in the solution, are mediated by pairwise Zn ribbon interactions. NrdR tetramerization and further higher multimerization based on the ATP-cone dimerization is conditioned by interactions occurring between helix-loop-helix motifs of C-terminal regions of the both cone domains, which together generate a specific four helix bundle, previously found in the crystal structures of *P. aeruginosa* NrdA class Ia RNR [46, 62].

These findings prove that NrdR acts as a nucleotide sensor, repressing the ribonucleotide reductases depending on changes in the ATP/dATP pool. When DNA synthesis decreases, a concentration of dATP in the cell rises. Therefore, at the ATP-cone domains of an ATP-bound NrdR dodecamer, which cannot bind to DNA, dATP replaces ATP in the “outer” site of each cone domain and in consequence forms dATP/ATP bound octamer being in equilibrium with a tetramer. Subsequently, the dATP/ATP bound tetramer binds to the DNA NrdR boxes, thereby inhibiting the expression of RNR genes. All these studies showed that NrdR is a transcriptional regulator with a complex nucleotide binding mechanism that tunes the expression of *S. coelicolor* RNR genes in response to the cellular concentration of nucleotides [62]. Interestingly, a previously suggested mechanism of action of NrdR by Torrents in 2014, assumed two oligomeric triphosphate nucleotide-bound forms of the protein: inactive ATP-bound NrdR, with abolished DNA binding properties, and active dATP-NrdR, which binds to NrdR-boxes by the zinc finger domain and inhibits *nrd* genes transcription [29]. Several authors have suggested that the levels of ATP and dATP bound to NrdR may provoke differences in the NrdR oligomerization state that would be in turn responsible for increasing the degree of repression when the amount of dATP in the cell rises [55, 60, 61]. In 2013, McKethan and Spiro proposed a more complex model of RNR allosteric regulation, in which NrdR selectively binds (deoxy)nucleoside di- and triphosphates, which are subsequently hydrolysed to monophosphates so as to regulate NrdR oligomeric state and specific DNA binding [63].

Objectives

Main objective of this thesis

The overall objective of the thesis is to analyze the structure of *E.coli* NrdR by crystallography, biophysical methods and mutagenesis.

To fulfill this purpose, we established the following subobjectives:

1. To subclone the *E. coli nrdR* gene into a vector system for high scale protein production.

2. To optimize the previously established purification protocol to increase the final NrdR yield for crystallographic purposes.

3. To crystallize *E.coli* NrdR alone or in complex with AMP/ATP/dATP and solve the crystal structure by molecular replacement (MR) or by single-wavelength anomalous diffraction (SAD) experimental phasing.

4. Verify the multimerization mechanism by mutants based on residues identified by analysis of the crystal structure, and optimize the corresponding purification protocols.

5. To analyze the oligomeric state analysis of NrdR and mutagenic proteins in the presence or absence of nucleotides by biophysical methods.

6. Interpretation of *in vivo* assays for wild-type and single point mutants to identify which residues are essential for protein functionality and stability in vitro and in vivo.

Materials and methods

M1. Production of NrdR

M1.1 Theoretical aspects of DNA cloning and gene expression

In order to determine X-ray protein crystal structure, it is necessary to obtain a protein crystal, which requires a significant amount of highly pure protein. Hence, for that purpose protein expression systems have been developed in order to produce recombinant proteins in a relatively fast, cheap and efficient manner. This is particularly true for protein expression in bacteria. Plasmids are circular molecules of double stranded DNA, naturally occurring in the cytoplasm of bacterial cells, which by means of the origin of replication *ori* sequences show the ability to replicate independently of the bacterial chromosome. Commercial vectors, created from bacterial plasmids, allow for transcription of encoded genetic information and protein translation by using the enzymatic machinery of the host. Expression vectors are equipped with selective markers (i.e. antibiotic resistance genes), which guarantee growth only for the cells that contain the appropriate marker (for example in an environment with antibiotic). The expression plasmids used in this work consisted of vectors containing the sequence of the gene encoding the protein of interest, so-called DNA insert, whose expression was activated in the bacterial host, specifically different *Escherichia coli* strains used as expression systems). The preparation of an expression plasmid containing the protein of interest requires a variety of DNA manipulation methods.

First, the *in vitro* replication or amplification of the DNA fragment encoding the protein of interest by the *Polymerase Chain Reaction* (PCR). Second, deletion of the unwanted DNA fragment from the expression vector via a double-digestion with restriction enzymes. Finally, joining both protein-encoding DNA and expression vector by the enzymatic ligation reaction. A powerful and widely used collection of plasmids to express recombinant proteins in *Escherichia coli* is the pET vector system, in which the expression is controlled by activation of the bacteriophage T7 transcription and translation mechanism that have been introduced in the bacterial gene. When the expression system is induced, activation of expression of T7 RNA polymerase within the cell occurs. The heterologous gene introduced into the pET vector is under the control of a T7 promoter, as it contains a lactose operator sequence (lac O) specifically recognized by the T7 RNA polymerase. The pET vector also has a gene coding for the repressor protein *LacI* that binds between the T7 RNA polymerase gene and its promoter, preventing the expression of the polymerase. Under non-induced conditions, the Lac repressor from pET vectors provides double control of expression, by blocking the transcription of the cloned gene from the T7 promoter, and by preventing the expression of the T7 RNA polymerase itself. In the T7 phage system transcription activation, an allolactose molecule binds to the Lac repressor to inhibit it and launch T7 RNA polymerase expression Thus, the T7 RNA polymerase gene is under control of the lacUV5 promoter in the bacterial genome of an specific expression system, which is carried out by λDE3 lysogen (from the λ phage genome). To activate the T7 RNA polymerase gene in the bacterial cells containing the plasmid, a specific inducer is required, Isopropyl β-D-1-thiogalactopyranoside (IPTG) that mimics allolactose and that also represses LacI, so that expression of the gene cloned in the vector starts. For all this to happen, special bacterial strains are used. To introduce the expression vector into the bacterial cell, the bacterial cell is modified as a competent cell by chemical transformation or by heat shock, so that pores are formed and the DNA can cross the cell wall and membranes.

Secondly, these cells are then grown on a nutrient-rich medium for not only proper cell growth, but also protein folding, that contains antibiotic(s) to select only those cells that incorporated the expression vector. Sometimes an additional antibiotic is added to select specific strains that are modified with an additional plasmid that confers special aptitudes (expression of human-type tRNAs for human genes heterologous expression, or chaperones, etc). The bacterial culture grows until it reaches a logarithmic growth phase and at optical density (OD600) = 0.6, the gene expression is induced by IPTG addition to the grown cell culture. Time of expression, OD of induction, temperature and IPTG concentration are the parameters that can be optimized in order to obtain a high amount of properly folded protein. Bacterial cells containing recombinant protein are harvested by centrifugation, flash-frozen by liquid nitrogen and stored in -80°C.

M1.2 Theoretical aspects of protein purification

Recombinant proteins expressed in *E. coli* are released via physical or chemical (or a combination of both) cell lysis and subjected to a purification process to isolate the protein of interest from unwanted cellular elements i.e. cell membranes, nucleic acids, carbohydrates, natural E. coli proteins, etc. and obtain a highly homogeneous of abundant protein solution (order of mg/ml) suitable for crystallization. There are many chromatographic methods used to purify recombinant proteins, based on various physical and chemical protein properties, hence mentioned are only those used during this research.

Histidine-tagged proteins have a high selective affinity for metal ions, e.g. Ni^{2+} which can be immobilized using chelating ligands such as nitrilotriacetic acid (NTA) on the chromatographic stationary phase, such as sepharose resin. In a bacterial lysate, a recombinant protein containing a 6-histidine tag is the strongest binder to this type of chromatography resin while other cellular proteins will not bind or will bind weakly. For that reason, this type of chromatography can be used as a first step in most of the cases of recombinant proteins purification procedures. Histidine tags are attached at the N- or C-terminal end of the protein, either by adding during amplification of the gene by PCR or during cloning in a plasmid that is designed to add this tail to the cloned protein gene. In principle, addition of the His-tag is less disruptive of the protein properties than other tags, yet it might be interesting to consider that it can affect protein folding (by invading internal cavities) or, in complexes, affect protein-protein interactions. In many cases (by 50% of the PDB deposited structures) histidine tag removal may not always be necessary, which is a big advantage, because the tag cleavage reaction in most cases reduces the final concentration of soluble protein.

It can also be possible to use additional tags, such as fused proteins that help folding (e.g. SUMO tag, Thioredoxin tag), or serve to stabilize and purify the protein of interest (POI) by using appropriate chromatographic resins (GST, glutathione binding protein; or MBP, maltose binding protein). The construct with these additional tags usually follow the scheme His_6- Stabilizing Protein Tag – cleavage site - POI. With these additional tags the Ni-NTA chromatography can be also used in the "reverse mode" upon cleavage of the POI, since the Stabilizing Protein Tag (SPT) contains the His-tag whereas the POI does not. More specifically, the purification consist of two steps: the fusion of SPT-POI is first bound to the column, it follows a cleavage of the fusion with the appropriated protease (see below) and loading the mixture of His-SPT, POI and uncleaved remains of SPT-POI to a second NI-NTA - the tag is bound to Ni^{2+}-NTA column whereas the non-tagged protein collection runs out in the so-called "flow-through" fraction. Ater analysis of the fractions by SDS-PAGE, these fractions that contain the pure POI are concentrated to obtain a highly concentrated protein solution suitable for experimental procedures, for example protein crystallization, SAXS studies, etc.

Regarding the proteases that cleave the fusion SPT-POI, the cysteine protease from Tobacco Etch Virus (TEV protease) is widely used. This enzyme belongs to chymotrypsin-like proteases and shows high sequence specificity. The TEV protease prefers for recognition ENLYFQ|S(G,A) sequence, but it was shown that it can also cleave other motifs with substitutions in the positions P2, P4, and P5 (EP5P4YP2Q|S(G,A)) which corresponds with the cleaving variability of the natural enzyme in these sites. The cleaving is always performed between Q (Gln) and S (Ser), which can be substituted with G (Gly) or A (Ala) with a similar efficiency.

The TEV protease cleaving mechanism is based on the „catalytic triad“ Cys-His-Asp where histidine works as the base donating protons to activate the cysteine nucleophile. The cysteine within the triad instead of serine may be the reason for its resistance to commonly used protease inhibitors such as PMSF and AEBSF (1mM), TLCK (1mM), Bestatin (1mg/ml), pepstatin A (1mM), EDTA (1mM), and E-64 (3mg/ml), and „complete” protease inhibitor cocktail (Roche). Zinc ions inhibit the activity of this enzyme at concentrations of 5 mM or greater. On the other hand, reagents that react with cysteine, like iodoacetamide or NEM, belong to efficient inhibitors of the TEV protease.

Of note, TEV protese is highly specific and active for its seven-amino acid sequence with minimal off-target effects. It has activity more than 10,000 units per 1 mg of protein. The activity depends on the type of target protein. The optimal amount of enzyme should be tested for each target fusion protein. The standard reaction buffer of TEV protease contains 50 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, and 1 mM DTT. The cleavage reaction should be incubated overnight in order to get a maximal cleavage, although most substrate cleavage happens in the first few hours. However, three hours of cleavage reaction are mostly sufficient as prolonged incubation times do not lead to proportional increases of cleavage. The TEV protease optimum temperature is 30°C, although it is active in a relatively wide temperature range of 4-34°C.

Size exclusion chromatography (SEC) known also as a gel filtration (GF) is a chromatographic purification technique that depends on the relative size or hydrodynamic volume of a macromolecule with respect to the average pore size of the packing resine (stationary phase). SEC resine consists of spherical particles with pores of different sizes where molecules small enough to enter the pores are retarded as compared to larger molecules. This method of purification can be performed in a wide temperature range and fractions are eluted isocratically with a single buffer (together with the protein, mobile phase) in which conditions can be varied depending on the requirements and parameters of the protein such as the protein isoelectric point for pH selection, salt concentration, presence of metal ions, detergent, etc, and the buffer used might be critical for the next experiment or the storage step.

During GF the biomolecules are under a constant pH and the run is done at a constant low pressure. Thus, purified proteins can be collected in any chosen buffer. Interestingly, SEC can be also approached as an analytical technique as it reveals the state of the sample, such as if it aggregates in the buffer conditions used or in the previous storage conditions, including vitrification with glycerol and liquid nitrogen or freezing that may alter the protein stability. Thus, GF reveals the quality and suitability of the sample for the next experimental procedures. Usually it is applied as a last, polishing step of the purification protocol, but also can be used to assess for thawed protein samples it is also a method to get rid of glycerol prior to crystallization.

Protein concentration can be determined by the colorimetric Bradford protein assay (BioRad). This method is based on absorbance measurement at 595 nm measurement of the protein solution mixed with the Bradford reagent (red, double protonated, cationic Coomassie Brilliant Blue G-250), which binds to proteins and undergoes a color change in the visible spectrum (blue, unprotonated, basic form). In the absence of protein, when the dye is red, Bradford reagent has an absorbance maximum (A_{max}) of 470 nm. In the presence of protein, the change to the anionic blue form of the dye shifts the A_{max} to 595 nm. Therefore, the measured absorbance value at 595 nm, either in a spectrophotometer or a microplate reader, is directly proportional to the amount of protein bound to the dye. The neutral green form of Coomassie G-250 that has donated an electron but is not bound to protein does not interfere with the measurement in the assay since it has an A_{max} of 650 nm. The concentration range of standards in the kits cover the linear range of the assay. Since the curve flattens at high concentrations of dye, the amount of protein in the sample will be underestimated when the concentration of protein is higher than the range of the linear portion of the curve, that is, at saturation conditions. Samples that have protein concentrations higher than the concentrations in the linear range must therefore be diluted and re-assayed to obtain a more accurate estimation of the protein concentration.

The basic quality, but also quantity verification method of protein preparations is Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE). Thus, a polyacrylamide contained in a solution is polymerized in a porous gel. The SDS, in its turn, develops a polypeptide chain that binds and denatures the protein thereby giving a uniform negative charge proportional to the size of the molecule. The protein loaded at the top of the gel (which is placed in a vertical position) is subjected to a differential potential with a negative charge at the top (attracts positively charged molecules, cathode) and a positive charge at the bottom (anode). During the run of the electrophoresis, separation of denatured charged molecules takes place in this electrical field, and the rate of migration through the gel pores is determined by their mass. By using an appropriate mass marker, an estimation of the size of monomeric proteins or complexes is estimated, while determining the purity of the preparation (more or less proteins accompanying the POI will show the level of purity).

M1.3 Subcloning and expression of 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR

Analysis of the constructs as well as the design of the oligonucleotides for PCR reactions were performed using GENTLE software. The construct *SUMO-nrdR* gene generated by our collaborators was subcloned from *pETite-SUMO (NN´)::gNrdR* plasmid (Lucigen) to pCri11a expression vector (pET-26b derivative). Note that pCRI vectors are intended to produce high protein amounts (in milligrams) for crystallographic purposes. During cloning, a TEV protease cleavage site and a 5aa linker that optimized the cleavage was added downstream to the existing SUMO protease cleavage site (the SUMO protein encodes a cleavage site in its own sequence). In this manner, we were able to use our in-house produced TEV protease. The SUMO-*nrdR* construct was amplified by standard Polymerase Chain Reaction (PCR) using primers: 5´ TAT ACC ATG GGC GAG AAC CTT TAC TTT CAA GGA TCC GGA TCC GGA TCC ATG CAT TGC CCA TTC TGT TTC GC 3´ (sense); 5´TAT ATA CTC GAG TTA GTC CTC CAG GCG CGC GAT CT 3´ (antisense). Specifically, the amplified construct encoded a 6-histidine tag, the SUMO protein (with its own cleavage site), the proteolytic site for the TEV protease (TEVcs, GluAsnLeuTyrPheGlnGly), a linker SerGlySerGlySer that optimized the cleavage, and residues 1-149 corresponding to full-length NrdR (uniprot ID. P0A8D0). *E. coli* DH5α strain cells were transformed with the product of ligation. DNA was extracted from the cells and purified by using the NucleoSpin Plasmid QuickPureTM Kit (Fisher Scientific). DNA concentration was determined by measuring the absorbance at 260 nm with BioDrop (Biochrom Ltd). Absorbance ratios at A260/A230 and A260/A280 were also measured to detect the presence of unwanted organic compounds such as trizol, phenol or guanidine thiocyanate.

The pCri11 plasmid with the 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR construct was used to transform *E.coli* BL21 (DE3) strain (Merck Millipore), which was the most efficient for homologous expression. It followed scaling-up the protein production to 4 liters of cultures of LB. These started with preparation of a pre-culture of 40ml with 30 µg/ml kanamycin grown o.n. until saturation. The next day, this preculture was used to inoculate a culture of 500 ml of LB containing 30 µg/ml kanamycin and 200µM $ZnSO_4$. After 3h of incubation at 37˚C, once the culture reached an optical density at 600 nm of 0.6, the bacterial expression was induced with 0.2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) and grown for 16h at 18˚C at 200 rpm. Cells were spun down by centrifugation for 40 minutes at 4.000 rpm, bacterial pellets flash frozen in liquid nitrogen and stored at -80°C.

M1.4 Purification protocol of 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR and 3(GlySer)-NrdR

M1.4.1 Purification of 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR and TEV protease cleavage

Cells were resuspended in 30 ml of lysis buffer containing 50 mM Tris-HCl pH 9.0 (isoelectric point of the construct is 6.38), 1M NaCl (to break non-specific protein-DNA interactions), 50µM $ZnSO_4$ (to maintain the Zn-finger domain), 1mM DTT (to prevent disulfide bonds formation), 10 U/ml DNase I (Thermo Fisher Scientific; to degrade traces of bacterial DNA bound to protein), 10% v/v glycerol (to destabilize protein-protein interactions with contaminants or prevent aggregation), 1 mM Phenylmethylsulfonyl fluoride (PMSF) (Thermo Fisher Scientific) and complete, EDTA-free tablets (Roche) to avoid proteolysis. Thereafter, lysate was sonicated for 3 minutes (cycles 3s ON, 2s OFF at 20% of the maximum power), which not only breaks the cells but also shears the DNA. The lysed mixture was stirred for 45 min at 4 ºC, after addition of streptomycin sulfate at a final concentration of 0.1% v/v to precipitate DNA traces. Afterwards, lysate was centrifuged for 25 minutes at 20.000 rpm.

The resultant supernatant was filtered with a syringe with membrane of a pore size of 0.22 µm (Merck KGaA©) and loaded onto a HisTrap™ 5 ml nickel column (Cytiva) installed to an FPLC Åkta system, and washed with buffer A (1M NaCl, 50mM TRIS-HCl pH 9.0, 20mM imidazole) for 5 column volumes (CV). For protein elution, a 10 ml step at 200 mM imidazole was performed with a gradient of buffer B (same content as A and additionally containing 500mM imidazole). The eluting fractions (2 ml) were collected and analyzed by 15% SDS-PAGE, and protein concentration by BioRad assay). The final yield was very high (6 ml at 9 mg/ml from 750 ml culture). For a successful subsequent digestion and next purification step, the imidazole from the protein sample was extracted by a buffer exchange step with 50 mM Tris-HCl at pH 9.0 and 500mM NaCl using a disposable ultrafiltrator (3000 MW cut-off, Vivaspin®).

The TEV-digestion trials were performed for 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR construct under various protease:protein ratios, at 4˚C o.n. in buffer containing 50 mM Tris-HCl pH 9.0, 500mM NaCl and 1mM DTT. The nature of the bands was identified by MALDI-TOF Mass Spectrometry (The Proteomics and Genomics Service CIB Margarita Salas, CSIC, Madrid).

M1.4.2 Isolation, concentration and quality check of 3(GlySer)-NrdR

Samples from cleavage of 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR expected to contain 6His-SUMO-GluAsnLeuTyrPheGln, 3(GlySer)-NrdR, traces of TEV protease and traces of undigested 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR, were loaded on a HisTrap HP 5 ml nickel column (Cytiva). The loaded sample bound to the column was washed with buffer 500mM NaCl, 50mM TRIS-HCl at pH 9.0, and eluted with a gradient with a buffer additionally containing 500mM imidazole. 3(GlySer)-NrdR was collected in the flow through. Traces of undigested 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR, 6His-SUMO-GluAsnLeuTyrPheGln (main component), and traces of TEV protease remained bound to the Ni-NTA resin and coeluted in a single 200 or 500mM imidazole elution step. 3(GlySer)-NrdR was concentrated to 16 mg/ml without significant losses using a disposable ultrafiltrator (3000 MW cut-off, Vivaspin®). Sample purity was assessed by 15% SDS-PAGE, purest fractions pooled and the corresponding concentration determined by BioRad assay.

Gel filtration was performed at RT in the presence of ATP, otherwise the protein was lost. 100 µl of native NrdR at a concentration of 2mg/ml were injected into Superdex® 200 10/300 GL column (Cytiva), previously equilibrated with buffer containing 50 mM

Tris at pH 9.0, 500 mM NaCl, 0.5 mM ATP and 1 mM DTT. Run was performed at 0.5 ml/min flow rate.

M2. Crystallization of NrdR

M2.1 Theoretical aspects of protein crystallization

Protein crystallization is a multi-parameter process, in which a biological macromolecule in solution is subjected to a dehydration process that induces an ordered precipitation in the form of regular arrays, thus crystals. The set up includes a protein drop (micro or nanoliter scale) that is mixed with a crystallization solution drop of similar volume, and the resulting drop is exposed to a large volume (a fraction of ml) of crystallization solution contained in a well. The whole set is sealed so the water from the protein in the small drop evaporates to equilibrate the large volume of crystallization solution, thereby dehydrating the protein. The successful crystallization solution cannot be anticipated for any protein, so different conditions that differ in their chemical composition (protein concentration, pH, type and concentration of precipitant and buffer, presence of metal ions, cofactors, additives, ligands, etc) are screened.

In addition several, different physical factors, i.e. temperature, pressure, viscosity of the solution, time, impurities also influence the formation of crystallization nuclei, which, depending on the conditions, may evolve as crystals. The transition between crystallization nuclei to crystals can be induced by slightly changing the chemical and physical parameters. A particularly important aspect limiting the success of protein crystallization is also the purity, homogeneity and stability levels of the protein sample. Note that structure similarity between the studied protein and an homologous one (or the same protein in complex with different factors) already crystallized does not guarantee positive results despite using the same crystallization conditions.

In order for crystals to be formed, that is for nucleation to occur, an energetic barrier must be overcomed). In order to visualize the successful crystallization conditions, a phase diagram can be constructed varying the precipitant concentration on the x-axis and the protein concentration on the y-axis. In the phase diagram are distinguished various areas. If the amorphous protein precipitation has occurred, then it is a clear indication for the precipitation zone, in which both variables have too high values to grow a crystal. On the other hand, if concentration of both protein and precipitant is too low, then the crystallization system is at the unsaturated zone that results in no nucleation, no precipitation and a clear crystallization drop. It is in the nucleation zone where the protein nucleates, but as the process of crystal growth a simultaneous decrease in protein concentration in the drop occurs, thus the primary value of the protein/precipitant ratio changes, a gradient of depleted concentration surrounds the crystal, which can attract more protein from the rest of the drop, or stay as it is. Therefore, the crystallization solution effectively moves to a new part of the phase diagram, and may stay at the metastable zone or shift to the unsaturated region Therefore, when the when the protein transits from diluted to solid state and the drop reaches the metastable zone, the crystals start to grow bigger. Otherwise, if the phase transition cannot be reached, the protein molecules will continue to create more crystal seeds rather than increasing the volumes (Figure M1).

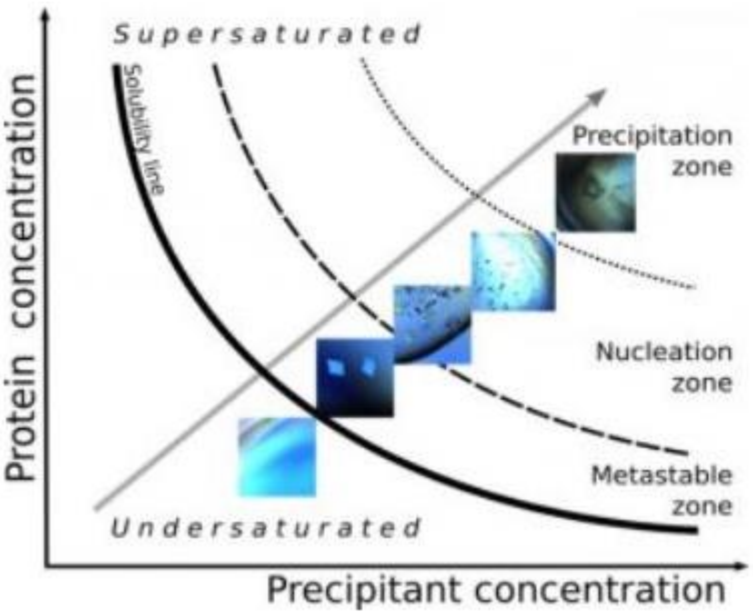


Figure M1. Crystallization phase diagram. Concentration of crystallizable protein plotted against the concentration of a precipitant shows several areas related to crystal growth and formation.

Currently, different crystallization methods are developed to grow crystals from saturated solutions. During this research, the hanging drop method based on the vapor diffusion technique was used, which involves preparing a crystallization solution (consisting of a protein precipitant or crystallizing agent, a buffer to provide a stable pH and additives) and placing it on the bottom of a well. A protein solution mixed with the crystallization solution is applied as a drop onto a coverslide (hanging drop vapour diffusion method) or left on a surface close to the well (sitting drop method). In the case of the coverslide, this is turned updown placing the protein drop facing the well containing the crystallization solution. Borders of the wells are covered with silicone for the sitting drop method, thus forming a closed system. As a result of the difference in concentration in both solutions, the water from the protein buffer slowly evaporates from the drops, which results in a gradual concentration increase of the precipitant in the solution. Once the solution has reached saturation, followed by supersaturation, at appropriate precipitant concentration nucleation begins in the drop. Other methods such as the micro-batch, capillary counter-diffusion and dialysis are also available.

Because of the large number of factors on which crystallization depends, a collection of crystallization solutions developed by different groups and companies are available to screen the type of buffer and pH, precipitants (PEG, MPD, ammonium sulfate, etc), salts and other compounds. All together, the presence of these substances promote protein crystallization. Many years of observations have shown that the most popular precipitates are polyethylene glycols of different lengths and ammonium sulfate. Optimization of crystallization conditions is achieved by changing parameters such as protein or precipitate concentration (for example change PEG concentration by 5 units in percentage (0, 5%, 10%,up to 40 5%), salts (in steps of 50 mM difference, such as 50, 100, 150, 200, 250 mM), buffer and pH (Tris, Hepes, Citrate, cacodylate, MES, etc, by increasing 0.5 units around the pK), temperature (4ºC, 20ºC, 18ºC, etc) and drop size 100 -200 nl in automated screenings, or 0.5, 1 or 2 microliters in manual screenings). The ultimate and decisive factor determining the quality of a crystal is the ability to diffract X-rays.

M2.2. Crystallization of 3(GlySer)-NrdR

3(GlySer)-NrdR at a concentration of 16mg/ml in buffer containing 500mM NaCl, 50mM TRIS-HCl at pH 9.0, 1mM DTT was used for initial crystallization trials at the Automated Crystallography Platform (PAC) in the Science Park, Barcelona (PCB). Crystallization screens were done at room temperature (RT) with 96-well plate sparse matrix screens such as PAC1, which includes Crystal Screens I and II from Hampton Research grouped into 96 conditions in the format of sitting-drop vapor-diffusion; PAC2, which contains Wizard Screens I and II, PAC10 also called *protein-DNA screen*; and PAC19, which consists of different type of salts and precipitants.

M3. NrdR crystal structure solution, refinement and validation

M3.1 Theoretical aspects of protein structure solution in crystallography

M3.1.1 Crystal arrangement, X-ray diffraction and Fourier synthesis

A crystal consists in three-dimensional repetition of a unit cell. Inside this, multiple copies of an asymmetric unit are arranged following a symmetry pattern, the crystal symmetry. Note that in an asymmetric unit molecules can be related with each other by non-crystallographic symmetry operations, which are not compatible with the periodicity of a crystal pattern in two or three dimensions. To this type of symmetry belong rotations other than 1, 2, 3, 4, 6.

X-ray diffraction in protein crystals is the effect of an electromagnetic wave interacting with electrons present in a highly ordered set of protein molecules in a crystal lattice. The diffraction of the X-ray waves is a convolution between the dispersion by the atoms of each protein in the crystal (millions of proteins) and the symmetry arrangement of the crystal lattice. The dispersed waves interact with each other by constructive or destructive interference. The constructive interaction results in X-ray intensities that are collected by a detector plate and which, all reflections altogether, reflect both the symmetry of the crystal and the structure of the protein in the crystal. This only happens when Bragg's law is satisfied. From the Bragg's equation, it is known that fixed-wavelength (λ) X-rays with a positive integer (n) are reflected or deflected by a set of crystal lattice planes (hkl) that are characterized by an interplanar spacing distance (d_{hkl}). But such a reflection (or deflection) occurs only at a specific angle of θ of the reflected X-ray:

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

Thus, constructive interference (that in a detector is collected as a diffraction spot) can only happen, when the multiple of the X-ray wavelength is consistent with the path length difference between two diffracted waves ($2d\sin\theta$) (W.L. Bragg and W.H. Bragg, 1913). As said above, the diffraction experiment images contain spots of the reflections whose geometrical distribution results from the symmetry of the crystal lattice, while the intensities encode information about the position of the protein atoms in the crystal. Solving the crystal structure involves determination of the positions of all atoms in the crystal. Therefore, the Fourier transform of the amplitudes ($|F(hkl)|$) and phases ($\varphi(hkl)$) of the X-rays scattered by each atom allows to calculate the distribution of the electron density ($\rho(xyz)$) in a crystal unit cell (with a volume of V):

$$\rho(xyz) = \frac{1}{V} \sum_{\substack{hkl \\ -\infty}}^{+\infty} |F(hkl)| \, . \, e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

In which V is the volume of the crystal unit cell; x,y,z, the position of a point in the unit cell; ρ is the electron density at x,y,z; h,k,l are the three integers defining parallel sets of the crystal lattice planes.

F_{hkl} is the structure factor of a reflection, which is calculated from the intensity (I_{hkl}) of the diffracted X-rays that is collected by a detector, since $|F_{hkl}|^2$ is proportional to I_{hkl}. However, the phases of the scattered X-ray φ_{hkl} defined as $F_{hkl} = |F_{hkl}|e^{-i\varphi(hkl)}$ are not recorded, thus the so-called crystallographic ¨phase problem¨ using different approaches described below need to be overcomed.

M3.1.2 X-ray diffraction data collection and processing

While processing diffraction images, the parameters of the unit cell and the space group are determined, and a list of an independent reflections defined by indicators h, k, l and intensity (I) together with their standard deviations (σI) is created. The standard deviation is the result of scaling together tall measurements of symmetry-related reflection intensities *I* that belong to symmetrically-related points h,k,l (hkl, -hkl, h-kl, hk-l, -h-kl, -hk-l, k-k-l, -h-k-l) and the value of these intensities should be identical. In addition, a reflection might be collected several times, and the mean of all values is assumed as the value of the intensity I, with the associated standard deviation. X-ray datasets are processed with dedicated program packages, such as XDS, which averages the intensities of the redundant reflections to the intensity corresponding to a single, unique reflection. The package includes three programs: XDS, XSCALE, XDSCONV. XDS produces a list of corrected integrated intensities of the reflections collected in a sequence of adjacent, non-overlapping images corresponding to the rotation of the crystal so that all reflections diffracted in all directions are collected. XSCALE places the reflections of a data set to a common scale and reports their completeness, and also provides information about the quality of the integrated intensities. The data is corrected for absorption effects and radiation damage by using the appropriate commands at XSCALE. XDSCONV converts the XSCALE outputs data files that contain the reflections into mtz format, which is required in the subsequent steps to determine the crystal structure by using the software package

CCP4. Statistical parameters obtained during data processing (Table M1) provide valuable information on the quality of the data set.

Estimator	Description
R_{merge}	This parameter is calculated with the formula: $\Sigma_{hkl}\,\Sigma_i \lvert I_i(hkl) - \langle I(hkl)\rangle \rvert \,/\, \Sigma_{hkl}\,\Sigma_i\, I_i(hkl)$, which is the difference estimator between the intensities of the symmetric reflections, theoretically identical. Its value should be as low as possible.
$\langle I/\sigma I\rangle$	An averaged signal to noise ratio. Reflections with $\langle I/\sigma I\rangle$ value equal or higher than 1 are considered reliable.
mosaicity (°)	A measure of the spread of crystal plane orientations, mean deviation from the ideal crystalline arrangement.
completeness (%)	A number of collected reflections in comparison to the number of theoretically possible reflections, unique for the given resolution and crystal symmetry. Total completeness of the data should remain at least 85%.
multiplicity	The average number of measurements per individual, symmetrically unique reflection.

Table M1. Data quality estimators.

As a final step, the intensities are converted to structure factor amplitudes, a step called “truncation”. To be able to calculate an electron density map by a synthesis of Fourier, not only the amplitude but the phase of the structure factor is required, which is achieved by the following methods.

M3.1.3 Molecular Replacement (MR)

A method that is widely used for solving the “phase problem” is Molecular Replacement. In this method, the phase information of the protein structure to be solved is obtained from a model, which closely resembles the structure to be solved i.e. is derived from a known structure of a protein similar to the structure of the protein under investigation. This means, the position of the protein molecule inside the unit cell has to be defined, which requires specification of three rotational angles regarding the orientation, and three translational vectors regarding the molecule's position inside the crystal cell. Therefore, a Patterson map, which is solely based on amplitude squares of structure factors, but omitting phases is calculated. The Patterson map is centrosymmetric and contains maxima/peaks representing intra- and intermolecular interatomic vectors. The heights of these peaks are proportional to the multiplication of the number of electrons of every pair of atoms that are related to the interatomic vectors. Intra-molecular interatomic vectors are shorter and proportional to the molecule size, and are related to the orientation of the molecule, so that by selecting those vectors initially a rotational search in Patterson space is performed. Once several orientations of the molecule are obtained and ranked, the position of the molecule in the crystal is calculated from every orientation. The position of the molecule is found from the longer Intermolecular interatomic vectors that relate one molecule with its neighbor in the crystal. By knowing the values of three rotation angles and three translation vectors it is possible to define the position and orientation of the molecule in the unit cell. An algorithm for the Molecular Replacement calculations involves four steps. The Patterson map calculation for an unknown structure is performed by generating the Patterson map of a known model structure. The structure is placed at coordinates 0,0,0. From this first position, the corresponding theoretical structure factors F_{calc} are calculated and compared with the F_{exp} from the experimental data. This process is iterated with thousands of rotations of the initial model to find the best fit between the theoretical, calculated values and the experimental ones, and the different results are ranked. Once the ranking list is available, the same is done for the translations. From every rotation ranked in the list, thousands of translations inside the crystal cell are calculated, and the corresponding F_{calc} are compared to the F_{exp}. The best fit gives the orientation (by rotation) and position (translation) of the model relative to the origin of the system of coordinates, indicating the position of the unknown structure. The best fit allows to assign the phases of the searching model to the experimental data. These phases allow the calculation of the electron density map, from which a model of an unknown protein is built. Maximum-likelihood based functions for the rotation and translation searches including criteria of crystal unit cell packing are used by MR programs such as *Phaser*.

A minimum of 25% identity of protein sequences offers real hope for a solution, but does not guarantee it. The main disadvantage of this method is that a successful solution of new, unknown protein families can be difficult to obtain without existing a model with highly conserved structural homology.

M3.1.4 Anomalous Scattering and Single-wavelength Anomalous Diffraction Experimental Phasing (SAD EP)

Anomalous Scattering occurs when an X-ray photon with an energy close to the characteristic absorption edge(s) of an atom hits an electron in this atom with simultaneous photon absorption and re-emission a photon with a lower energy level, which causes a shift in amplitude and phase of the scattered radiation, and triggers electronic transition in the atom. Thus, the atom excitation causes the breakdown of Friedel's law, which states that the intensity of symmetrically-related reflections show equal intensity, I(hkl) = I(-h-k-l). These reflections are the so-called Friedel pairs. Thus, the analysis of the differences in intensities I(hkl) and I(-h-k-l) on the diffraction images allows to calculate the positions of special anomalous dissipating atoms. By means of the Patterson function, these atoms are localized in the real space (x,y,z) inside the crystal, generating a substructure of "anomalous atoms". With these positions at the real space, the phase of the reflections affected by the anomalous scattering can be calculated and expanded to the rest of reflections ultimately leading to the calculation of an electron density map.

In an anomalous diffraction experiment, the Friedel pairs, as they have different intensities, the corresponding structure factors $(F_{PH})^{+}$ and $(F_{PH})^{-}$, (given that I is proportional to F^2), also have different values. This occurs because of a cumulative normal scattering effect of ordinary atoms (F_P) and scattering of excited "anomalous atoms" (F_H). This effect is exploited to determine the position of the heavy atoms inside the crystal unit cell. Nowadays, given the quality of the data produced by highly collimated beams and excellent detectors, the X-ray data of crystals diffracted at the heavy atom absorption edge is sufficient to determine the positions of the anomalous scatterers.

Performance of anomalous diffraction experiments and solving a protein structure via SAD (Single Anomalous Diffraction) Experimental Phasing method, requires presence of heavy atoms in the protein crystal, heavier than C, N or O (which have X-ray absorption edges at energies that can be reached by synchrotrons or in-house sources). These atoms can be either naturally belonging to the protein (Mn in active sites, or as atoms stabilizing the structure such as Zn in the Zn-finger proteins) or that intentionally have been introduced into the protein in two ways. By addition of Selenium bound to methionines (Se-Met) to bacterial cell culture medium, in order to incorporate Se instead of S at protein methionines, and obtain Se-Met protein and the corresponding crystal derivative. Another potential (but not that popular) derivatives are those from cysteines (Se-cysteine). Metals are also added by soaking the crystals in a harvesting solution containing the metal of choice (Au, Pt etc.).

Finally, both methods can be combined. Partial solutions obtained by MR can be complemented by searches of atomic positions of atoms, and with the combined phases calculate an electron density map.

M3.1.5 Structure refinement and validation

With the initial phases obtained by the above methods, a Fourier synthesis (see section M3.1.1) is calculated and the electron density distribution in the crystal is obtained, into which an initial structure of the protein at atomic level throughout the crystallographic asymmetric unit (and the unit cell by symmetry operation) is built by automated and manual model building methods.

The initial model consists of protein molecules with associated ligands and solvent water molecules. Quite often compounds form the solution of the crystallization condition and/or protein maintenance buffer. Once built as much as the density allows and no more than this, the macromolecular model is submitted to atomic position refinement, which is performed automatically and whose aim is to modify the atomic positions, yet restrained to limit the atomic displacements. The chemical bonds, angles, charge interactions, van der Waals contacts, etc. are subjected to restrictions to not achieve unreasonable values. With these modifications, the corresponding amplitude and phase of every theoretical diffracted reflection is calculated. The calculated amplitudes are compared to those of the experimental data, and when the differences are minimal, the theoretical phases are then assigned to the experimental amplitudes. The basic criteria for determining the correspondence between calculated (model-based) and observed (data-based) structure factors is the crystallographic *R* residual indicator:

$$R = \frac{\sum_{(hkl)} w_{(hkl)} \left| |F_o(hkl)| - k|F_c(hkl)| \right|}{\sum_{(hkl)} |F_o(hkl)|}$$

In which, $w_{(hkl)}$ stands for weight of the *hkl* observation, F_0 and F_C are structure factor amplitudes of the diffraction data and built model respectively, k is a scale factor.

A standard structure solution trial combines cycles of model building and refinement until the disagreement between the experimental data and obtained model reaches the point in which it cannot be reduced any further or stays in the area of an acceptable range.

The final purpose of refinement is to achieve the best fitting between the generated model and the diffraction data. As the R factor (R_{work}) is affected by overfitting, an additional estimator called R_{free} factor is generated from a pool of test reflections, randomly selected from the entire diffraction data set once they are scaled and truncated. This set of reflections are excluded from the refinement process. During automatic refinement, the values of R_{work} and R_{free} factors, but also the difference between them, are monitored. This allows us to estimate if the refinement is going in the right direction, which happens by the simultaneous decrease in the value of both factors. Overfitting is detected when the introduction of additional parameters optimized for refinement is not justified by a sufficient number of reflections. Thus, the increase in the R_{free} value associated with simultaneous decrease in R_{work} value is evidence for over-interpretation that leads to overfitting. Overfitting is also observed when the difference between R_{work} and R_{free} factors exceeds 5%, while acceptable overfitting is achieved below 5% difference. At the final refinement stage, the R_{work} and R_{free} factors have the lowest possible value, although it is strongly dependent on the resolution and quality of the experimental data.

Among the refinement strategies, there are the following ones. The model parameters are refined both at the reciprocal space, against the X-ray data (the calculated F_{calc} amplitudes are refined against the experimental ones F_{obs}, the theoretical phases φ_{calc} associated to the F_{obs}), and at the real space against the calculated electron density map (atomic positions circumscribed at the electronic density peaks with a certain degree of freedom), at each refinement cycle. More specifically, with the Fourier transform, the model-derived calculated amplitudes of structure factors F_{calc} are calculated from the coordinates xyz that correspond to every atom from the model, and compared to the experimental amplitudes F_{exp}. In the case that the R_{free}, R_{work} indicators are high, as it typically occurs right after the structure is solved, a first strategy is to refine the whole molecule or the domains as a rigid-body. To this end, the coordinates are grouped (for the whole structure of in domains), the F_{calc} calculated and refined in the reciprocal-space against F_{exp}. This entails moving the protein or the domains inside the asymmetric unit, which modifies the phases and also the F_{calc} values and systematically compared to F_{exp}. When the position of the protein or the protein regions are close to what is present in the crystal, the residuals R decrease in value. In the next step, the position of the atoms, xyz, is refined. This is combined with the refinement of the B-factors, which are the temperature values that describe the vibration of the atoms and whole protein regions, and the ADPs (atomic displacement parameters) that are related to the degree of uncertainty of the atomic position due to the experimental error. At low resolution, the B-factors are refined isotropically, while at very high resolution (say, beyond 1.0 Å resolution), at which the asymmetric displacement of the atoms can be detected, the ADP refinement is anisotropic.

After several successful cycles of automated refinement the residuals R_{free} and R_{work} factors decrease. Visual inspection of the fitting of the structure within the calculated map at the real space allows to manually place regions inside the density, delete regions that do not fit and build the missing parts of the model. This can be combined with automated model building. Manual and automated model building can be combined with automatic refinement, in several cycles until the R_{free} reaches a value close or below 0.3. This indicates the refinement is satisfactory and the space group or the sequence of the protein traced in the crystal corresponds to the real one (it may happen that another protein crystallized, or the space group is not correct). During the next refinement step the values are decreased, yet the success depends both on the resolution of the diffraction data (determined during the data scaling, the amount of data increasing exponentially with resolution) and the degree of disorder within the crystal that will limit the amount of atoms added in defined sites.

Special situations may occur, such as alternate conformations of side chains or of a whole protein region (such as a loop, or an entire domain), or partial presence of a ligand, an ion or a solvent atom. Given that the model corresponds to an average of all diffracting molecules in the crystal, the disorder is due to the fact that in every unit cell the crystallized protein has different conformations or the ligand is bound to a fraction of the crystallized protein molecules. When the positions of the atoms are partial, this reflects the parameter occupancy, which in general is set to a value of 1, but is set to 0.5 in the alternating conformations (the fraction of the occurrences the atom is present in either position in the crystal). With data at very high resolution the occupancy of an atom can be refined.

Refinement of an atom requires 4 parameters: the coordinates x,y,z and a vibrational factor B that accounts for the uncertainty of this position (see below). However, the number of reflections, which is the experimental data, might not account for the number of parameters to be refined during model building (protein and DNA atoms, ligands, water molecules, etc). Therefore, there are the optimization targets, which impose assumptions and criteria during the refinement to diminish the number of parameters to be determined during automated refinement. This is especially useful when the resolution (thus, the number of reflections) is low, which can compromise the agreement of the model amplitudes with the experimentally measured amplitudes. One restrain is applicable when several almost identical copies of the same structure coexist in the asymmetric unit may. By applying a non-crystallographic symmetry (NCS) restrain between such almost-identical molecular copies, we assume that the relative positions of the atoms or the side chains between one and another molecule are maintained (thus, their movements in xyz are restrained to be the same in all copies). Regions with different conformations between molecules must not be included in the NCS restraints.

Another restraint is the knowledge about the presence of secondary structure motifs (α-helices, β-sheets), which imposes distance restraints for hydrogen bonds in α-helices and β-sheets during refinement. In case of poor geometry or overfit of the structure at the final stage of refinement, an optimization of the X-ray/stereochemistry weight can be applied to determine the appropriate relation between the X-ray target function and the geometry; in cases the data is poor and thus not enough experimental parameters are available, the stereochemistry will prevail to preserve the physical integrity of the molecule, whereas in case of exceptional data at very high resolution the stereochemistry restraints can be released. Another successful strategy is to impose fixed rotamers and conformers during refinement.

At the last steps of refinement, automatic addition of hydrogens to the structure will facilitate the maintenance of appropriate distances between atoms, yet the real position of these hydrogens is not refined (since they would increase the total number of parameters). As well, automated refinement at the real space by the refinement programs can also include automatic addition and update of water molecules coordinated by the protein (or DNA) structure, by automatically placing them into empty density blobs and deleting the incorrectly built ones. However, visual inspection of all water molecules must be done to ensure that they will be between 2.5 and 3.5 A distance from charged atoms (from the protein or another water molecule). Addition of hydrogens to the structure and placement of coordinating water molecules can have a strong impact on the geometry and greatly improve the quality of the model.

Before the interpretation of the final model, the proper validation is required. One of the criteria that determines the quality of the model is the root mean square deviation (r.m.s.d.) of individual parameters, such as bond length, angles, torsion angles, etc, from ideal values. Validation also includes an interpretation of the Ramachandran plot (which is a graphical representation of distribution of backbone torsion angles φ/ψ, allowing for identification of the incorrect conformations still remaining in the model), which may entail changing the main chain conformation while monitoring the correctness of side chain rotamers and selecting proper ones etc. Thus, once the validation process is satisfactory, the model can be interpreted to understand the functional mechanism.

M3.2 Crystallographic structure solution of NrdR

M3.2.1 X-ray anomalous diffraction data collection

NrdR orthorhombic crystals were diffracted at the P14 beamline at the German Electron Synchrotron (DESY, Hamburg). X-ray diffraction tests were performed with a transmission of 100%, exposure time of 0.05 sec and 1˚ oscillation at a beam flux of 2 x 10^{13} photons/sec. Data was collected on an EIGER 16M Single-Photon Counting (SPC) detector. Given that crystals were supposed to contain a Zn ion at the Zn-finger domain, a fluorescence scan was done to determine the zinc absorption edge, which was around 1.28Å (9.66 keV). Therefore, the X-ray wavelength was set at this energy. In order to obtain a data set with optimal Friedel pair differences, the crystal was reoriented to expose the Friedel pairs at the same image by using the mini-kappa goniostat. The crystals were of diffraction quality, highly stable, and ten X-ray datasets were collected, including three datasets collected at different orientations of one single crystal at nominal resolutions of 2.35 Å (2 datasets) and 2.5 Å resolution respectively. These data were collected with 360˚ rotation in an oscillation range of 0.1˚ with 100% of transmission at the peak wavelength, in all three diffraction points in the crystal, to generate a robust (with high completeness and multiplicity) single-wavelength anomalous data set (SAD). All collected datasets were processed with the program package XDS as described in point 3.1.2 of the previous paragraph. The positions of the zinc and sulfur atoms (from the methionines) were used to guide protein sequence assignment.

M3.2.2 Structure solution trials

Molecular Replacement trials including searches of Zn-finger and ATP-cone domains as well as the model of *E.coli* NrdR provided by ITASSER (Iterative Threading ASSEmbly Refinement) performed in PHASER were unsuccessful.

The structure was solved by SAD Experimental Phasing with a single X-ray data set, sufficient to define the positions of the anomalous scatterers, the Zn ions coordinated by the four cysteines of NrdR Zn finger domains.

At the beginning, a stepwise approach was tried, in which the anomalous signal of the Zn ions was enough to calculate the crystallographic phases corresponding to the Zn substructure, and could be automatically extended to the phases of the remaining reflections from the whole (protein + Zn) data set. A Fourier synthesis was calculated to obtain an initial experimental electron density map. However, improvement of the map by density modification employed as implemented in Phenix did not yield interpretable maps.

Despite this, we nevertheless combined the density modified phases with the initial experimental phases (phase combination, which preserves the original information) with PHASER, but since the anomalous signal was weak, the resulting electron density maps were neither successful and did not allow to solve the structure. As an alternative, we tried different options provided by the CRANK2 pipeline (ccp4i2 software package), which simultaneously combines the different steps of structural determination. In this approach, it combines the phase information with information from density modification and model-building in real space. Density modification was performed with PARROT, and automated model building with BUCCANEER. Trials with a combined algorithm using CRANK2 resulted in a clear structure solution as evaluated by visual inspection of the map and values of the residual indicators R_{work} and R_{free} equal to 0.46 and 0.50, respectively (Rfree by Brunger et al).

M3.2.3 Model building, refinement and validation

The output model from CRANK2 was used as an input in AUTOBUILD from the Phenix suite. This consistently improved the phases as assessed by the residual values of R_{work} and R_{free}. In the next step, the output from AUTOBUILD was used as an input in PHENIX Refine to refine the positions of the atoms in the space, an assign a thermal B factor that accounts for the uncertainty of the atom position, thereby contributing to the positional error and improving the fitting between the experimental data and the refined protein structure.

For model building of the NrdR structure and refinement, two cyclically repeated stages were alternated, automated refinement at the reciprocal space, and manual model building at the real space. Calculations with PHENIX Refine involved geometry refinement of the model by NCS restraints between the four molecules in the asymmetric unit at the early stages of refinement, positional refinement of individual atoms that was restrained as protein residues and water molecules, refinement of B-factors and ADPs. Manual model building included manual placement of protein regions and atoms inside the density, followed by real space refinement with COOT based on the electron density maps with 2Fo - Fc and Fo - Fc coefficients. After several cycles, the residual R Factors improved to R_{work}=0.22 and R_{free}=0.25, which are acceptable values.

M4. Subcloning, expression and purification of NrdR mutants E36A, E42A, Y131A, del132-149

Mutants NrdR E36A, NrdR E42A, NrdR Y131A and deletion NrdR del132-149 were generated by using the oligonucleotides listed in Table R1. The template in the pCri11a vector included the SUMO-*nrdR* gene preceded by the his-tagged SUMO protease cleavage site and TEV protease cleavage site (6His-SUMO-TEVcs-SerGlySerGlySer-NrdR construct) was used as a template, and mutated with the “round the horn” PCR method (Munteanu, 2012) with Herculase DNA polymerase (*Agilent*) to generate the corresponding constructs 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR E36A (termed SUMO-NrdR E36A hereafter), 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR E42A and 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR Y131A (SUMO-NrdR E42A and SUMO-NrdR Y131A, respectively). The deletion 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR del132-149 (SUMO-NrdR del132-149) construct was generated by amplification of the SUMO-*nrdR* gene fragment corresponding to SUMO protein and NrdR sequence aa 1-131 by PCR with Pfu Ultra DNA polymerase and cloned into pCri11a with TEV protease cleavage site and a linker (SerGlySerGlySer) that optimized the cleavage added downstream to the existing SUMO protease cleavage site. The quality of all constructs was verified by DNA sequencing.

Expression and purification followed the same protocol as for wild-type NrdR. To remove the imidazole from the HisTrap fractions, change the pH and decrease the salt concentration, a buffer exchange was performed by using a HiTrap^{TM} 5ml desalting column (Cytiva) equilibrated with 50mM Na_2HPO_4 at pH 9.5 (for SUMO-NrdR E36A and SUMO-NrdR E42A), or 500mM NaCl and 50 mM Tris-HCl at pH 9.0 (SUMO-NrdR Y131A). For SUMO-NrdR del132-149, the buffer exchange step was performed by using a HiPrep 26/10 desalting 53 ml column (Cytiva) equilibrated with 500mM NaCl and 50 mM Tris-HCl at pH 8.0. After buffer exchange, TEV protease digestion tests were performed O.N. at 4˚C, at a protease/protein ratios of 1:10, 1:15, 1:20, 1:25 (SUMO-NrdR E36A) or 1:10 in case of SUMO-NrdR E42A or 1:100, 1:50, 1:25, 1:10 (SUMO-NrdR del132-149) or 1:25 (SUMO-NrdR Y131A), 1mM DTT. Cleavage efficiency was assessed by SDS-PAGE and the corresponding bands confirmed by MALDI-TOF Mass Spectrometry (Proteomics and Genomics Service at CIB-Margarita Salas, CSIC, Madrid).

After digestion, the digestion products included the corresponding NrdR mutant (NrdR E36A, NrdR E42A, NrdR Y131A, or NrdR del132-149), the SUMO tag (6His-SUMO-GluAsnLeuTyrPheGln), traces of TEV protease and the undigested fusion, which were loaded on a HisTrap HP 5 ml nickel column (Cytiva) for a re-chromatography step. For NrdR E36A and NrdR E42A, the washing buffer contained 500mM NaCl and 50mM Na_2HPO_4 pH 9.5. Elution of digested NrdR E36A and NrdR E42A mutants was achieved with a gradient with buffer A (500mM NaCl, 50mM Na_2HPO_4 at pH 9.5) and buffer B (with additional 500mM imidazole). In case of NrdR Y131A and NrdR del132-149 the washing buffer contained 500mM NaCl, 50mM TRIS-HCl at pH 9.0 or 8.0, respectively, and the elution buffer contained additional 500mM imidazole.

Cleaved NrdR mutants were collected in the flow through. Traces of TEV protease, undigested fusion and 6His-SUMO-GluAsnLeuTyrPheGln tag coeluted at 500 mM imidazole. Cleaved NrdR E36A and NrdR E42A were lost during the re-chromatography step. Highly pure cleaved NrdR Y131A was concentrated to 5.5 mg/ml without significant losses. Purification of cleaved NrdR del132-149 required 0.25mM ATP and 0.005% non-ionic detergent NP-40 in washing and elution buffers, during the following concentration step required additional 0.25 mM ATP to reach 1.7 mg/ml protein, but with substantial losses and impurities.

A final polishing step consisted in a Size-exclusion chromatography of SUMO-NrdR and SUMO-NrdR mutants. 250 μl of sample at a concentration of 2 mg/ml were injected at RT into Superdex® 200 10/300 GL column (Cytiva) equilibrated with buffer 50 mM Tris at pH 9.0, 500 mM NaCl and 0.25 mM ATP (ratio protein:ATP, 1:5) and run at 0.5 ml/min flow rate. In all purification steps, the purity of digested mutagenic proteins was assessed by 15% SDS-PAGE. Final protein concentration was estimated by the BioRad protein assay.

M5. Oligomeric state analysis of NrdR and NrdR mutants E36A, E42A, Y131A, del132-149 by Size-exclusion Chromatography coupled to Multi-angle Laser Light Scattering (SEC-MALLS)

M5.1 Theoretical aspects of SEC-MALLS method

Analytical Size Exclusion Chromatography (SEC), commonly used to determine the molecular weight of proteins and protein-protein complexes in solution, is a relative technique that relies on the elution volume of the analyte to estimate the molecular weight based on the elution of a set of protein standards. When the protein is not globular or undergoes non-ideal column interactions, the calibration curve based on protein standards is invalid, and the molecular weight determined from elution volume is incorrect. Multi-Angle Laser Light scattering (MALLS) is a technique that determines the absolute molecular weight of an analyte in solution. When light passes through a solvent, interaction with the molecules leads to scattering of the light beam axis, which can be observed by a collection of sensitive detectors at different angles to catch the and different angles of the scattered beamThe presence in the solution of a substance characterized by different refractive index with respect to the solvent will cause an excess of scattering that will depend on the concentration and molar mass of this compound, and for large particles, on the scattering angle.

A MALLS static light detector allows the calculation of the absolute molecular weight (Mw) of isotropic scatterers (smaller than 10-15 nm in radius), which scatters light evenly in all directions. For anisotropic scatterers (radius longer than 10-15nm) both the absolute molecular weight (Mw) and the radius of gyration (Rg) can be calculated, as for larger macromolecules the amount of scattered light varies with the angle of scattering, and so is detected by part of the collection of detectors. The difference in scattering angle depends on the size (strictly, radius of gyration) of the molecule. By having both parameters measured, Mw and Rg, and knowing the theoretical molar mass (Mn), it is possible to determine the shape of the molecule (globular or rod shaped etc.), the oligomerization state, but also the polydispersity value (Mw/Mn), which determines sample homogeneity.

In the first step of sample characterization by this method, the molecules of a sample are separated by Size Exclusion Chromatography and the corresponding parameters measured with a coupled Multi-Angle Laser Light Scattering device (SEC-MALLS), so that the hydrodynamic radius of the measured sample is determined. A UV-light is placed at the end of the SEC run, so the scattering of the samples is measured as the protein molecules leave the column. A refractive index detector is coupled after the chromatography in order to determine the protein concentration of each eluted peak and, based on that, estimate molecular weight values. Since the measurement is performed at each elution volume, SEC-MALLS can determine if an eluting peak is homogeneous (monodisperse) or heterogeneous (polydisperse) and, thereby distinguishing between protein species such as monomers, native oligomers or aggregates, and heterocomplexes. Moreover, coupling MALLS and refractive index detectors to a chromatography system allows to estimate a a very accurate reference baseline, since the buffer is measured before and after the eluted peaks and therefore an elaborated buffer matching required for accurate measurement of samples in batch (measured directly in a cuvette) is avoided.

SEC-MALLS detector was used to determine the molecular weight of both digested and fusion forms of NrdR and the complexes formed with different nucleotides in solution. The same study was performed on fusion forms of NrdR mutants E36A, E42A, Y131A, del132-149. For all conducted experiments it was very important to determine the optimal conditions in which protein-nucleotide complexes were stable during the measurements, which included protein concentration, protein/nucleotide ratio in the sample and ratio of protein in the sample to nucleotide in the elution buffer. The sample preparation protocols are detailed in the following subsections.

M5.2 Absolute molecular weight estimation of NrdR multimers

Multiple Angle Laser Light Scattering (MALLS) experiments were performed at the Protein Purification Facility of the Barcelona Science Park. A DAWN-HELEOS-II detector (Wyatt Technology, USA), used for both scattering and refractive index measurements (required for calculations of molecular mass and protein concentration), equipped with a 664.3nm laser was coupled to a Superdex 200/10/300

Increase column (Cytiva) installed to an ÄKTA Pure Protein Purification System (Cytiva, IL USA) at a flow rate 0.5 ml/min. For NrdR the elution buffers contained 50 mM Tris-HCl at pH 9.0, 250 mM NaCl, 1mM DTT. In the presence of nucleotides, a first set of experiments were carried out in the same buffer additionally containing 0.25mM AMP, dATP or ATP. In a second series the protein was preincubated 3 h at RT with the NTDs at a protein:nucleotide molar ratio 1:20 which were also present in the elution buffer to achieve maximum nucleotide saturation. Injected volumes of the protein were: 80µl (NrdR); 85µl (NrdR +AMP); 60µl (NrdR +dATP), 60µl (NrdR +ATP). SEC data was analyzed using UNICORN 7 (Cytiva, USA). MALS data was analyzed with ASTRA 7 (Wyatt Technology, USA). Calculations were performed by Laura Company Sapiña from the PCB crystallography platform; she also equilibrated the columns and supervised the different SEC-MALLS experiments.

M5.3 Absolute molecular weight estimation of SUMO-NrdR and SUMO-NrdR mutants E36A, E42A, Y131A, del132-149

Estimation of the oligomeric state of the undigested forms of SUMO-NrdR and SUMO-NrdR mutants by SEC-MALLS was performed with Superdex 200 10/300 GL column (Cytiva) under the same conditions and instrumentation described above, and in the presence or absence of dATP and ATP. ADP was tested only for SUMO-NrdR. 200ul of each fusion protein at a concentration of 2mg/ml were injected onto the gel filtration column equilibrated with elution buffer containing 50 mM Tris-HCl pH 9.0, 500 mM NaCl or 50 mM Tris-HCl pH 9.0, 500 mM NaCl, with 0.25mM dATP or 0.25mM ATP for SUMO-NrdR and SUMO-NrdR mutants E36A, E42A, Y131A, del132-149. SUMO-NrdR was also tested with 0.25mM ADP. The proteins were preincubated 3h at RT with the nucleotide (protein:nucleotide molar ratio 1:20). The theoretical molecular weight (Mw) values were calculated by using the ProtParam tool from the expasy website (https://web.expasy.org/protparam/).

M6. Secondary structure analysis of NrdR, SUMO-NrdR and SUMO-NrdR mutants E36A, E42A, Y131A, del132-149 by Circular Dichroism (CD)

M6.1 Theoretical aspects of CD method

Chiral chromophores of the protein molecules will interact differently with right- and left-circularly polarized light (L-CPL and R-CPL) resulting in two related phenomena. The circularly-polarized light rays will travel through different optically active medium with different velocities due to the different indices of refraction, termed optical rotation or circular birefringence. Right- and left-circularly polarized light will also be absorbed to different extents at some wavelengths due to differences in extinction coefficients for the two polarized rays called circular dichroism (CD). Thus a CD spectrum of a molecule of interest typically consists of a scan across a wavelength range and measures the extent of dichroism as a function of wavelength. For proteins and peptides ultra violet (UV) CD spectra in the far UV region (180-260 nm) is typically done to analyze secondary structural features and UV CD spectra analysis in the near UV region (260-300 nm) for tertiary structural features. Hence, in a typical circular dichroism (CD) experiment the difference in absorbance of right- and left- circularly polarized light as a function of wavelength is measured for dichroic sample, which is irradiated with equal amounts of L-CPL and R-CPL. The difference in absorbance (ΔA) is measured as:

$\Delta A=\Delta \varepsilon Cl$

where $\Delta\varepsilon$ is the difference between L-CPL and R-CPL molar extinction coefficients, C stands for the molar protein concentration and l is the path length (in cm) for left- and right-circularly polarized light.

The molar ellipticity θ (milli degrees) is the unit of a single measurement and is defined as:

$\theta=3298.2\Delta\varepsilon$

where $\tan\theta=(ER-EL)/(ER+EL)$

where ER and EL are the L-CPL and R-CPL electric field vector magnitudes of the electromagnetic radiation, which undergo left and right-handed rotation around the wave propagation vector based on the classical electromagnetic theory.

M6.2 Data collection

Circular dichroism experiments were performed in the far UV region (260-195 nm) to check for stability and secondary structure content of wild type NrdR both digested and SUMO-fusion forms and SUMO-NrdR mutants E36A, E42A, Y131A, del132-149. All proteins were subjected to buffer exchange with a buffer containing 5 mM potassium phosphate at pH 7.5, 100 mM potassium chloride. In order to collect CD data a quartz sample cell of 1 mm path length was used for data collection on a Jasco J-815 CD spectrophotometer, with a scan rate of 100 nm/min, data pitch of 1 nm resulting in 100 data points between 260-195 nm. Protein concentration was kept constant at 1 mg/ml for all samples.

Results

R1. Production of NrdR

R1.1 Subcloning of SUMO-NrdR

Initially, the *E.coli nrdR* gene was cloned in *pETite-SUMO (NN´)::gNrdR* vector (Lucigen vector, henceforth) by L. Pedraz (from the group of E. Torrents, IBEC, Barcelona). Thus, the Lucigen vector fused NrdR to a SUMO protein tag that additionally contained an N-terminal six-histidine tag for purification with Ni-NTA resin. According to our collaborators, SUMO significantly stabilized NrdR. However, the cleavage of the tag by the SUMO protease (which recognizes and cleaves a sequence within the SUMO protein) was very inefficient (data not shown). According to the manufacture instructions, one of the advices for efficient cleavage with SUMO protease is to add a linker sequence between SUMO and the protein of interest, thus the construct had a Gly and Ser (GS) between SUMO and NrdR, which the collaborators found to improve the cleavage efficiency but this was not enough for our purposes.

As an alternative, since in the laboratory we regularly used the TEV protease, which usually yielded high amounts of digested protein, we added a TEV protease cleavage site (TEVcs, sequence ENLYFQG) at the N-terminus of the SUMO-NrdR construct. Interestingly, the SUMO protease cleaved this construct with higher efficiency, probably because the TEVcs served as a linker additional to the GS linker and favored the cleavage. However, TEV digestion was still low, from initial 5 mg of SUMO-NrdR, only 0.1 mg of NrdR were obtained. Given that, by increasing the linker length between SUMO and NrdR increased the efficiency, we decided to add a long SGSGS linker between TEVcs and NrdR (Figure R1), which substantially increased the efficiency of the cleavage (see section R1.3). This construct contained a 6 his-tag at the N-terminus, followed by the SUMO protein, the TEVcs, the SGSGS linker sequence and NrdR, and it will be termed SUMO-NrdR hereafter.

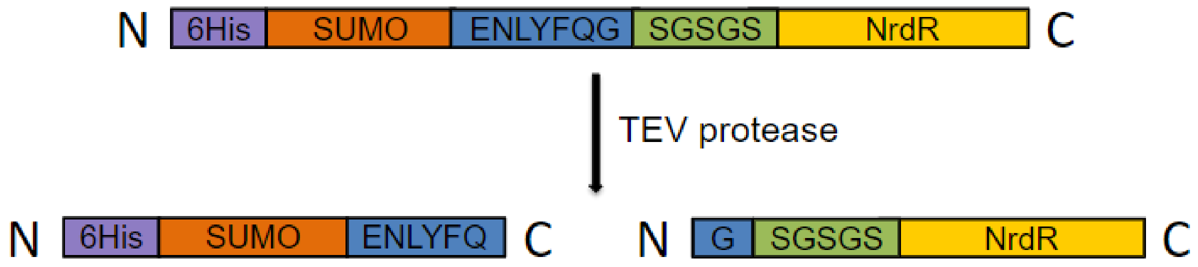


Figure R1. Schematic representation of fusion SUMO-NrdR protein. **Top**, *E.coli* NrdR (yellow box) was fused to a 6His-SUMO tag (in purple and orange, respectively), followed by the TEV cleavage site (in blue) and a 5 aa linker to enhance cleavage (in green). **Bottom**, products of the TEV digestion, note the 3(GlySer)-NrdR digestion product on the right.

R1.2 Expression and purification of SUMO-NrdR

SUMO-NrdR was expressed in *E.coli* BL21 (DE3) as previously done by our collaborators (see conditions in materials and methods, section M1.3), and solubilized in the lysis buffer consisting of 50 mM Tris-HCl pH 9.0. Note that the isoelectric point of the fused proteins was pI=6.38, and a buffer with a pH 1.5 points higher or lower was necessary. However, the pI of the NrdR construct alone was pI=7.85, so we followed carefully the behavior of the protein to make sure it was not precipitating. Likewise, because NrdR is a DNA binding protein, the buffer contained high salt content to prevent undesired non-specific protein/DNA complexes (see materials and methods and Figure R2A).

The first purification protocol step consisted in a Ni-NTA affinity chromatography. The protein was eluted with a single step of 200 mM imidazole that produced a rather narrow elution peak that showed a high optical density at 280 nm of 680 mAu, indicating high yields (Figure R2B). However, analysis by SDS-PAGE showed that these central fractions contained a mixture of fusion NrdR and other proteins or protein fragments (Figure R2C, MW of SUMO-NrdR, 31.23 kDa). Fractions containing the protein sample were pooled and the concentration measured, showing a final yield of 9 mg/ml in 6 ml (54 mg of the protein from 500ml culture), yet the samples were not highly pure. Besides, we observed that the SUMO-NrdR eluting fractions showed a ratio of absorbances A260/A280 of 0.8. In general terms, the ratio of absorbances A260/A280 of a protein in solution is equal or below 0.7 (or, A260 is 70% or less with respect to A280). A ratio higher than 0.7, is indicative of the presence of additional molecules that absorb at 260 nm, such as (poly-)nucleotides. NrdR binds DNA and the results of the collaborators suggested that dNTP/NTP ratio regulates the activity of the protein, so the high A260/2A80 could reflect the presence of an NTP or dNTP bound to NrdR. The possibility that the protein had bound to nucleic acids was low since such DNA or RNA binding results in much higher A260/A280 ratios and, in addition, the samples had been extensively sonicated, which breaks long DNA or RNA molecules, and exposed to high salt concentration, which disrupts protein/DNA complexes. The bands were extracted from the SDS-PAGE, and their content analyzed by fingerprint mass spectrometry, which confirmed the presence of NrdR at the two thickest bands indicated by the black arrow (full-length) and asterisk (degraded) in Figures R2A, C, R3A, B. Therefore, SUMO-NrdR was available but it showed partial degradation.

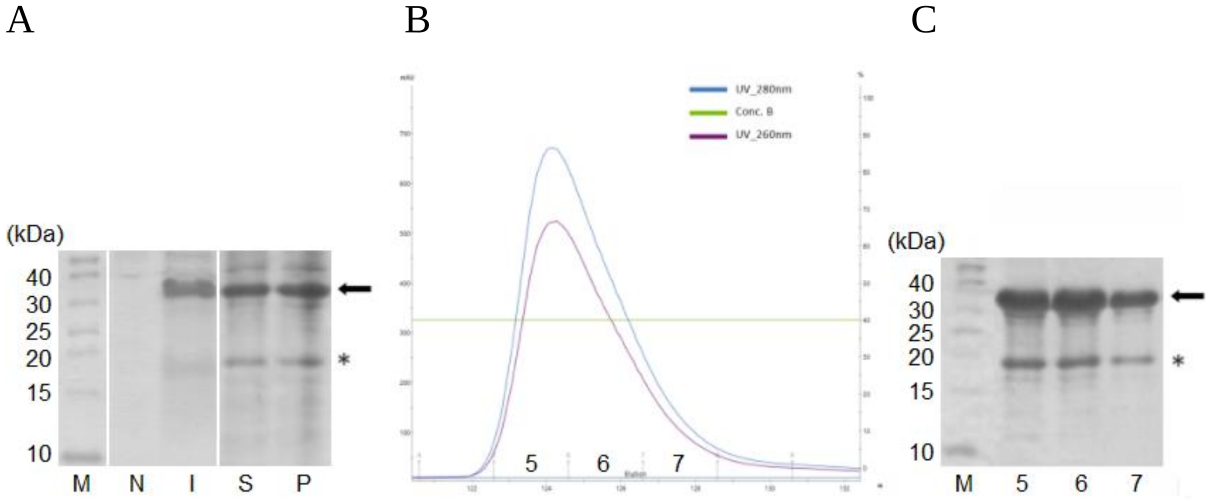


Figure R2. **High scale expression, solubilisation and purification of SUMO-NrdR analyzed by 15% SDS-PAGE**. **(A)** Induced (I), non-induced (N), soluble (S), pellet (P) samples of SUMO-NrdR expressed in *E.coli* BL21(DE3). The black arrow indicates the expected MW of SUMO-NrdR. The asterisk corresponds to bands of partially degraded protein. **(B)** Elution profile of SUMO- NrdR from a Ni^{2+}-NTA column (left axis, optical density (OD) at 280 nm, blue curve, and at 260 nm, purple curve; right axis, imidazole concentration, green line). Note the elution was done at a constant imidazole concentration of 200mM. The A260/A280 ratio equal to 0.8 suggests presence of nucleotides bound to the protein (see text). **(C)** 15% SDS-PAGE analysis of the IMAC elution peak showing the presence of SUMO-NrdR in fractions 5-7**.** Black arrow and asterisk as in **A**.

R1.3 TEV proteolytic cleavage, purification, concentration and quality check of NrdR

Next step consisted in digestion of the SUMO-NrdR fusion protein followed by a re-chromatography purification by Ni-NTA to separate NrdR from the rest of digested and undigested products. However, for an efficient cleavage reaction, it was necessary to extract the imidazole from the protein sample and decrease the NaCl concentration. For that purpose, a buffer exchange with buffer 50 mM Tris-HCl at pH 9.0 and 500mM NaCl was performed by using a disposable ultrafiltrator. The pH 9.0 was higher than the theoretical value (as calculated with ProtParam (https://web.expasy.org/protparam/), pI=6.38 for SUMO-NrdR, pI=6.15 for SUMO tag and pI=7.85 for digested NrdR, so that the protein stability was maintained throughout the whole purification procedure.

Several trials of digestion of SUMO-NrdR with TEV protease were performed overnight (O.N.) at 4˚C, at 500mM NaCl and 1mM DTT with increasing protease/protein ratio 1:1000, 1:500, 1:250 (Figure R3A), and 1:100, 1:50, 1:25 (Figure R3B). The resulting samples were run in 15% SDS-PAGE, and showed that protein was not cleavable at protease:protein ratio lower than 1:250, and at ratio higher than 1:100 the cleavage was directly proportional to the amount of protease, with the highest digestion efficiency obtained at a protease:protein ratio 1:25 (Figure R3B). To assess the content of the bands upon digestion, they were analyzed by MALDI-TOF Mass Spectrometry.

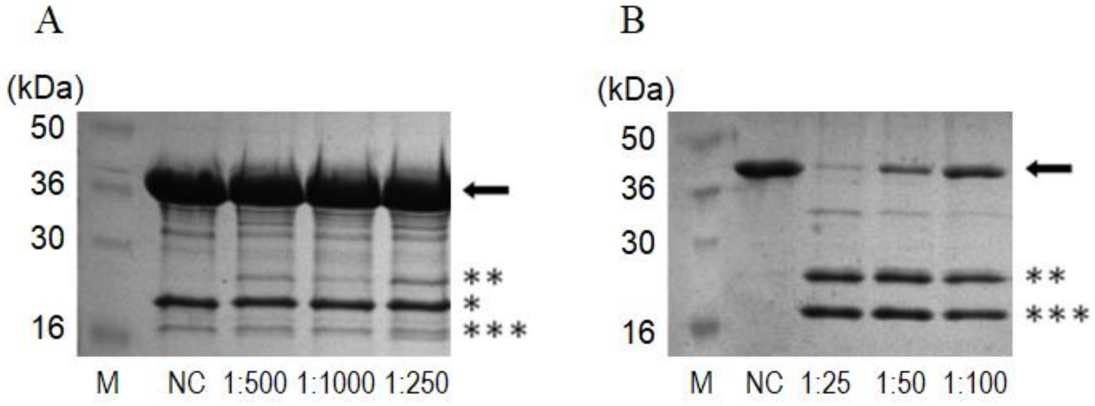


Figure R3. TEV protease digestion trials of SUMO-NrdR checked by 15% SDS-PAGE. Protease:protein w/w ratios tested are indicated below each line. Not successful trials are shown in panel **A**, while panel **B** shows positive results at higher ratios. The highest, almost fully efficient cleavage was observed at 1:25 protease:protein w/w ratio. The black arrow indicates the expected bands of SUMO-NrdR. An asterisk corresponds to partially degraded protein, double asterisk the His-tagged SUMO-GluAsnLeuTyrPheGln (approx. 20 kDa), and triple asterisk the 3(GlySer)-NrdR (approx. 17 KDa). Lanes NC (not cleaved), correspond to undigested SUMO-NrdR loaded as negative control.

After digestion (Figure R4A), a Ni-NTA re-chromatography was run. The digestion sample was expected to have 3(GlySer)-NrdR, 6His-SUMO-GluAsnLeuTyrPheGln tag, traces of TEV protease and traces of undigested SUMO-NrdR (Figure R4B). In this second chromatography, 3(GlySer)-NrdR eluted in the flow through during sample loading, therefore the flow-through was fractionated to avoid excessive dilution of the protein, and all fractions checked on 15% SDS-PAGE (Figure R4C). Such a reverse chromatography rendered highly pure 3(GlySer)-NrdR so that all flow through fractions were concentrated to 500 ul, which reached a concentration of 16mg/ml, optimal for further experiments.

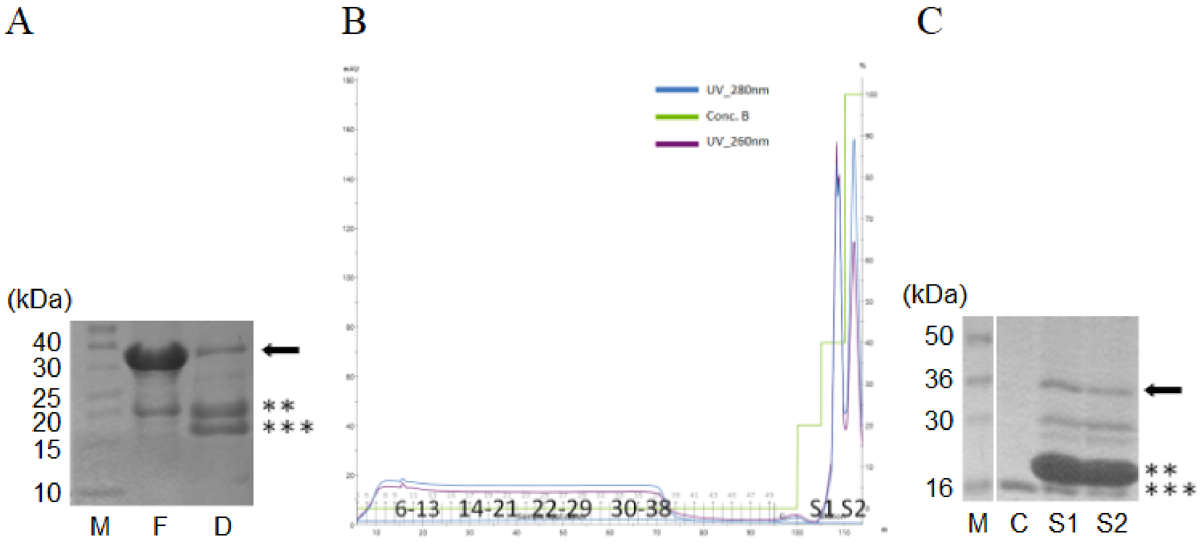


Figure R4. Ni-NTA affinity re-chromatography and concentration step of digested NrdR. **(A)** 15% SDS-PAGE showing the fusion SUMO-NrdR construct (approx. 35kDa, black arrow) in lane F, and the TEV digestion product in lane D. The two lower intense bands correspond to His-tagged SUMO-GluAsnLeuTyrPheGln (approx. 20kDa) (double asterisk) and 3(GlySer)-NrdR (approx. 17kDa) (triple asterisk) **(B)** Reverse His-Trap affinity chromatography of the digested protein sample. 3(GlySer)-NrdR was collected in the fractionated flow-through (fractions 6 to 38), which showed a A260/A280 ratio equal to 0.9, likely caused by nucleotide moiety(es) bound to the protein (see text). The 280 nm (in blue) and 260 nm (in purple) UV light curves and concentration (% volume) of buffer B containing imidazole (green line) are shown. His-tagged components (SUMO protein, TEV protease and non-digested fusion protein) eluted at the last chromatographic steps with 40 and 100% buffer B (200 and 500 mM imidazole, respectively). **(C)** 15% SDS-PAGE with the protein from the flow through (“to be concentrated”) shown in lane C and his-tagged components eluted at 200 and 500mM of imidazole are shown in lanes S1, S2 respectively. Asterisks indicate the same as in **A**.

The next step consisted in a gel filtration of NrdR. This required concentration of the sample, which was usually stored O/N at 4º since the previous chromatography. Both storage of highly concentrated samples and further concentration makes proteins highly susceptible to proteases and hence degradation. SDS-PAGE analysis showed that no degradation occurred during concentration of the protein, indicating the absence of proteases.

The following gel filtration was initially done without the presence of any nucleotide, which caused a complete loss of protein, which did not even elute, and thus revealing the instability of NrdR without the ligand. The same chromatography in the presence of excess of ATP (0.5 mM), both preincubated with the sample and in the equilibration buffer of the column, showed a peak with a retention volume of 8.8 ml (Figure R5A), much before than expected (15 ml) and inconsistent for a 17.66 kDa globular protein unless it was highly expanded or assembled in multimers. Interestingly, the peak showed an OD ratio 260/280nm higher than 2, while the same parameter in the flowthrough of the previous NrdR purification (Figure R4B) was equal to 0.9, further suggesting binding of the ATP to NrdR. Since this initial peak was very small (OD at 280 nm of 20 mAu, for an injection of 100 μl at 2 mg/ml), the collected fractions were concentrated and loaded on a 15% SDS-PAGE, which confirmed the presence of intact and pure NrdR (Figure R5B). The protein was present in all fractions, with high 260 nm absorbance, which pointed to a nucleotide strongly bound to the polypeptide.

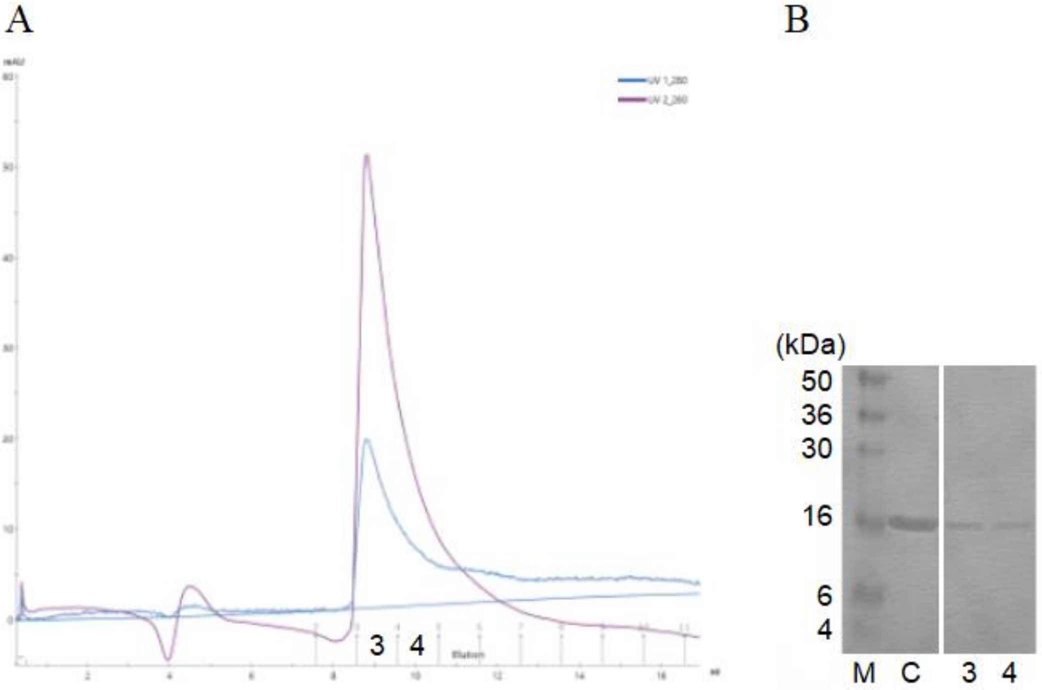
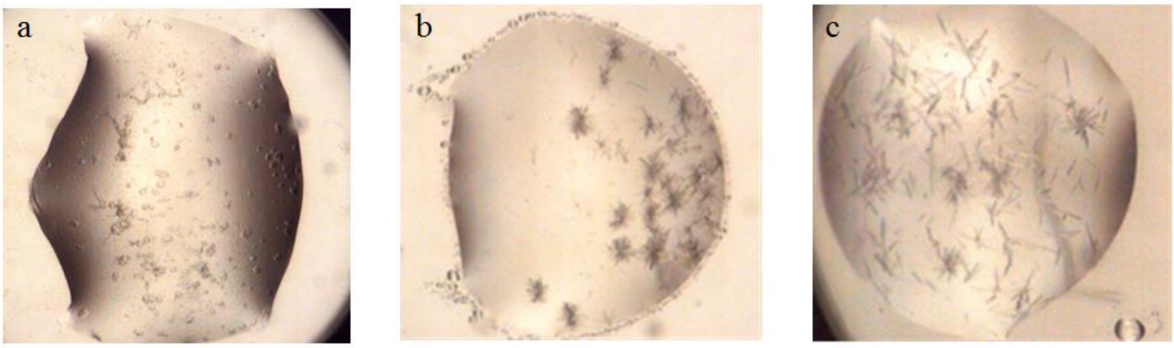


Figure R5. Size exclusion chromatography of NrdR in the presence of ATP . **(A)** Gel filtration in a Superdex 200 10/300 GL showed a retention volume of 8.8 ml, suggesting protein oligomerization, and a much higher absorbance value of 260 over 280 nm (A260/A280>2). **(B)** 15% SDS-PAGE analysis confirmed the presence and purity of digested NrdR in eluted fractions 3 and 4. C, concentrated sample from the collected peak.

R2. NrdR crystallization trials and diffraction tests of protein crystals

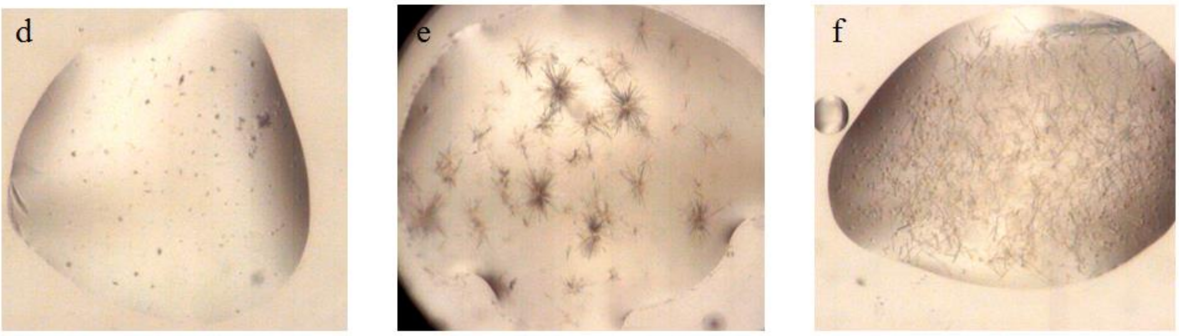
Initial crystallization trials of NrdR were performed by automated dispensing of the crystallization drops by using dedicated robots (see materials and methods). First crystallization conditions tested were the ones of the screens PAC1, PAC2, *Protein-DNA screen* PAC10 and PAC19. The drop volumes used in such an automated crystallization set up were 0.1µl:0.1µl of protein:crystallization solutions. Such small volumes facilitate testing large chemical conditions screenings by using very low amounts of samples. After three weeks, the analyses of the crystallization drops showed several conditions of screen PAC10 in which needle- and cubic-shaped crystals appeared. Examples are shown in Figure R6, cubic crystals in panel *a* and *d*, and needles in the rest of panels. Note that all conditions contained low PEG or alcohols, dimethyl sulfoxide (DMSO), sulfate or 2-methyl-2,4-pentanediol, which are typically crystallization conditions for DNA. Therefore, some doubts came across as the purified protein always had a high A260/A280 ratio, thus the sample could have contained DNA that crystallized in the analyzed drops.



PEG 400 10.0%v/v
MES 0.05M, pH 6.0
BaCl2 0.005M

Propanediol 30.0%v/v
TRIS 0.05M pH 7.5

DMSO 20.0%v/v
TRIS 0.05M pH 8.0



Li2SO4 1.5M
HEPES 0.1M pH 7.5

MPD 12.0%w/v
TRIS 0.1M, pH 8.5
MgCl2 0.05M

Isopropanol 15.0%v/v
KCl 0.2M, MgCl2 0.01M
TRIS 0.05 M pH 7.5

Figure R6. Crystallization of NrdR. The initial screens showed the tendency of NrdR to form needles (b, c, e, f) or cubic shaped crystals (a, d), which appeared after three weeks.

Crystal needles from conditions containing 50mM Tris-HCl pH 7.5, 30% v/v propanediol and 50mM Tris-HCl pH 8.0, 20% v/v DMSO (shown in panels *b* and *c* from Figure R6, respectively) were fished with small nylon loops and were vitrified in liquid nitrogen. For this, a cryo-protection step was required by adding a vitrification buffer containing 50mM Tris-HCl pH 7.5, 30% v/v propanediol, 20% v/v glycerol, or 50mM Tris-HCl pH 8.0, 20% v/v DMSO and 20% v/v glycerol, respectively. Crystals from the two initial PAC10 conditions were tested at the ALBA synchrotron (Cerdanyola del Vallès), and showed diffraction at 10 Å resolution with a diffraction pattern characteristic for protein or DNA crystals, so we could confirm they were not salt crystals (Figure R7).

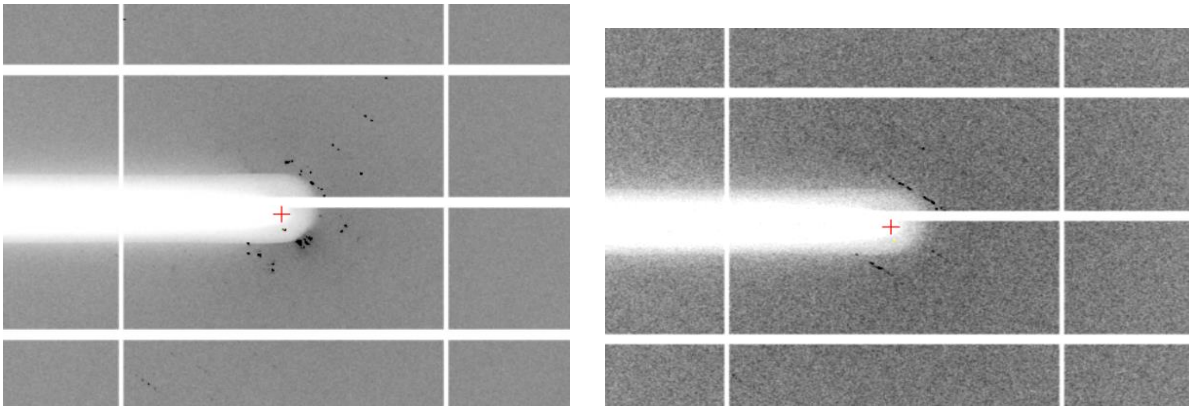


Figure R7. Diffraction pattern obtained from NrdR crystals. Crystals from crystallization conditions indicated at panel R6b **(on the left)** and R6c **(on the right)** showed low resolution around 10 Å, yet they confirmed X-ray diffraction of macromolecules, protein or DNA.

As the diffraction resolution was too low, we optimized the crystallization conditions by modifying the parameters of the protein amounts and the crystallization solution. Thus, the optimization of those two successful conditions at panels b and c from Figure R6 (Propanediol-Tris, and DMSO-Tris, respectively) was performed at larger volumes to stimulate the growth of bigger crystals at room temperature (RT) and at 4˚C, by the vapor-diffusion sitting-drop method using 48-well double drop plates (Hampton Research) which feature two drop containers arranged vertically, upper (hereby called first) and lower (second). Within the plates, the first drop contained equimolar volumes of protein and crystallization solutions (1µl and 1µl, respectively), whereas the second drop contained double volume of NrdR over volume of crystallization solution (2µl and 1µl, respectively) to increase the protein concentration in the final 3 ul volume of crystallization drop.

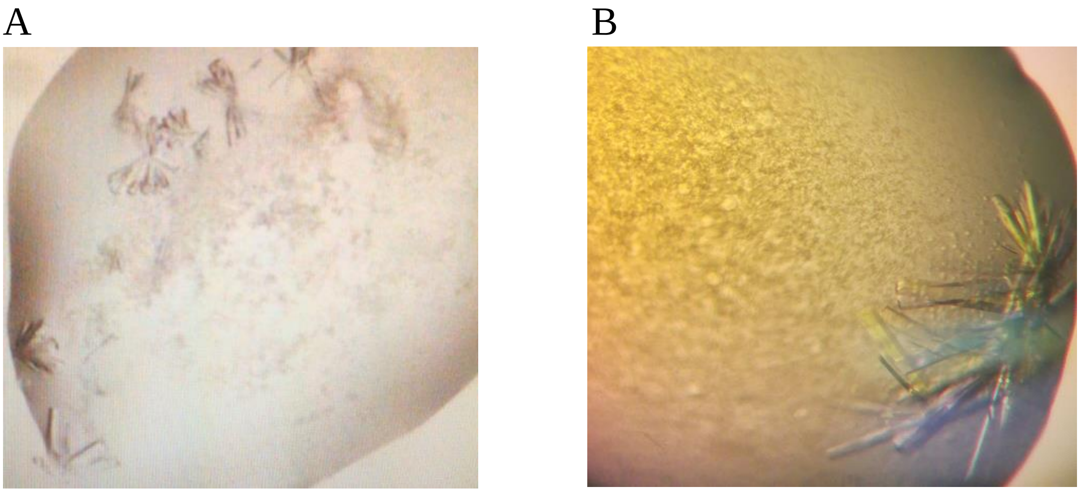
In the crystallization drop, the excess of water evaporates and is captured by the much bigger volume of the reservoir solution. Therefore, once all the water evaporated, the final volume is assumed to be the volume of the crystallization solution. And thus, if the initial protein solution volume was 2:1 with respect to the crystallization solution, the final protein concentration will be twice with respect to the initial one. With this approach we further explored more crystallization conditions.

For conditions Propanediol-Tris and DMSO-Tris, we set up optimization screenings by hand with bigger volumes. The parameters of the crystallization solution varied in a grid screen of 4 rows (top to bottom) and 6 columns (left to right). For condition Propanediol-Tris, in which the pH ranged from 7.5, 8, 8.5 to 9.0, ongoing from top to bottom of the plate, the concentration of the precipitant (%v/v of Propanediol) ranged from 20, 22, 24, 26, 28 to 30, ongoing from left to right. For condition DMSO-Tris, in which the pH also ranged from 7.5, 8, 8.5 to 9.0, ongoing from top to bottom of the plate, the concentration of the precipitant (%v/v of DMSO) ranged from 8, 10, 12, 14, 16 to 18, ongoing from left to right. Surprisingly, this optimization caused precipitation to all drops.

Unexpectedly, after six weeks, micro crystals appeared from the different initial high throughput screening condition from (PAC19) in 0.15M KSCN, 0.1M Tris-HCl at pH 7.5, 15% v/w PEG6000 (Figure R8A). Optimization screens of this condition were performed at room temperature (RT) and at 4˚C, by the vapor-diffusion sitting-drop method using 48-well double drop plates (Hampton Research). Within the plates, the first drop contained equimolar volumes of protein and crystallization solutions (1µl and 1µl, respectively), whereas the second drop contained double volume of NrdR over volume of crystallization solution (2µl and 1µl, respectively) to increase the protein concentration in the final 3 ul volume of crystallization drop.

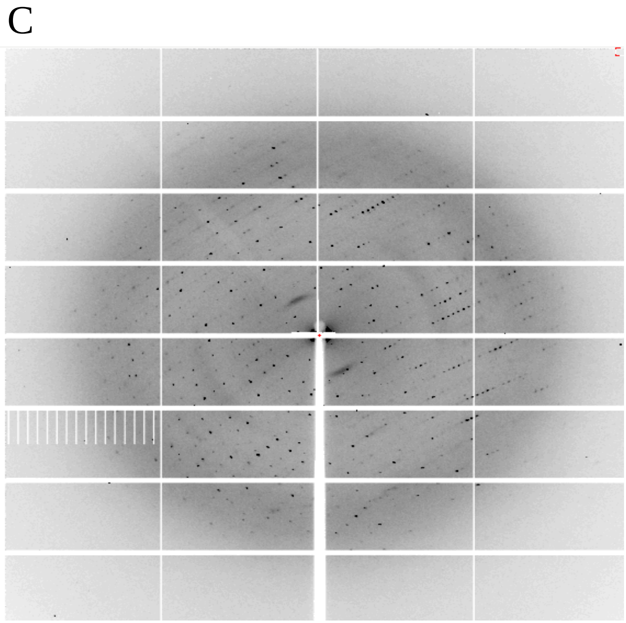
Optimization screenings of PEG6000-Tris condition were set up by hand with bigger volumes. The parameters of the crystallization solution varied in a grid screen of 4 rows (top to bottom) and 6 columns (left to right) and were designed as following: i) screen, in which the pH ranged from 7.5, 8, 8.5 to 9.0, ongoing from top to bottom of the plate, with salt concentration (molarity of KSCN) varying from 0.05, 0.10, 0.15, 0.20, 0.25 to 0.3, ongoing from left to right, at a constant 15% v/w PEG6000 concentration; ii) two screens, in which the pH ranged from 7.5, 8, 8.5 to 9.0, ongoing from top to bottom of the plate, with precipitant concentration (%v/w PEG6000) varying from 7, 9, 11, 13, 15 to 17 (first screen) and from 19, 21, 23, 25, 27 to 29 (second screen), ongoing from left to right, at a constant 0.15M KSCN concentration.

Optimization screenings resulted in three new successful conditions. One of them, consisting in 0.15M KSCN, 0.1M Tris-HCl at pH 9.0, 7% v/w PEG6000 (Figure R8B) rendered crystals with a bar shape, which appeared after three weeks. Crystals were fished with small nylon loops and were vitrified in liquid nitrogen. For this, a cryo-protection step was required by adding a vitrification buffer containing 0.15M KSCN, 0.1M Tris-HCl at pH 9.0, 7% v/w PEG6000, 20% v/v glycerol. Crystals were tested at the German Electron Synchrotron (DESY, Hamburg).



15% v/w PEG 6000, 0.1M Tris-HCl pH 7.5, KSCN 0.15M

7% v/w PEG 6000, 0.1M Tris-HCl pH 9.0, KSCN 0.15M



diffraction up to 2.35Å resolution

Figure R8. Optimization of crystals and X-ray diffraction pattern. **(A)** Crystallization drop of the indicated condition in a nano-scale plate. Note the presence of short bars and crystalline precipitation. **(B)** Optimization of the crystallization condition in A in bigger volumes set up manually resulted in much bigger and larger bars. **(C)** Diffraction pattern of crystals from crystallization drop in panel (B). This is an image collected at the characterization step of crystals during data collection.

R3. X-ray data collection of Single-wavelength Anomalous Diffraction (SAD) and data processing

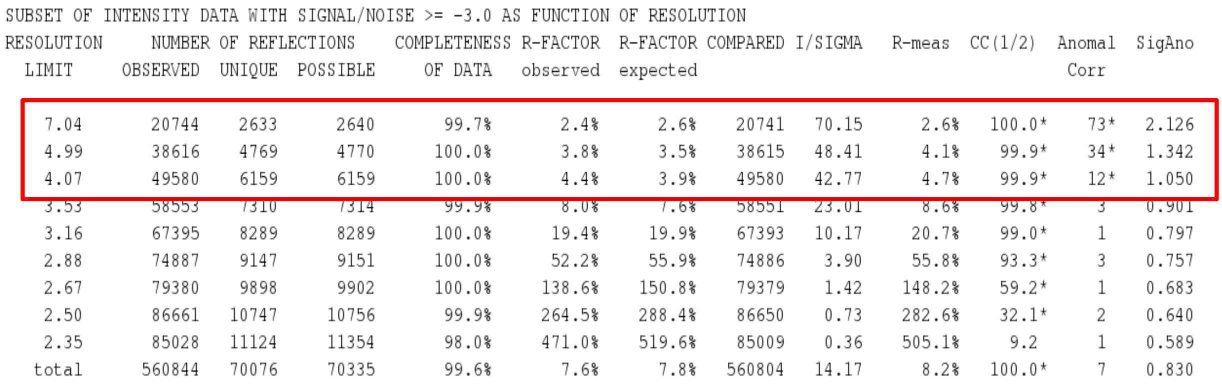
X-ray diffraction data from the optimized *E. coli* NrdR crystals mentioned in the previous section (condition 0.15M KSCN, 0.1M Tris-HCl at pH 9.0, 7% v/w PEG6000, Figure R8B) were collected at the zinc absorption edge (1.28Å), since NrdR contains a Zn-finger domain that suggested a Zn bound to it, which was backed up by the fact that Zn was required during protein expression in *E. coli*. Ten X-ray datasets were collected from six crystals, including three datasets collected from different points of the same crystal. Crystals diffracted at nominal resolutions ranging from 2.35 to 2.55 Å. Those crystals that rendered highest resolution and whose data was subsequently used for structure determination, the data collection was for 360º in an oscillation range with 0.1º.

All datasets were processed with the XDS package (described in Materials and Methods). During processing, the first step was to determine the maximal resolution at which the data was useful, by following several criteria. The correlation coefficient CC1/2 compares the correlation of the reflections in two data set halves, but also the intensity of reflections at the last resolution shell (minimum I/s/>1.5, or close to it). Another parameter to take into account is the overall residual factor R_f. Processing of different data sets showed similar resolutions for a CC1/2 of 50, ranging from 2.7 to 2.8 Å resolution (Table R1), except for a dataset collected at nominal resolution of 2.35 Å that yielded a CC1/2 of 50 at 2.6 Å resolution. At this step, the resolution range was not yet fixed as this would be done in the next scaling step with the CCP4 suite, so the data was processed until 2.35 Å resolution. During indexing, all data sets showed that crystals belonged to the orthorhombic symmetry system $P2_x2_x2_x$. All crystals showed a crystallographic unit cell with highly similar parameters that ranged such as a = 51.098-51.661 Å, b = 255.291-257.773 Å and c = 132.785-134.500 Å, and α = β = γ = 90º (Table R1).

The data processed with XDS was scaled with XSCALE. Scaling of the data indicated that data set 7_1 showed the highest resolution. Interestingly, this data set showed an anomalous signal above 1 at 4.07Å, as indicated in Table R2. This data was then converted with XDSCONV to the mtz format, which is appropriate for the CCP4 suite used to solve the structure. In order to increase the completeness of the data set, three data sets 7_1, 7_2, 7_3 collected from a single crystal (Table R1) were merged all together and scaled by XSCALE (Table R3). The other seven datasets were not used for structure determination or further model improvement.

Name	Resolution [Å] at CC1/2=50	Space group	Unit cell parameters: axes a, b and c [Å]; angles α, β, γ [˚]
4_1	2.7	C 2 2 21	51.541 257.311 134.172 90.000 90.000 90.000
4_2	2.8	C 2 2 21	51.547 257.773 134.500 90.000 90.000 90.000
5_1	2.75	C 2 2 21	51.661 256.913 133.951 90.000 90.000 90.000
6_1	2.75	C 2 2 21	51.106 255.291 132.785 90.000 90.000 90.000
7_1	2.6	C 2 2 21	51.185 255.368 132.940 90.000 90.000 90.000
7_2	2.7	C 2 2 21	51.098 255.765 133.053 90.000 90.000 90.000
7_3	2.8	C 2 2 21	51.457 257.005 133.771 90.000 90.000 90.000
8_1	2.7	C 2 2 21	51.631 257.152 134.163 90.000 90.000 90.000
10_1	2.7	C 2 2 21	51.443 256.932 134.027 90.000 90.000 90.000
10_2	2.8	C 2 2 21	51.454 257.288 133.955 90.000 90.000 90.000

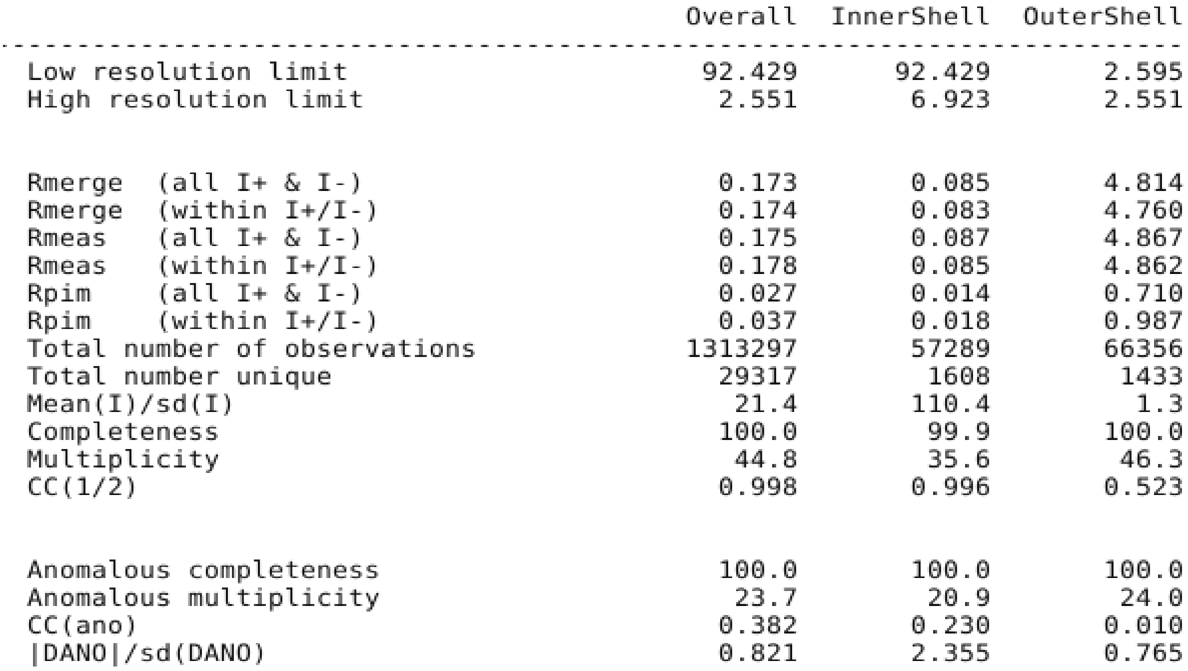
Table R1. List of all collected datasets processed by XDS. The resolution corresponds to the one at which the correlation coefficient value ($CC_{1/2}$) at the last resolution shell equals 50. The assigned space group, and unit cell parameters (axes a, b and c, and angles α, β, γ) are indicated. The orange cells indicate the best dataset with a $CC_{1/2}$ equal to 50 at 2.6 Å resolution.



```
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION
RESOLUTION     NUMBER OF REFLECTIONS    COMPLETENESS R-FACTOR  R-FACTOR COMPARED I/SIGMA   R-meas  CC(1/2)  Anomal  SigAno
  LIMIT     OBSERVED  UNIQUE  POSSIBLE     OF DATA   observed  expected                                      Corr

    7.04       20744    2633      2640       99.7%       2.4%      2.6%    20741   70.15      2.6%   100.0*    73*   2.126
    4.99       38616    4769      4770      100.0%       3.8%      3.5%    38615   48.41      4.1%    99.9*    34*   1.342
    4.07       49580    6159      6159      100.0%       4.4%      3.9%    49580   42.77      4.7%    99.9*    12*   1.050
    3.53       58553    7310      7314       99.9%       8.0%      7.6%    58551   23.01      8.6%    99.8*     3    0.901
    3.16       67395    8289      8289      100.0%      19.4%     19.9%    67393   10.17     20.7%    99.0*     1    0.797
    2.88       74887    9147      9151      100.0%      52.2%     55.9%    74886    3.90     55.8%    93.3*     3    0.757
    2.67       79380    9898      9902      100.0%     138.6%    150.8%    79379    1.42    148.2%    59.2*     1    0.683
    2.50       86661   10747     10756       99.9%     264.5%    288.4%    86650    0.73    282.6%    32.1*     2    0.640
    2.35       85028   11124     11354       98.0%     471.0%    519.6%    85009    0.36    505.1%     9.2      1    0.589
   total      560844   70076     70335       99.6%       7.6%      7.8%   560804   14.17      8.2%   100.0*     7    0.830
```

Table R2. XSCALE output from the dataset collected at the resolution limit of 2.35Å. At each resolution shell, the total number of collected reflections, number of unique reflections, completeness of data, R-factors (R-Factor, observed and expected), signal to noise ratio (I/SIGMA), residual R between equivalent measurements (R-meas), the correlation coefficient between two halves of the data set CC(1/2), the anomalous correlation (Anomal Corr), and anomalous sigma (SigAno) are indicated. The shells with anomalous signal are indicated with a red frame.



```
                                           Overall  InnerShell  OuterShell
-------------------------------------------------------------------------
 Low resolution limit                       92.429     92.429      2.595
 High resolution limit                       2.551      6.923      2.551

 Rmerge  (all I+ & I-)                       0.173      0.085      4.814
 Rmerge  (within I+/I-)                      0.174      0.083      4.760
 Rmeas   (all I+ & I-)                       0.175      0.087      4.867
 Rmeas   (within I+/I-)                      0.178      0.085      4.862
 Rpim    (all I+ & I-)                       0.027      0.014      0.710
 Rpim    (within I+/I-)                      0.037      0.018      0.987
 Total number of observations              1313297      57289      66356
 Total number unique                         29317       1608       1433
 Mean(I)/sd(I)                                21.4      110.4        1.3
 Completeness                                100.0       99.9      100.0
 Multiplicity                                 44.8       35.6       46.3
 CC(1/2)                                     0.998      0.996      0.523

 Anomalous completeness                      100.0      100.0      100.0
 Anomalous multiplicity                       23.7       20.9       24.0
 CC(ano)                                     0.382      0.230      0.010
 |DANO|/sd(DANO)                             0.821      2.355      0.765
```

Table R3. XSCALE output from the data set obtained by merging and scaling three data sets 7_1, 7_2, 7_3 collected from a single crystal at the resolution limit of 2.55Å. Overall, at inner and outer resolution shell, the total number of collected reflections, number of unique reflections, completeness of data, multiplicity, signal to noise ratio (Mean (I)/ sd (I)), agreement between multiple measurements of a given reflection (R-merge), residual R between equivalent measurements (R-meas, Rpim), the correlation coefficient between two halves of the data set CC(1/2), the anomalous completeness, the anomalous multiplicity, the anomalous correlation (CC (ano)), and anomalous signal to noise ratio (|DANO|/ sd (DANO)) are indicated. Note the high multiplicity for the anomalous data.

R4. NrdR crystal structure solution and refinement

R4.1 Calculation of Matthews coefficient

The Matthews coefficient corresponding to the unit cell for a protein of 17.66 kDa indicated the following result:

Cell volume: 1753822.000

For given protein molecular weight: 17660 Da

Nmol/asym	Matthews Coeff.	%solvent	P(2.60)	P(tot)
1	12.41	90.10	0.00	0.00
2	6.21	80.20	0.00	0.00
3	4.14	70.29	0.01	0.02
4	3.10	60.39	0.14	0.15
5	2.48	50.49	0.48	0.47
6	2.07	40.59	0.34	0.34
7	1.77	30.68	0.01	0.02
8	1.55	20.78	0.00	0.00
9	1.38	10.88	0.00	0.00
10	1.24	0.98	0.00	0.00

Table R4. Matthews Coefficient calculations performed with ccp4 for dataset 7_1. Number of molecules in the asymmetric unit (Nmol/asym), the crystal volume per unit of protein molecular weight (Matthews Coeff), solvent content in the crystal (%solvent), probability at 2.6 Å (P(2.60)) and probability P(tot) are indicated.

Therefore, in the asymmetric unit four to seven different molecules related by non-crystallographic symmetry could coexist, as typically the solvent content in protein crystals ranges from 27% to 65%, with an average of 43%. However, seven molecules would have a very low solvent content, of 30%, making it improbable, which suggested a more realistic number of molecules in the asymmetric unit up to six.

R4.2 Structure solution by Molecular Replacement

Please note that, at the time, the structure of *S. coelicolor* NrdR (Grinberg et al. 2022) was not available. Another possibility would have been with the recent advent of the AlphaFold algorithm (REF), which provides almost all structures of domains of human proteins and model organisms. However, at that time AlphaFold was not available either. A cone domain was predicted as a C-terminal domain of NrdR, similar to the one found at RNRs. However, similarity between cone domains of RNRs and NrdR is low. Searches of similar sequences with Blast throughout the Protein Database to find the most similar ATP-cone domains structures, yielded *P.aeruginosa* and *E.coli* NrdA molecules. The similarity was 15.3% and 22.2% for ATP-cone domains 1 and 2, of *P.aeruginosa* NrdA, respectively (https://www.uniprot.org/blast/?about=Q9I4I1[32-132]&key=Domain, https://www.uniprot.org/blast/?about=Q9I4I1[148-237]&key=Domain, PDBid: 5IM3), and 30.2% for ATP-cone domain of *E. coli* NrdA (https://www.uniprot.org/blast/?about=P00452[5-95]&key=Domain, PDBid: 4ERM).

On the other hand, similarity between the N-terminal zinc finger domain of NrdR and randomly chosen proteins was lower than 5%. This was reflected by unsuccessful trials of molecular replacement searches that included cone domains (PDBid: 5IM3, 4ERM) and zinc finger domains (PDBid: 5IJ4, 2ELY) as searching models. Searches were done with the domains independently. Moreover, another molecular replacement strategy was based on predicted three-dimensional structural model of *E. coli* NrdR (https://www.uniprot.org/uniprot/P0A8D0.fasta) by I-TASSER (Iterative Threading ASSEmbly Refinement), (https://zhanggroup.org/I-TASSER/). Therefore, a full-length model predicted by the provided algorithm (Figure R9A), but also the zinc finger or the ATP-cone domain separately were used for MR in search for structure solution. However, this approach was also unsuccessful.

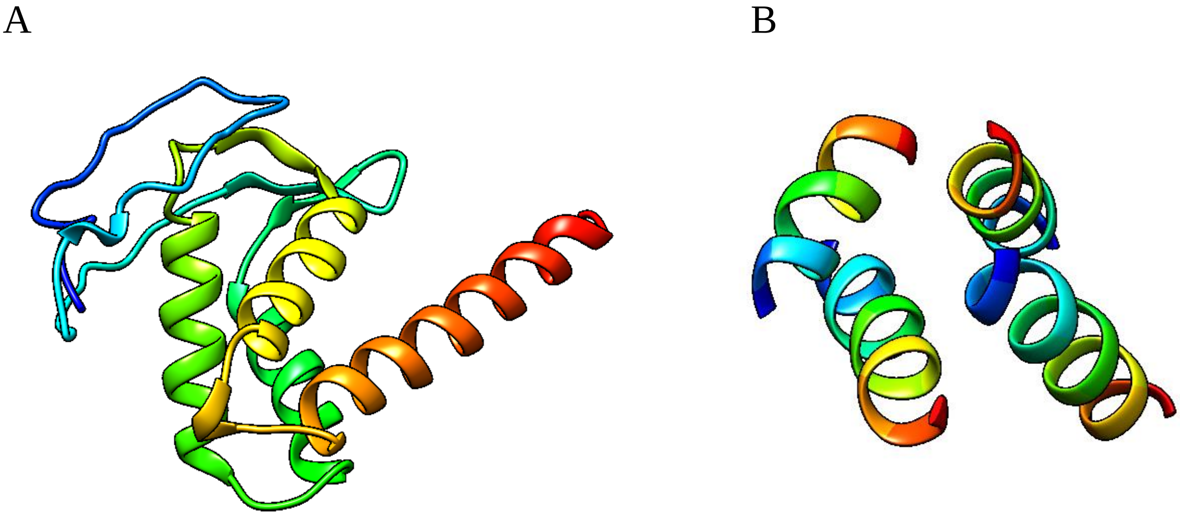


Figure R9. ***E. coli*** **NrdR predicted models of the domains. (A) NrdR A-cone domain predicted by I-TASSER.** This model was used for MR in search of a structure solution, but it was unsuccessful. **(B) Shredded model of ATP-cone (only alpha helices).** Alpha helices of ATP-cone domain used as an input in AUTOBUILD improved the initial model from 284 up to 497 residues.

Therefore, the only alternative to solve the structure was by finding positions of heavy atoms present in the crystal, by the SAD Experimental Phasing method (see materials and methods paragraph 3.1.4), as SAD data sets were collected at the Zn absorption edge.

R4.3 Structure solution by Single-wavelength Anomalous Diffraction Experimental Phasing

The crystal structure of *E. coli* NrdR was solved by SAD (single anomalous diffraction) experimental phasing - The single dataset from crystal 7_1 was sufficient to find the positions of the anomalous scatterer, in our case the Zn ions supposed to be bound to Zn fingers of NrdR. Initially, the stepwise approach was performed with PHASER, which consists in determining first the position of the metal atoms, followed by assignment of the phases to all reflections, and secondly once the map is calculated a density modification to discard spurious density at regions that should correspond to the solvent (which should be flat) is performed. Usually, at this point the map is not very clear, since due to the phase error the Fourier synthesis depicts density at regions that do not correspond to the protein, but to solvent in the crystal. This is due to the fact that phases contain a high error that can be reduced by the solvent flattening method, which imposes flatness at solvent regions (detected as disconnected density), and improves the quality of the electron density map. Usually, at this point an interpretable electron density map suitable for macromolecular model building and refinement is obtained. However, this approach was unsuccessful as the final phases resulted in an unsatisfactory Fourier Synthesis of maps, so the structure of NrdR was not solved with it.

Alternatively, since the anomalous signal was too weak, trials with CRANK2 pipeline, which simultaneously combines the initially estimated phases with density modification and model-building in real space, resulted in 5 substructure positions corresponding to the Zn ions, and a clear structure solution consisted of 284 residues, in which the solvent was clearly differentiated from continuous density that could be interpreted as secondary structure elements such as α-helices and β-strands (Figure R10A). The values of the quality indicators were R_{work}=0.46 and R_{free}=0.50, which were not unreasonable given that the model was incomplete. CRANK2 partial model showed regions corresponding to α-helices and β-strands, together with other regions traced as an elongated main chain with no secondary structure, which could contribute to phasing, but they were clearly incomplete as a secondary structure was intuitively predicted there. Moreover, some regions were traced out of the density.

Therefore, we manually discarded these regions and used the resulting model, together with the I-TASSER shredded model of the ATP-cone (only alpha helices, Figure R9B) as an input in AUTOBUILD from the PHENIX suite. This consistently improved the phases as assessed by the residual values of R_{work} and R_{free}, and indicates a better match between the experimental and the model-derived, calculated data. As a result, the model was further completed up to 497 residues and improved the indicators to R_{work}=0.37 and R_{free}=0.40 (Figure R10B), indicating a better match of the theoretical data derived from the model with the experimental data.

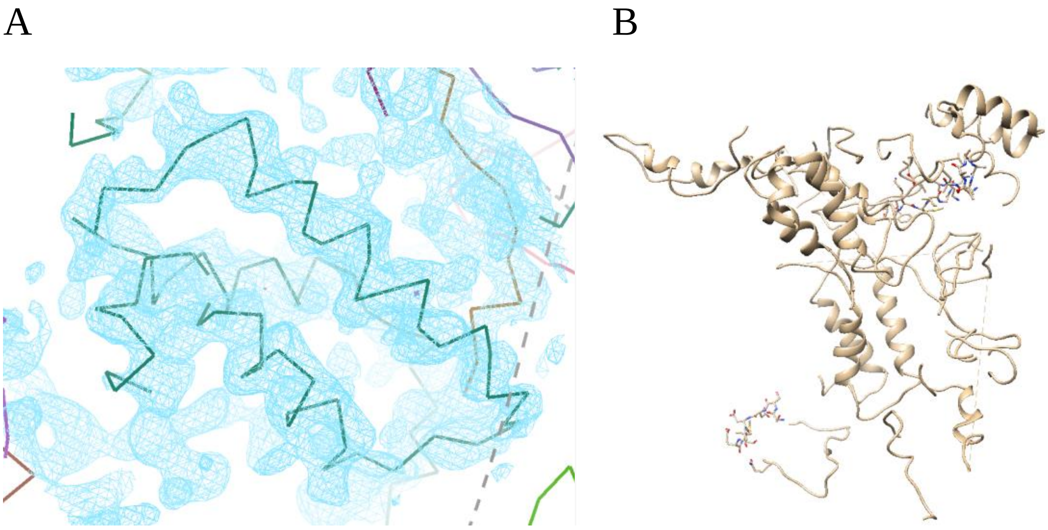


Figure R10. Structure solution and model building. (A) Pipeline CRANK2 found 5 Zn^{2+} ions that could phase the whole data set and by using the best estimated phases calculate a Fourier synthesis (electron density map) into which it could trace 284 residues. **(B)** Output model from CRANK2 was an input for AUTOBUILD together with the I-TASSER ATP-cone shredded model with only alpha helices and no loops. The polypeptide chain fragments out of density were deleted by the program and additional ones placed inside density instead, completing the model to 497 correctly placed residues.

In the next step, the output from AUTOBUILD was used as an input in PHENIX Refine for automatic refinement at the reciprocal space of the atomic positions. The structure released by PHENIX was visually inspected with COOT for manual model building at the real space and real space refinement of the molecule inside the electron density. The automated model building minimized the discrepancy between calculated and measured amplitudes of the structure factors. The structure contained several chains of aminoacids that needed to be connected and expanded with more aminoacids to form the full polypeptide.

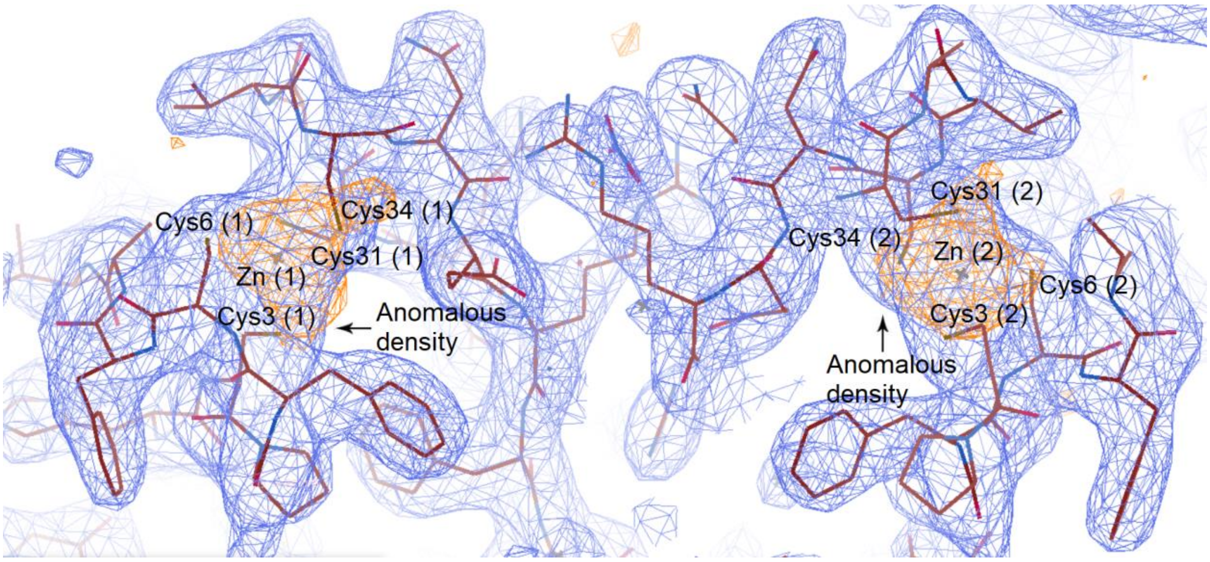
Model building inside the empty density was done with caution considering the number of potential molecules in the asymmetric unit (from four to six, see point 4.1), trying to not build where another molecule should be. A guide to connect the different spread amino acid chains was based on the complete *E. coli* NrdR model predicted by I-TASSER. With the anomalous data present in the unmerged data set, a Fourier transform was calculated with the F^+ and F^- Bijvoet pairs as coefficients, so that an anomalous density map was obtained, which showed clear discrete peaks at 3.0 r.m.s.d. (as the parameter in COOT, see Materials and Methods) that corresponded to the positions of zinc atoms and sulfur atoms of the four cysteine residues (Cys3, Cys6, Cys31, Cys34) of the zinc finger domain (Figure R11A, B), but also to the positions of the fifth methionine residues (Met48, Met70, Met86, Met107, Met113) of the ATP-cone domain (Figure R11A, C). Thus, the NrdR sequence register could be ascertained, and the molecules inside the asymmetric unit enlarged.

A

1 3 6
MHCPFCFAVDTKVIDSRLVGEGSSVRRRRQ
31 34 48
CLVCNERFTTFEVAELVMPRVVKSNDVREP
70 86
FNEEKLRSGMLRALEKRPVSSDDVEMAINH
107 113
IKSQLRATGEREVPSKMIGNLVMEQLKKLD

KVAYIRFASVYRSFEDIKEFGEEIARLED

B



C

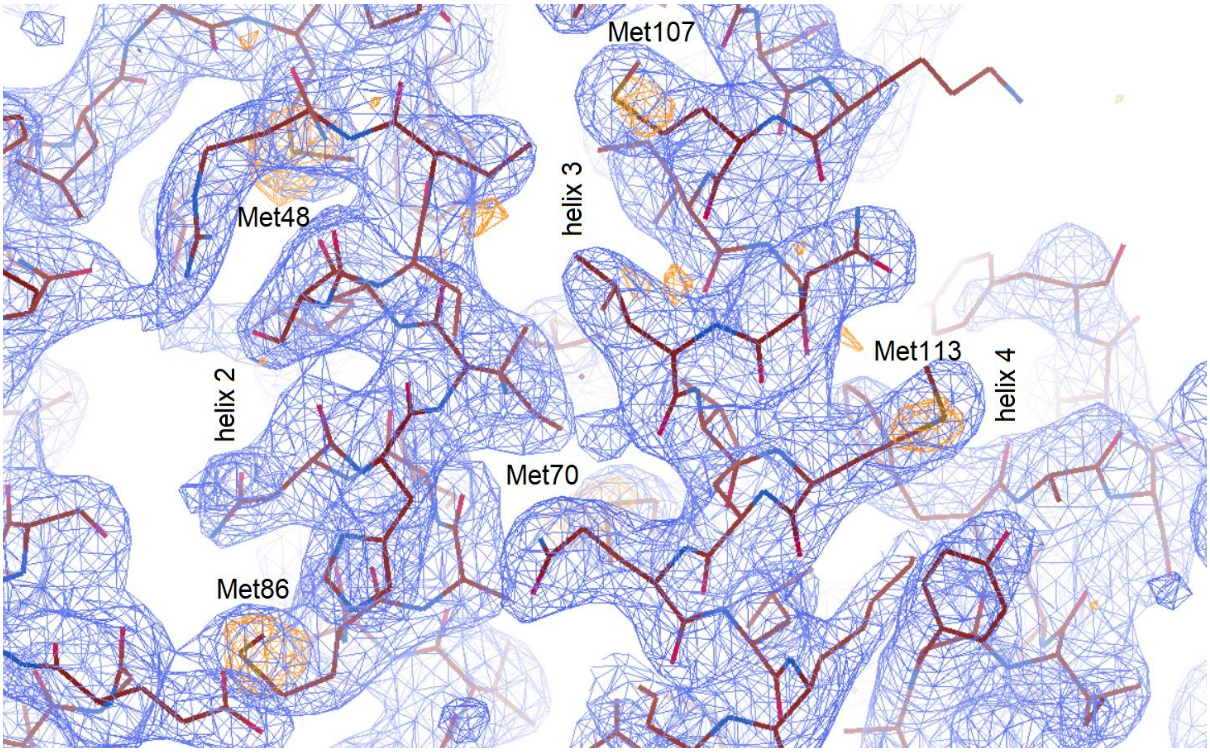


Figure R11. NrdR sequence assignment. (A) The positions of anomalous scatterers in the sequence of NrdR are shown in orange. **(B)** The 2mFo-DFc Fourier synthesis (in blue; 1.0 r.m.s.d) along with a partially built protein model (shown as a dark red sticks). The anomalous density is shown in orange (3.0 r.m.s.d) and corresponds to zinc atoms (Zn, represented as gray crosses) together with sulfur atoms of the four cysteine residues (Cys3, Cys6, Cys31, Cys34) of the two zinc finger domains interacting with each other. The zinc atoms and cysteine residues are indicated. **(C)** Electron density around the ATP-cone domain is shown (in blue), together with the anomalous map (in orange) that surround the sulfur atoms of the methionine residues (Met48, Met70, Met86, Met107, Met113; the S-C bond is shown as a dark yellow stick). The recognizable secondary structure elements of the NrdR ATP-cone are indicated.

The resulting first model built manually was again submitted to AUTOBUILD followed by PHENIX refine. By these cycles of manual building alternated with automatic model building, the dubious regions were discarded, and new regions were built in clearly defined density that appeared with the phases improved in each cycle. In addition, regions with similar structure within the asymmetric unit served to identify possible mates related by non-crystallographic symmetry, eventually resulting in four molecules A, B, C and D. At the final steps, manual model building alternated only with automatic refinement with PHENIX. At this stage, amino acid rotamers were relocated and water molecules added.

The analysis of the crystal packing of the molecules, which had been built from scratch (without a molecular replacement contribution), indicated that no more than four molecules were present in the asymmetric unit (Figure R12A, B). In addition, the crystal packing revealed the interactions of the two last helices of the ATP-cone domains between different asymmetric units. Interestingly, these two last helices of the cone domain showed variable conformations, the density being more or less complete depending on the molecule. In molecule A, the density was more complete so that more aminoacids could be built than in > C > D > B, and in this latter it showed severe disorder.

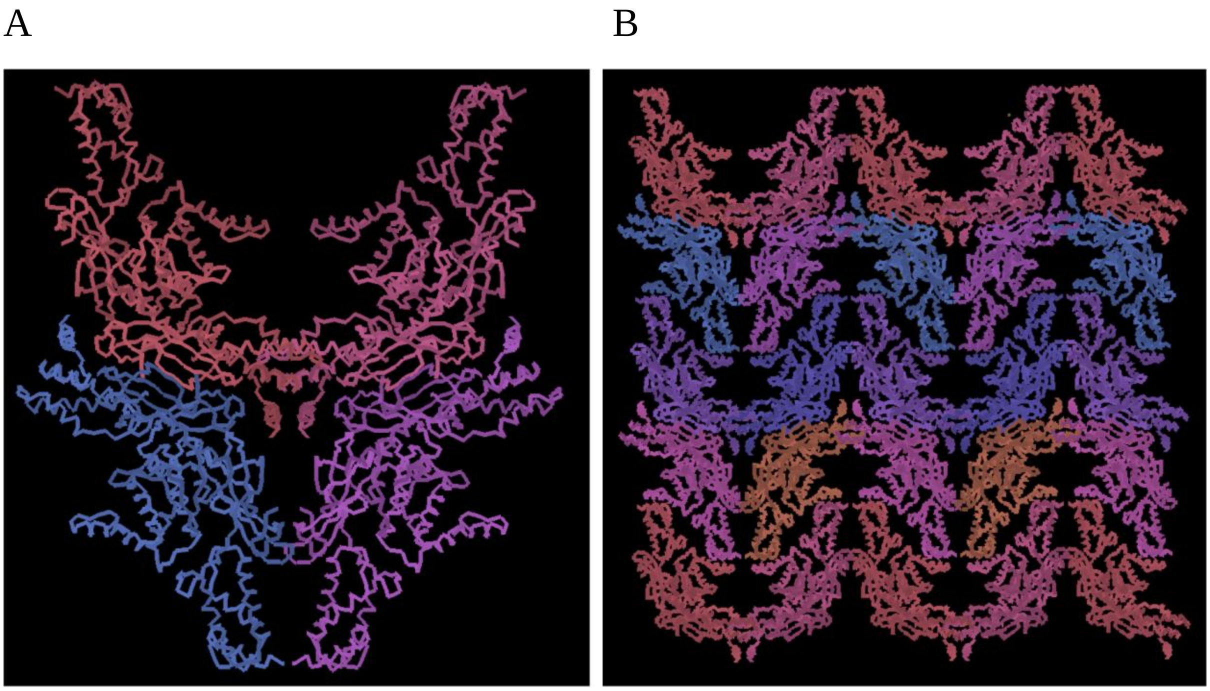


Figure R12. Crystal packing for *E. coli* NrdR. (A) A unit cell contains four interacting tetramers, shown in different colors. **(B)** Crystal packing of several unit cells are shown, every subunit is represented in a different color. Note the orientation of the molecules are the same as in (A).

In addition, analysis of the electron density in molecules A, B, C and D showed clear positive difference density at the inner cavity of the cone domain that suggested the presence of a nucleotide, which is consistent with the allosteric regulation nature of cone domains from ribonucleotide reductases (RNRs). Nucleotides were tentatively placed and refined, and after several cycles of atomic positional refinement, visual inspection and model building confirmed the presence of an adenosine, yet the number of phosphates was variable but clear from one molecule to another. ATP in chains B and D, and ADP in chain C, whereas in chain A the defined nucleotide is

AMP. The model was built, and positions of atoms refined to an $R_{work}=0.22$ and $R_{free}=0.25$, which are acceptable values (Table R5).

Wavelength	1.2835
Resolution range	92.421 - 2.600 (2.693 - 2.600)
Space group	C 2 2 21
Unit cell parameters a, b and c in Å; α, β, and γ in degrees	51.19 255.37 132.94 90 90 90
Unique reflections	27674 (2708)
Multiplicity	14.9 (15.4)
Completeness (%)	99.93 (99.96)
I/sigma(I)	14.17 (1.5)
CC1/2	1 (0.505)
Reflections used in refinement	27674 (2708)
Reflections used for R-free	1372 (112)
R-work	0.2229 (0.3640)
R-free	0.2494 (0.4165)
CC(work)	0.924 (0.654)
CC (free)	0.928 (0.577)
Number of non-hydrogen atoms	4789
Macromolecule atoms	9490
Protein residues	582
RMS (bonds)	0.003
RMS (angles)	0.65
Ramachandran favored (%)	94.84
Ramachandran allowed (%)	4.80
Ramachandran outliers (%)	0.36
Rotamer outliers	0.39
Clashscore	2.002

Average B-factor	109.0
B-factor macromolecules	109.0

Table R5. Data collection parameters, processing, scaling and refinement statistics. Last resoution shell values are indicated in brackets.

R4.4 Stereochemistry analysis of the final NrdR structure

The final model showed good stereochemistry. The Ramachandran plot, which represents energetically allowed regions for backbone dihedral angles ψ against φ of aminoacid residues in protein structure, showed all residues in preferred or allowed regions (Figure R13).

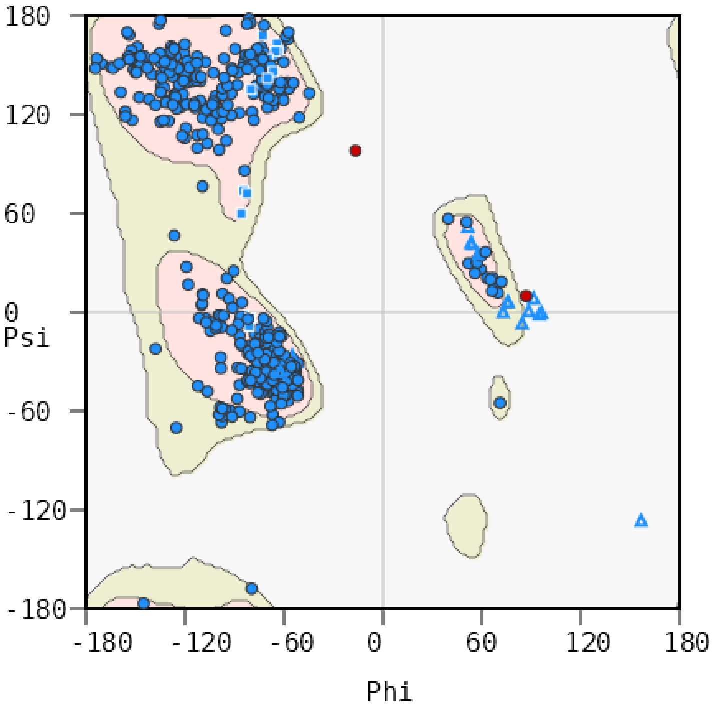


Figure R13. Ramachandran plot for *E. coli* NrdR crystal structure. 94.84% of torsional (Phi) and dihedral (Psi) angles in the polypeptide chain are in preferred regions (blue circles), 4.80% in allowed regions (blue squares and triangles) and 0.36% stands for outliers (red circles). This representation was obtained from COOT.

The analysis of the stereochemistry as provided by the Protein data bank showed overall good statistics except for the B-factors (Figure R14).

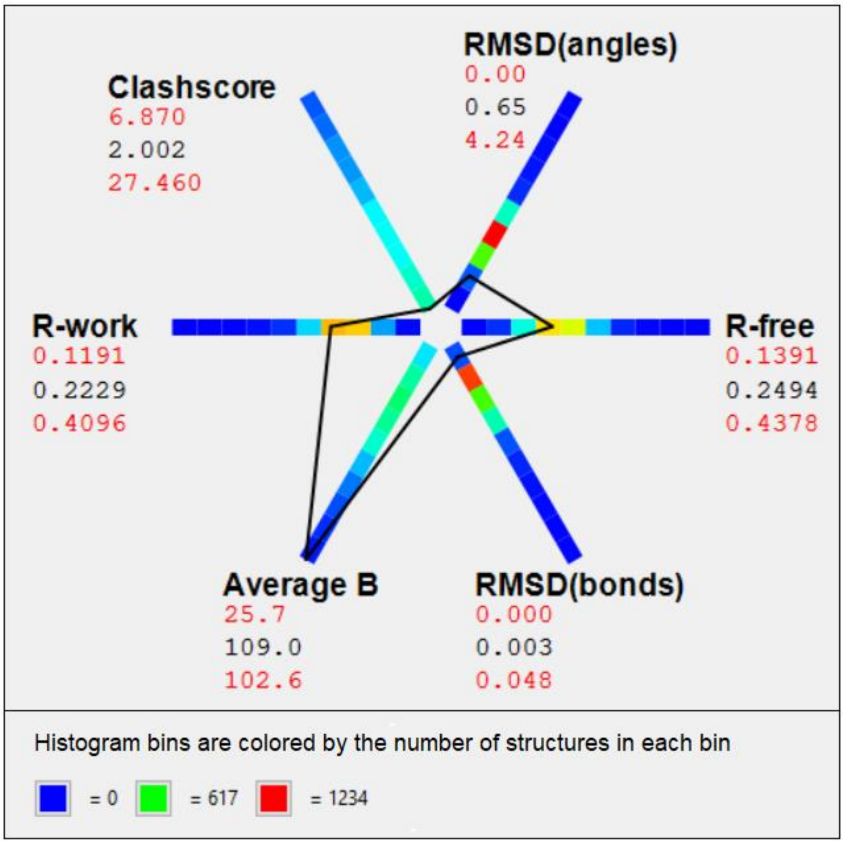


Figure R14. Statistics of the refined model. Validation statistics of the model versus available PDB structures with similar resolution. The graphics show a blue to red scale defining value ranges for each parameter. The closer to red, the more structures report that value for the corresponding parameter.

R5. Overall structure of NrdR

R5.1 Tetramerization of NrdR triggered by interactions between domains

The final model of *E.coli* NrdR crystal structure (Figure R13, R14, Table R5) shows four protein molecules (A, B, C and D) in the crystallographic asymmetric unit (a. u.) (Figure R15).

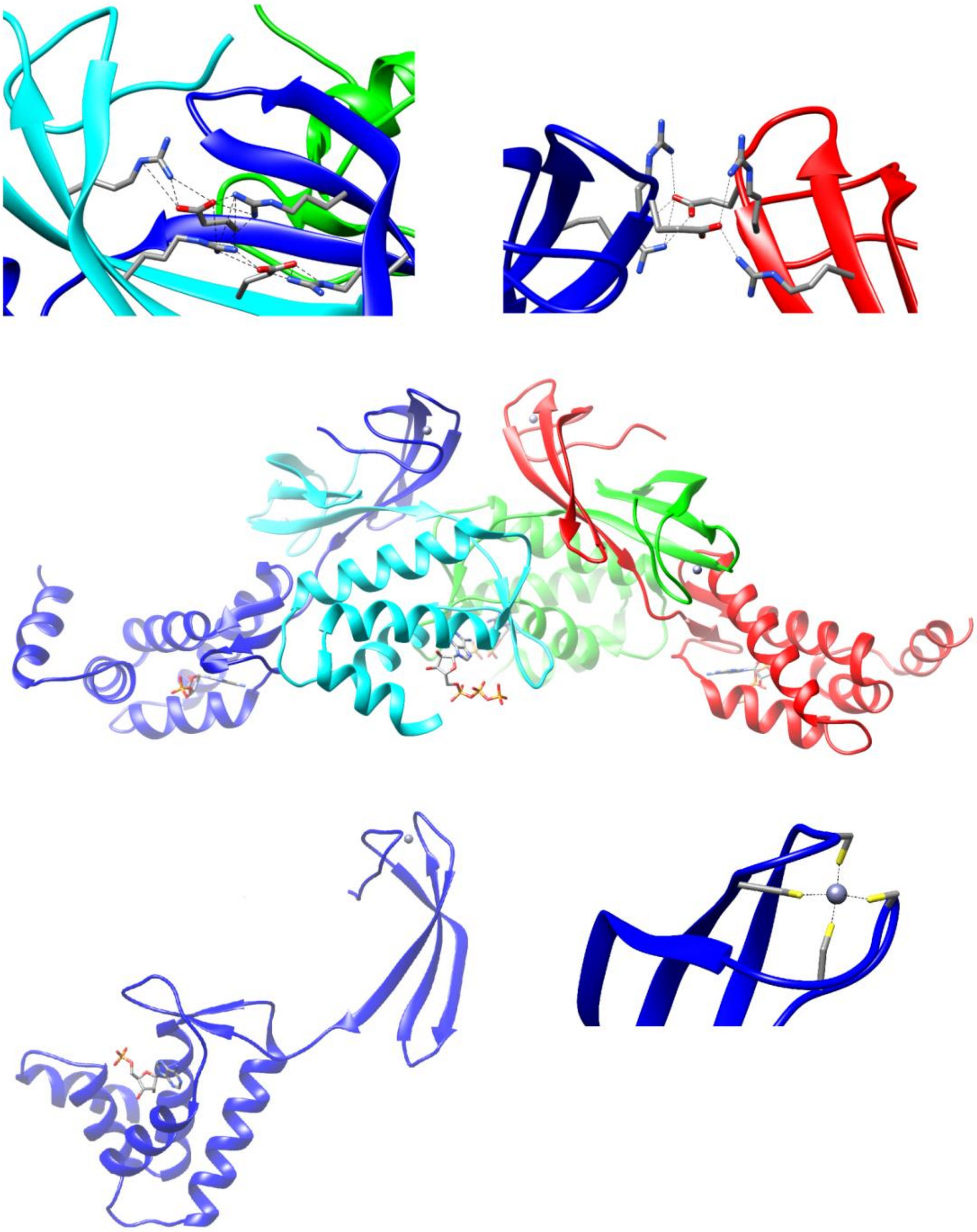


Figure R15. Overall structural organization of *E. coli* NrdR. The central tetramer is formed by molecules A (dark blue), B (turquoise), C (green) and D (red). Molecule A, as the most complete, is taken as a reference. Close-up views of the zinc ion coordination by cysteine residues in the zinc-finger domain, interactions between glutamate and arginine residues within Zn-fingers A and B, and A and D are shown. Aminoacids are represented as sticks.

At the very N-terminal end, the crystal structure shows part of the linker added to the protein that facilitated digestion (see sections M1.3, M1.4.1, R1.1, R1.3). This is followed by an N-terminal zinc-finger domain of three β-strands (β-strands 1 to 3) in which the four cysteines, Cys3 and Cys6 (at the loop preceding strand β1), and Cys31 and Cys34 (at the loop between β2 and β3), coordinate a Zn ion at the tip of the domain (Figure R15). The ZFD is followed by an ATP cone domain (ACD), which is characteristic of RNRs, and consists of a β-hairpin and a β-strand (β-strands 4, 5 and 6), and five α−helices (α−helix 1 to 5). The β-hairpin covers the widest part of the cone, which forms a cavity that hosts a nucleotide. The last α-helix of the cone domain shows different orientations and variable degrees of flexibility, being the best defined one in molecule A, but with weaker density in molecules C and D and being B the one with faintest electron density, suggesting higher disorder. The stability of the helix is likely induced by packing interactions with symmetry molecules, as this region in molecule B does not perform any interaction with symmetry mates, also suggesting high flexibility for this region.

The four molecules in the asymmetric unit make extensive contacts. The ZFDs of molecules A and B intertwine with each other in an antiparallel manner, while the corresponding cone domains are arranged head-to-tail, hiding a total surface of 1335.01 Å^2 (which is the addition of the surface covered in A and in B). The same occurs between molecules C and D, which hide a total surface of 1530.97 Å^2. Between A and B ZFDs, Glu42 (mol A) salt bridges both Arg27 (mol A) and Arg29 (mol B), and the converse also occurs, Glu42 (mol B) interacts with Arg27 (mol B) and Arg29 (mol A) (Figure R15). Such interactions are also present between molecules C and D ZFDs. These two interactions A/B and C/D apparently define two ZFD dimers that, in addition, interact with each other such as B/A:D/C, being A:D the two Zn dimers interface. At the A:D interface, the central regions of respective domains hide a total surface of 175.32 Å^2, in which Glu36 (mol A) that salt-bridges Arg28 and Arg37 (mol D), and vice versa, Glu36 (mol D) interacts with Arg28 and Arg37 (mol A) (Figure R15). Therefore, electrostatic interactions stabilize the contacts between ZFDs within a dimer and between dimers.

The ACDs also make interactions with each other within a dimer (A/B and C/D) yet, in this case, the contacts are not antiparallel but in tandem, face-to-tail. Also here, a cone domain from one dimer interact with a cone domain from the other dimer (but, since the subunits within a dimer are intertwined, the ACD from molecule B contacts: molecule C ACD (and not A and D as for the ZFDs). Between ATP-cone domain ‘dimers’, the involved B:C contact surface is of 373.33 Å^2. In addition, ACD B contacts a crevice formed by ZFD C and D, from the other dimer (and vice versa, cone C interacts with the ZFD crevice between A/B).

R5.2 Comparison of NrdR domains from NrdR structure

Molecules show higher similarity between specific pairs, such as molecules A and D (r.m.s.d. A/D= 2.68Å), and B and C (r.m.s.d. B/C= 1.56Å), while AB and CD are dissimilar (r.m.s.d between AB= 14.21Å; C/D= 14.12Å, A/C= 14.74Å; B/D= 14.21Å), which suggests a couple of dimers in which the monomers in a dimer are significantly different. Differences between A and B, and between C and D, are localized at the loop between β-strands 3 and 4 (35-48aa). However, the domains of the molecules are overall similar and, therefore, molecule A is taken as the reference as it is the most complete.

The four Zn fingers are not structurally identical, as indicated by their divergence > 0.5 Å upon superposition (below 0.5 Å, differences are considered negligible, and due to experimental error). However, the similarity does not follow the overall similarity shown above for the full-length molecule.

Superposition of ZFD A and D show the lowest r.m.s.d, is of 1.00 Å between ZFD A/D and A/C. 1.1Å between B/D, and 1.1 Å for A/B; Å; 1.6 Å for A/C, B/C and C/D show the highest differences, 1.8 Å and 1.7 Å, respectively. Superposition of the loop between β-strands 3 and 4 (35-48aa) shows the highest structural similarity between molecules A/D and B/C pairs (r.m.s.d. 1.7 Å and 1.1 Å respectively), but despite being “more similar”, the structural differences are significant (i.e. >> 0.5 Å). The other superpositions show 2.65Å between loops A/B, 2.69Å between A/C, 3.1Å between B/D, and 3.01 Å between C/D.

Superposition of the four cone domains A, B, C and D, discarding the last loop and the last helix (aa 132-149), so from aa 49 to 131, shows similar variability between all molecules. r.m.s.d of 1.47Å between molecules A and B, 1.58Å between A and C, 1.60Å between A and D, 1.34Å between B and C, 1.60Å between B and D, 1.54Å between C and D. The highest differences between ATP cone domains are due to the last loop and the last helix (aa 132-149) since they are differently oriented.

Superposition of the loop between β-strands 4 and 5 (aa 53-57) shows the lowest r.m.s.d. of 0.59Å between B/D, and the highest 1.32Å between B/C, 1.31Å between C and D, but also 1.18Å between A/ B, 1.16Å between A/C, 1.08Å between A/D, suggesting that the orientation of the nucleotide is slightly different in case of ATP-cone mol B.

R5.3 The cone domain active site

At the wider part of the cone domain, a well-defined density was found in molecules A, B, C and D, which suggested the presence of a nucleotide. Addition of the nucleotide into the structure improved the phases cycle by cycle so the electron density was every time clearer. And eventually strong enough to trace all four ligands with confidence. Interestingly, the nucleotides consisted of an ATP in molecules B and D, an ADP in molecule C, and an AMP in molecule A (Figure R16).

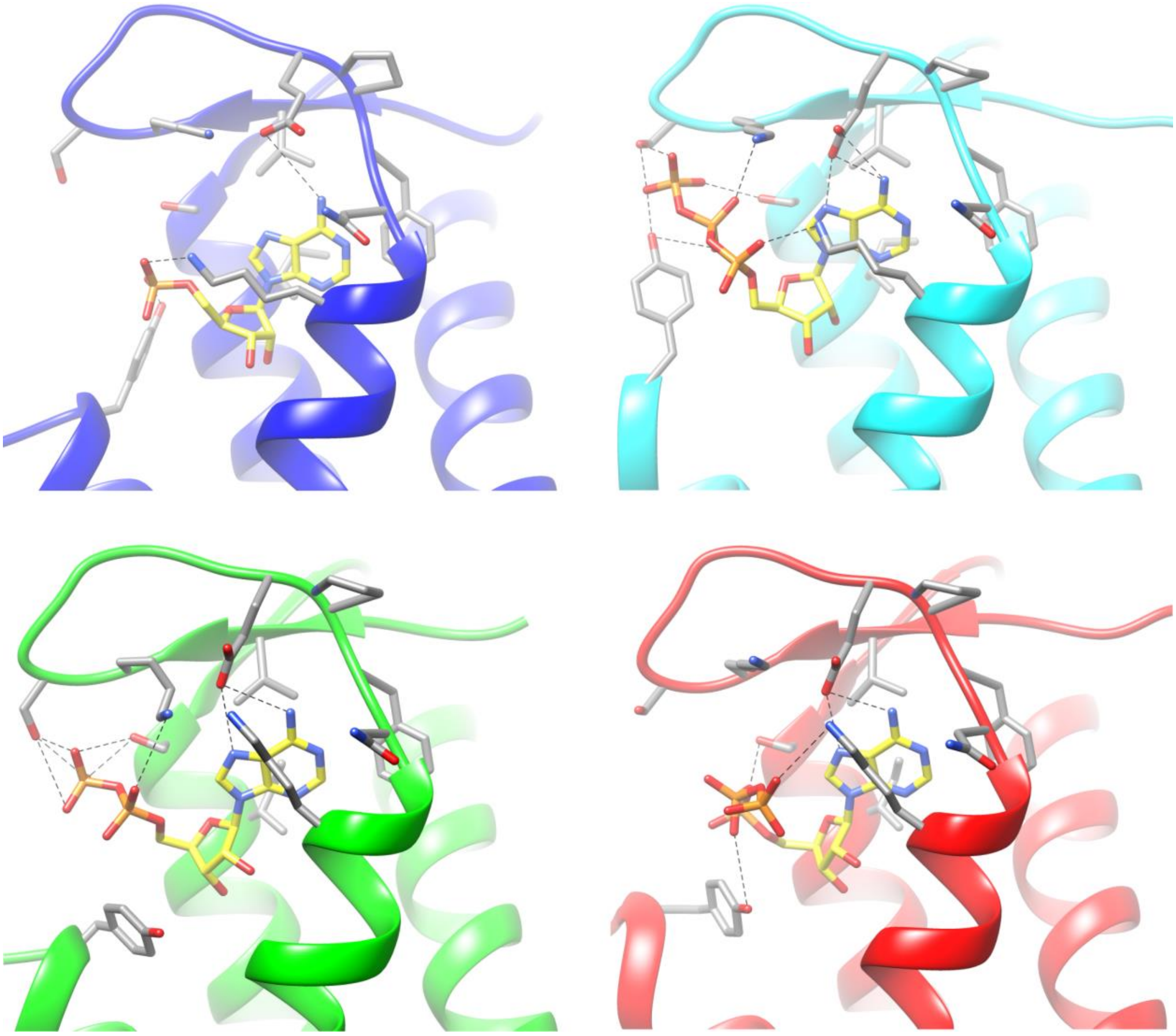


Figure R16. Presence of a nucleotide at the cone domain. **(A)** At molecule A conde domain (in blue), an AMP was found at the active site. **(B)** In molecule B (in cyan) an ATP was found, involving additional residues that stabilize the nucleotide, compared to molecule A (A). **(C)** In molecule C (in green), an ADP was found. **(D)** In molecule D (in red) an ATP was found that shows the phosphates in different orientations with respect to molecule B (B). In all panels, note the different orientations of the side chains. Aminoacids are represented as sticks.

On top of the ATP-cone cavity, the lid consists of a β4-loop-β5-loop motif (from Arg50 to Asp62) that covers and contacts the nucleotide. The residues from following α-helix 1, α-helix 3 and α-helix 4 also contact the nucleotide. All residues involved in coordination of the four nucleotides show variable orientations, especially depending on the presence of the full chain of triphosphates or its shorter variants. Specifically, the adenine ring is in a hydrophobic pocket made by Val51, Phe61, Lys65, Leu66 and Ile108 side chain methylenes, whereas the charged atoms from Glu59, main chain carbonyl of Pro60, and Asn62, interact with the bound nucleotide.

In molecules B, C and D Ser105 stabilizes the tail of nucleotide phosphates but, in contrast to the adenine ring that is positioned almost identically in all four A, B, C and D molecules, the phosphates tail has different orientations from one molecule to another. Ser54 also participates in triphosphate stabilization, but only in the case of molecule B. Therefore, in molecule B Ser54 and Ser105 contact the γ-ATP phosphate, in molecule C Ser105 contacts the β-phosphate and in D the α-phosphate. The absence of phosphates β and γ in AMP from molecule A reduces the number of contacts of the nucleotide with the protein. In mol A Ser105 does not make any contact with the phosphate of the nucleotide.

One residue that presents an important variation of the rotamers is Tyr131 from helix α4. Interestingly, residue Tyr131 has the same rotamer in molecules C and D. This occurs, because Tyr131 just precedes the last helix of the cone, which in molecules C and D these helices make contacts with the same helices from symmetry partners D’ and C’, respectively. Note that molecule C cone domain contains an ADP and molecule D an ATP. In both cases they show the same rotamer m-85 for Tyr131, but the main chain is displaced outwards in molecule D, towards the last two helices and the symmetry partner, without contacting the nucleotide.

However, in molecules A and B, Tyr131 is oriented differently. In molecule A adopts the rotamer t80 conformation and interacts with the O from the Ser105 main chain at helix α3. Instead, in B, Tyr131 adopts the m-85 conformation as in molecules C and D, but here the main chain is dragged even more outwards so that the Tyr131 OH reaches and interacts with the α phosphate from the ATP. All these rotamer reorientations suggest a potential role of residue Tyr131 in nucleotide recognition. Possibly, Tyr131 rotamers depend on the overall configuration of this region, in particular in the orientation of the last two helices α4 and α5.

R5.4 Crystal contacts correspond to functional interfaces triggering higher level multimerization of NrdR

Besides the contacts within the a.u., other relevant interactions are also present between ´tetramers´ from different a.u. This involves the last two α-helices α4 and α5 and the loop Lα4-α5 in between from the cone domain from molecules A, C and D (Figure R17, Table R6).

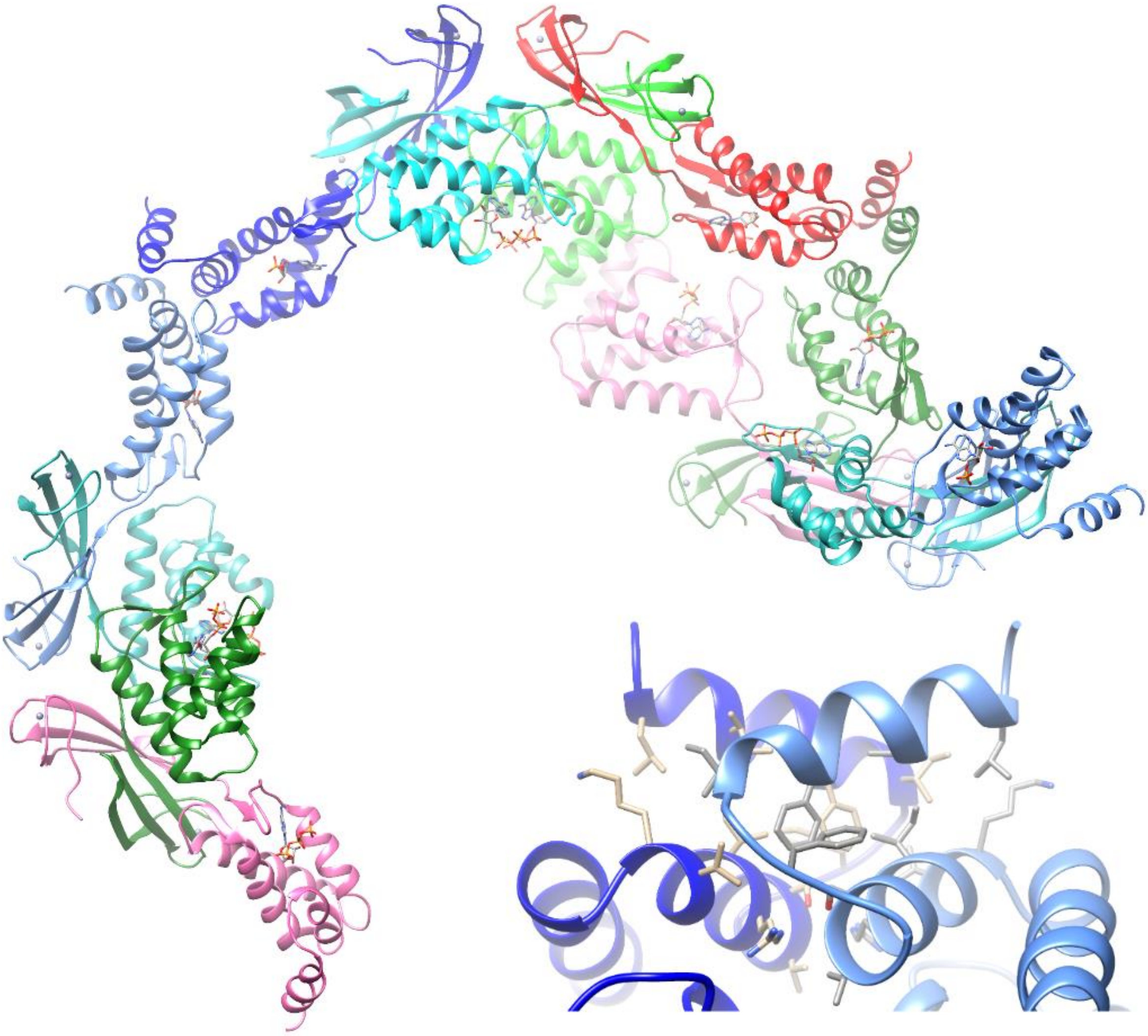


Figure R17. Crystallographic neighbors of NrdR tetramer. The central tetramer is formed by molecules A (dark blue), B (turquoise), C (green) and D (red). The neighboring molecules are labeled as A’ (sky blue), B’ (light blue), C’ (dark green) and D’ (pink). Close-up view of the interaction between the last two helices α4 and α5 from symmetry partners A and A’ is shown in the bottom right part. Aminoacids are represented as sticks.

Interestingly, aa 132-149 that involves the last loop Lα4-α5 and helix α5 is clearly visible only in molecule A, which makes an extensive contact with the same region from symmetry partner A', hiding together a total surface of 919.70 Å^2. Similar contacts, but with no density for side chains from loop Lα4-α5 and helix α5 region are also found between molecules C and D, which contact symmetry partners D' and C', respectively (Figure R17). Note that molecule B does not make interactions through this segment, and the electron density map is extremely weak at this specific region.

Mol A	Mol A'	Type of interaction
Lys121	Ile137	Hydrophobic
Val122	Phe134, Ile137, Phe140	Hydrophobic
Ile125	Ile137, Phe140	Hydrophobic
Arg126	Ser129	Hydrophilic
Arg126	Val130	Hydrophobic
Ser129	Arg126	Hydrophilic
Val130	Arg126	Hydrophobic
Phe134	Val122	Hydrophobic
Ile137	Lys121, Val122, Ile125, Ile144, Leu147	Hydrophobic
Phe140	Val122, Ile125, Phe140, Ile144	Hydrophobic
Ile144	Ile137, Phe140,	Hydrophobic
Leu147	Ile137	Hydrophobic

*Hydrophobic interactions = 91,3 %; Hydrophilic interactions = 8,7%

Table R6. Main residues implicated in helix-helix interactions established between helices α4 and α5 **segments (aa 121-149)**. Hydrophobic interactions are in white background, whereas hydrophilic interactions are marked in green.

To explore whether the interactions above described reflect the multimerization ability of the protein, we performed an analysis of the interfaces with the PDBePISA software. A favorable, negative Gibbs Free Energy for the total interfaces between dimers A /B or C/ D was found (Table R7), suggesting that they are functional and essential for protein dimerization, and not merely crystal contacts. However, the interfaces between

ZFDs from mol A and D, and between ACDs from mol B and C, appeared less significant, suggesting crystal packing contacts.

Analysis of the interactions by PISA revealed favorable, negative Gibbs Free Energy (see Table R7) for the interfaces between the cone domains of symmetry partners, involving the last two helices α4 and α5 and the intermediate loop Lα4-α5. This suggests a functional interface that causes large multimerization.

Chain ID	Chain ID	Symmetry operation	Interface area [Å2]	Gibbs Free Energy [kcal/mol]
D	C	X, Y, Z	1530.97	-10.59
B	A	X, Y, Z	1335.01	-6.71
A	A	-X+3, Y, -Z+3/2	919.70	-3.76
D	C	X, -Y+1, -Z+1	883.18	-9.19
B (ATP)	B	X, Y, Z	373.72	-1.02
C	B	X, Y, Z	373.33	-1.89
D (ATP)	D	X, Y, Z	365.18	-3.77
C (ADP)	C	X, Y, Z	360.13	-3.81
A (AMP)	A	X, Y, Z	282.41	-1.69
D	B	X, Y, Z	209.59	-0.40
C	A	X, Y, Z	207.28	0.01
D	A	X, Y, Z	175.32	3.32

Table R7. Analysis of the interfaces of the NdR crystallographic tetramer with the PDBePISA software. Molecules from symmetry partners are indicated with their corresponding symmetry operation, (BATP), D(ATP), C(ADP) and A (AMP) indicate interfaces between nucleotides and ATP-cone domains.

R6. Production of NrdR mutants E36A, E42A, Y131A, del132-149

R6.1 Subcloning, expression and purification of SUMO-NrdR mutants E36A, E42A, Y131A, del132-149

The crystal structure shows several subunits that interact by different large interfaces, some of them predicted as biologically relevant (see section R5, Figure R15, R16, R17). Based on the crystal structure and PISA prediction (Table R7), a series of mutants were generated to evaluate the effect of disruption of the protein-protein interactions observed in the crystal (Figure R18). Therefore, we designed PCR primers (Table R8) to obtain SUMO-NrdR mutants E36A, E42A, Y131A, and deletion of the C-terminal helix α5 and the previous loop Lα4-α5, mutant del132-149, by using 6His-SUMO-TEVcs-SerGlySerGlySer-NrdR as the template.

MHCPFCFAVDTKVIDSRLVGEGSSVRRRRQ
36 42
CLVCNERFTTFEVAELVMPRVVKSNDVREP

FNEEKLRSGMLRALEKRPVSSDDVEMAINH

IKSQLRATGEREVPSKMIGNLVMEQLKKLD
131 132-149
KVAYIRFASVYRSFEDIKEFGEEIARLED

Figure R18. Single-site aminoacid mutations (green) and deletion (red) of NrdR. These mutations were designed to test the stability of the interactions described in sections R5.2 and R5.3.

Primer	DNA sequence
E36A_for	5’ - CTGGTGTGTAATGCACGTTTCACCACC - 3’
E36A_rev	5’ - GGTGGTGAAACGTGCATTACACACCAG -3’
E42A_for	5’ -TTCACCACCTTTGCAGTGGCGGAGCTG - 3’
E42A_rev	5’ - CAGCTCCGCCACTGCAAAGGTGGTGAA - 3’
Y131A_for	5’ -TGCCTCTGTCGCCCGCAGTTTCGAA - 3’
Y131A_rev	5’ - GAAACTGCGGGCGACAGAGGCAAA - 3’
del132-149_for	5’ -TATACCATGGGCGAGAACCTTTACTTTCAA - 3’
del132-149_rev	5’ - ATACTCGAGTTAGTAGACAGAGGCAAAACGGATATA GGCGAC - 3’

Table R8. DNA sequences of primers used for subcloning of the indicated mutants.

All mutants were confirmed by gene sequencing (see figure R19). Fusion SUMO-NrdR mutants were expressed in *E.coli* BL21 (DE3) strain (Figure R19), as previously done for the SUMO-NrdR fusion.

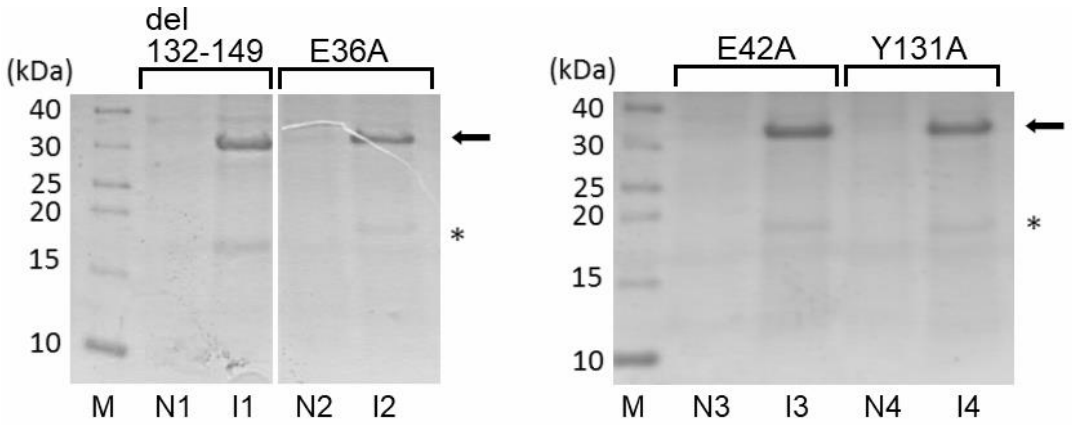
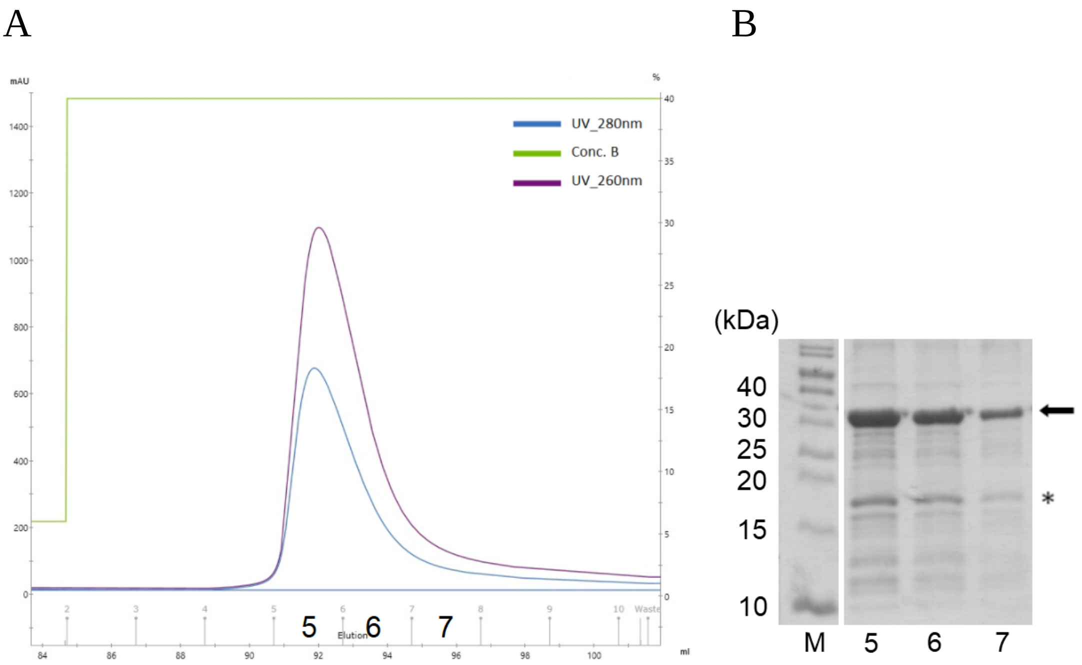


Figure R19. High scale expression of SUMO-NrdR del132-149, E36A, E42A and Y131A (from left to right) analyzed by 15% SDS-PAGE. Induced **(I1-4)** and non-induced **(N1-4)** fractions of all expressed NrdR mutants. The arrows indicate bands of expressed proteins. Asterisks correspond to bands of partially degraded proteins.

The first purification chromatography step consisted in a NI-NTA affinity chromatography, in which the mutant fusion proteins eluted in a 200 mM imidazole step in a peak collected in three elution fractions (SUMO-NrdR E36A, Figure R20A; SUMO-NrdR del132-149, Figure R21A; SUMO-NrdR E42A and SUMO-NrdR Y131A, not shown), as the WT. In all cases the A260/A280 ratio was in all cases much higher than 1, suggesting the presence of a nucleotide bound to the proteins, as for the WT.



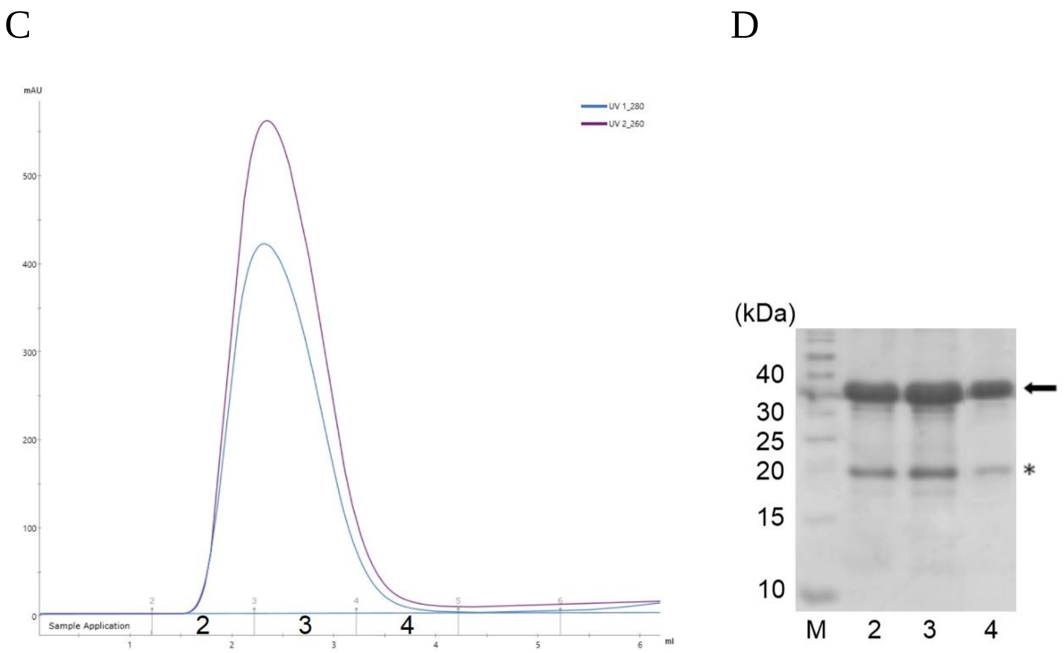


Figure R20. Purification steps of SUMO-NrdR E36A. (A) Elution profile of SUMO-NrdR E36A from a Ni-NTA column (optical density (OD) at 280 nm, represented by the blue curve; OD at 260 nm, purple curve; right axis, imidazole concentration, which corresponds to the green line). A A260/A280 ratio higher than 1 suggests presence of nucleotides bound to the protein. Note the protein eluted in a single imidazole concentration step **(B)** 15% SDS-PAGE analysis of the IMAC elution peak fractions showing a band corresponding to SUMO-NrdR E36A in fractions 5-7. The black arrow indicates the expected MW of SUMO-NrdR mutants. The asterisk corresponds to a partially degraded mutant**.** Note that due to the SUMO tag the proteins run at higher molecular weight. **(C)** A 5 ml (HiTrap) column was used to exchange the buffer of SUMO-NrdR E36A to extract the imidazole. **(D)** 15% SDS-PAGE of the desalting elution peak of SUMO-NrdR E36A. Note the higher purity compared to the previous IMAC. The black arrow and asterisk as in (B). Note that due to the SUMO tag the proteins run at higher molecular weight.

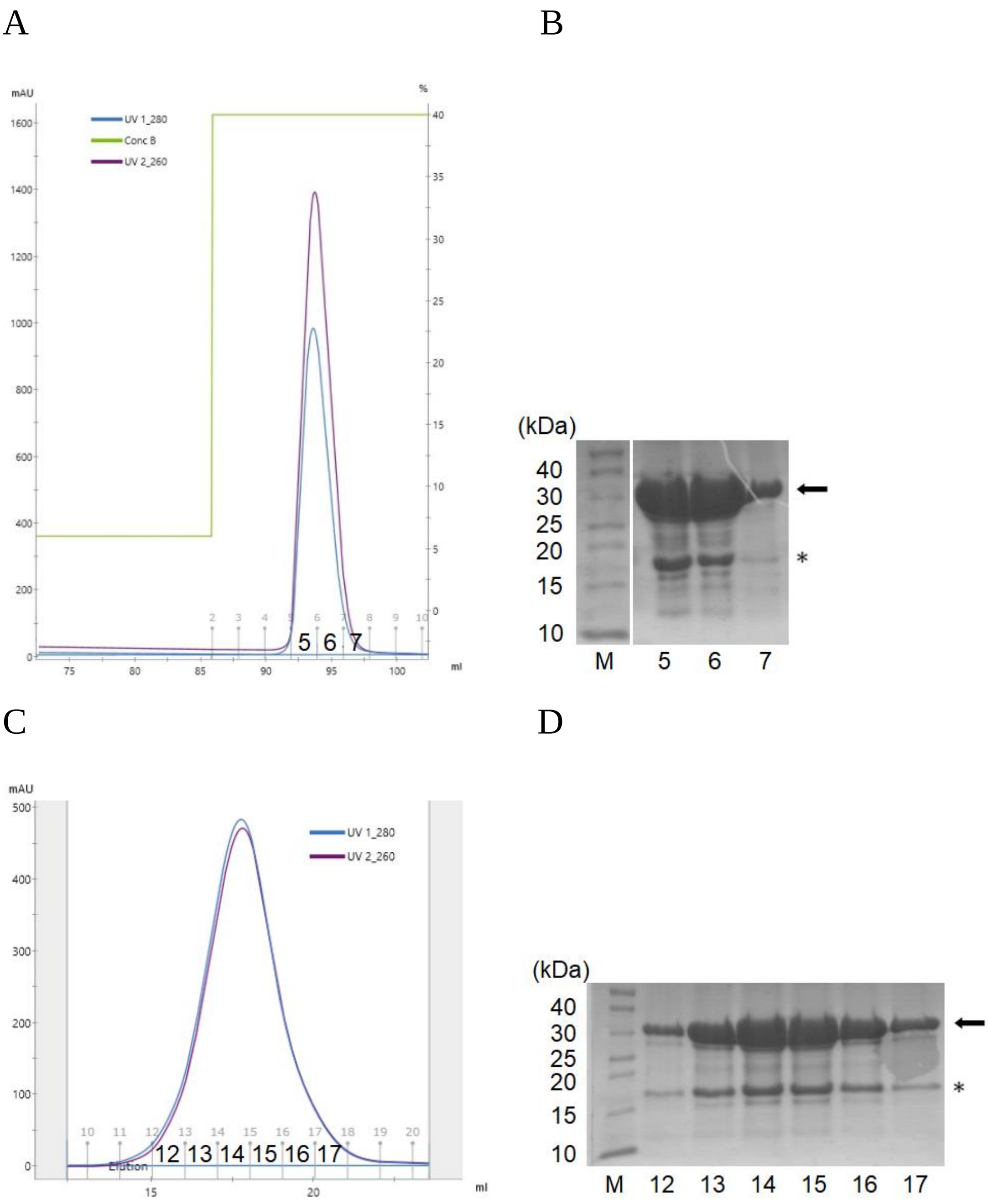


Figure R21. Purification procedures of SUMO-NrdR del132-149. (A) Elution profile of SUMO- NrdR del132-149 from a Ni-NTA column (optical density (OD) at 280 nm, blue curve, and at 260 nm, purple curve; imidazole concentration, green line). The A260/A280 ratio higher than 1 suggests presence of nucleotides bound to this deletion mutant. **(B)** 15% SDS-PAGE IMAC elution peak showing the presence of SUMO-NrdR del132-149 in fractions 5-7. The black arrow indicates the expected MW of SUMO-NrdR del132-149. The asterisk corresponds to partially degraded mutant protein**. (C)** Buffer exchange of SUMO-NrdR del132-149 to extract the imidazole. The absorption at 280 nm (blue line) and at 260 nm (purple line) is shown. Note that A260/A280 ratio has a lower value than the one registered during IMAC. **(D)** The desalting elution peak of SUMO-NrdR del132-149, checked by 15% SDS-PAGE revealed higher sample purity compared to eluted IMAC fractions. Black arrow and asterisk as in **B**.

All mutants, showed that the fractions contained mixtures of fusion NrdR mutants and other proteins from *E. coli*, as assessed by by SDS-PAGE (SUMO-NrdR E36A, Figure R20B; SUMO-NrdR del132-149, R21B; SUMO-NrdR E42A, SUMO-NrdR Y131A, not shown). Fractions containing the protein sample were pooled and the concentration measured by Bradford protein assay, showing a final yield of >50 mg of each fusion mutant per 500 ml of culture. The ratio of absorbances A260/A280 was in all cases >1.0, suggesting the presence of nucleotides bound to all protein mutants, and similar to the WT fusion protein. The nature of the bands was analyzed by fingerprint mass spectrometry, which confirmed that the identified bands in the SDS-PAGE corresponded to the NrdR mutants and their degraded forms, which is indicated by black arrow and asterisk, respectively in Figures R19, R20B, D, R21B, D.

R6.2 TEV cleavage, isolation, concentration and quality check of NrdR mutants E36A, E42A, Y131A, del132-149

Next step consisted in digestion of the fused proteins followed by a re-chromatography purification by Ni-NTA to separate digested 3(GlySer)-NrdR mutants from the rest of digested and undigested products (all bearing a His-tag). For an efficient cleavage reaction, it was necessary to previously extract the imidazole from the protein samples. On the other hand, we noticed that there was a difference between the theoretical values of isoelectric points between the SUMO-NrdR fusion mutants and their digested forms. Specifically, pI values of the SUMO-NrdR E36A and E42A fusions were 6.28, whereas pI values of 3(GlySer)-NrdR E36A and E42A were 8.41. For mutant SUMO-NrdR del132-149 pI was 6.84, whereas pI of 3(GlySer)-NrdR del132-149 was 9.21. The theoretical values of the isoelectric points estimated for SUMO-NrdR Y131A and 3(GlySer)-NrdR Y131A were 6.15 and 7.85, respectively, the same as for SUMO-NrdR and NrdR.

Therefore, during the buffer exchange imidazole was extracted, the NaCl concentration was decreased to 500mM and the pH of the buffer was changed. Specifically, we changed to buffer 50mM Na_2HPO_4 at pH 9.5 for both SUMO-NrdR E36A (Figure R20C) and SUMO-NrdR E42A (not shown), 50 mM Tris-HCl at pH 9.0 for SUMO-NrdR Y131A (not shown), and pH 8.0 for SUMO-NrdR del132-149 (Figure R21C). The theoretical pI values were calculated with ProtParam (https://web.expasy.org/protparam/). Interestingly, the analysis of the eluted samples from the buffer exchange column by SDS-PAGE showed that exchanging the buffer worked also as an additional washing step as it increased the purity of the mutants fractions (SUMO-NrdR E36A, Figure R20D; SUMO-NrdR del132-149, R21D; SUMO-NrdR E42A, SUMO-NrdR Y131A, not shown).

TEV protease digestion trials were subsequently performed in small-scale tests overnight (O/N) for all mutants at 4˚C, 500mM NaCl, buffer and pH as indicated above, and 1mM DTT. The tested protease/protein ratios were 1:10, 1:25, 1:50, 1:100 for SUMO-NrdR del132-149. For SUMO-NrdR E36A and SUMO-NrdR E42A 1:10, 1:15, 1:20, 1:25. The resulting samples were run in 15% SDS-PAGE, which showed that cleavage was directly proportional to the amount of used protease, and that the highest cleavage efficiency was at a protease:protein ratio 1:10 for both constructs (Figure R22A, B, R23D). In the case of SUMO-NrdR Y131A, a protease:protein ratio 1:25 was enough for sufficient cleavage (Figure R23G). The nature of the bands was verified by MALDI-TOF mass spectrometry.

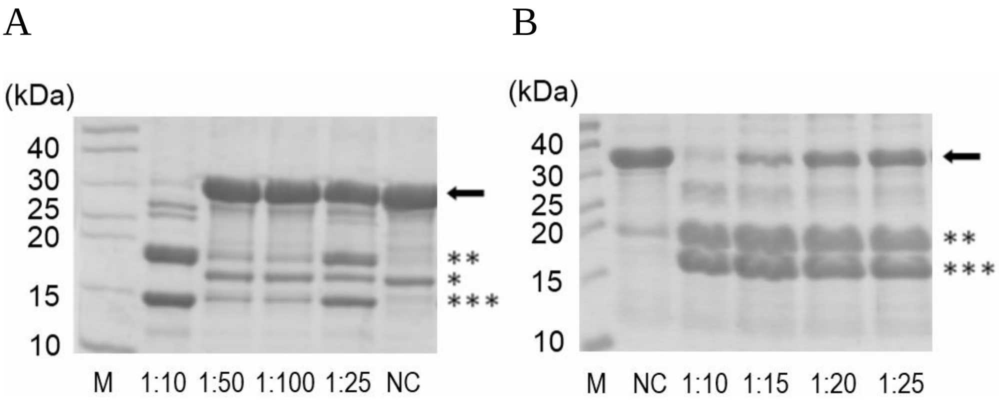
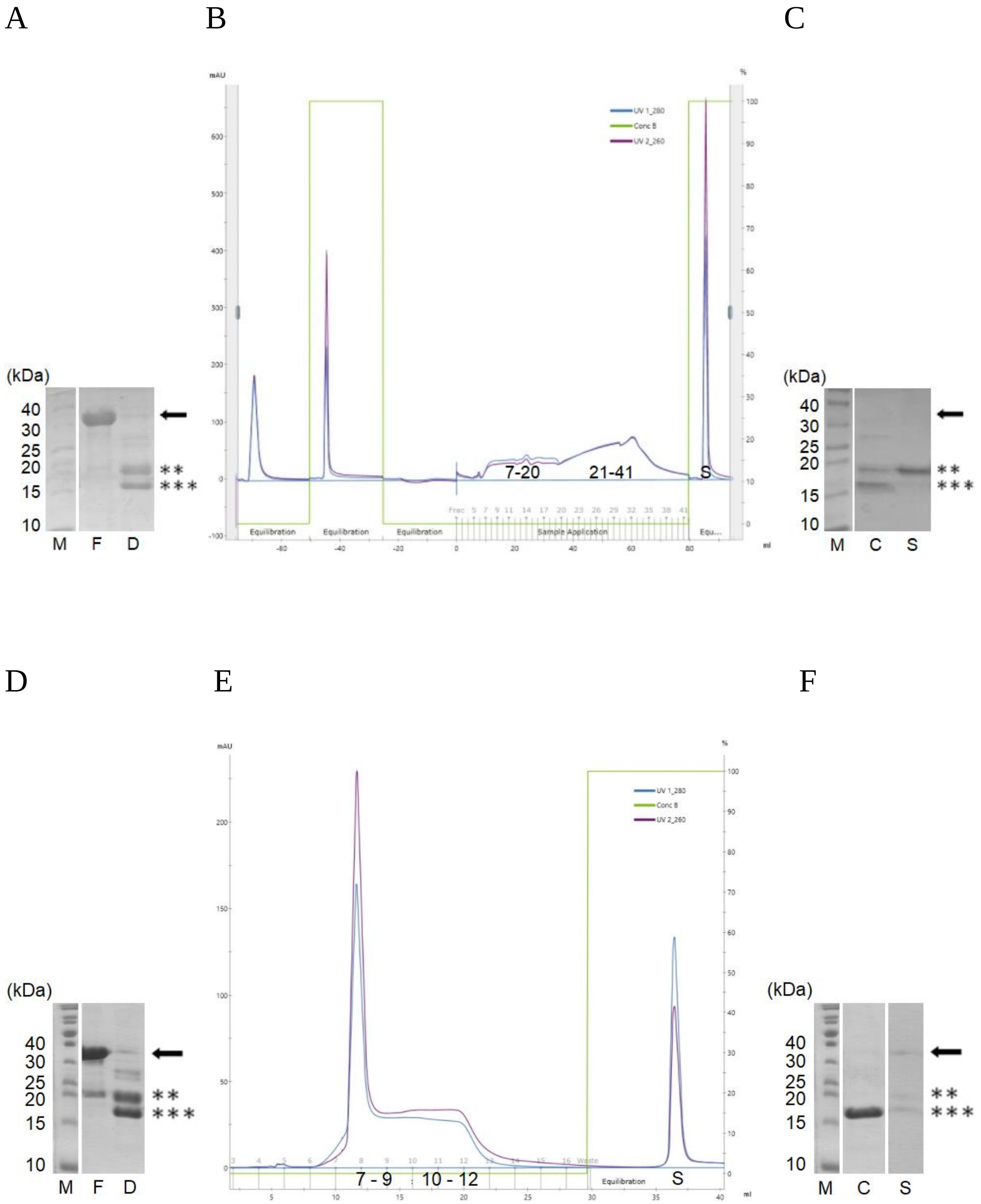


Figure R22. TEV digestion trials of SUMO-NrdR del132-149 and SUMO-NrdR E36A. (A) An asterisk corresponds to partially degraded proteins, double asterisk to the digested SUMO-TEVcs (the band at 20 KDa) and triple asterisk to digested NrdR del132-149 (approx. 15 KDa). **(B)** Digested NrdR E36A (approx. 17 KDa). Protease:protein w/w ratios tested are indicated below each lane on the 15% SDS-PAGE. The highest, almost fully efficient cleavage was observed at 1:10 protease: protein w/w ratio for both NrdR mutants. In all lanes, the double band between 25-30 kDa corresponds to the excess of TEV protease. The black arrow indicates the expected bands of SUMO-NrdR mutants. Lanes NC (not cleaved), correspond to undigested SUMO-del132-149 and E36A mutants loaded as negative controls.

After digestion, a re-chromatography with Ni-NTA affinity was run with the digested sample containing one of the corresponding 3(GlySer)-NrdR mutants, the SUMO tag (6xHis-SUMO-GluAsnLeuTyrPheGln), traces of TEV protease and undigested fusion SUMO-NrdR mutant (Figure R23B, E, H, K). In this second chromatography, the corresponding NrdR mutants were collected from the flow through during sample loading, and all fractions checked on 15% SDS-PAGE (Figure R23C, F, I, L). The remaining proteins bound to the column were eluted with a single step of 500mM imidazole.



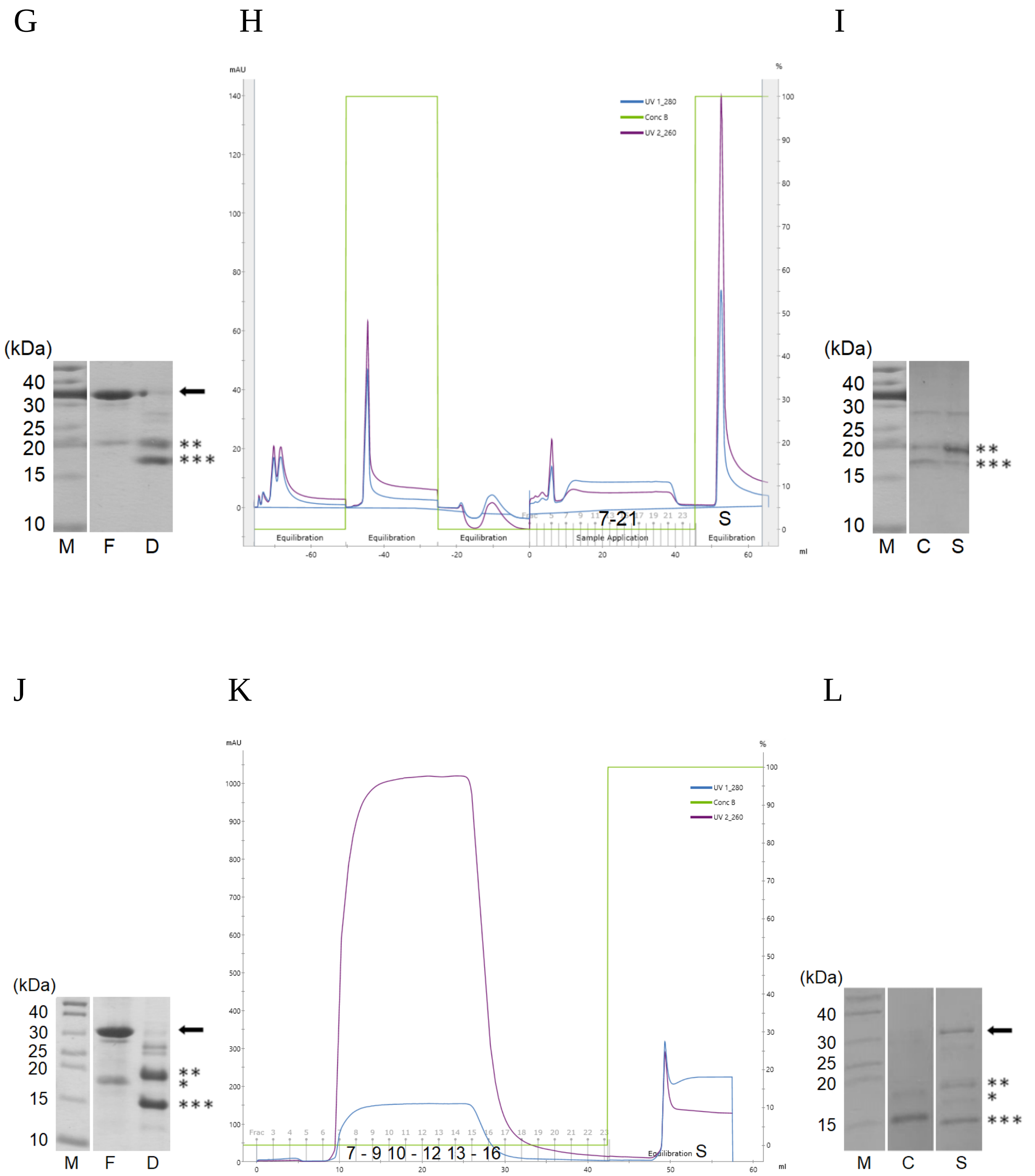


Figure R23. Reverse IMAC and concentration step of digested NrdR E36A (A-C), E42A (D-F), Y131A (G-I) and del132-149 (J-L). (A, D, G, J) 15% SDS-PAGE showing fusion constructs in lane F, and TEV digestion of SUMO-NrdR mutants in lane D. The two weak bands between 25-30 kDa correspond to the TEV enzyme, whereas the two lower intense bands correspond to His-tagged SUMO-TEVcs (20 kDa, two asterisks) and digested NrdR mutagenic constructs (three asterisks) **(A)** E36A (17kDa) **(D)** E42A (17kDa) **(G)** Y131A (17kDa) **(J)** del132-149 (15kDa). The black arrow corresponds to the fusion protein, and the single asterisk to partially degraded mutants. **(B, E, H, K)** Reverse His-Trap affinity chromatography of digested NrdR mutant samples. The absorbance at 280 nm (in blue) and 260 nm UV light (in purple) curves are shown, together with the concentration (% volume) of buffer

B containing imidazole (green line). **(B, E)** NrdR E36A and E42A flow-through fractions 7 to 20 (21-41 were discarded) and 7 to 21 were collected respectively, which showed a value of A260/A280 less than 1. His-tagged components (SUMO protein, TEV protease and non-digested fusion protein) eluted at the last chromatographic step with 100% buffer B (500 mM imidazole). **(H)** NrdR Y131A construct was collected from the fractionated flow-through (10 to 22 ml), which showed a A260/A280 ratio > 1, likely caused by nucleotide moiety(es) bound to the protein. **(K)**, NrdR del132-149 was collected in fractions 7-16, note that the A260/A280 ratio > 5, which was caused by the excess of ATP and the presence of the nonidet p-40 detergent, which strongly absorbs at 260nm.. **(C, F, I, L)** 15% SDS-PAGE with His-tagged components eluted at 500mM of imidazole were loaded in lane S, and concentrated protein from the flow through in lane C. Indicators are the same as for **(A, D, G, I).** Pure, highly concentrated protein represented by a single band, was observed only for mutant Y131A as shown in panel **I**.

All samples showed great variability in the ratio of absorbances A260/A280. The mutant samples showed an important variety of this ratio that, considering the NrdR function, we attributed to a different degree of binding to single nucleotides. Cleaved NrdR E36A and NrdR E42A, affected at the ZFD, showed a A260/A280 ratio lower than 1 (Figure R23B, E), which we interpreted as partial loss of bound nucleotides at the ACD. Indeed, the collected flow-through fraction containing cleaved NrdR E36A and E42A was not pure, as assessed by SDS-PAGE (Figure R23 C, F), and the protein was almost lost during concentration. Also during collection of the flow-through, which contained cleaved NrdR E36A (Figure R23B) fractions 21-41 were discarded, as the Abs ratio 260/280 started to rise up in unexpected manner.

We suspected that loss of nucleotides by these mutants could have created the instability of the protein, therefore we repeated the isolation and concentration procedures of NrdR E36A and NrdR E42A in the presence of triphosphates and/or higher salt content, which did not result in any improvement (not shown).

Re-chromatography of digested NrdR Y131A showed that the sample eluted with an A260/A280 ratio higher than 1 (Figure R23H), suggesting the presence of a nucleotide moiety(es) bound to the protein. The sample showed high stability and, therefore, the whole flow-through fraction was concentrated to 1ml at a concentration of 5.5 mg/ml and highly pure NrdR Y131A was kept for further experiments (Figure R23I).

The digested NrdR del132-149 mutant showed a A260/A280 ratio higher than 5 at the flow-through fraction (Figure R23K), likely due to the excess of ATP and the presence of the nonidet p-40 detergent (0.172 mM) that had to be added to rescue the sample that showed high instability. Yet, the obtained NrdR del132-149 samples (1.7 mg/ml, after concentration) were not completely pure as assessed by SDS-PAGE (Figure

R23L). The nature of the bands was analyzed by fingerprint mass spectrometry, which confirmed that the identified bands in the SDS-PAGE corresponded to the NrdR mutants.

As a final step we included a size exclusion chromatography (SEC) to investigate if the mutants were suitable for a MW analysis by SEC-MALLS (see section R7). This required concentrating the samples to inject them to the column as this type of column dilutes the sample considerably. However, digested mutants NrdR E36A, NrdR E42A and NrdR del132-149 were highly unstable if not lost during concentration, as it was the case for E36A and E42A mutants, which did not elute in SEC in the presence of ATP and del132-149 also did not elute (not shown). Therefore, their elution profiles obtained by SEC could not be compared to that of WT NrdR.

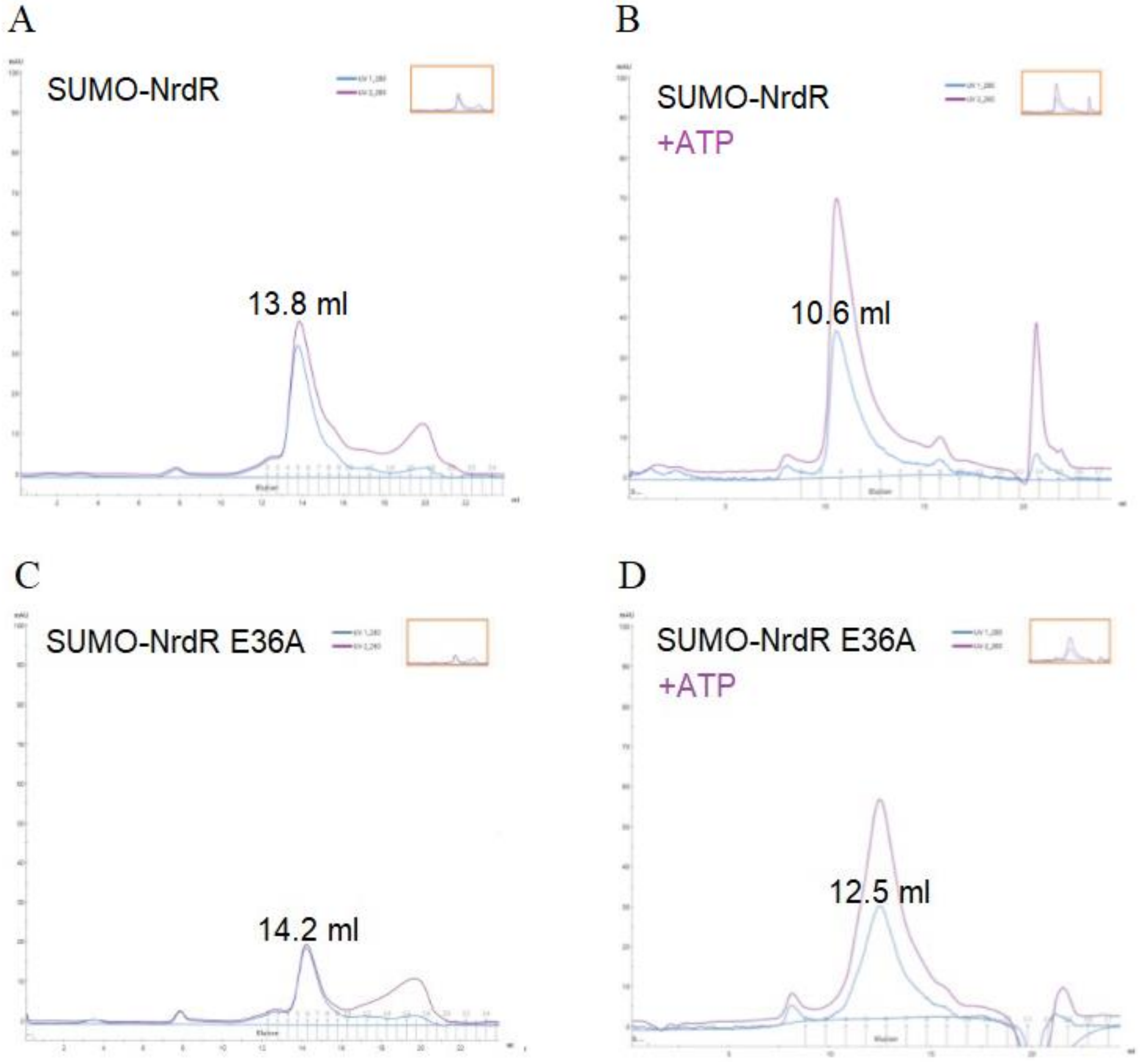
Since the NrdR mutants showed high instability, we performed gel filtration runs with the fused version of SUMO-NrdR E36A, SUMO-NrdR E42A and SUMO-NrdR del132-149 as an alternative, and compare them with SUMO-NrdR. These experiments were done in the presence of ATP (with incubation step), to check if this nucleotide induced highly organized assemblies as observed for WT NrdR (Figure R5A; R25). For a test, we also performed a second round of SEC runs of SUMO-NrdR and mutants without the nucleotide, as absence of nucleotide during SEC caused complete loss of WT NrdR, but here with the fusion protein, presumably more stable, could give some information about multimerization in this condition.

In the absence of ATP, SEC runs of SUMO-NrdR E36A and E42A resulted in longer retention, as they eluted at 14.2 and 15.6 ml, respectively (Figure R24C, E,), compared to SUMO-NrdR (13.8 ml, Figure R24A). This suggested smaller multimers. Instead, SUMO-NrdR del132-149 eluted before (12.5 ml) in spite of its smaller molecular weight (Figure R24G).

In the presence of ATP (Figure R24B), the retention volume of SUMO-NrdR was 10.6 ml, whereas the retention volumes of SUMO-NrdR E36A, SUMO-NrdR E42A and SUMO-NrdR del132-149 were 12.5 ml, 14.8 ml and 15.0 ml, respectively (Figure R24D, F, H). Such a delay in a gel filtration is indicative of a smaller hydrodynamic radius than the native protein, consistent with a more compact conformation, or due to multimer disruption here likely due to impairment of oligomerization by the point mutation at protein-protein interfaces.

Thus, in the presence of ATP (with incubation step), SUMO-NrdR (Figure R24A, B), SUMO-NrdR E36A (Figure R24C, D) and SUMO-NrdR E42A (Figure R24E, F) reduced their retention volumes compared to the non-bound respective forms, indicating an increase of the hydrodynamic radius. Intriguingly, NrdR del132-149 did the opposite.

The elution profile of SUMO-NrdR del132-149 with ATP (with incubation step) in the elution buffer, revealed a number of peaks that suggested the presence of aggregates, high oligomers but also forms smaller than SUMO-NrdR (Figure R24H). In the absence of ATP, SUMO-NrdR del132-149 also several peaks appeared (Figure R24G), which indicated the importance of aa132-149 fragment in WT NrdR stability.



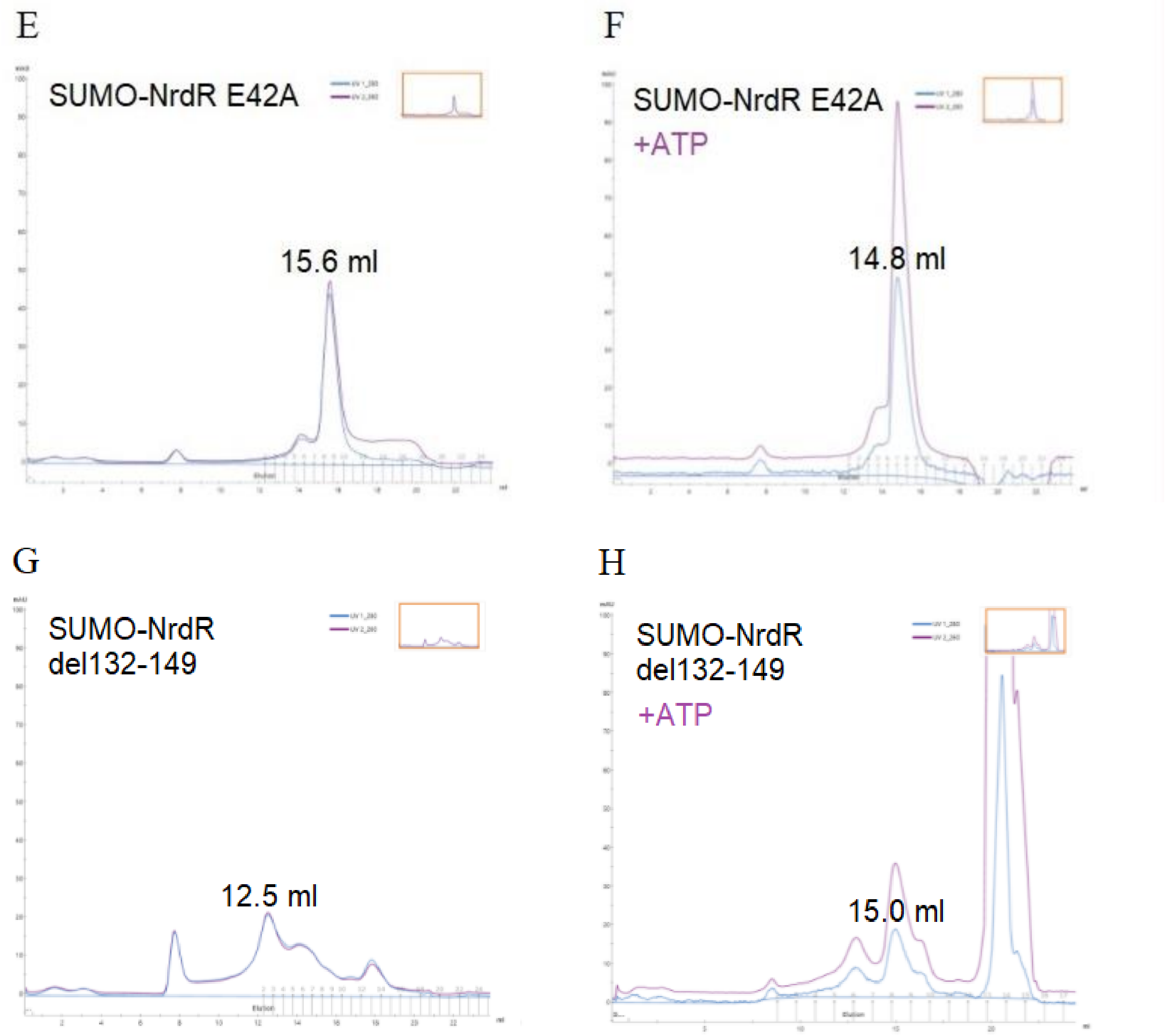


Figure R24. Size exclusion chromatography of SUMO-NrdR and mutants SUMO-NrdR E36A, SUMO-NrdR E42A, SUMO-NrdR del132-149 in the presence (B, D, F, H) and absence (A, C, E, G) of ATP showed the influence of the mutated residues in NrdR oligomerization. **(A, B)** Gel filtration of SUMO-NrdR in the absence and presence of ATP, respectively. The elution profile is shown at 280 (blue curve) and 260 nm (purple curve). The ratio A260/A280 >1 was consistent with ATP binding. The peak at 10.6 ml corresponds to SUMO-NrdR, whereas the peak at 20 ml is due to ATP in excess. **(C, D)** SEC of SUMO-NrdR E36A in the presence of ATP accelerated the elution (12.5 ml) compared to the sample without ATP (14.2 ml). **(E, F)** SUMO-NrdR E42A mutant in the presence of ATP eluted at 14.8 ml, sooner than without ATP (15.6 ml). **(G, H)** SEC of SUMO-NrdR del132-149 in the presence of ATP showed several peaks, the biggest one at 15 ml. In the absence of ATP, the number of peaks increased, the two major peaks at 12.5 and 15ml respectively, suggesting multiple forms.

R7. Multimerization analysis of NrdR and NrdR mutants E36A, E42A, Y131A, del132-149 by Size-exclusion Chromatography coupled to Multi-angle Laser Light Scattering (SEC-MALLS)

R7.1 Absolute molecular weight estimation of NrdR multimers reveals protein oligomerization dependence on type of bound nucleotide

The size exclusion chromatography showed that NrdR in complex with ATP eluted much sooner than the unbound protein, indicating a large Stokes radius consistent with either a wide protein conformation or oligomerization, or both phenomena (Figure R5A). Further, RNRs show different oligomeric states if ATP or dATP is bound to their cone domain, and a similar effect could occur in NrdR. To discern the actual oligomeric state, we determined the absolute MW of apo NrdR, and NrdR in the presence of AMP, dATP or ATP by SEC coupled to Multi Angle Laser Light Scattering (SEC-MALLS).

The nucleotides were included in the elution buffer, whilst the injected samples either contained only protein, or protein preincubated with one of the nucleotides. NrdR ran as a dimer (Figure R25, yellow curve) with an absolute MW of 36.65 and 33.80 kDa (theoretical MW of 17.66 kDa including the N-terminal 3(GlySer), Table R9, yellow), and showed high sample homogeneity as the polydispersity value was of 1.00 (a value above 1.02 indicates a polydisperse sample; see tables R9, R10). NrdR run in the presence of AMP only in the buffer showed a MW of 49.05 and 38.75 kDa, whereas preincubated with AMP a MW of 49.24 and 38.70 kDa (Table R9; blue and dark blue respectively). Values of both conditions, preincubated with the nucleotide or exposed to the nucleotide only in the buffer, were highly consistent and suggested the presence of dimers and trimers within the corresponding eluted peaks (Figure R25; blue and dark blue curves, respectively).

NrdR in the presence of dATP revealed a slight shift between elution peaks of protein without and with preincubation. The non-preincubated protein showed a MW of 129.16 kDa (Table R9 green; Figure R25, green curve) indicating the presence of heptamers, while the protein preincubated with dATP showed a MW of 131.17 kDa suggesting a slight increase towards 8 subunits, probably indicating an equilibrium between heptamers and octamers (Table R9 dark green; Figure R25, dark green curve).

The presence of ATP in the buffer triggered formation of bigger multimers, consisting of 9-10 subunits with an average Mw of 168.17 kDa (Table R9, red; Figure R25, red curve). Interestingly, NrdR preincubation with ATP revealed an even bigger shift corresponding to an absolute MW of 254.80 kDa, indicating formation of 14-mers (Table R9 dark red; Figure R25, dark red curve). These data clearly indicated a dynamic increase of the oligomerization state that depended on the exposure of the protein to the ATP.

In conclusion, NrdR multimerized as dimers-trimers in the presence of AMP, heptamers in the presence of dATP, and 14-mers if ATP was present.

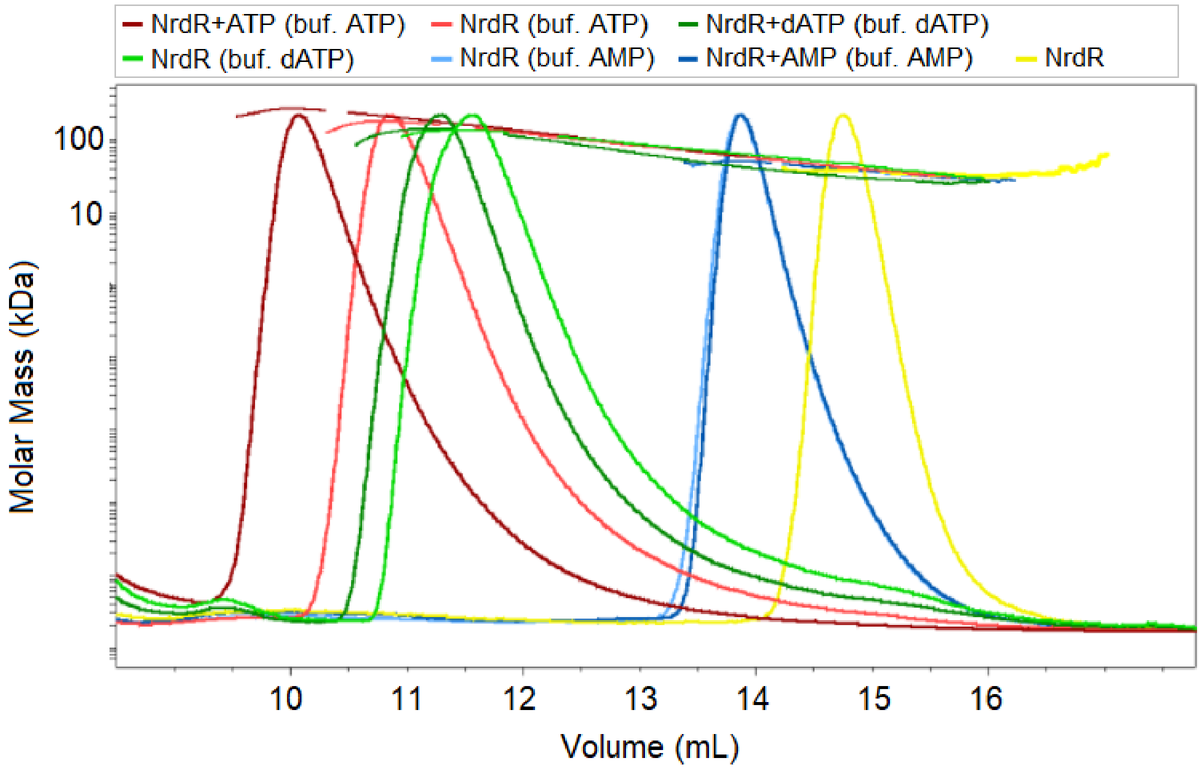


Figure R25. **Nucleotide-dependent oligomerization state of 3(GlySer)-NrdR**. SEC-MALLS scattering profiles of apo NrdR and exposed to nucleotides in the elution buffer with (dark colors) or without preincubation step (lighter colors) revealed differences in elution volumes (shifts between peaks). The estimated MW was, for NrdR (yellow), Mw= 36.65, 33.80 kDa; NrdR+AMP (blue), Mw= 49.05, 38.75 kDa (dark blue), Mw= 49.24, 38.70 kDa; NrdR+dATP (green), Mw= 129.16 (dark green), Mw= 131.17kDa; NrdR+ATP (red) Mw =168.17, (dark red) Mw= 254.80kDa. MALLS detection data was normalized for all obtained peaks. Left axis represents weight-average molar mass (Mw). The horizontal lines along each peak correspond to the measured MW at each point and their curved shape indicates that the proteins form heterogenous mixtures of oligomers staying in a dynamic equilibrium.

Sample	Mn (kDa)	Mw (kDa)	Polydispersity (Mw/Mn)	Calculated mass (µg)	Mass fraction (%)
NrdR	36.65±0.08	36.65±0.08	1.00±0.00	39.09	32.82
	33.70±0.30	33.80±0.31	1.00±0.01	30.59	25.60

NrdR (buf. AMP)	48.98±0.07	49.05±0.07	1.00±0.00	38.28	49.59
	38.14±0.13	38.75±0.11	1.02±0.00	38.90	50.41
NrdR+AMP (buf. AMP)	49.19±0.08	49.24±0.07	1.00±0.00	37.20	48.13
	38.02±0.18	38.70±0.14	1.02±0.01	40.10	51.87
NrdR (buf. dATP)	128.93±0.22	129.16±0.23	1.00±0.00	26.76	37.51
NrdR+dATP (buf. dATP)	129.94±0.23	131.17±0.23	1.01±0.00	28.37	31.10
NrdR (buf. ATP)	167.39±0.33	168.17±0.32	1.00±0.00	23.72	39.30
NrdR+ATP (buf. ATP)	253.96±0.46	254.80±0.45	1.00±0.00	34.40	32.34

Table R9. Absolute molecular weight estimation of NrdR oligomers in the presence of nucleotides . Estimated absolute MW for every multimer. reveals differences in MW of NrdR in complexes with different (d)NTP, but also between oligomers formed with the same type of nucleotide with and without pre-incubation step. Calculations were done based on MALLS signal (Figure R25) collected during SEC runs. Mn, number-average molar mass; Mw, weight-average molar mass; Mw/Mn, polydispersity index related to sample homogeneity. ±standard deviation values are indicated. The last column reveals the protein mass for each elution peak.

R7.2 Absolute molecular weight estimation of SUMO-NrdR and SUMO-NrdR mutants E36A, E42A, Y131A, del132-149 shows impairment in oligomerization mechanism

Since the NrdR mutants showed important instability, we decided to analyze the multimerization state of the different variants fused to the small SUMO protein (12,53 kDa). As shown in section R6.2, the SEC elution profiles of fusion forms of SUMO-NrdR and of mutants SUMO-NrdR E36A, E42A, del132-149 showed retention volumes that varied from one protein to another (Figure R24). In order to clarify if this phenomenon was due to different assembling states caused by disruption of the protein-protein interfaces, or due to different elongated forms that would also alter the Stokes radius, we aimed at determining the corresponding absolute molecular weight by SEC-MALLS.

Therefore, the non-digested forms SUMO-NrdR E36A, E42A, Y131A and del132-149 were compared with SUMO-NrdR in the presence or absence of a nucleotide. Preincubation of NrdR with dATP triggered formation of higher oligomers (Figure R25, Table R9; dark green, dark red) than without preincubation (Figure R25, Table R9;

green, red), so we decided to analyze SUMO-NrdR E36A, E42A, Y131A and del132-149 mutants preincubated with one of the nucleotides, which was also present in the running buffer during SEC-MALLS experiments. The effect of ADP was also analyzed, but only for undigested SUMO-NrdR, as this nucleotide was well defined by the electron density in ATP-cone of molecule C in the protein crystal structure (Figure R16C) but it is likely an oxidation product of ATP bound in *E. coli* than a an ADP with a functional meaning.

SEC-MALLS of SUMO-NrdR in the absence of nucleotide (Figure R26A, light pink curve; Table R10, light pink) rendered three peaks that corresponded to 136.31, 72.99 and 42.16 kDa, consistent with a tetramer, dimer and monomer, respectively (theoretical MW of 31.23 kDa). In the presence of ADP it showed a MW of 97.40 and 70.65 kDa (Table R10, red), suggesting the presence of trimers and dimers within eluted peak, respectively (Figure R26A, red curve).

Interestingly, in the presence of dATP, SUMO-NrdR showed a MW of 176.80 and 139.54 kDa (Figure R26A, brown curve; Table R10 brown), consistent with the presence of hexa- and pentamers, respectively. However, consistent with the digested protein, the biggest assembly of SUMO-NrdR occurred in the presence of ATP, eluting as a multimer of 9 subunits (275.41 kDa, Figure R26A, dark red curve, Table R10, dark red). All these elution profiles changed due to mutations at the interfaces identified in the crystal.

Interestingly, SUMO-NrdR E36A and Y131A showed similar behavior. In the absence of nucleotide, dimers appeared accompanied by a small peak of monomers (MW of 59.27 and 50.58 kDa for E36A, and 61.12, 54.13 kDa for Y131A, vs 31.20 and 31.17kDa for respective monomers) (Figure R26B, light blue curve; R26D, pink curve; Table R10, light blue and pink respectively), which was consistent with non-mutated SUMO-NrdR. However, in the presence of dATP, no higher order multimers than trimers and dimers appeared (Mw of 92.02, 68.38 kDa for E36A, and 79.62, 59.90 kDa for Y131A) or probably a mixture of these forms for Y131A (Figure R26B, dark blue curve; R26D, purple curve; Table R10; dark blue, purple). In the presence of ATP, the elution peaks consisted in multimers with a MW between two and four subunits (Mw of 102.84 and 79.33 kDa for SUMO-NrdR E36A, and 96.78, 69.79 kDa for SUMO-NrdR Y131A, respectively) (Figure R26B, blue curve; R26D, magenta curve; Table R10; blue, magenta). In both mutants, thus, dATP systematically induced smaller multimers than ATP.

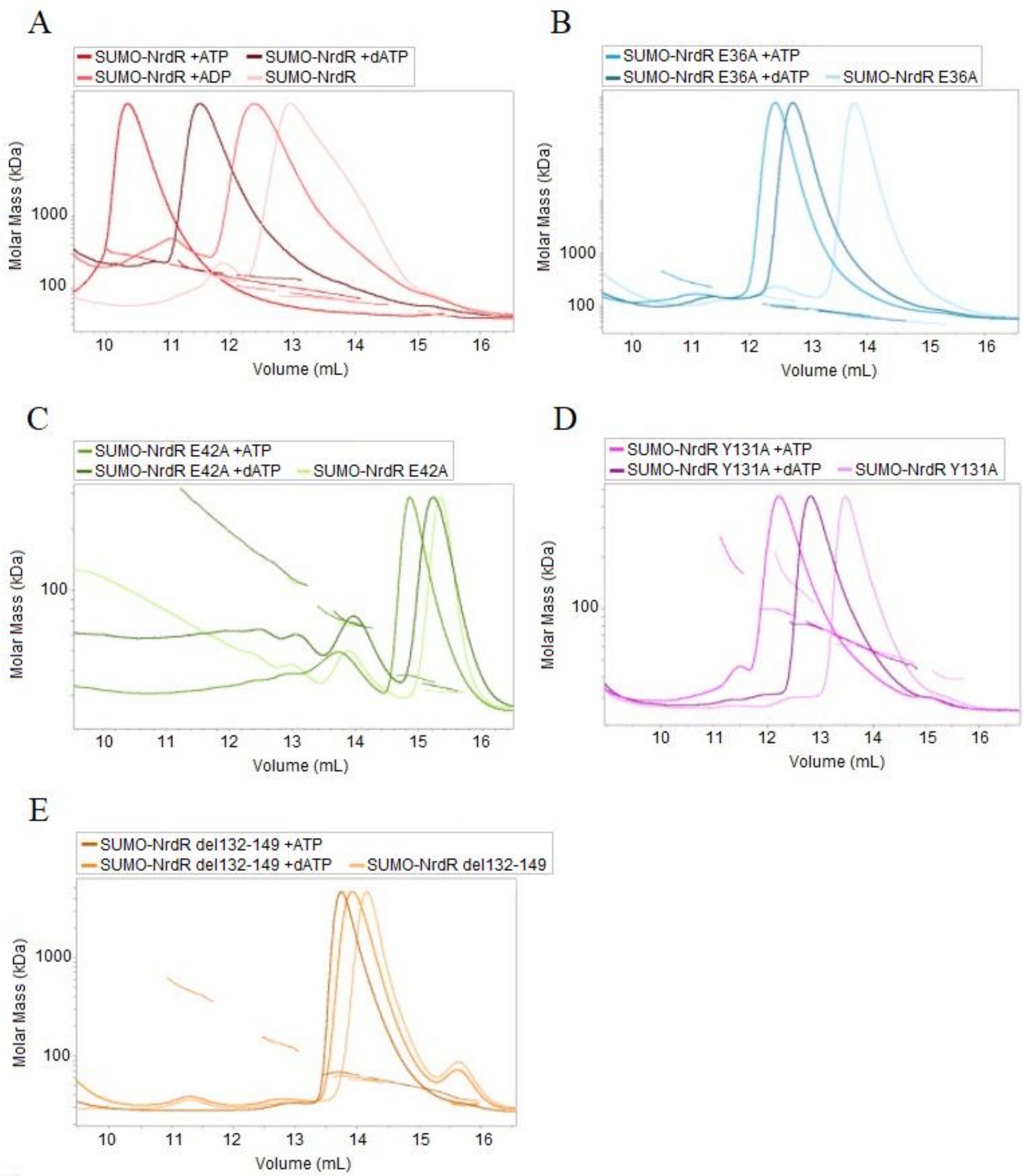


Figure R26. **Nucleotide-dependent oligomerization state of fusion SUMO-NrdR and the corresponding mutants E36A, E42A, Y131A and del132-149**. SEC-MALLS scattering profiles of SUMO fusion WT protein and mutants (light colours), and exposed to different nucleotides during the preincubation step (dark colours), revealed a shift in the corresponding elution volume. **(A)** SUMO-NrdR (light pink), Mw= 136.31, 72.99 and 42.16 kDa; SUMO-NrdR +ADP (light red), Mw= 97.40, 70.65 kDa; SUMO-NrdR +dATP (dark brown), Mw= 176.80, 139.54 kDa; SUMO-NrdR +ATP (red), Mw= 275.41kDa. **(B)** SUMO-NrdR E36A (light blue), Mw= 59.27, 50.58 kDa; SUMO-NrdR E36A +dATP (dark blue), Mw= 92.02, 68.38 kDa; SUMO-NrdR E36A +ATP (blue), Mw= 102.84, 79.33 kDa. **(C)** SUMO-NrdR E42A (light green), Mw= 67.29, 31.39 kDa; SUMO-NrdR E42A +dATP (dark green), Mw= 68.87, 33.21 kDa; SUMO-NrdR E42A +ATP (green), Mw= 72.61, 36.88 kDa. **(D)** SUMO-NrdR Y131A (pink), Mw= 61.12, 54.13 kDa; SUMO-NrdR Y131A +dATP (purple), Mw= 79.62, 59.90 kDa; SUMO-NrdR Y131A +ATP (magenta), Mw= 96.78, 69.79 kDa. **(E)** SUMO-NrdR del132-149 (light orange), Mw= 55.08, 32.52 kDa; SUMO-NrdR del132-149 +dATP (dark orange), Mw= 59.24, 31.44 kDa; SUMO-NrdR del132-149 +ATP (brown), Mw= 65.87, 51.39 kDa. Curved horizontal lines along each peak refers to the Mw and the fact that they show a slope indicates heterogenous mixtures of oligomers presumably staying in a dynamic equilibrium. MALLS data was normalized for all peaks. Left axis represents weight-average molar mass (Mw).

Importantly, SUMO-NrdR E42A showed a very low level of oligomerization, which was illustrated by predominant presence of monomers and, in a lesser amount, dimers, in the absence of nucleotide (MW of 67.29 and 31.39 kDa), but also in the presence of dATP (68.87 and 33.21 kDa) and ATP (72.61 and 36.88 kDa), which did not rescue the effect of the mutation;. (Table R10, light green without nucleotide, dark green with dATP, green with ATP; Figure R26C, light green, dark green, green curves, respectively). These results clearly indicated that the interface between Zn fingers dimers involving Glu42 was the most important, seeding contact to build up higher oligomeric states.

SEC-MALLS of SUMO-NrdR del132-149, (Figure R26E, light orange curve; Table R10, light orange) rendered two peaks that corresponded to 55.08 and 32.52 kDa, consistent with a dimer and monomer, respectively (SUMO-NrdR del132-149, 29.22kDa). In the presence of dATP, dimers and monomers were detected (59.24, 31.44 kDa), but not higher multimers. Interestingly, ATP induced dimers accompanied by a small peak of monomers, but not higher order multimers (65.87, 51.39 kDa) (Figure R26E; Table R10; orange, light brown).

In conclusion, the difficulties in purifying the digested NrdR mutants indicated the importance of the interfaces for protein stability. In addition, the mutations reduced the degree of NrdR multimerization strongly induced in the WT protein, revealing the importance of these residues and the corresponding protein surfaces in NrdR multimerization. Notably, E42A was the most drastic mutation as it almost completely abolished dimerization and any higher multimerization, suggesting that the antiparallel contacts through the ZFD are crucial for the interactions in other regions to occur.

Sample	Mn (kDa)	Mw (kDa)	Polydispersity (Mw/Mn)	Calculated mass (µg)	Mass fraction (%)
SUMO-NrdR	135.24±1.29	136.31±1.30	1.01±0.01	39.12	6.88
	72.95±0.69	72.99±0.69	1.00±0.01	461.74	81.20
	42.06±0.45	42.16±0.45	1.00±0.02	63.82	11.22
SUMO-NrdR +ADP	97.23±0.19	97.40±0.19	1.00±0.00	61.59	38.60
	69.65±0.32	70.65±0.30	1.01±0.01	97.95	61.40
SUMO-NrdR +dATP	175.55±2.38	176.80±2.44	1.01±0.02	107.65	50.65
	139.27±2.79	139.54±2.77	1.00±0.03	100.67	47.36
SUMO-NrdR +ATP	273.13±3.17	275.41±3.19	1.01±0.02	241.04	44.02
SUMO-NrdR E36A	59.24±0.61	59.27±0.61	1.00±0.01	318.33	60.16
	50.45±0.90	50.58±0.88	1.00±0.03	183.87	34.75
SUMO-NrdR E36A +dATP	91.91±0.11	92.02±0.11	1.00±0.00	184.11	48.14
	67.22±0.17	68.38±0.16	1.02±0.00	196.91	51.49
SUMO-NrdR E36A +ATP	102.70±1.11	102.84±1.11	1.00±0.02	203.16	54.91
	78.58±1.00	79.33±1.00	1.01±0.02	150.57	40.70
SUMO-NrdR E42A	67.25±1.28	67.29±1.28	1.00±0.03	15.16	11.15
	31.39±0.43	31.39±0.43	1.00±0.02	115.81	85.13
SUMO-NrdR E42A +dATP	68.70±0.15	68.87±0.15	1.00±0.00	34.06	15.43
	33.19±0.07	33.21±0.07	1.00±0.00	139.83	63.37
SUMO-NrdR E42A +ATP	72.38±0.99	72.61±0.99	1.00±0.02	46.93	16.54
	36.85±0.47	36.88±0.47	1.00±0.02	236.83	83.46
SUMO-NrdR Y131A	61.12±0.05	61.12±0.05	1.00±0.00	274.44	47.86
	53.99±0.11	54.13±0.11	1.00±0.00	241.86	42.17
SUMO-NrdR Y131A +dATP	79.57±0.11	79.62±0.11	1.00±0.00	214.55	44.14
	59.00±0.16	59.90±0.15	1.02±0.00	271.51	55.86
SUMO-NrdR Y131A +ATP	96.69±0.07	96.78±0.07	1.00±0.00	142.51	44.45
	68.52±0.08	69.79±0.08	1.02±0.00	168.48	52.54

SUMO-NrdR del132-149	55.06±0.61	55.08±0.61	1.00±0.02	206.54	72.76
	32.50±0.41	32.52±0.41	1.00±0.02	77.32	27.24
SUMO-NrdR del132-149 +dATP	59.17±0.70	59.24±0.70	1.00±0.02	232.79	78.87
	31.44±0.67	31.44±0.67	1.00±0.03	53.23	18.03
SUMO-NrdR del132-149 +ATP	65.84±0.08	65.87±0.08	1.00±0.00	869.49	52.24
	50.40±0.14	51.39±0.15	1.02±0.00	783.59	47.07

Table R10. **Absolute molecular weight of NrdR fused to SUMO (SUMO-NrdR) and fusion SUMO-NrdR mutants E36A, E42A, Y131A, del132-149 in the absence and presence of nucleotides.** Calculations were done based on the MALLS signal (Figure R26). Mn, number -averaged molar mass; Mw, weight average molar mass; Mw/Mn, polydispersity index, ± indicates the standard deviation. The last columns indicate the protein mass in each elution peak.

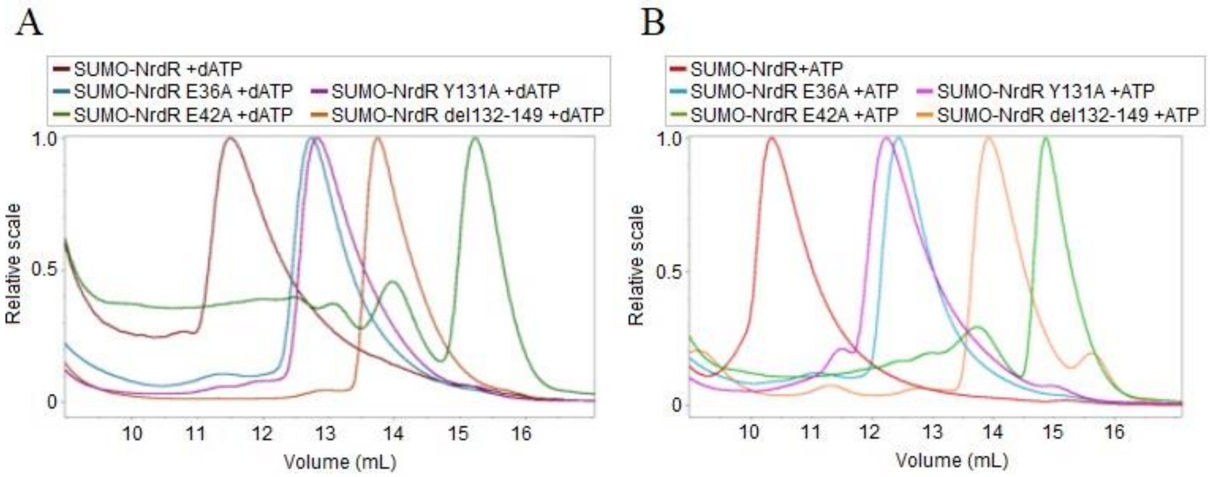


Figure R27. **Superposition of nucleotide-dependent oligomers of SUMO-NrdR and SUMO fusions of E36A, E42A, Y131A, del132-149 mutants**. SEC-MALLS scattering profiles of SUMO-WT protein and mutants exposed to dATP (dark colors) or to ATP (bright colors). **(A)** SUMO-NrdR +dATP (dark brown), SUMO-NrdR E36A +dATP (dark blue), SUMO-NrdR E42A +dATP (dark green), SUMO-NrdR Y131A +dATP (purple), SUMO-NrdR del132-149 +dATP (brown). **(B)** SUMO-NrdR +ATP (red), SUMO-NrdR E36A +ATP (blue), SUMO-NrdR E42A +ATP (green), SUMO-NrdR Y131A +ATP (magenta), SUMO-NrdR del132-149 +ATP (orange). MALLS detection data was normalized for all obtained peaks. Left axis represents a relative scale.

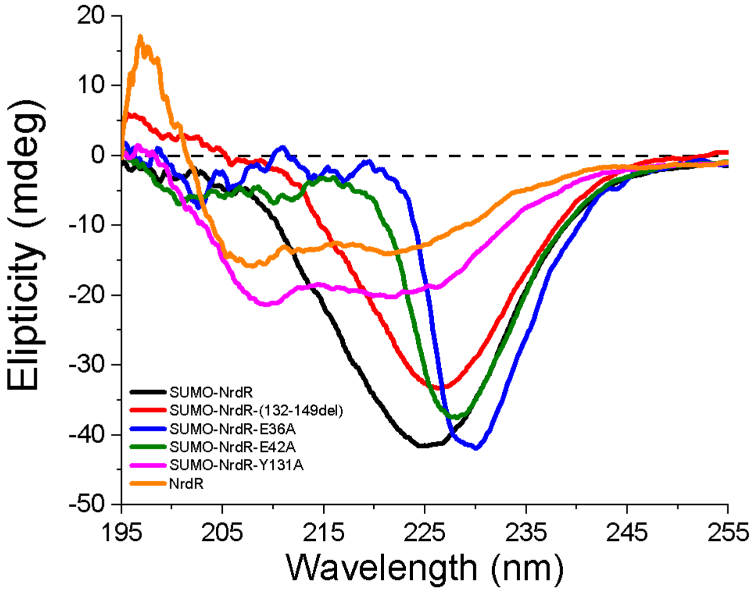


Figure R28. Stability of WT digested NrdR (orange), Y131A-NrdR mutant (pink) and SUMO-NrdR (black), SUMO-NrdR E36A (blue), SUMO-NrdR E42A (green), SUMO-NrdR-del132-149 (red). CD spectra for full-length NrdR and mutants used in SEC-MALLS experiments confirmed proper folding of all studied proteins. The spectra show that the digested NrdR Y131A have a profile similar to the wild-type digested protein, while SUMO-NrdR E36A, E42A, del132-149 mutants, have a profile similar to the SUMO-NrdR.

The CD spectra registered for digested NrdR and NrdR Y131A show dominant alpha-helical fold due to ATP-cone domain over the βetha-fold of the Zn-finger domain. The spectra registered for SUMO-NrdR and SUMO-NrdR E36A, E42A, del132-149 mutants reveal a curve shape typical for β-folds due to the zinc finger domain and SUMO protein (Figure R28). Shifts and width between fusion mutants are possibly related with the differences in the oligomeric state of mutants, as shown by SEC and SEC-MALLS techniques, while depth of all obtained curves is a concentration related parameter. Circular dichroism experiments confirmed the stability of WT NrdR and Y131A NrdR mutants, but also SUMO-NrdR, and all SUMO-NrdR mutants E36A, E42A, and del132-149.

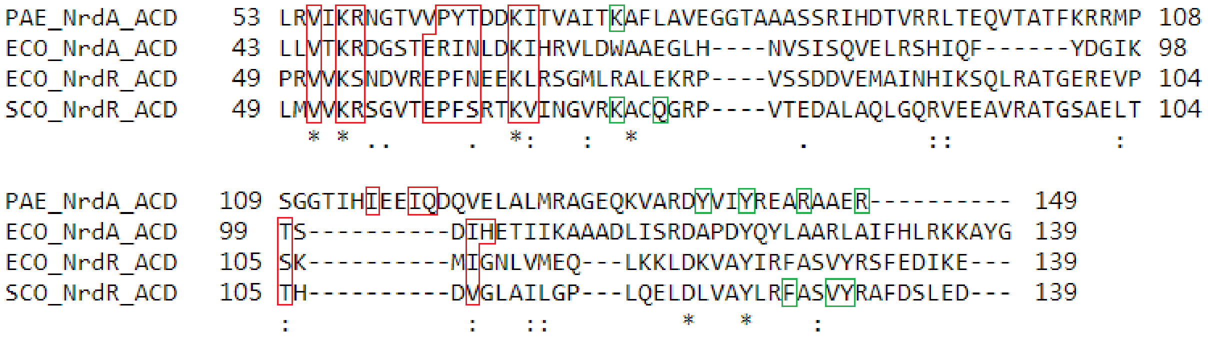
R8. Relationship between NrdR function and overall multimerization mechanism

R8.1 Structural and sequential comparison of NrdR and NrdA proteins

There are four structures of ACD available, the two RNRs *P. aeruginosa* NrdA (PDB ID 5IM3), and *E. coli* NrdA (4ERM); and the two NrdRs from *E. coli* (our structure) and NrdR from *S. coelicolor* (7P37). The sequence alignment of the ATP-cone domains of these proteins RNR NrdA from *P. aeruginosa*, www.uniprot.org/uniprot kb/Q9I4I1 /entry), NrdA from *E. coli* (www.uniprot.org/uUniprotkb /P00452/ entry) and *E. coli* repressor NrdR (www.uniprot.org/uniprotkb/P0A8D0/entry) and NrdR from *S. coelicolor* (www.uniprot.org/uniprotkb/O69980/entry) showed similarities in the aminoacid from the ATP-cone domains participating in NTPs binding (Figure R30). This alignment shows that binding of one NTP cavity is considerably conserved (Figure R30 red frames), but also that sequences of *P. aeruginosa* NrdA and *S. coelicolor* NrdR contain residues coordinating second nucleotide (Figure R30A green frames).

However, some drastic aminoacid changes are found. The more abundant long side chain of Arg54 in S. coelicolor NrdR (Arg58 and Arg48 in P. aeruginosa NrdA and *E. coli* NrdA, respectively) is substituted by a Ser54 in *E. coli* NrdR. In the former organisms, this residue contacts α-ATP and β-ATP phosphates in *E. coli* NrdA, which binds one NTP, but also interestingly contacts γ-ATP phosphate in *P. aeruginosa* NrdA, and in *S. coelicolor* NrdR, binding two NTPs. Phosphate coordination variety by mentioned arginine residue is caused by differences in phosphate chain conformation due to bound number of NTP, as binding of two NTPs in the ATP-cone domain requires different position of phosphates, than in case of binding one NTP. In *E. coli* NrdR structure, in which ATP-cone binds one NTP, Ser54 contacts γ-ATP phosphate in molecule B and β-ADP phosphate in mol C (see Figure R16B, C), but does not preserves any contacts in case of AMP and ATP, bound in molecules A and D, respectively (see Figure R16A, D).

Another drastic change is rigid Pro60 in *E. coli* NrdR (Pro60 and Pro64 in *S. coelicolor* NrdR and *P. aeruginosa* NrdA, respectively), which in *E. coli* NrdA is a long and charged Arg54 side chain. Notably, the ATP cone structures of *P. aeruginosa* NrdA and *S. coelicolor* NrdR show the binding of a second nucleotide. The aa's contacting such a second NTP are framed in green in Figure R30. Interestingly, not all aa contacting the second NTP in *P. aeruginosa* NrdA are conserved in aligned sequence of *S. coelicolor* NrdR. However, aa contacting the second NTP in *S. coelicolor* NrdR are present in *E. coli* NrdR, but absent in *E. coli* NrdA. This is not strange, as the cone domains between RNRs and NrdR are dissimilar.



```
PAE_NrdA_ACD      53 LRVIKRNGTVVPYTDDKITVAITKAFLAVEGGTAAASSRIHDTVRRLTEQVTATFKRRMP 108
ECO_NrdA_ACD      43 LLVTKRDGSTERINLDKIHRVLDWAAEGLH----NVSISQVELRSHIQF------YDGIK 98
ECO_NrdR_ACD      49 PRVVKSNDVREPFNEEKLRSGMLRALEKRP----VSSDDVEMAINHIKSQLRATGEREVP 104
SCO_NrdR_ACD      49 LMVVKRSGVTEPFSRTKVINGVRKACQGRP----VTEDALAQLGQRVEEAVRATGSAELT 104
                       * * ..     .  *:   :  *           .        ::           :

PAE_NrdA_ACD     109 SGGTIHIEEIQDQVELALMRAGEQKVARDYVIYREARAAER----------  149
ECO_NrdA_ACD      99 TS----------DIHETIIKAAADLISRDAPDYQYLAARLAIFHLRKKAYG  139
ECO_NrdR_ACD     105 SK----------MIGNLVMEQ---LKKLDKVAYIRFASVYRSFEDIKE---  139
SCO_NrdR_ACD     105 TH----------DVGLAILGP---LQELDLVAYLRFASVYRAFDSLED---  139
                     :            :   ::         *   *    :
```

Figure R30. Struture-based alignment of *P.aeruginosa* NrdA, *E.coli* NrdA, *E.coli* NrdR, *S.coelicolor* NrdR. The structures show that aa found to interact with the NTP also found in NrdR ("first" NTP) are similar (aa framed in red). Instead, the "second" nucleotide found in *S.coelicolor* NrdR and *P.aeruginosa* NrdA is recognised by the nucleotides framed in green.

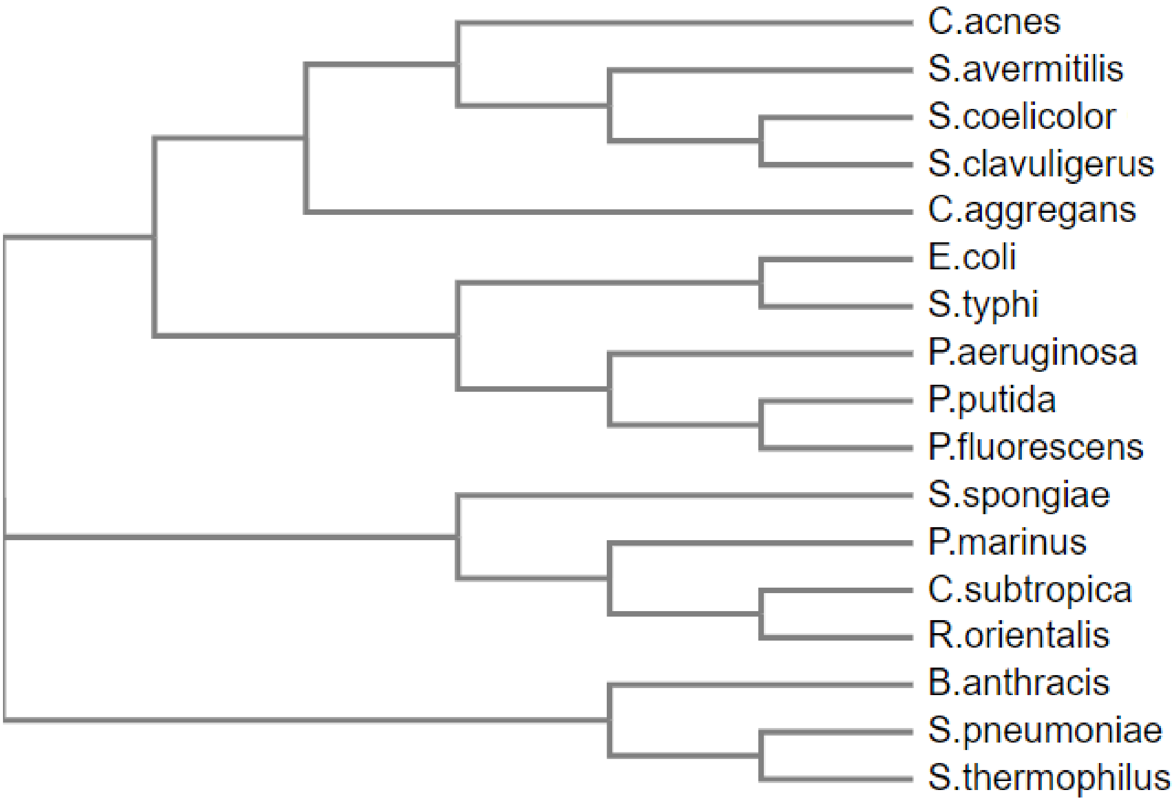


Figure R31. Phylogenetic relationship between NrdR proteins from different bacterial species.

```
                                 Zn-finger                          ATP-cone

C.acnes             MRCPFCQHDDSRVLDSRVCEDGTAIRRRRQCPMCERRFTTIER---MQMTVHKRNGIEEP      57
S.avermitilis       MHCPFCRHPDSRVVDSRTTDDGTSIRRRRQCPDCSRRFTTVET---CSLMVVKRSGVTEP      57
S.coelicolor        MHCPFCRHPDSRVVDSRTTDDGTSIRRRRQCPDCSRRFTTVET---CSLMVVKRSGVTEP      57
S.clavuligerus      MHCPFCRHPDSRVVDSRTTDDGTSIRRRRQCPDCSRRFTTVET---ASLMVIKRSGVTEP      57
S.spongiae          MKCPRCGKQEIRVLESRSAEGGQSVRRRRECMSCGYRFTTYERIEFMPIMVIKRDGSRES      60
P.marinus           MQCPTCQNTDSRVLESRSADSGKSVRRRRECLNCSFRFTTYERVETMPVSVLKKDGGREL      60
C.subtropica        MQCPYCQHTNSRVLESRSSEGGQSIRRRRECLNCKHRFTTYERIEFVPITVIKHDGKKES      60
R.orientalis        MQCPYCQHTNSRVLESRSSEGGQSIRRRRECLCCKHRFTTYERIEFVPITVIKHDGKKES      60
C.aggregans         MRCPICHFPETDVIDTRKLYEGEVIRRRRKCRACGRRFTTYERIESVSLMVVKKDGTREP      60
E.coli              MHCPFCFAVDTKVIDSRLVGEGSSVRRRRQCLVCNERFTTFEVAELVMPRVVKSNDVREP      60
S.typhi             MHCPFCFAVDTKVIDSRLVGEGSSVRRRRQCLVCNERFTTFEVAELVMPRVIKSNDVREP      60
P.aeruginosa        MHCPFCGAHDTKVIDSRLVAEGDQVRRRRECLACGERFTTFETAELVMPRLIKQDGSRQP      60
P.putida            MHCPFCGANDTKVIDSRLVAEGEQVRRRRECVACGERFTTFETAELVLPRLIKQDGTRQP      60
P.fluorescens       MHCPFCGANDTKVIDSRLVAEGEQVRRRRECLACGERFTTFETAELVMPRLIKTDGSRQP      60
B.anthracis         MRCPSCSHNGTRVLDSRPVDEGRSIRRRRECESCLSRFTTFERVEESPLIVVKKEGTREE      60
S.pneumoniae        MRCPKCGATKSSVIDSRQAEEGNTIRRRRECDECQHRFTTYERVEERTLVVVKKDGTREQ      60
S.thermophilus      MRCPKCQHNKSNVIDSRQAEDGNTIRRRRECDACHARFTTFERVEEVPLLVVKKDGTREQ      60
                    *:** *      *:::*    *  :****:*  *  **** *        : * ..  :


C.acnes             FSRDKVIRGVRNACKGRPVTDADLALLGQRVEDGLRARGVAEIPSEEIGLAILGPLRELD      117
S.avermitilis       FSRTKVINGVRKACQGRPVTEDALAQLGQRVEEAVRATGSAELTTHDVGLAILGPLRELD      117
S.coelicolor        FSRTKVINGVRKACQGRPVTEDALAQLGQRVEEAVRATGSAELTTHDVGLAILGPLQELD      117
S.clavuligerus      FSRTKVISGVRKACQGRPVTEDALAQLGQRVEEAVRATGSAELTTHDVGLAILGPLQELD      117
S.spongiae          FNRNKILQGVMRACQKTQVSVKQMEELVNEIEEKLQLEDVQEVTTLRIGEMVLERLQRLS      120
P.marinus           FDKQKLFTGISRACEKTNFSSEEIINFVDGIESQIVQDSNKDIKSSHIGELILKNLRKEN      120
C.subtropica        FDASKLLRGMVRACEKTGISHQRLETIVDDIEAYLQQRPQREVTTHEIGQLVLKYLREEN      120
R.orientalis        FDPSKLLRGMVRACEKTGISYQKLETIVDDIEAQLQQRTQREVTSQEIGQLVLKYLRQEN      120
C.aggregans         YDREKIARGVRTACYRRPVSADAIEQLVNEVETAVMATDQQEISSQFIGDAVMRRLRDLD      120
E.coli              FNEEKLRSGMLRALEKRPVSSDDVEMAINHIKSQLRATGEREVPSKMIGNLVMEQLKKLD      120
S.typhi             FNEDKLRSGMLRALEKRPVSADDVEMALNHIKSQLRATGEREVPSKMIGNLVMEQLKKLD      120
P.aeruginosa        FDEDKLRAGMQRALEKRPVSVERLEAAIGHIKHQLRATGEREIKSRVLGELVMAELQKLD      120
P.putida            FDEEKLRAGMQRALEKRPVSVERLEAALAHIKSRLRATGEREVKSLVVGEMVMAELRKLD      120
P.fluorescens       FDEEKLRAGMQRALEKRPVSVERLEAALVHIKHKLRATGEREVKSLVVGELVMTELQKLD      120
B.anthracis         FNKEKILRGLIKACEKRPVSLRQLEEVTQSVERELRNLGISEVKSDMIGEIVMEELRDID      120
S.pneumoniae        FSRDKIFNGIIRSAQKRPVSSDEINMVVNRIEQKLRGRNENEIQSEDIGSLVMEELAELD      120
S.thermophilus      FSRDKIFNGILMSAQKRPVSSEDIENAITRIEQNIRRNHDGEVDSEVIGNLVMKELADLD      120
                    :.  *:  *:  :     .:   :      ::  :      :: :  :*  ::  *   .


C.acnes             SIAYLRFASVYLNYNTIEDFATEIDRLRDESSDKTKSSSRKRLS--------------KQ      163
S.avermitilis       LVAYLRFASVYRAFDSLEDFEAAIAELREEQRERPAVDDEDHEDAGAERQGTDRGSGGTV      177
S.coelicolor        LVAYLRFASVYRAFDSLEDFEAAIAELRETTGHPG-----EEDDTGAGSQENDRGPTGAG      172
S.clavuligerus      LVAYLRFASVYRAFDSLEDFEAAIAELREHRLPA---------------ERSGGGTCGTG      162
S.spongiae          EVAYVRFASVYRKFQGIKDFVTELEQLERLETHLRRDLERPLRNSPPSESESTASPDWVG      180
P.marinus           EVAYIRYASVYRKFNGVKDFISTLESLKGSSKNQLASIS---------------------      159
C.subtropica        EVAYIRFASVYGRFKGIKDFVETLDQLQEETISSPMSQWSKS------------------      162
R.orientalis        EVAYIRFASVYGRFQGIKDFVDTLKQLQEEDLVPSETIWKPA------------------      162
C.aggregans         EVAYIRFASVYRSFADIGKLREAVDELLEREGTRNGHSSAATTDQGTTDNHSRM------      174
E.coli              KVAYIRFASVYRSFEDIKEFGEEIARLED-------------------------------      149
S.typhi             KVAYIRFASVYRSFEDIKDFGEEIARLQD-------------------------------      149
P.aeruginosa        EVAYIRFASVYRRFQDLNEFREEIERLAREPAKE--------------------------      154
P.putida            EVAYIRFASVYRRFQDLDEFREEIDRLAREPAKE--------------------------      154
P.fluorescens       EVAYIRFASVYRRFQDLNEFREEIDRLAREPAKQ--------------------------      154
B.anthracis         DVAYVRFASVYRQFKDLNVFIEELKDILQKERE---------------------------      153
S.pneumoniae        EITYVRFASVYRSFKDVSELESLLQQITQSSKKKKER-----------------------      157
S.thermophilus      EITYVRFASVYRSFKDVDEIEELLQEITKTVRAKKESKK---------------------      159
                     ::*:*:****  :  :  :   :  :
```

Figure R32. Sequence alignment of 17 NrdR proteins showed similarities within zinc finger and ATP-cone domains, which allows to distinguish key residues responsible for DNA and nucleotide binding. Stars indicate identical residues, semicolons indicate similar residues.

Phylogenetic comparison of putative NrdR proteins from 17 bacterial species performed by the maximum likelihood algorithm and LG (Le-Gascuel) amino acids substitution model implemented in the PhyML program classifies NrdR proteins in three main groups (Figure R31). Alignment of these 17 phylogenetically related NrdR sequences (Figure R32) shows conservation of the zinc finger domain (aa 3-34) and in particular of residues Cys3, Cys6, Cys31, Cys34 that coordinate the Zn, but also Glu42, Arg27 and Arg29, which are crucial for protein dimerization. Interestingly, residue Glu36 necessary for protein tetramerization is only present in *E. coli*, *S. typhi* and *Pseudomonas* species, which suggests possible differences in the multimerization mechanism in case of *C. acnes*, *S. avermitilis, S. clavuligerus* and *S. coelicolor* NrdRs, they include an insertion of three aa at the linker between the ZFD and the ATP-cone domain (Figure R32, R33).

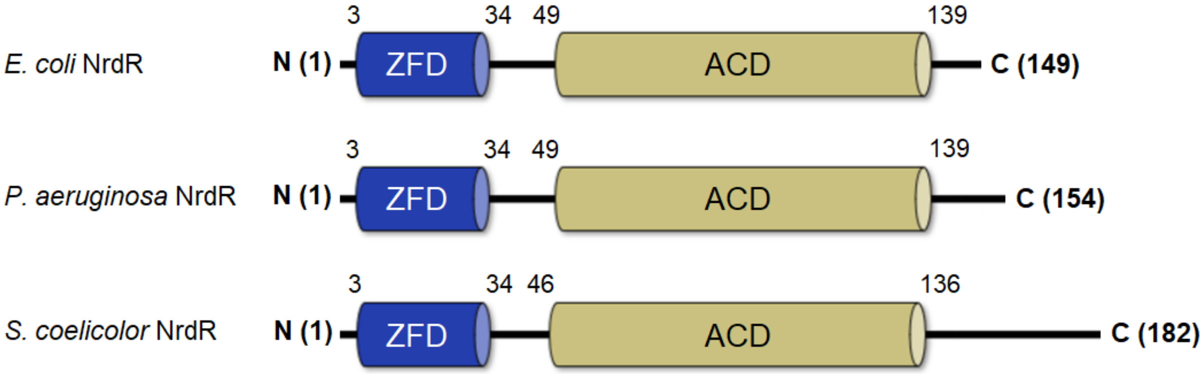


Figure R33. Predicted domains for the three NrdR bacterial proteins from *E. coli, P. aeruginosa* and *S. coelicolor.*

Alignment of the NrdR sequences also shows that the aa participating in the binding of the NTP molecules are conserved. This includes Val51 (or Leu in case of *Pseudomonas* species), Lys53, Glu59 (or Gln in case of *Pseudomonas* species), aromatic Phe61 (or Tyr61 in case of *C.aggregans*), hydrophilic X62 (X=Ser, Asn, Asp), Lys65, hydrophobic X66 and X108 (for both X=Val, Ile, Leu), Ser105 (or Thr105) and Tyr131 within the fully conserved ASVY motif (note the mentioned aa have a minus three units shift due to the deletion in *C. acnes, S. avermitilis, S. coelicolor, S. clavuligerus*). Binding of the second NTP molecule in case of *S. coelicolor* NrdR ATP-cone is mediated by residues Lys69, Gln72, Phe124, Val127, Tyr128 (Grinberg et al., 2022), which correspond to Arg72, Glu75, Phe127, Val130, Tyr131 in *E. coli*. So, hypothetically, *E. coli* NrdR may bind a second NTP molecule but, interestingly, in the crystal structure of *E. coli* NrdR we found only one nucleotide in each ATP-cone. However, protein was crystallized without ATP addition, which could have limited binding of a second ATP molecule. Residues that perform interactions by the C-terminal helixes (aa 121-149) with crystallographic neighbors in *E. coli* NrdR structure, i.e. Lys121, Val122, Ile125, Arg126, Ser129, Val130, Phe134, Ile137, Phe140, Ile144, Leu147 are conserved (similar or identical) throughout NrdRs from different bacterial species, suggesting that oligomerization mechanism occurs in the same manner.

R8.2 Structural comparison of NrdR from E. coli and S. coelicolor

It has recently been published the EM structure of *S. coelicolor* NrdR (Grinberg et al., 2022; pdb id: 7P37), which shows a closed ring of a 12-mer in which the ACD of each subunit binds to two ATP molecules simultaneously. The structural superposition of the ZFD shows the conservation of the interacting interface between ZFDs from molecules A and B, that is, the antiparallel ZFD dimer. However, the relative orientations between ZFD and ACD are different between the two structures, so that the ATP cone domains are reorganized to perform the 12-mer interactions. Note that between the two domains there is a linker that in *E. coli* is longer and it could also account for the different orientations between domains.

S. coelicolor NrdR ACD 12-mer shows two ATP molecules bound to the same cavity as the ATP in *E.coli*. The “first” ATP (ATP1) coincides with the position of the ATP in E. coli NrdR. The second ATP (ATP2) is far away from the lid, and its adenine base makes a π-π interaction with the adenine from another ATP2 of another ACD in the 12-mer. Interestingly, the interaction between these two ATP-cones in *S. coelicolor* is highly similar to the one of molecules A and A’ in our crystal structure, involving aa 121-149, suggesting that such an interaction also may exist in the other bacterial species due to the relevant sequence similarity.

The *S. coelicolor* NrdR report also shows a complex bound to both ATP and dATP, which forms an octamer (pdb id: 7P3Q). The overall arrangement is different from the one in *S. coelicolor* 12-mer or the one observed in our *E. coli* NrdR crystal structure. Notwithstanding, the antiparallel interaction of the ZFD is conserved. The ACD is in a third different relative orientation to the ZFD (different from the 12-mer and from the one in E. coli), but it also makes contacts by helices α4 and α5 with a second ACD, as found in the 12-mer and in *E. coli*. At the ACD, an ATP1 molecule is positioned in the same orientation as in *E. coli* NrdR ATP and *S. coelicolor* ATP1, whereas the dATP is positioned as ATP2. However, the dATP2 moyeties do not perform the π-π interactions observed between the ATP2s in the 12-mer.

Another, third structure reported for *S. coelicolor* NrdR includes a tetramer bound to the NrdR-boxes of *nrdRJ* promoter region (see section I3.2 Figure I5), and in which each ACD binds an ATP1 and a dATP2 (pdb id: 7P3F, Grinberg et. al in 2022). The overall arrangement is the same as a tetramer in the octamer, but the second tetramer in this latter is substituted by the DNA. Very interestingly, two ZFDs are again arranged in the recurrent antiparallel dimer, with each subunit fitting into two consecutive major grooves and inserting the more distal β-strand deep into the corresponding DNA groove and imposing a DNA bend. Arg12, Asp15, Arg17, Arg26-29 and Arg37, highly conserved, participate in DNA binding suggesting the same type of DNA recognition throughout species, including *E. coli*.

Our structure does not contain DNA and, as above mentioned, two pairs of ZFD (pairs AB and CD), form the BA:DC tetramer in the a.u., where the contact A:D occurs. As explained, at this interface, Arg28 and Arg37 from A (or D) molecule, are involved in interactions with Glu36 from molecule D (or A, respectively). The same surface in molecules B and C is facing the solvent. Note that in *S. coelicolor* NrdR, an Arg36 is found at the position equivalent to Glu36, so the salt bridges established by this latter are not possible in the former. Instead, Arg36 is close to Glu99 from the ACD of the second ZFD dimer subunit.

Regarding the ACD, the interactions observed in *E.coli* NrdR crystal structure between a.u., which involve helixes α4 and α5 (aa 121-149) between molecules CD and D’C’ respectively, are also observed in *S. coelicolor* NrdR bound to DNA. However, these helical contacts of the two ACDs are with the ACDs from a second ZFD dimer that also fits into the major groove but three DNA helical turns away from the first ZFD pair. This second ZFD pair also bends the DNA. Therefore, despite the ACDs performing similar contacts in both structures, the overall arrangement is different, the ACDs of a ZFD are separated to contact the other ACDs pair. The dATP2 moieties do not perform the π-π interactions with the ones of the contacted ACD, as occurred in the octameric arrangement. The *S. coelicolor* DNA-bound tetramer fits into the concave surface of a DNA bend by two distant ZFD pairs. Since the residues in the ZFD that contact the DNA are conserved, and the contacts between ACD are observed in all structures, a similar arrangement is expected for *E. coli* NrdR bound to the DNA.

Discussion

This doctoral thesis reveals the crystal structure of *E. coli* of the transcriptional repressor NrdR that inhibits the expression of the *RNR* operon, and deciphers four types of key interactions necessary for the formation of complexes. The NrdR structure consists of four molecules found in the asymmetric unit that interact extensively with each other. Interactions include the N-terminal ZFD that makes two types of contacts. Two of these domains (molecules A and B) make a stable antiparallel contact that involves salt bridges between Glu42 (mol A) with Arg27 and Arg29 (mol B; and the reverse contact). A second contact is between the two ZFD antiparallel pairs (sort of tetramer), is less stable, and involves Glu36 (from mol A) with Arg28 and Arg37 (from mol D) and the reverse interactionAt the cone domain, different nucleotides are found (AMP in molecule A, ADP in molecule C and ATP in molecules B and D). Interestingly, close to the nucleotide binding site, Tyr131 is found. The cone domains also interact in pairs in the above A, B, C and D molecules but, in a more stable manner, with crystallographic mates by the C-terminal helixes (aa 132-149). From these observations, one question that arises was whether these contacts were important for the function and stability of NrdR.

NrdR and the mutants show different stabilities

Based on the above interactions, mutants E36A, E42A, Y131A, and del132-149 (deletion of the C-terminal helix from the cone domain to the end of the structure) were designed and fused to the SUMO protein, as native NrdR. All native and mutant fusion forms SUMO-NrdR, SUMO-NrdR E36A, SUMO-NrdR E42A, SUMO-NrdR Y131A, and SUMO-NrdR del132-149 were expressed in *E. coli* BL21 (DE3). The addition of $ZnSO_4$ to the protein expression cultures was crucial for high protein production, probably because it stabilized a correct folding of the ZFD present at the N-terminus. Zn (II) is coordinated by the -SH sulfhydryl groups (also termed thiol group) of four cysteine residues, thereby locking the ββα fold of the Zn finger domain and stabilizing the small hydrophobic core of this domain. Regarding the mutants, the fused forms showed good yields, yet the purification conditions of both SUMO-E42A and SUMO-NrdR Y131A required to be optimized.

After the TEV-digestion step, the theoretical isoelectric point of the undigested and digested forms (pI_{SUMO} and pI_{digest}, respectively) of the mutagenic proteins differed. The digestion efficiency was higher for SUMO-NrdR and SUMO-NrdR-Y131A than for

SUMO NrdR-E36A, NrdR-E42A and NrdR-del132-149, suggesting a strong influence of residues Glu36, Glu42 and aa132-149 segment for protein stability. Cleaved NrdR and Y131A mutant were recovered in good amounts upon the re-chromatography step after digestion and under the same experimental conditions. Further, highly pure NrdR and Y131A proteins were easily concentrated without significant losses. Interestingly, digestion of fusion SUMO-NrdR E42A to NrdR-E42A caused a complete loss of the bound nucleotide as evidenced by the comparison of the absorbance ratio A_{260}/A_{280} value between WT NrdR and NrdR-E42A. This is indicative of the inter-relationship between the ZFD, which contains the mutated residue, and the ACD, which contains the nucleotide. These results showed that excision of the SUMO fusion from NrdR-E42A unveiled a destabilized ZFD that maybe due to the impossibility of this domain in making appropriate contacts, or by transmission of the instability to the cone domain, or both reasons simultaneously, severely affected the capability of the cone domain to bind nucleotides.

Regarding NrdR-E36A, which disrupts the interface between ZFD antiparallel pairs, the bound nucleotides were presumably maintained as the A_{260}/A_{280} ratio did not change upon digestion. In any case, despite E36A maintaining the nucleotide, both NrdR-E36A and E42A mutants were completely lost after digestion, reflecting high inability for both. The crystal structure of WT NrdR shows that Glu36 and Glu42 salt-bridge respective pairs of arginine residues (Glu36 contacts Arg28 and Arg37; Glu42 contacts Arg27 and Arg29). Therefore, the mutants probably collapse as the charges of the arginines are not anymore compensated by the ablated glutamates from a second mutated Zn-finger domain, making this effect more drastic when the interface between the antiparallel pair is affected. The low value of the absorbance at 260 nm in NrdR-E42A points to the loss of NTPs at the cone domain triggered by the Zn finger domain. However, it is not clear from the NrdR structure how the stability of the Zn finger domain is translated to the stability of the cone domain. One possibility could be a not complete multimerization that impairs contacts between cone domains, which were required to stabilize nucleotide binding.

A successful re-chromatography of cleaved NrdR-del132-149 required buffer optimization. Note this deletion affects the last C-terminal helix, whose weak density in the electron density maps suggests high flexibility. This deletion was designed based on the fact that both helices α4 and α5 contact the symmetry partners in molecules A, C and D. In molecule B, the residues after Tyr131 could not be traced so that the loop Lα4-α5 and the following helix α5 were not visible due to high flexibility, whereas the rest of the structure remained intact.

Therefore, in order to break the interactions of helices α4 and α5 with the symmetry partners, the deletion of residues beyond Tyr131 (Lα4-α5 and the following helix α5) was an option. Once the mutant NrdR-del132-149 was at hand, DTT did not stabilize it as it did with digested WT and Y131A mutant NrdRs (probably by compensating the reactivity of the cysteines at the ZFD), which can be explained by the long deletion that could had unstructured the ACD. Likewise, the presence of ATP did not help either, suggesting that the instability was not due to an absence of ATP that could had destabilized the cone domain.

The structural analysis suggests that excision of the last C-terminal 3aa loop and 15aa helix, is expected to expose Ile125, Ala128 and Phe134, thus a group of apolar amino acids to the solvent. Addition of a non-ionic non-denaturing detergent together with DTT and ATP rescued NrdR-del132-149. Probably, the hydrophobic portion of Nonidet P-40 covered the non-polar protein surface masking the exposed hydrophobic residues of NrdR-del132-149, thus preventing aggregation and precipitation. However, the final concentration step resulted in substantial losses and impurities and when injecting the sample to a gel filtration column the protein did not elute. Therefore, the construct was highly unstable despite the presence of the detergent that initially helped.

In conclusion, the SUMO tag showed to be a very good, small but highly stable fusion partner for expression of NrdR mutants E36A, E42A, del132-149 that otherwise were lost. Such an instability suggests that the highly solubility of the 6His-SUMO-TEVcs sequence compensated for the altered mutants and, as a consequence, the stabilization of the whole protein. On the other hand, mutation at the ZFD interfaces disturbed binding of a nucleotide to the ATP cone, suggesting an interdependence between the two domains and also with multimerization.

The NrdR E36A and E42A mutations at the Zn-finger domains, and Y131A and del132-149 at the ATP-cone domain, interfere with NrdR oligomerization

The absolute MW of NrdR in the absence of nucleotides showed a highly homogeneous sample that eluted as a dimer in a gel filtration when compared to the

MW standards. However, gel filtration is not conclusive to determine the absolute molecular weight of the sample as elution is governed by the Stokes radius, which cannot distinguish between expanded forms of a protein and multimers. Indeed, MALLS analyses confirmed multimerization. In the presence of AMP, NrdR formed a more expanded sample that could correspond to a mixture of dimers and trimers. Interestingly, the presence of dATP in the gel filtration running buffer induced the appearance of apparently heptamers, who's preincubation with the nucleotide before gel filtration pushed to octamers.

Regarding ATP, it triggered formation of even larger forms consistent with 9-10-mers, which preincubation with ATP pushed to larger shapes, apparently 14-mers. In conclusion, NrdR dimer multimerized as dimers-trimers in the presence of AMP, hepta- and octamers in the presence of dATP, and 14-mers if ATP was present. It was interesting to see that the two nucleotides ATP and dATP induced important differences in the multimer content.

Instability of NrdR E36A, E42A and del132-149 impaired a precise analysis of the effect of nucleotides with these mutants. This led us to try with the fusions SUMO-NrdR -E36A, -E42A and -del132-149 and compared with SUMO-NrdR, without nucleotide or in the presence of ATP. In the absence of ATP the retention volume of SUMO-NrdR E36A was similar to that of WT SUMO-NrdR, suggesting a dimer in both cases. Instead, SUMO-NrdR del132-149 appeared sooner in gel filtration indicating a bigger Stokes radius (more expanded or bigger multimers for this mutant). In contrast, SUMO-NrdR E42A surprisingly appeared as a monomer. In the presence of ATP, we could observe different oligomerization states, SUMO-NrdR del132-149 and SUMO-NrdR E42A eluted as monomers, whereas SUMO-NrdR E36A, SUMO-NrdR formed tetramers, dimers and monomers.

ADP induced dimers and trimers. Interestingly, in the presence of dATP penta- and hexamers were present. However, the biggest assembly of SUMO-NrdR occurred in the presence of ATP, eluting as a multimer of nine subunits. SUMO-NrdR E36A and Y131A showed mostly dimers in the absence of nucleotides similar to the non-mutated SUMO-NrdR. However, in the presence of dATP, only trimers and dimers appeared respectively, or probably a mixture of these forms for Y131A. ATP induced multimers between two and four subunits, but never reached the 9-mers shown by the WT. The most striking result was the inability of SUMO-NrdR E42A, to multimerize even in the presence of dATP and ATP, clearly indicating that Glu42 is at the most important interface between Zn fingers to build up higher oligomeric states. Finally, SEC-MALLS of SUMO-NrdR del132-149 rendered two peaks consistent with a dimer and monomer either without nucleotide or in the presence of dATP, while ATP induced mostly dimers but not higher order multimers, which also shows the importance of aa 132-149 fragment for higher oligomer formation.

Both E36A and Y131A mutants suggest a similar size of oligomers (dimers and trimers, or tetramers in the presence of ATP). Yet this does not imply the same type of multimer. Y131A mutation does not disrupt any interface, but participates in nucleotide binding and, more importantly, in recognition between ATP and dATP as shown for *S.coelicolor* NrdR by I. Grinberg et al. in 2022. As shown by these authors, binding to one or another nucleotide induces important but different rearrangements. In case of E36A mutation, intermolecular contacts between ZFD dimers AB and CD are broken, but it can still form oligomers that plausibly involve CD dimers at the ZFD that interact by aa 132-149 segment with additional C’D’ molecules, generating tetramers similar to the ones between crystallographic asymmetric units.

Strikingly the Glu42 mutation had the most devastating effect, resulting in only monomers and posing the question of how alteration of the ZFD AB interface can modulate the interactions by the cone domain. Regarding aa 132-149 deletion, SEC-MALLS showed monomers or a mixture of monomers and dimers, indicating that it also irrevocably broke NrdR oligomerization, and not only canceled C-terminal helix-helix interactions corresponding to the ones observed between different crystallographic a.u., but also broke the interactions between ZFD-pairs dimers (Glu36 interface) and between the ZFD-pair monomers (Glu42 interface)observed in the a.u. The results regarding this deletion mutant must be taken with caution, since the deletion induced destabilization due to the exposition of hydrophobic residues to the solvent, which could also have altered the overall structural stability. But the results still can suggest that the NrdR contacts formed between C-terminal helixes (121-149aa) observed in the crystal structure of *E.coli* NrdR between symmetry mates are also essential for protein multimerization from dimers to higher oligomers.

In summary, the mutations reduced the degree of NrdR multimerization upon nucleotide binding, suggesting the importance of these residues and the corresponding protein surfaces in the multimerization of activated NrdR. The stimulation of oligomerization by nucleotide binding varied depending on the residue mutated. ATP or dATP modulated oligomerization of E36A and Y131A so that dimers or trimers were detected but never bigger multimers. For E42A almost no oligomerization was observed, underlying its crucial role. The deletion 132-149aa affects the last helix of the cone domain, and thus no activation by nucleotides is not surprising in this mutant.

Regarding this latter, the formation of the NrdR/ATP enzyme-ligand complex could trigger the stabilization of the C-terminal helices at the ACD, allowing the formation of intermolecular interactions, whereas the absence of nucleotide would result in an increase of flexibility of this regions thereby inhibiting protein multimerization. Interestingly, NrdR also multimerizes in a complex with ADP, which can also be explained by the stabilization of the C-terminal helices. Regarding the NrdR-AMP complex found in the crystal, SEC-MALLS showed that it did not stimulate formation of large multimers, only dimers and trimers, suggesting that presence of AMP in the crystal might be rather a result of dephosphorylation of ADP or ATP bound to cone domain in mol A at the beginning of the crystallization process.

Indeed, our collaborators performed Ion-pair reverse-phase High Pressure Liquid Chromatography (HPLC) experiments with samples 4 months old, from which successful crystallization trials were performed, without addition of nucleotides. The only nucleotides present in the sample analyzed by HPLC were ADP and AMP in the ratio (2:1), yet the same collaborators showed that NrdR binds exclusively ATP or dATP. This suggested degradation of ATP over time during the crystallization process, which took by about 3 weeks to an ADP (in molecule C) and an AMP (in molecule A). Ion-pair reverse-phase performed with digested NrdR preincubated with (d)ATP confirmed 96% occupancy for both triphosphates in protein samples.

Our collaborators also determined which of the identified amino acids are essential for NrdR function by measuring the expression level of the RNR *nrdA* operon in different *E. coli* strains that had ablated the *nrdR* gene. Δ*nrdR* were transformed with plasmids containing the WT protein or the mutant forms. In exponential and stationary phases, K12 Δ*nrdR* cells showed higher levels of (un-repressed) *nrdA* gene compared to both the WT strain and Δ*nrdR* transformed with WT NrdR (which restored the expression to WT levels in both growth phases). A similar response to this latter was observed for K12 Δ*nrdR* complemented with pLG338-E36A and pLG338-Y131A. Very interestingly, *nrdA* expression in K12 Δ*nrdR* + pLG338-E42A was higher than in K12 WT and even higher than the expression in K12 Δ*nrdR* strain.

These results indicate that among mutations Glu36Ala, Glu42Ala and Tyr131Ala, lack of residue Glu42 not only abolishes NrdR ability to repress transcription of RNRs, but exacerbates the expression of the operon, which tries by all its means to synthesize dATP to inhibit NrdR and itself, which is not achieved due to the inactivating Glu42 mutation. Instead, WT NrdR repression activity level was maintained when residues Glu36 and Tyr131 were mutated. Therefore, amino acid Glu42 is critical for the correct repressive function of RNRs and it may be essential for the right folding of NrdR.

S. coelicolor NrdR (Grinberg et al., 2022; pdb id: 7P37) shows a closed ring 12-mer of proteins. Probably, the 12-mer of *E. coli* NrdR observed in solution in the presence of ATP has the same structural organization as the ATP-bound *S. coelicolor* NrdR. The structural superposition of *S. coelicolor* and *E. coli* NrdR monomers evidences the high structural conservation of the Zn-finger domains and ATP-cone domains, which can be also assumed for the other bacterial species due to the high overall similarity. In principle the short insertion of three aa at the interdomain loop should not introduce important differences in an overall 12-mer arrangement in *C. acnes, S. avermitilis, S. coelicolor, S. clavuligerus*. The fact that the ZFD antiparallel dimer is conserved suggests that it is biologically relevant for protein function through all NrdR proteins.

Given the sequence conservation, the dATP moiety could also bind similarly in *E. coli* NrdR and induce a similar overall octameric structural arrangement. Yet, it is important to keep in mind that the crystal contacts also can induce crystal-specific arrangements. We also crystallized NrdR in another crystal form, which also showed the same tetramer explained in this thesis, with the same interactions inside the a.u., but the ACDs performed different contacts between a.u. that do not resemble the ones in the Cryo-EM structure and in the structure specifically described in this thesis. The second *E. coli* NrdR structure was incomplete and at lower resolution. Therefore, the same tetramer present in the a.u of both *E. coli* NrdR structures suggests that it is biologically relevant for function of *E. coli* NrdR.

In our crystal, the ZFD are arranged in two pairs AB and CD, and the two pairs contact each other by the A and D molecules, thus forming the BA:DC arrangement inside the a.u. This “tetramer” is not compatible with DNA binding as the geometry does not fit the DNA *NrdR boxes*. The high conservation with *S. coelicolor* NrdR suggests that the actual contact with the DNA involves one ZFD pair contacting two consecutive major groove cavities, and that the BA:DC interaction between the two ZFD pairs is broken to place the AB and CD pairs at three DNA helical turns distance. The A:D contact in the BA:DC arrangement is not found in any *S. coelicolor* NrdR structure, suggesting an irrelevant (crystallographic) interaction.

However, the A:D interaction is mediated by Glu36, which salt bridges Arg28 and Arg37 from the ZFD. In *S. coelicolor* Glu36 corresponds to Arg36, which could not make such contact. Note that Arg28 and Arg37 equivalents in *S. coelicolor* contact the DNA. Therefore, the potential existence of a functional A:D contact in which Glu36 stabilizes Arg28 and Arg37 or inhibits their DNA binding ability is not implausible. Note that our studies in solution of *E.coli* NrdR showed that when is mutated to alanine (NrdR E36A the oligomerization is impaired either in the presence of ATP or dATP, revealing an intriguing role of this residue in *E.coli* NrdR multimerization. Note that this interaction is not crucial *in vivo* as shown by our results in which the NrdR E36A mutant kept the ability to repress RNR operon.

One possibility of its potential relevance might be participation in a transition state during multimerization and/or activation. Our SEC-MALLS data combined with *in vivo* studies showed that the amino acid that plays a key role in protein multimerization, and which is present in all NrdR bacterial species, is Glu42, which interacts with Arg27 from its own molecule and Arg29 from the other ZFD subunit. In the *S. coelicolor* NrdR/DNA EM structure, Arg27 and Arg29 also interact with Glu42, further supporting a role in keeping the functional AB or CD ZFD pairs in all species.

Therefore, important residues that alter multimerization in *E. coli* are not performing contacts in *S. coelicolor*, suggesting that the high conservation point to similar interactions throughout species but specific mechanisms among NrdR are also plausible.

Conclusions

1. Functionality of transcriptional repressor NrdR is strictly regulated by multimerization. The crystal structure of *E. coli* NrdR revealed key interactions that altered the multimerization in this protein.
2. Biophysical analyses in solution (SEC and SEC-MALLS) revealed that the NrdR multimer is unstable when not bound to nucleotides. The WT protein alone or in the presence of nucleotides AMP, ADP, ATP found in the crystal structure of *E. coli* NrdR, but also in the presence of dATP resulted in elution of different assemblies. In the absence of nucleotides only dimers are present for WT NrdR.
3. The same studies performed with SUMO-NrdR and SUMO-NrdR E36A, E42A, Y131A, and del132-149 mutants revealed the importance of residues Glu36, Glu42, Tyr131 and of segment aa 132-149 in oligomerization, with the highest impact noted for SUMO-NrdR E42A and del132-149, resulting in monomers or dimers. Therefore, the interactions between Glu42 with Arg27 from its own molecule and Arg29 within the neighboring Zn-finger, but also those between aa 121-149 between ATP-cones are fundamental for higher oligomeric state formation.
4. *In vivo* assays performed for NrdR E36A, E42A and Y131A indicate that lack of residue Glu42 completely abolishes NrdR ability to repress transcription of RNRs, while ablation of residues Glu36 and Tyr131 did not cause a decrease of the repression level. Therefore, amino acid Glu42 is essential for the repressive function of RNR.
5. The abundance of NrdR protein in bacteria and extrapolation of the results obtained for *E. coli* and *S. coelicolor* NrdR points to an ATP/dATP-guided mechanism in which the type of multimers change and coordinates the repression activity of the RNR operon.

References

1. WHO. Global action plan on antimicrobial resistance. World Health Organization. Geneva, Switzerland. 2015.

2. Van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin SA, Laxminarayan R. Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. Lancet Infect Dis. 2014;14(8):742-750.

3. Hawkey PM. The origins and molecular basis of antibiotic resistance. BMJ. 1998;317(7159):657–660.

4. Taylor NG, Verner-Jeffreys DW, Baker-Austin C. Aquatic systems: maintaining, mixing and mobilizing antimicrobial resistance? Trends Ecol Evol. 2011;26(6):278-284.

5. Perry JA, Wright GD. The antibiotic resistance “mobilome”: searching for the link between environment and clinic. Front Microbiol. 2013;4:138.

6. Jiang X, Ellabaan MMH, Charusanti P, Munck C, Blin K, Tong T, Weber T, Sommer MOA, Lee SY. Dissemination of antibiotic resistance genes from antibiotic producers to pathogens. Nat Commun. 2017;8:1-7.

7. Stokes HW, Gillings MR. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiol Rev. 2011;35(5):790-819.

8. Humeniuk C, Arlet G, Gautier V, Grimont P, Labia R, Philippon A. Beta-lactamases of Kluyvera ascorbata, probable progenitors of some plasmid-encoded CTX-M types. Antimicrob Agents Chemother. 2002;46(9):3045-3049.

9. Poirel L, Liard A, Rodriguez-Martinez JM, Nordmann P. Vibrionaceae as a possible source of Qnr-like quinolone resistance determinants. J Antimicrob Chemother. 2005;56(6):1118-1121.

10. Patel R, Piper K, Cockerill FR, Steckelberg JM, Yousten AA. The biopesticide Paenibacillus popilliae has a vancomycin resistance gene cluster homologous to the enterococcal VanA vancomycin resistance gene cluster. Antimicrob Agents Chemother. 2000;44(3):705-709.

11. WHO. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. World Health Organization. Geneva, Switzerland. 2017b.

12. Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol. 2007;5(3):175-186.

13. Tao Y, Yue Y, Wang J. Abundance and diversity of antibiotic resistance genes possibly released to ambient air by experiments in biology laboratories. Science of The Total Environ. 2021; 797:149147.

14. D'Costa VM, McGrann KM, Hughes DW, Wright GD. Sampling the antibiotic resistome. Science. 2006;311(5759): 374-377.

15. Zhang XX, Zhang T, Fang HH. Antibiotic resistance genes in water environment. Appl Microbiol Biotechnol. 2009;82:397-414.

16. Wright GD. Antibiotic resistance in the environment: a link to the clinic? Curr Opin Microbiol. 2010;13(5):589-594.

17. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother. 2010;54(3): 969-76.

18. Cho H, Uehara T, Bernhardt TG. Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. Cell. 2014;159(6):1300-1311.

19. Wright GD. Molecular mechanisms of antibiotic resistance. Chem Commun (Camb). 2011;47(14):4055-4061.

20. Blair JM, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol. 2015;13(1):42-51.

21. Cantón R, Novais A, Valverde A, Machado E, Peixe L, Baquero F, Coque TM. Prevalence and spread of extended spectrum β-lactamase-producing Enterobacteriaceae in Europe. Clin Microbiol Infect. 2008;14:144-153.

22. Ambler RP. The structure of β-lactamases. Philos Trans R Soc Lond B. 1980;289:321-331.

23. Bush K, Jacoby GA, Medeiros AA. A Functional classification scheme for β-lactamases and its correlation with molecular structure. Antimicrob Agents Chemother. 1995;39(6):1211-1233.

24. Naas T, Poirel L, Nordmann P. Minor extended-spectrum beta-lactamases. Clin Microbiol Infect. 2008;1:42-52.

25. Jacoby GA, Bush K. β-Lactam Resistance in the 21 Century. W: White DG, Alekshun MN, Mcdermott PF, Levy SB. (eds) Frontiers in Antimicrobials Resistance, ASM Press, USA. 2005;53-65.

26. Poirel L, Naas T, Nordmann P. Genetic support of extended-spectrum β-lactamases. Clin Microbiol Infect. 2008;14:75-81.

27. Marcadé G, Deschamps C, Boyd A, Gautier V, Picard B, Branger C, Denamur E, Arlet G. Replicon typing of plasmids in Escherichia coli producing extended-spectrum β-lactamases. J Antimicrob Chemother. 2009;63(1):67-71.

28. Paterson DL, Bonomo RA. Extended-spectrum beta-lactamases : a clinical update. Clin Microbiol Rev. 2005;18(4): 657-686.

29. Torrents E. Ribonucleotide reductases: essential enzymes for bacterial life. Frontiers in cellular and infection microbiology. 2014;4(52).

30. Reichard P, Ehrenberg A. Ribonucleotide Reductase - a Radical Enzyme. Science. 1983;221(4610):514-9.

31. Crespo A, Pedraz L, Torrents E. Function of the Pseudomonas aeruginosa NrdR Transcription Factor: Global Transcriptomic Analysis and Its Role on Ribonucleotide Reductase Gene Expression. PLoS One. 2015;10(4).

32. Jiang W, Yun D, Saleh L, Barr EW, Xing G, et al. A manganese(IV)/iron(III) cofactor in Chlamydia trachomatis ribonucleotide reductase. Science. 2007;316(5828):1188-91.

33. Eliasson R, Pontis E, Jordan A, Reichard P. Allosteric control of three B12-dependent (class II) ribonucleotide reductases. Implications for the evolution of ribonucleotide reduction. The Journal of Biological Chemistry. 1999;274(11):7182-9.

34. Nordlund P, Eklung H. Structure and function of the Escherichia coli ribonucleotide reductase protein R2. Journal of Molecular Biology. 1993;232(1):123-64.

35. Nordlund P, Reichard P. Ribonucleotide reductases. Annu Rev Biochem. 2006;75:681-706.

36. Sintchak MD, Arjara G, Kellogg BA, Stubbe J, Drennan CL. The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. Nat. Struct. Biol. 2002;9:293-300.

37. Larsson KM, Jordan A, Eliasson R, Reichard P, Logan DT, Nordlund P. Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase. Nat. Struct. Mol. Biol. 2004;11:1142–49.

38. Uhlin U, Eklund H. Structure of ribonucleotide reductase protein R1. Nature 1994;370:533–39.

39. Uppsten M, Farnegardh M, Jordan A, Eliasson R, Eklund H, Uhlin U. Structure of the Large Subunit of Class Ib Ribonucleotide Reductase from Salmonella typhimurium and its Complexes with Allosteric Effectors. J. Mol. Biol. 2003;330:87–97.

40. Logan DT, Andersson J, Sjoberg BM, Nordlund P. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. Science 1999;283:1499–504.

41. Matthews CK. DNA precursor metabolism and genomic stability. The FASEB Journal. 2006;20(9):1300-14.

42. Eriksson M, Uhlin U, Ramaswamy S et al. Binding of allosteric effectors to ribonucleotide reductase protein R1: reduction of active sites cysteines promotes substrate binding. Structure. 1997;5(8):1077-92.

43. Zimanyi CM, Ando N, Brignole EJ, Asturias FJ, Stubbe J, Drennan CL. Tangled up in knots: Structures of inactivated forms of E. coli class Ia ribonucleotide reductase. Structure. 2012;20(8):1374-83.

44. Zimanyi CM, Chen PY-T, Kang G, Funk MA, Drennan CL. Molecular basis for allosteric specificity regulation in class Ia ribonucleotide reductase from Escherichia coli. eLife. 2016;12(5).

45. Torrents E, Westman M, Sahlin M, Sjoberg BM. Ribonucleotide reductase modularity: Atypical duplication of the ATP-cone domain in Pseudomonas aeruginosa. The journal of biological chemistry. 2006;281(35):25287-96.

46. Johansson R, Jonna VR, Kumar R, et al. Structural Mechanism of Allosteric Activity Regulation in a Ribonucleotide Reductase with Double ATP Cones. Structure. 2016;24(6):906-17.

47. Jacobson BA, Fuchs JA. Multiple cis-acting sites positively regulate Escherichia coli nrd expression. Molecular microbiology. 1998;28(6):1315-22.

48. Herrick J, Sclavi B. Ribonucleotide reductase and the regulation of DNA replication: an old story and an ancient heritage. Molecular microbiology. 2007;63(1):22-34.

49. Augustin LB, Jacobson BA, Fuchs JA. Escherichia coli Fis and DnaA proteins bind specifically to the nrd promoter region and affect expression of an nrd-lac fusion. Journal of bacteriology. 1994;176(2):378-87.

50. Han JS, Kwon HS, Yim JB, Hwang DS. Effect of IciA protein on the expression of the nrd gene encoding ribonucleoside diphosphate reductase in E. coli. Molecular and general genetics. 1998;259(6):610-4.

51. del Mar Cendra M, Juarez A, Madrid C, Torrents E. H-NS Is a Novel Transcriptional Modulator of the Ribonucleotide Reductase Genes in Escherichia coli. Journal of bacteriology. 2013;195(18):4255-63.

52. Vassinova N, Kozyrev D. A method for direct cloning of fur-regulated genes: identification of seven new fur-regulated loci in Escherichia coli. microbiology. 2000;146(12):3171-82.

53. Lizewski SE, Schurr JR, Jackson DW et al. Identification of AlgR-regulated genes in Pseudomonas aeruginosa by use of microarray analysis. Journal of bacteriology. 2004;186(17):5672-84.

54. Sjöberg B-M, Torrents E. Shift in Ribonucleotide Reductase Gene Expression in Pseudomonas aeruginosa during Infection. Infection and Immunity. 2011;79(7):2663–9.

55. Boston T, Atlung T. FNR-mediated oxygen-responsive regulation of the nrdDG operon of Escherichia coli. Journal of bacteriology. 2003;185(17):5310-3.

56. Grinberg I, Shteinberg T, Gorovitz B, Aharonowitz Y, Cohen G, Borovok I. The Streptomyces NrdR Transcriptional Regulator Is a Zn Ribbon/ATP Cone Protein That

Binds to the Promoter Regions of Class Ia and Class II Ribonucleotide Reductase Operons. J Bacteriol. 2006 Nov;188(21):7635–44.

57. Borovok I, Gorovitz B, Yanku M et al. Alternative oxygen-dependent and oxygen-independent ribonucleotide reductases in Streptomyces: cross-regulation and physiological role in response to oxygen limitation. Molecular microbiology. 2004;54(4):1022-35.

58. Rodionov DA, Gelfand MS. Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling. Trends Genet. 2005;21(7):385-9.

59. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research. 2000;28(1):33-6.

60. Torrents E, Grinberg I, Gorovitz-Harris B et al. NrdR controls differential expression of the Escherichia coli ribonucleotide reductase genes. Journal of bacteriology. 2007;189(14):5012-21.

61. Grinberg I, Shteinberg T, Hassan AQ, et al. Functional analysis of the Streptomyces coelicolor NrdR ATP-cone domain: role in nucleotide binding, oligomerization, and DNA interactions. Journal of Bacteriology. 2009;191(4):1169-79.

62. Grinberg I, Martínez-Carranza M, Bimai O, et al. A nucleotide-sensing oligomerization mechanism that controls NrdR-dependent transcription of ribonucleotide reductases. Nature communications. 2022;13(2700).

63. McKethan BL, Spiro S. Cooperative and allosterically controlled nucleotide binding regulates the DNA binding activity of NrdR. Molecular microbiology. 2013;90(2):278-89.

64. Naveen V, Hsiao CD. NrdR Transcription Regulation: Global Proteome Analysis and its Role in Escherichia coli Viability and Virulence. PloS one. 2016;11(6):e0157165.

65. Aravind L. Wolf YI, Koonin EV. The ATP-cone: an evolutionarily mobile, ATP-binding regulatory domain. Journal of Molecular Microbiology and Biotechnology. 2000;2(2):191-4.