A crystallographic study of effector and substrate binding for the class III ribonucleotide reductase from *Thermotoga maritima*

Viktoria Bågenholm
KEMT30: Master's Degree Project, 60 hp
Protein Science
Autumn 2012 – Spring 2013
Department of biochemistry and structural biology

Supervisors: **Oskar Aurelius**
                      Renzo Johansson
Examiner: **Derek Logan**
Presented: 2013-05-17

Abstract

The protein ribonucleotide reductase catalyses the *de novo* synthesis of deoxyribonucleotides (dNTPs) from ribonucleotides (NTPs), using radical chemistry. It is under strict allosteric regulation, partly from the dNTPs it synthesises that regulate what substrate should be synthesised, but also from dATP/ATP, which regulates overall activity. It is a protein that is found in almost every living organism and it has been subdivided in to three main classes: class I is strictly aerobic, class II indifferent to the presence of oxygen and class III is anaerobic. They all share a similar fold and reaction mechanism and most of them, with the exception of class II, require two different sets of homodimers to function: one generating the radical, the other carrying out the synthesis.

This project continues previous work on the class III ribonucleotide reductase from *Thermotoga maritima* (tmNrdD), a thermophile first discovered in the vicinity of sea vents off the coast of Italy. The protein was expressed in *Escherichia coli*, purified and the purification optimised. Crystallisation was carried out and optimised for substrate and effector binding. X-ray diffraction data were collected to obtain structures of the various effector substrate complexes, to study their differences in structure for further insight into the protein's allosteric substrate specificity regulation.

tmNrdD was successfully purified and several structures with effector, as well as with effector and substrate, were obtained. The effector-substrate complex dATP-CTP bound in accordance with structures of the same effector substrate complex from class II, while the substrate of the dTTP-GTP complex was seen to bind in a previously unknown position. Here the phosphates had shifted, in comparison to CTP, and the ribose had flipped, pointing away from important catalytic residues. Substrate binding with ATP was attempted, but it was only possible to see the presence of pyrophosphate in the active site. Taken together with the unusual conformation of GTP this hints at a mobile substrate, possibly due to the protein being crystallised in an inactive state.

List of Abbreviations

BT4	Bacteriophage T4
CAT	Chloramphenicol acetyl transferase
dNDP	Deoxyribonucleotide diphosphate
dNTP	Deoxyribonuceltide triphosphate
DTT	Dithiothretiol
ESRF	European Synchotron Radiation Facility
MES	2-(N-morpholino)-ethanesulfonic acid
MS	Mass spectrometry
NDP	Ribonucleotide diphosphate
NMR	Nuclear magnetic resonance
NTP	Ribonucleotide diphosphate
PEG	Polyethylene glycol
RNR	Ribonucleotide reductase
tmNrdD	*Thermotoga maritima* α_2 class III ribonucleotide reductase

Table of Contents

Introduction
The first life forms on earth were strictly anaerobic and their genetic material was based on RNA (Robertson and Joyce 2012). However, since RNA is not so stable and prone to mutations, life evolved into using the reduced and more stable form of RNA, DNA, as the basis for its genetic material (Nordlund and Reichard 2006). Part of this evolutionary process involved developing a new enzyme, capable of reducing the building blocks of RNA, ribonucleotides (NTPs), into the building blocks for DNA, deoxyribonucleotides (dNTPs). This protein, which is the only enzyme to catalyse the *de novo* synthesis of dNTPs known to date, was ribonucleotide reductase (RNR) (Jordan and Reichard 1998).

This class of enzymes was first discovered in *Escherichia coli* in 1961 by Peter Reichard (Reichard *et al.* 1961) and have since then been found in almost every type of living organism. The exceptions are a few parasites, including *Giardia lamblia* and *Tritrichomonas foetus*, which rely on their hosts and salvage pathways for dNTPs (Adam 2001). RNR hydrolyses NTPs or NDPs into dNTPs/dNDPs using complex free radical chemistry with a thiyl radical at the active site. To maintain balanced dNTP pools RNR is strictly allosterically regulated at a site called the specificity site by the dNTPs themselves, since the same enzyme catalyses the reduction of all four NTPs. In addition, some RNRs also have an ATP cone, called the activity site, that binds ATP or dATP and regulates the overall activity of the protein: ATP activates and dATP deactivates it (Nordlund and Reichard 2006).

The RNRs have been subject to much divergent evolution and are today divided into three classes, based on how they interact with oxygen and how they generate their radical (Torrents *et al.* 2001; Larsson *et al.* 2004). Evolution is assumed to be divergent, and not convergent, primarily because of the very high similarity in allosteric regulation between the different classes (Poole *et al.* 2002) and their common overall structure (Nordlund and Reichard 2006). The three classes share structural similarities, even if their sequence identity can be very low. From an evolutionary perspective it is commonly agreed that class I is the newest form of the enzyme: it is strictly aerobic, unlike the other two classes, and would have evolved after the appearance of oxygen, while DNA, and thus RNR, evolved before this. Class III is generally considered to have been the first type of RNR. However, the matter is still under some debate (Torrents *et al.* 2002) and there is a possibility class II might instead have been first (Poole *et al.* 2002).

Class I
Class I is, as already mentioned, strictly aerobic and is split into the subgroups Ia, Ib and Ic. It is the most studied and the most widespread: Ia can be found in all higher organisms, as well as many microorganisms, while Ib is only found in bacteria (Johansson *et al.* 2010) and Ic mostly in parasites (Hogbom 2011). It normally consists of two homodimeric proteins: α_2 and the smaller β_2 (Rofougaran *et al.* 2006). While α_2 contains the active, activity and specificity sites, β_2 contains an Fe-O-Fe centre used to generate a tyrosyl radical. This radical is then transferred to a cysteine at the active site on the α subunit, generating the thiyl radical required for the enzyme to function. Ic instead uses an Mn-Fe centre for radical generation and storage, lacking the otherwise strictly conserved tyrosine entirely (Hogbom 2011). The electrons required for the reduction of the NTPs are supplied by thioredoxin for class Ia, and the additional protein NrdH for class Ib (Jordan *et al.* 1997). All class Ia and some Ib enzymes contain the ATP cone activity site on the α_2 protein. For mice it has been shown that when ATP or dATP binds to hyperactivate or deactivate the protein, respectively, class I proteins enter a higher oligomeric state, forming $\alpha_6\beta_2$ oligomers, instead of the normal $\alpha_2\beta_2$ complex (Rofougaran *et al.* 2006). In *E. coli* the RNR instead forms the $\alpha_2\beta_2$ complex when active, and an $\alpha_4\beta_4$ complex when inhibited by dATP (Rofougaran *et al.* 2008).

Class II
Class II RNRs are indifferent to oxygen and can be found in prokaryotes. They lack the β_2 protein, existing as homodimers or monomers of the α subunit. Those that are monomeric contain an additional 130 residue insertion to mimic the dimer interface (Sintchak *et al.* 2002), which is
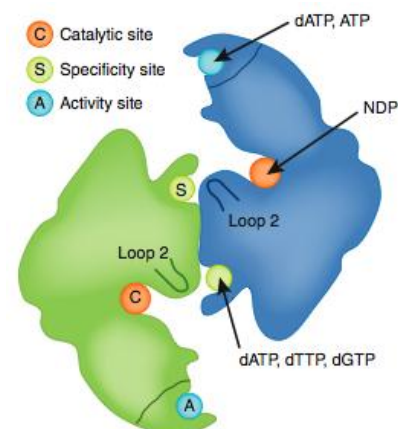
important for allosteric regulation. Since class II lacks the β_2 protein, which normally generates the radical required for enzymatic activity, they instead use the additional cofactor adenosylcobalamin to create a transient 5'-deoxyadenosyl radical, which it transfers to the α subunit to give the same active site thiyl radical as for class I (Larsson *et al.* 2004). They use thio- or glutaredoxin as electron donors (Nordlund and Reichard 2006).

Class III

Class III is present only in prokaryotes and is strictly anaerobic. It initially generates a stable glycyl radical, being one of only four enzymes with known structure that use them (Luttringer *et al.* 2009). If the active glycyl radical is exposed to oxygen a cleavage occurs adjacent to it, removing the 25 C-terminal residues, including the radical itself and rendering the protein non-functional (King and Reichard 1995). Though class III requires an α_2 and a β_2 protein, the latter only functions as an activase. This activase generates the glycyl radical on α_2 by cleaving S-adenosylmethionine with the help of an Fe-S cluster, generating an adenosyl radical, which is transferred to the glycine. The generated glycyl radical is then stable for several cycles of catalysis, and so the activase dissociates. This radical is presumably shuttled between the glycine, where it is stored, and the active site cysteine, where it enables NTP reduction (Logan *et al.* 1999). α_2 uses formate as the electron donor in the catalysed reaction (Mulliez *et al.* 1995). Class III RNRs contain an additional metal site not found in the other classes: a Zn site co-ordinated by 4 cysteines, or sometimes 3 cysteines and a histidine, containing one Zn, in the C-terminal region. This metal site is believed to assist in interactions with the β_2 protein, meaning it is indirectly important for enzymatic activity (Logan *et al.* 2003).

Overall structure

The first RNR structure to be solved was the Class I enzyme in *E. coli* (Nordlund and Eklund 1993; Uhlin and Eklund 1994) and since then structures from every class have been solved, with the exception of the β_2 protein from class III. All three classes share important structural features and a common overall fold of the α subunit, which consists of a 10-stranded α/β barrel with two sets of five parallel β-sheets, arranged in an antiparallel manner. A connecting finger loop protrudes into the cavity created in the centre (Uhlin and Eklund 1994; Logan *et al.* 1999; Sintchak *et al.* 2002). At the tip of this loop the cysteine harbouring the thiyl radical is located, meaning this region contains the active site. The specificity site is instead located by the interface of the α_2 dimer and a loop, commonly referred to as loop 2, connects it with the active site. The activity site, present in only some RNRs, is located far from either of these sites,

Figure 1 The general structure of an RNR (Logan 2011).

being situated in the N-terminal region of the protein in a small α-helical bundle (Figure 1) (Nordlund and Reichard 2006).

In this setup the interactions of the dimers plays an important part, since the specificity site consists of residues from both subunits. It is for this reason that monomeric class II RNRs contain a region mimicking the dimer interface. Dimeric class II are organised in a similar way to class I. The α subunits of class III, however, dimerise with the β-barrel pointing at a 90° angle compared to classes I and II, making the distance between the active and specificity sites longer. Despite this loop 2 still bridges the distance between the 2 sites (Figure 2) (Nordlund and Reichard 2006).
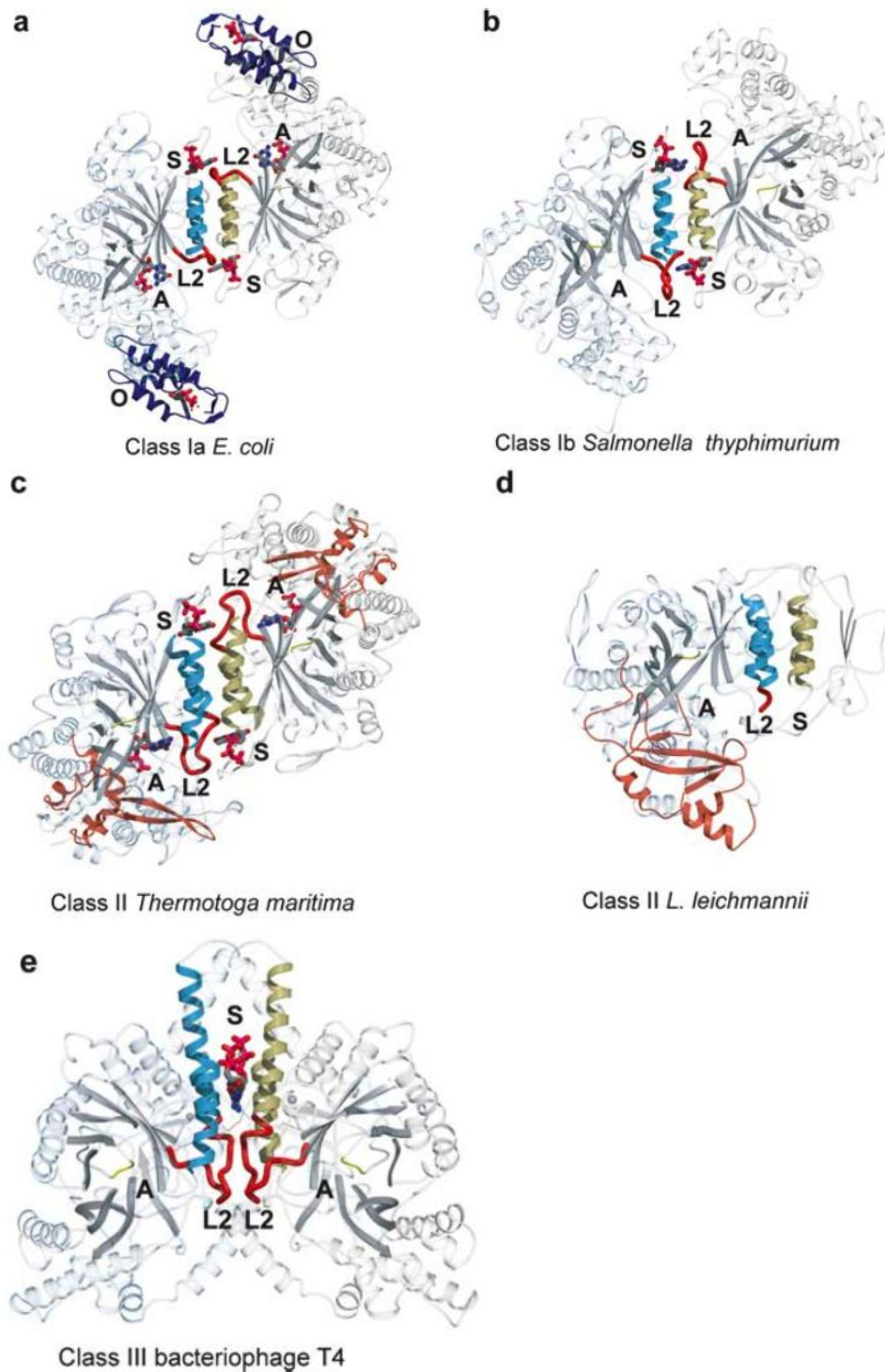
Figure 2 Overview of the three classes of RNR, showing how the dimers are oriented. The ATP-cone is blue, the binding site for adenosylcobalamine is brown. Loop 2 and bound dNTP are coloured red. Abbreviations used: L2 = Loop 2, S = Specificity site and A = Activity site. From: (Nordlund and Reichard 2006)

Reduction mechanism

The reduction of NTP/NDPs by RNR is similar in all three classes, though class III differs the most. Class I use diphosphates as substrates, class III use triphosphates and in class II there are examples of both being used (Kolberg *et al.* 2004). The stable tyrosyl radical on the β_2 protein in class I, or the transient adenosyl radical generated by adenosylcobalamin for class II, are transferred to a universally conserved cysteine at the tip of the finger loop, at the active site of the α_2 protein. This cysteine, corresponding to Cys439 in Class I *Escherichia coli* RNR, creates a transient thiyl radical (Reece *et al.* 2006). When an NTP binds, the ribose is positioned so that the thiyl radical cysteine is

close to the 3´-carbon, from which it removes the hydrogen atom and creates a substrate radical intermediate. The radical is then moved, through the 2´C and O to Cys225. This removes the oxygen atom, which becomes protonated and leaves as water. The 2´C is reprotonated again by the sulphide on Cys462, which creates a disulphide bridge with Cys225 and the two become a disulphide radical. This radical is transferred back the 3´C of the ribose, causing the reverse of the initial reaction step: the 3´C is protonated and the Cys439 thiyl radical is regenerated. From here it is transferred back to the cysteine on the α_2 protein where it was generated, where it is stable enough to await the next round of catalysis. The disulphide bond formed between Cys225 and Cys462 is regenerated using thio- or glutaredoxin (Nordlund and Reichard 2006). Cys225 and Cys462 are generally conserved and located on 2 adjacent barrel β-strands in the active site (Logan *et al.* 1999).

The crystal structure of a class III RNR with bound substrate is yet to be published, but a reduction mechanism has still been suggested. The stable glycyl radical (Gly580, Bacteriophage T4 (BT4)) is located on the α_2 protein, instead of β_2, and is brought into contact with the active site cysteine (Cys290, BT4 numbering, corresponds to Cys439 in *E. coli*) through a long loop from the C-terminal region (Andersson *et al.* 2000). The initial and final steps of the reduction are similar to that of class I and II. However, only Cys225 is conserved in class III (Cys79, BT4 numbering), there is no Cys462 equivalent, and since there is only one cysteine no disulphide bridge is formed. Cys79 still harbours the radical intermediate and the reduction and generation of water is instead performed by formate, which binds directly at the active site. After one round of reduction the radical is transferred back to Gly580, similarly to what happens in class I (Figure 3) (Eklund and Fontecave 1999).

Figure 3 Proposed reduction mechanism of the class III RNR. The residues are numbered according to BT4 class III RNR (Andersson *et al.* 2001).

Allosteric regulation
RNR is allosterically regulated at the specificity and activity sites. The specificity site on the α subunit regulates the production of different dNTPs: binding dTTP stimulates reduction of GTP. Production of the pyrimidines dUTP (later converted to dTTP by a thymidylate synthase) and dCTP are upregulated by dATP and ATP. dGTP binding stimulates dATP reduction (Figure 4) (Kolberg *et al.* 2004). All four dNTPs that regulate RNR bind in a similar manner at the dimer interface: for class I and II the phosphate is anchored in one monomer, being neutralised by a lysine and an

arginine. The 3´OH of the deoxyribose binds directly to an aspartate, and a glycine through a water molecule, with the base sticking out towards loop 2 on the other monomer (Figure 5 and 6) (Larsson *et al.* 2004). While the effector binds across the dimes interface axis for classes I and II, it instead binds along the dimer interface axis in the class III RNRs (Larsson *et al.* 2001).

Each base affects the conformation of loop 2 differently, in turn affecting the specificity of the active site. In class I and II this is primarily dependant on three amino acid residues: Lys202, Gln203 and Arg207 (*Thermotoga maritima* class II RNR numbering). When dTTP-GDP is bound loop 2 adopts a β-hairpin structure, making 4 hydrogen bonds to dTTP. One is direct, by Lys202 to the CO (O4) of the base, and three are bridged by water: the carbonyl of Gln203 and the amide nitrogen of Gly204 to CO (O4), and the carbonyl on Val200 to N3 (Figure 5). The conformation of loop 2 gives space for GDP, which is very bulky,

Figure 4 Overview of the allosteric regulation of RNR. (Larsson *et al.* 2001)

in the active site and projects Arg207 into the binding pocket, clamping the substrate down into it. Here the adenosine guanidinyl moiety makes π-π stacking interactions with the base and water bridges the residue with the ribose and phosphates of the substrate. Interestingly, when no substrate is bound, dTTP does not hydrogen bond with loop 2, indicating that the substrate is also required to stabilise this conformation (Table 1 and Figure 6) (Larsson *et al.* 2004).

Loop 2 also forms a β-hairpin when binding dATP-CDP/UDP, but it is truncated with increased regions of disorder. This is stabilised by NH (N6) and N1 in dATP, which hydrogen bond to the backbone carbonyl and amide groups of Lys202, respectively. Loop 2 also binds CDP/UDP: Gln203 projects into the active site, hydrogen bonding with the NH and CO groups of the CDP base, but only to CO for UDP (Table 1, Figure 5) (Larsson *et al.* 2004).

Figure 5 The four different nucleotides, here shown as monophosphates. The ribose and base rings are numbered.
(www.tulane.edu/~biochem/nolan/lectures/rna/frames/nucs.htm)

For dGTP/ADP loop 2 is disordered, with the exception of an additional turn at the end of the helix that it is attached to, and an α-conformation of residues 201-203. Val200 and Val201 hydrogen bond to the NH (N2) and NH (N3) of dGTP to stabilise this. Lys202 is projected into the active site, hydrogen bonding to the N1 atom on the ADP base (Table 1, Figure 5 and 6) (Larsson *et al.* 2004).

For class III there are only structures with bound effector. Here the phosphate and ribose of all dNTPs interact with the same residues. The phosphates are co-ordinated by a Mg ion and interact with several lysines, as well as Gln176 (BT4 class III numbering), while the 3´-OH of the ribose hydrogen bonds to Glu100. The two main residues responsible for effector recognition are Gln114 and Glu181. Gln114 hydrogen bonds to N1 for dATP, N3 and O4 in dTTP and N3 and N4 in dCTP. dGTP does not interact directly with Gln114, but this residue is part of a hydrogen bonding network that includes Glu181 and dGTP. Glu181 hydrogen bonds to N6 in dATP and N1 and N2 in dGTP, but does not directly interact with dTTP or dCTP, instead hydrogen bonding with Gln114 in those cases. Loop 2 retains the conformation of the native state when dATP is bound and undergoes the largest changes when affected by dTTP (Larsson *et al.* 2001).

The activity site is generally found in class Ia and several class III RNRs (Logan 2011). It requires higher levels of ATP/dATP, since they otherwise only associate to the active and effector sites, respectively (Nordlund and Reichard 2006). The activity site allows regulation of overall dNTP levels, important since too high levels may increase the mutation rate (Chabes *et al.* 2003).
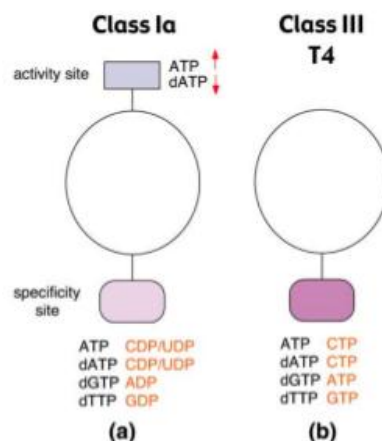
Table 1 Overview of the different interactions between RNR and its substrates and effectors. *T. maritima* class II RNR numbering. For numbering of the atoms in the effector, see figure 5.

Effector	Substrate	Loop 2 conformation	Regions that bind from:		RNR Residues that bind to:	
			Effector	Substrate	Effector	Substrate
dTTP	GDP	β-hairpin	CO (O4), N3	π-π stacking interactions from the base	Val200, Lys202, Gln203, Gly204	Arg207
dGTP	ADP	Disordered, except a turn at the end of helix αB and α-conformation of residues 201-203.	NH (N2), NH (N3)	N1	Val200, Val201	Lys202
dATP	CDP	Truncated β-hairpin with increased regions of disorder	NH (N6), N1	NH, CO	Lys202	Gln203
	UDP			CO		
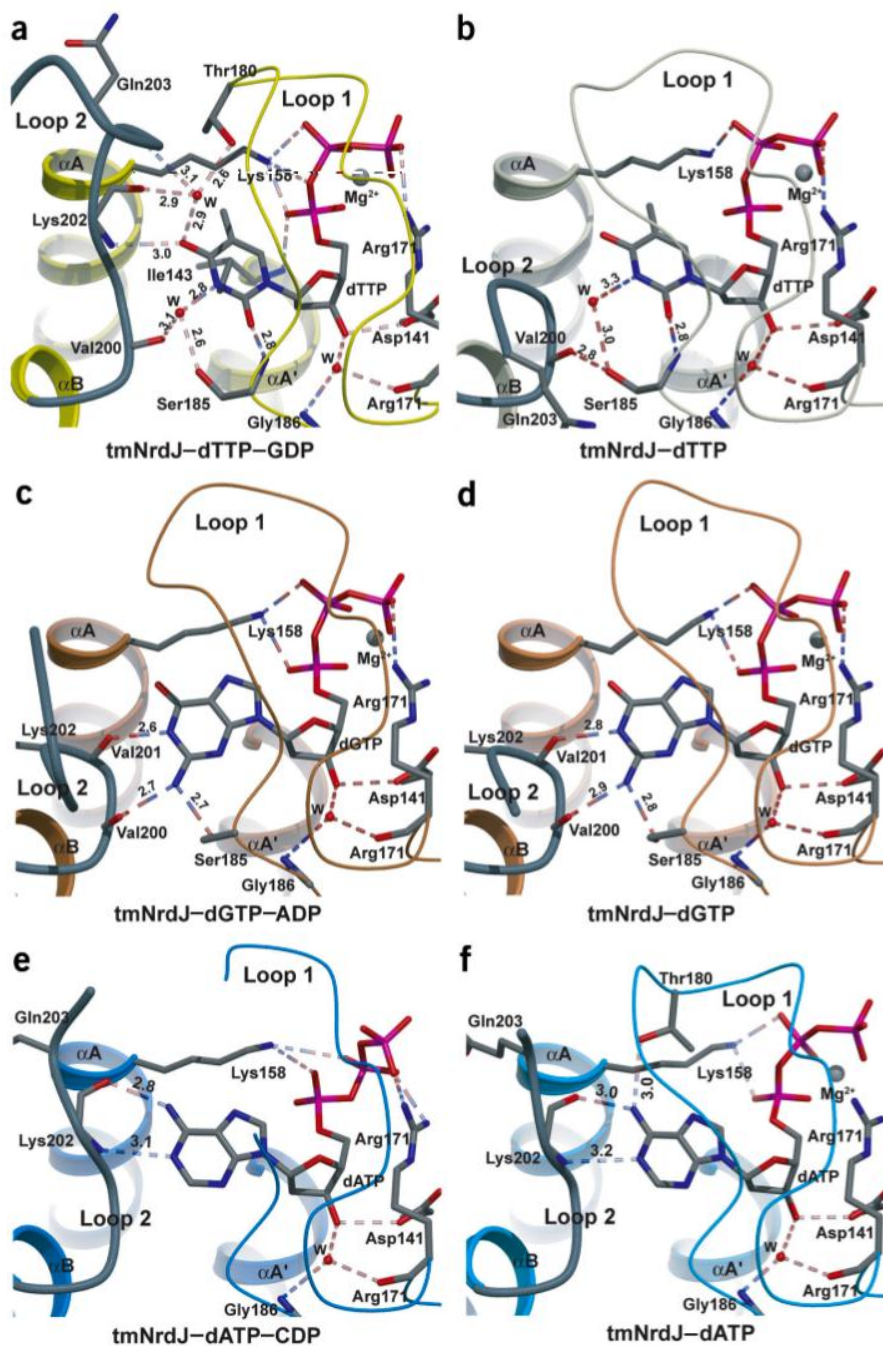
Figure 6 Specificity site structure for tmNrdJ, a class II RNR from *T. maritima*, showing the different effectors bound with and without substrate. Some constant interactions to the ribose have been omitted for clarity (Larsson *et al.* 2004).

The studied class III enzymes
The presence of a third, anaerobic class of RNRs was first found in cell extracts from *Methanobacterium thermoautrophicum* (Hogenkamp *et al.* 1987). Since then the class III RNRs have been primarily studied in *E. coli* (Fontecave *et al.* 1989)*,* BT4 (Young *et al.* 1994) and, to a somewhat lesser extent, *Lactococcus lactis* (Torrents *et al.* 2000) and *T. maritima* (Aurelius 2011). The latter is an anaerobic eubacteria, which can be found by sea vents on the ocean floor round Italy and the Azores (Huber *et al.* 1986). The first class III enzyme to be structurally solved and characterised was from BT4 (Logan *et al.* 1999). Since then only the structure of the *T. maritima* class III RNR has also been solved (Aurelius 2011). The three RNRs from *L. lactis, E. coli* and BT4 share about 50% sequence identity*,* while *T. maritima* only has about 30-25% sequence identity with these three.

Homologues to class III RNR
The class III RNR activases are part of the radical S-adenosylmethionine (SAM) superfamily and here several homologues can be found. These proteins all use SAM to generate an adenosyl radical, most commonly for cleaving C-H bonds. A common motif, CxxxCxxC, together with SAM, binds four irons to create a 4Fe-4S cluster which catalyses the reaction, present in β_2 for class III RNR. While SAM is mostly used only as a coenzyme and regenerated after each round of catalysis, class III RNRs are part of a group which actually use it as a substrate. Examples are biotin synthase and lipoyl synthase, which use SAM to insert a sulphur into a C-H bond. Class III RNR, pyruvate formate lyase 1 (PFL, along with pyruvate formate lyase 1 activating enzyme), benzylsuccinate synthase and glycerol dehydratase use SAM for the generation of a glycyl radical. Of these, PFL was the first enzyme to be discovered to use glycyl radicals (Wang and Frey 2007). PFL and its activase are the ones most commonly mentioned as homologues to class III RNR's α_2 and β_2, respectively (Torrents *et al.* 2001). The two proteins are thought to have a common origin and are both anaerobic, PFL being a part of anaerobic metabolism (Poole *et al.* 2002).

Previous and current work
This project focused on purification and structural studies of the α_2-subunit of class III RNR from *T. maritima* (tmNrdD) using X-ray crystallography, which has already been studied within the group. All purification and initial preparation of tmNrdD was done by Etienne Mulliez in Grenoble. He also produced Se-Met versions of the protein to assist in solving the phase problem. The initial crystallisation screens were done by Alexander Balhuizen and Renzo Johansson, focusing on a different crystal form to the current one. Tobias Beck set up commercial crystallisation screens and one possible hit from these was followed up by Oskar Aurelius, producing the current crystal form in citrate buffer. He also produced Se-Met crystals for multiple wavelength anomalous dispersion data collection and phase determination, solved the structure and collected some datasets of dNTP-RNR complexes, produced mostly by soaking. Renzo Johanson later collected a data set on a native crystal in 2-(N-morpholino)-ethanesulfonic acid (MES) buffer and did some surface plasmon resonance trials to see if the substrate interacts with tmNrdD.

The current project focused on determining the difference in structure of the protein tmNrdD while bound to its various effectors and substrates. To facilitate this, production and purification of tmNrdD was carried out, based on a protocol from Etienne Mulliez. The protein has a confirmed metal site, which in *E. coli* and BT4 is known to contain Zn (Logan *et al.* 2003; Luttringer *et al.* 2009). It also co-binds Mg with the effector to coordinate the phosphates and the exact position of this metal ion in tmNrdD had not been confirmed. Thus, metal characterisation of these two sites was carried out, substituting Mg for Mn. The protein was crystallised in MES buffer and crystallisation optimised in the new conditions. The produced crystals were used to do the metal characterisation and to study the difference in structure with bound effectors and substrates through X-ray crystallography.

Methods

Protein purification

pRSFduet-1 plasmids containing the tmNrdD gene were transformed into chemocompetent cells from the *E. coli* strain BL21-CodonPlus(DE3)-RIL (Agilent Technologies) using heat shock at 42°C for 30 s. The transformants were stored as glycerol stocks at -80°C, which were used for all subsequent cultures. Some of these cells were streaked on to agar plates and single colonies inoculated in 10 ml LB-medium overnight at 37°C. These 10 ml were used to inoculate 1 l of LB-medium, which was grown at 37°C to an OD_{600} of 0.5-0.7. Protein expression was induced by adding IPTG to a final concentration of 250 μM, and the culture continued for 3 hours. The cells were harvested by centrifugation at 6000 rpm for 12 min. All growth media contained $ZnCl_2$ (150 μM), chloramphenicol (100 μg/ml) and kanamycin (30 μg/ml), the agar plates contained kanamycin and chloramphenicol to the same concentrations.

The amount of time necessary to reach optimal protein expression was investigated by growing the cultures for 5 hours after the addition of IPTG. 1 ml was removed at regular time intervals and used to measure OD_{600}, as well as for running on an SDS-PAGE gel. The presence of the contaminant chloramphenicol acetyl transferase (CAT) was used as a guide to determine relative concentrations of the samples on the gel, and the amount of protein in each relevant band analysed to determine optimal growth time. The intensities of the bands on the gel were analysed with the ImageQuant TL software (GE Healthcare). To determine whether CAT, the chloramphenicol resistance protein, affected growth or expression, cultures were grown in the presence of kanamycin, but without chloramphenicol. The CAT gene is only important for the RIL plasmid and not essential to maintain the tmNrdD plasmid in the cells. The difference in growth was compared through their OD_{600} measurements and by running both types of culture on an SDS-PAGE gel.

The cell pellets were dissolved in about 25 ml lysis buffer (50 mM Tris•HCl, 250 mM KCl, 200 μM PMSF, 5 mM dithiothretiol (DTT), pH 8) and the cells were lysed using a French press. The cell debris were removed by centrifugation at 20 000g for 20 minutes and the pellet saved as P1. Since tmNrdD is thermophilic, contaminating proteins were removed by heating the supernatant to 75°C for 5 minutes. This incubation was initially performed for 20 min, but to minimise the loss of protein the time was reduced. The sample was centrifuged (10 000 g, 10 min, pellet saved as P2), before the supernatant was saturated with 60% ground ammonium sulphate, to precipitate the remaining protein, including tmNrdD. This was centrifuged at 10 000 g, 10 min, the supernatant saved as S1 and the pellet saved at -20°C as P3. To further optimise the purification protocol it was investigated weather sonication improved the protein yield compared to French press. The dissolved cell pellets were sonicated at 45 W with 20 s pulses and 60 s rest, run for 6 cycles.

The P3 pellet was dissolved in Buffer I (1M ammonium sulphate, 50 mM TRIS•HCl, 50 mM KCl, pH 8) and the concentrations of protein and nucleotide were measured with Nanodrop (Thermo Scientific) before it was centrifuged at 10 000 g for 10 min. The supernatant was run on a Butyl Sepharose high trap FF column, 5 ml (GE Healthcare), where it was eluted with a gradient from 0%-100% Buffer II (50 mM Tris•HCl, 50 mM KCl, pH 8) at a flow rate of 5 ml/min. All fractions from the appropriate protein peak were pooled. After measuring protein concentration the samples were concentrated in 15 ml concentration tubes with a molecular weight cutoff of 30 kDa (Millipore centrifugal filters). As a final step the tube was filled with elution buffer (5 mM DTT, 50 mM Tris•HCl, 50 mM KCl, pH 8), concentrated to the desired volume and the concentration measured again. An SDS-PAGE gel and mass spectrometry (MS, method described below) were run on the concentrated fractions. The protein was transferred to a different tube and stored at 4°C.

As a final step the concentrated protein was run with the elution buffer on a 23 ml Superdex 200 column (GE Healthcare), with a flow rate of about 0.5 ml/min. TmNrdD appeared as a large peak at about 10 ml after addition of the concentrate and fractions were collected and pooled in this area. The concentration of the pooled fractions was measured, they were concentrated to about 25 mg/ml and dynamic light scattering (DLS) was used to check homogeneity. The protein was aliquoted into 20-30 μl portions, flash frozen in liquid nitrogen and stored at -80°C for later use.

A detailed protocol for the purification can be found in appendix I.

NMR and Mass spectrometry

To investigate the binding of tmNrdD to its effectors and substrates it was analysed with nuclear magnetic resonance spectrometry (NMR), together with various nucleotides. Solutions of protein in D_2O were run with 100 µM effectors and 50 µM substrate in 20 mM Tris•HCl, 50 mM NaCl, 2 mM DTT, 1 mM $MgCl_2$, at 25°C and 50°C. Proton 1D NMR experiments were conducted using a Varian DirectDrive 600 MHz (14.1 T) spectrometer. Water suppression was achieved by pre-saturation and WATERGATE methods. After an initial run the sample was incubated in the NMR tube for varying concentrations of NrdD up to 12 µM, to observe a possible shift change or broadening of the peaks for the hydrogens of C1 and C5 for the pyrimidines and C8 for the purines (figure 5), which would be indicative of binding. Two different batches of protein were used. Both batches, NrdD1 and NrdD2, were sent by Etienne Mulliez, and were purified using similar protocols to the one described above, but using sonication to lyse the cells instead of French press. NrdD1 was sent in spring 2009 and NrdD2 in spring 2012 and arrived in 50 mM Tris•HCl, 50 mM KCl, 5 mM DTT, pH 8.0. NrdD2 was mostly used for the NMR experiments, and the different nucleotides run with it were dTTP, dATP, ATP, GTP, GDP or CTP in different runs and combined in presumed effector/substrate pairs. For NrdD1 it was instead dATP, ATP or dGTP.

To test protein stability the NMR samples were run on an SDS-PAGE gel, along with proteins left in the fridge for several weeks. In-gel trypsin digestion was performed on appropriate bands from these gels and analysed using matrix-assisted laser desorption ionisation – time of flight / time of flight MS (MALDI-TOF/TOF MS). For the in-gel digestion the bands were cut out of the gel and washed with 50 mM NH_4HCO_3 50% ethanol (v/v) until the gel pieces were no longer blue. They were then dehydrated in 50% ethanol (v/v) for a few minutes until white and opaque, before reduction by incubation at 37°C for 30 min in 10 mM DTT 50 mM NH_4HCO_3. After another dehydration step they were alkylated with 10 µl 0.55 mM iodoacetamide in 50 mM NH_4HCO_3 in the dark for 30 min. The gel was washed with 50 mM NH_4HCO_3 50% ethanol (v/v) for five minutes and then dehydrated again. The liquid was removed, the gel pieces dried and digested with 8 µl 12 ng/µl trypsin in 50 mM NH_4HCO_3 for five minutes. Excess liquid was removed before 8 µl 50 mM NH_4HCO_3 was added and the gel pieces incubated at 37° over night. 8 µl 0.5% TFA was used to extract the peptides by incubation for 30 min at room temperature, before spotting on a MALDI plate. A MALDI matrix solution consisting of 5mg/ml hydroxycinnamic acid in 50% acetonitrile, 0.1% (v/v) phosphoric acid (Kjellström and Jensen 2004), was added afterwards to each position. The Matrix solution contained two peptide standards [des-arg-bradykinin (m/z 904.468) and ACTH 18-39 (m/z 2465.199)] that were used for internal mass calibration in every analyte/matrix position. MALDI-MS and MS/MS analyses of the fractions were performed on a 4700 Proteomics Analyzer MALDI-TOF/TOF™ mass spectrometer (Applied Biosystems, Framingham, MA). MS/MS data acquisition was performed in a two-step process as described previously (Gonzalez *et al.* 2009). Database searching was done using Mascot (Matrix Science) with Swissprot as the database and with a peptide mass tolerance of 50 ppm and a fragment mass tolerance of 0.2 Da.

Crystallisation

TmNrdD, both from NrdD2 and newly produced batches, was crystallised with MES buffer, based on conditions that had previously been successful. Most of the initial screens were set up at the MAX-lab crystallisation facility. Table 2 shows a general overview of the different screens conducted. All screens contained, unless otherwise stated, 5 mM DTT and 100 mM MES and were crystallised using vapour diffusion with sitting drop in MRC plates (Jena Bioscience) or EasyXtal plates (Qiagen). The nucleotide screens also contained 10 mM $MgCl_2$, except in a couple of cases when Mn was used instead. Mn will bind in a similar manner to Mg, allowing metal characterisation of that site, and was used since it has an flourescence edge wavelength that is possible to measure with an X-ray light source. All proteins were centrifuged for 10 min at 10 000 g before pipetting into the screens, to remove aggregates. The later screens used protein that had been dialysed with MES buffer, pH 7.5, for 20 min with slow stirring to remove the Tris it was in.

Table 2 Overview of the different crystallisation screens performed with tmNrdD.

Screen	[Protein] (mg/ml)	Added nucleotides and other variables	pH	%PEG
Original native	20	-	4.9-5.9	9-17
Microseeding	10 and 5	-	5.7-6.1	6-12
dNTP	10 and 5	All dNTPs	5.7-5.9	8-12
dNTP+NTP	10	All dNTPs, GDP	5.7-5.9	10-12
Native, larger drops	10	100 or 80 mM MES	5.7-5.9	10-12
dNTP, heated[3]	10	All dNTPs	5.7-5.9	10-12
Heated vs unheated[3]	10	100 or 80 MES	5.5-5.9	10-13
Mg vs Mn	10	dTTP, 10 mM $MgCl_2$ or $MnSO_4$	5.7-5.9	10-12
First new batch[1]	6.87	-	4.9-5.9	10-15
Second batch[1]	20, 15 and 4.5	-	4.9-5.9	8-20
Test peak 1 and 2[1]	20	-	5.3-5.9	14-20
3rd batch conc. test[1]	20, 10 and 5	-	5.3-6.3	11-22
Higher pH[1]	20	-	5.5-6.5	10-19
Zn and MES dialysis[1]	20	1 mM Zn or dialysed with MES	5.5-6.1	10-19
Co-crystallisation[1]	20	All dNTPs	5.5-6.1	10-19
Oskar's new batch[1,2]	20	-	5.5-6.5	11-19
dNTP screen[1,2]	20	dATP, dGTP and dCTP	5.7-6.1	12-17
dNTP screen, heated[1,2,3]	20	dATP, dGTP and dCTP	5.7-6.1	12-17
Larger drops[1,2]	20	-	5.7-6.3	17-19
Pipette test[1,2]	20	By hand	5.9-6.1	14-19
High-throughput 1[1,2]	20	By hand, hanging drop	5.9	16-17
High-throughput 2[1,2]	20	By hand, hanging drop	6.1	16-17

[1] = Used protein from batches produced as part of this project.

[2] = The screens were performed with protein dialysed in MES-buffer for 20 min at room temperature.

[3] = Protein and nucleotide (where applicable) were heated to 75°C for 20 min, then cooled for 1 h in the heat block before pipetting into the screen.

Data collection

Data collection on promising crystals was carried out at the I911-3 beamline station at MAX-lab with a 225 mm marMosaic CCD detector and the ID23-1 beamline at the European Synchotron Radiation Facility (ESRF) with a PILATUS 6M-F detector. The cryoprotectant for the older batch (NrdD2) contained 20% polyethylene glycol (PEG) 400, 12% PEG3000, 100 mM MES pH 5.7 and 0.5 mM dNTPs or/and 2 mM NTPs, where applicable. For the crystals from newer batches the cryoprotectant was the same, except the pH was 5.9 and it also contained 5 mM DTT. Most of the data was collected at wavelengths around 0.9-1.0 Å. Some crystals were used for metal characterisation of Zn, Fe and Mn, and for these crystals the most appropriate wavelength for anomalous data collection of each metal was determined by fluorescence scans. For Fe a wavelength that was known to be on the high energy end of the theoretical absorption edge was used: the fluorescence scan readout was of so low intensity that an optimal wavelength could not be determined. The wavelengths used for the different metal characterisations were 1.28 Å for Zn, 1.89 Å for Mn and 1.60 Å for Fe. Appendix II gives a more comprehensive overview of the crystals tested for data collection.

*Data processing, Modelling and Refinement*

Indexing, integration of the diffraction images, scaling of the data and generation of an mtz file was done using the XDS suite of programs (Kabsch 2010). Using data from a previously solved native structure (Aurelius 2011), the Rfree flags were imported using CCP4 (Winn *et al.* 2011), after which followed rigid body refinement with refmac5 through CCP4 (Murshudov *et al.* 2011). In a couple of cases the models were not solvable by rigid body refinement, in which case molecular replacement was done with the same native structure, using the Phaser software through CCP4 (McCoy *et al.* 2007). Restrained refinement was then carried out, again with refmac5 through CCP4, before manual refinement in Coot (Emsley *et al.* 2010), with any effectors or substrates added early in the refinement process. Several cycles of alternating refmac5 restrained refinement and Coot manual

editing was then carried out to tidy up the models. The web based program MolProbity was used for further validation after each of the later refinement cycles (Chen *et al.* 2010). To evaluate effector and substrate binding the average B-factors, as well as B-factors of all bound phosphate groups, were compared.

Results

Protein purification

OD_{600} was reached after roughly 3.5 h and there was no clear difference in cell yield between the cultures grown with or without chloramphenicol (Table 5). Chloramphenicol also did not seem to affect the presence of CAT. It was only necessary to continue protein expression for 3 h after induction, since expression plateaued after this time (Figure 7, Table 3).
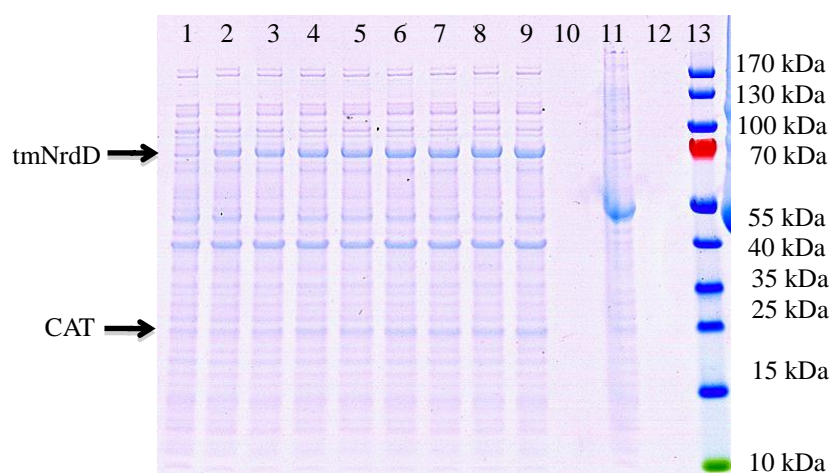
Figure 7 SDS-PAGE gel determining optimal length of time for protein expression. Lane 1-9 are samples taken at the following time points, in order: at inoculation (time 0), 1 h, 2 h, 2.5 h, 3 h, 3.5 h, 4 h, 4.5 h and 5 h. Lane 11 is not related.

Table 3 Overview of the intensity analysis of the bands in figure 7. The intensity of the bands at 70 kDa (tmNrdD) and 25 kDa (CAT) in figure 7 were plotted graphically, and the area under the peaks used for this table. To compare the relative intensities of various lanes the tmNrdD peak area was first divided by the CAT peak area. These values were then all divided by this value for time 0 to determine the relative increase in intensity.

Lane	1	2	3	4	5	6	7	8	9
NrdD (intensity peak area)	1700	2900	3100	3100	3100	3700	3700	3600	3700
CAT (intensity peak area)	3800	3000	3200	3000	3300	3300	3400	3400	3400
NrdD/CAT	0.447	0.967	0.969	1.033	0.939	1.121	1.088	1.059	1.088
(NrdD/CAT)/(t0(NrdD/CAT))	1	2.161	2.165	2.310	2.100	2.506	2.433	2.367	2.433
Time (min)	0	60	120	150	180	210	240	270	300

Sonication gave a higher P1 weight than French pressing, but not enough to be significant, since the pellet weight between different French pressed batches varied greatly. P1 was always the smallest of the pellets generated during purification, weighing about 0.5 g per litre culture. Manual inspection of SDS-PAGE gels run when the solubilised P1 pellet was incubated for 20 min, compared to 5 min, showed a clear decrease in the relative intensity of the tmNrdD band (Figure 8). The P2 pellet was always the largest and weighed about 1.7 g per litre culture. The final pellet, which was precipitated with ammonium sulphate, had an average weight of 0.8 g per litre culture (Table 5).
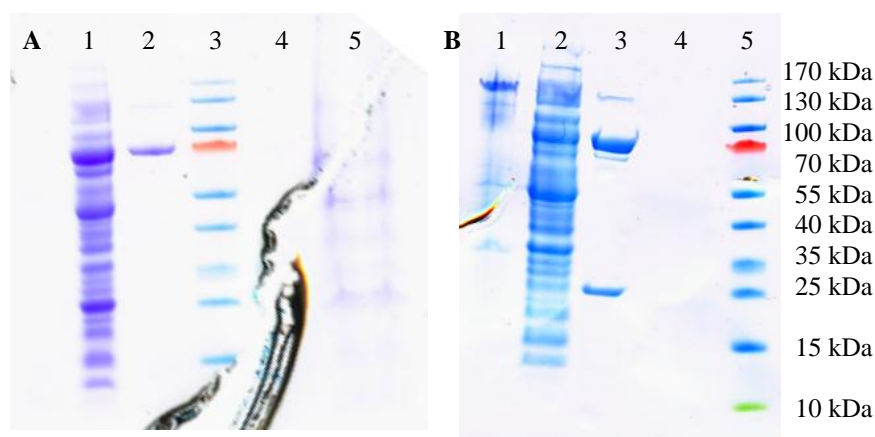
Figure 8 A comparison of the pellets from purification when the sample had been heated for 20 min (A) or 5 min (B). tmNrdD is located at 70 kDa (the red band in the ladder). Lanes in A (1-5, respectively): P2, P3, Ladder, S1, P1. Lanes in B (1-5 respectively): P1, P2, P3, S1, ladder.
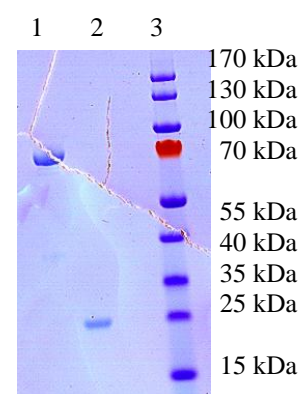
Figure 9 The final SDS-PAGE gel of the two butyl sepharose peaks. tmNrdD is 70 kDa while CAT is 25 kDa. Lanes 1 and 2 are butyl sepharose peaks 1 and 2, respectively.

Two peaks were visible at the end of the butyl Sepharose elution profiles and collected separately. The first peak contained predominantly tmNrdD, while the second one contained both tmNrdD and CAT, which was determined with MS (results not shown) and SDS-PAGE (Figure 9). It was possible to separate CAT and tmNrdD form the second peak with gel filtration, but the tmNrdD yield was so low that this peak was not used in subsequent purification batches. For the tmNrdD peak the gel filtration mainly showed one large peak at about 10 ml from injection (Figure 10). The final protein yield increased by roughly a factor of two for each batch of expression, with the final batch reaching a yield of about 4.3 mg wet weight of protein for 1 l liquid culture (Table 4).

Table 4 Overview of protein purification, showing the weights of each cell pellet and the protein yield per litre of bacterial culture. All weights are the wet weight of the pellet.

	Cells (g)	P1 (g)	P2 (g)	P3 (g)	Yield (mg)
121207-1	4.44		2.32	0.49	0.34
121207-2	4.30	0.67	3.20	1.08	
121207-3	3.48				
121207-4	2.94				
130118-1	3.61	0.53	1.98	1.02	0.80
130118-2	3.08	0.35	1.85	0.83	0.80
130118-3[1]	3.40	0.73	1.17	0.71	-
130118-4[2]	3.51	0.28	1.61	0.78	2.39
130118-5[2]	3.20	0.28	1.61	0.78	2.39
130118-6[2]	3.61	0.28	1.61	0.78	2.39
130228-1km[3]	3.04	0.69	1.47	0.71	4.33
130228-2km[3]	2.62				
130228-1ch[3]	3.18				
130228-2ch[3]	3.51				
Average	3.42	0.47	1.87	0.80	

[1]=Used to test different homogenisation methods
[2]= 3 l culture pooled into two tubes, so an average is shown here
[3]=Each of these pellets contained 4 l of bacterial culture and all of them have not been purified yet.
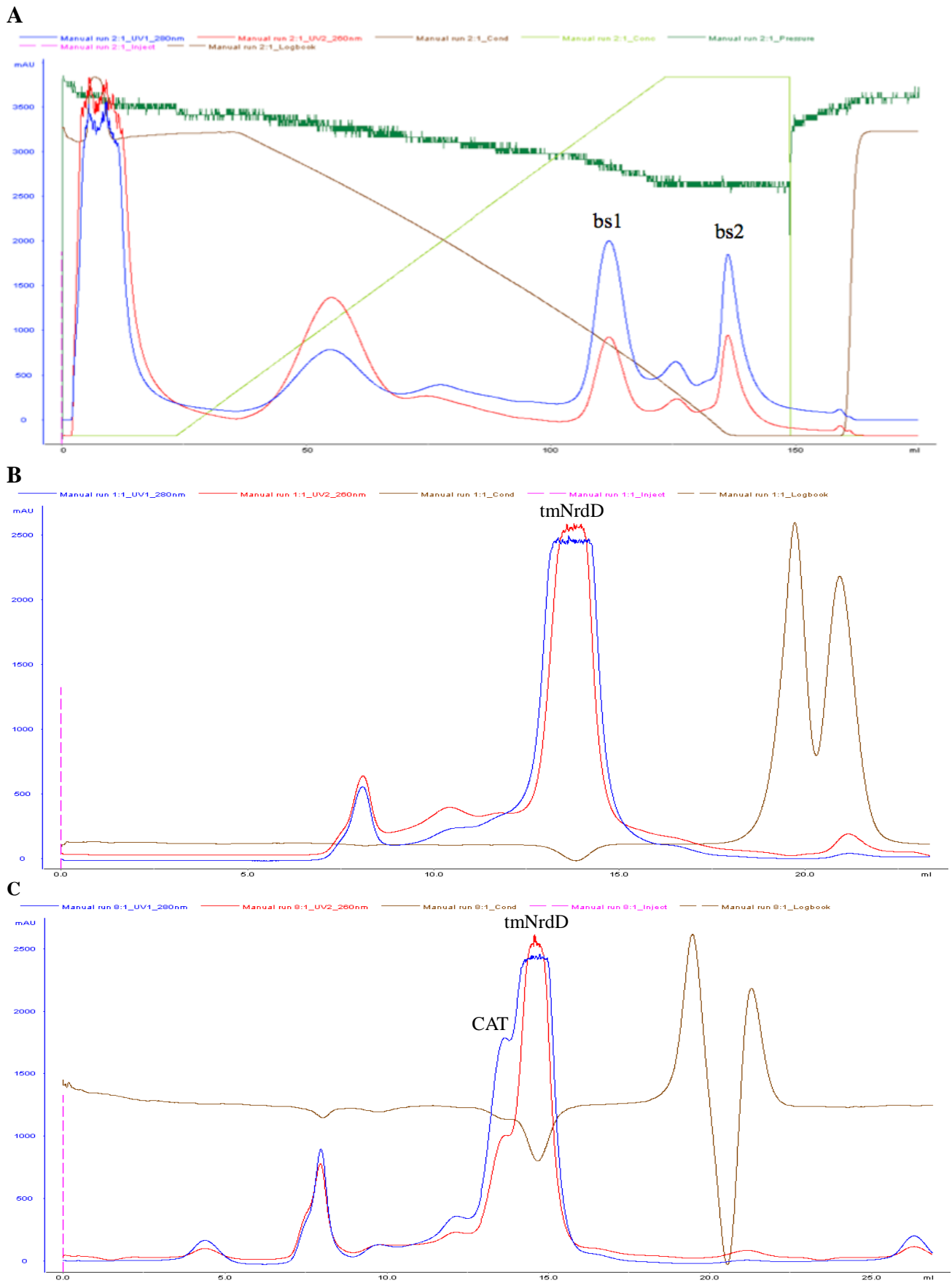
Figure 10 Chromatography spectra from the later stages of protein purification. Red: UV 260 nm, blue: UV 280 nm, brown: conductivity, light green: elution gradient, dark green: pressure. A: spectra from the butyl sepharose column. The two larger peaks after 10 ml are, first, tmNrdD (bs1) and then tmNrdD + CAT (bs2). B: gel filtration spectra of bs1, where tmNrdD elutes between 10 and 15 ml. C: gel filtration spectra of bs2, where CAT is the shoulder peak of the larger tmNrdD peak. The final two peaks in conductivity in spectra B and C are unidentified contaminants, possibly nucleotides.

NMR and Mass Spectroscopy

The NMR experiments did not show any nucleotide binding, but were instead indicative of breakdown of all nucleotides into diphosphates, except dTTP, when using the NrdD2 batch (Figure 11). This breakdown was not detectable for NrdD1, nor for nucleotides without the presence of NrdD2 protein (Figure 13). The absolute shifts were around 7.7 ppm and 1.9 ppm for the C1 and C5 pyrimidine hydrogens respectively, and around 8.1 for the C8 purine hydrogen. The only detectable protein in all samples was tmNrdD, both on the gel (Figure 12) and with MS (data not shown). The SDS-PAGE gel revealed that in the NrdD2 batch not only were the nucleotides were being broken down, but so was tmNrdD. This got worse over time for protein samples stored at 4°C (Figure 12). Because of the breakdown of both protein and nucleotides in NrdD2, and the NrdD1 batch being almost used up, the decision was made to purify more protein.
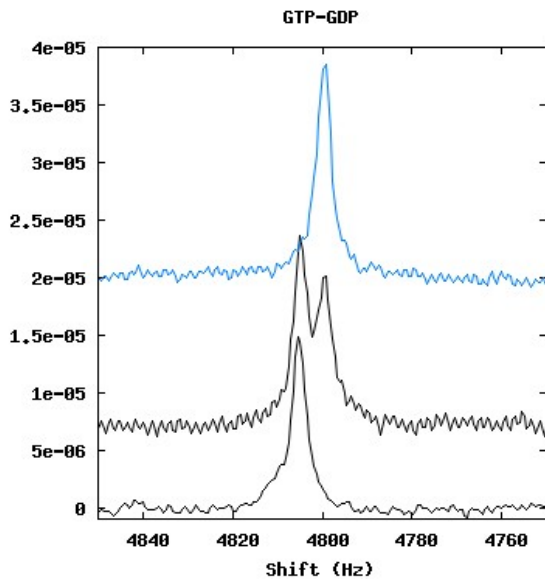
Figure 11 Three spectra from one NMR experiment: the top one contains 50 µM GTP, the middle one has added 4 µM NrdD and the final one has added 15 µM GTP. The change in shift is assumed to correspond to GDP, since this peak is present in NMR runs with GDP. The shift can not be explained by binding of GTP to tmNrdD, since 50 µM GTP should saturate 15 µM tmNrdD, thus some GTP should the left in solution.
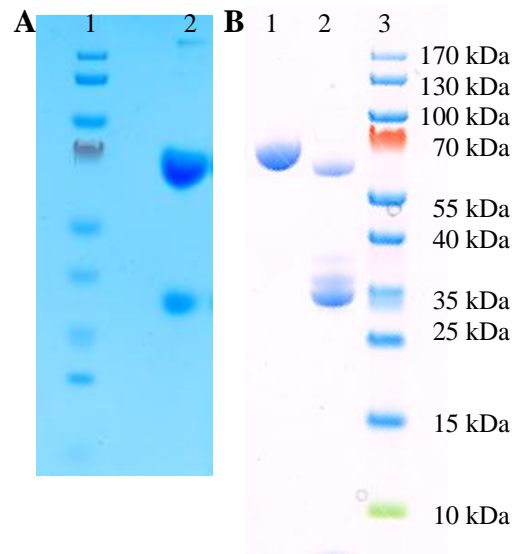
Figure 12 Two SDS-PAGE gels showing the degradation of the NrdD2 batch over time. A: gel of NrdD2 after running NMR. B: samples of NrdD1 (lane 1) and NrdD2 (lane 2) left in the fridge for several weeks.
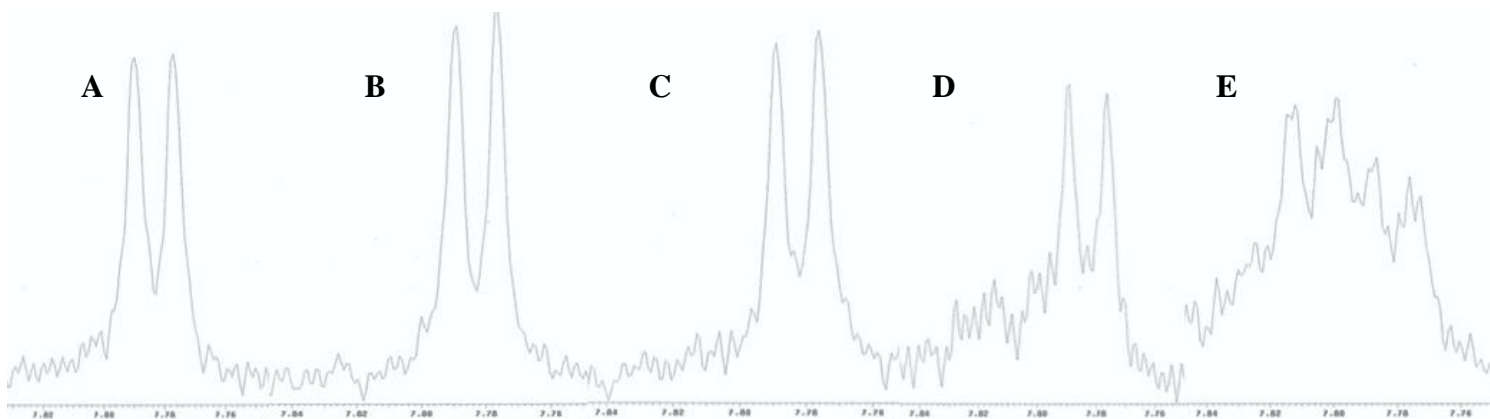

Figure 13 A: NMR spectra of 50 µM CTP. B: The same sample as in A after 2 h, showing no breakdown of nucleotide. C: B with 4 µM added tmNrdD. D: B with 8 µM added tmNrdD. E: B with 13 µM tmNrdD after 1 h. The nucleotide is stable until the addition of protein, when it slowly starts to degrade.

Crystallisation and data collection

Optimal crystallisation conditions for NrdD2 were 10-12% PEG3000, with a pH of about 5.7 for the native protein and slightly higher for co-crystallisation with nucleotides. For the newly produced batches the pH optimum was instead 5.9-6.1, with higher PEG3000 requirements of 16-17%. To obtain crystals of tmNrdD with bound nucleotides co-crystallisation was mainly used for NrdD2, while for the new batches soaking was mostly used. The crystals were visible after one day of growth and reached their maximum size after 2-3 days. Precipitation in the drops was common, with a lot of nucleation and microcrystals. When the protein was dialysed with MES buffer before crystallisation the results improved greatly, with significantly larger crystals more frequently (Figure 14). Setting up manual hanging drops, rather than use the pipetting robots at the MAX-lab crystallisation facility, also improved the size of the crystals.

The crystals using protein in Tris buffer generally diffracted to about 3 Å, while the MES-dialysed crystals diffracted to about 2 Å. No successful data sets with co-crystallised nucleotide could be obtained with the MES-dialysed protein, however these crystals diffracted well more often than those with protein in TRIS buffer. For crystals soaked with nucleotides the soaking needed to be carried out for at least 1 h for binding to occur. The data sets were generally anisotropic to varying degrees. The space group was always $P2_1$, with two different sets of unit cell parameters, seen below. For the smaller of these the b unit cell parameter was slightly more varied (Table 5). The smaller unit cell contained one dimer of tmNrdD and was known from before. The larger unit cell contained two dimers and had not been seen in previous work.
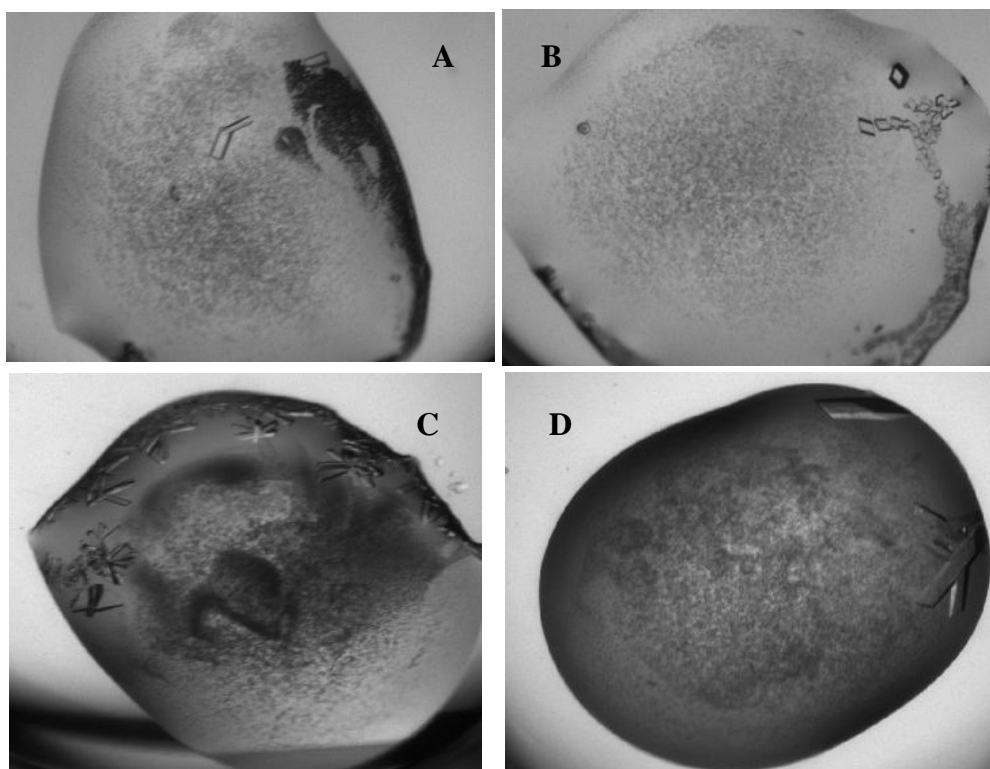

Figure 14 Representative crystallisation drops of good crystals before and after MES-dialysis. A and B are drops using protein in Tris buffer. C and D show crystals of MES-dialysed protein. The crystals in C are mostly flat triangles, with one of the sides facing upwards towards the top of the drop, making them look like rods. D is a drop set up with the robots at the MAX-lab crystallisation facility, but is also representative of good drops that were set up manually.

Table 5 Summary of the different types of crystal for which data collection was successful. No combination of effector and/or substrate is repeated in this table, for a more comprehensive list see appendix II. All NTPs are 2 mM, the dNTPs are 0.5 mM and all contain 10 mM Mg, unless otherwise stated.

Effector	Substrate	Soak time	a	B	c	α	β	γ	Resolution	Notes
-	GTP	50min	79.5	92.5	87.3	90	112.7	90	2.2	
-	ATP	2h 20min	77.6	93.2	86.5	90	112.2	90	2.6	0.2 mM ATP
dTTP	GDP	2h 55min	78.2	95	87.1	90	112.4	90	2.2	
dTTP	ATP	2h 45min	79.9	99.5	88	90	112	90	2.1	
-	ATP	2h 40min	78.1	98.1	87	90	111.9	90	1.8	
dGTP	-	1h	77.9	93	86.6	90	112.7	90	2.8	
dCTP	-	1h 5 min	77.7	92.5	86.8	90	112.4	90	2.3	
dATP	-	1h 20 min	77.7	94.7	87.1	90	111.6	90	2.5	
dATP	CTP	2h 45 min	78.1	95.5	87.3	90	111.8	90	2.5	
dTTP	-	-	77.4	87.1	86.1	90	112.1	90	3	Zn anomalous, Mn used instead of Mg
dTTP	-	-	78.1	93.7	86.8	90	112.2	90	2.8	Mn anomalous, Mn used instead of Mg
dTTP	-	-	77.6	87.2	86.4	90	112.1	90	3.4	Fe anomalous, Mn used instead of Mg
dGTP	ATP	3h 15min	92.6	94.6	137.2	90	96.7	90	1.9	
dGTP	ADP	3h 25min	93.7	95.5	138.6	90	96.9	90	1.9	
dGTP	CTP	2h 45min	92.5	94.8	137.1	90	96.7	90	1.9	
dTTP	GTP	1h 25 min	92	95.5	137.4	90	97	90	2	4 mM GTP

Effector binding

Crystal structures of all three effectors (dATP, dTTP and dGTP) bound to the effector site were obtained. The only successful effector binding achieved with NrdD1 was co-crystallised with dTTP. Soaking with dATP, dGTP and dCTP was attempted with the newer batches and successful datasets for dATP and dGTP effector binding were achieved in this way. dCTP, however, did not bind, regardless of soaking time. The phosphates of all effectors bound in the same positions, co-ordinated by the Mg ion, Lys150, Thr154, Trp203, Gln207 and Gln210. Both the pyrimidine bases interacted with similar residues, while dTTP had fewer interactions. All three were co-ordinated by His158 and Glu147 (Figure 15). Loop 2 behaved as seen in previous work with tmNrdD: there was a clear conformational change when dATP was bound, with dTTP it resembled the native structure and with dGTP it was disordered (Aurelius 2011).
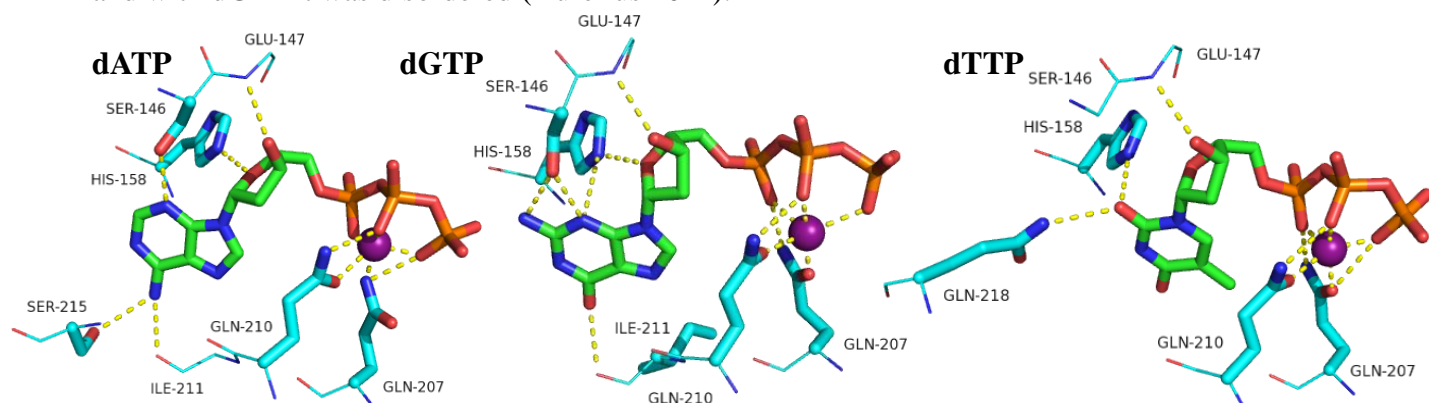
Figure 15 The tmNrdD effector site with the three different effectors bound: dATP, dGTP and dTTP. Lys150, thr154 and trp203 that co-ordinate the phosphates, as well as a few side chains, have been omitted for clarity. The protein residues are shown as cyan lines, with the side chains in bold. The purple sphere represents the Mg.
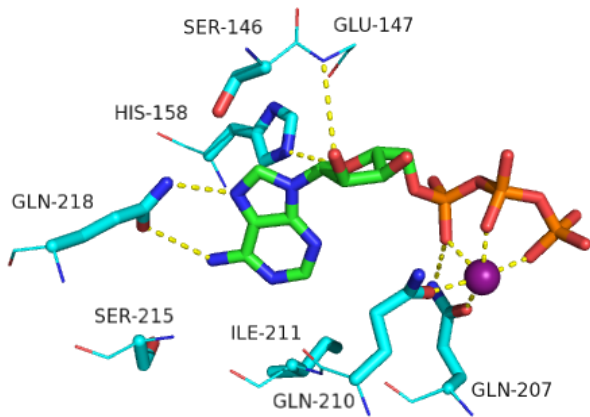
Figure 16 The tmNrdD effector site with bound ATP, shown in the same way as for the effectos in figure 15. Residues that interact with dATP have been included for comparison.
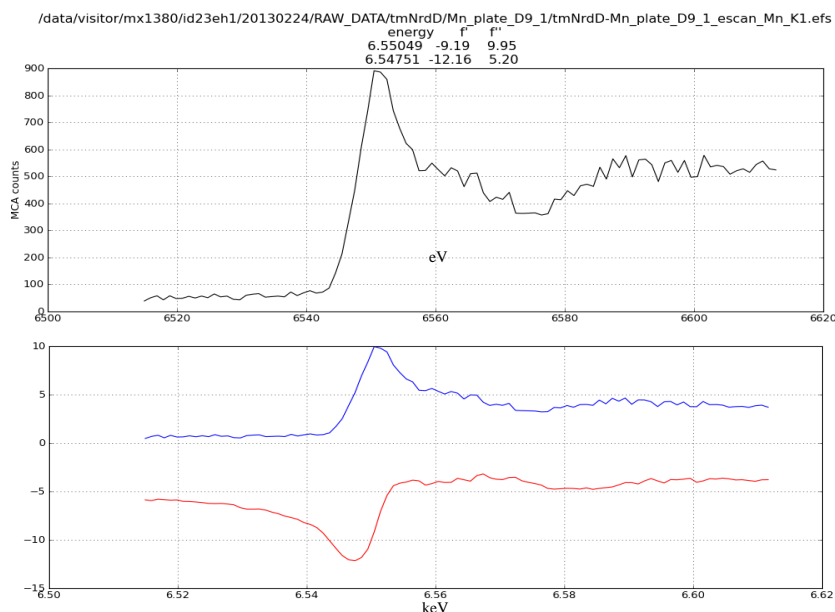
ATP also bound the effector site, when crystals were soaked with it in the absence of dNTPs. The phosphates co-ordinated in the same way as the other effectors and the ribose interacted with similar residues as for dATP (Figure 16). However, the ribose was shifted so that the 2´-OH took part in these interactions, while in dATP this hydroxyl group is missing and it is the 3´-OH that performs this role. In addition to this the ATP base was flipped compared to dATP, resulting in different and less possible interactions. ATP binding did not seem to induce the same conformational changes as dATP, but remained very similar to the native and dTTP models (Figure 17).
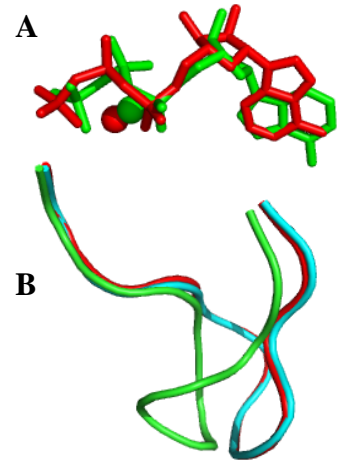

Figure 17 A: comparison of ATP (blue) and dATP (red) binding in the effector site. The ball by the phosphates represents the Mg. B: superposition of loop 2 when dATP (green), ATP (red) and dTTP (cyan) are bound.

Metal characterisation

The absorption edge scan for Zn showed a clear presence of a metal ion. In the same scan for Fe, however, it was not possible to detect any metal presence (Figure 20). An anomalous density map was generated for each dataset and in the case of Zn there was an anomalous signal at 30 σ in the expected location of the metal site at the optimal wavelength. No such peak could be seen in the anomalous density map for Fe.
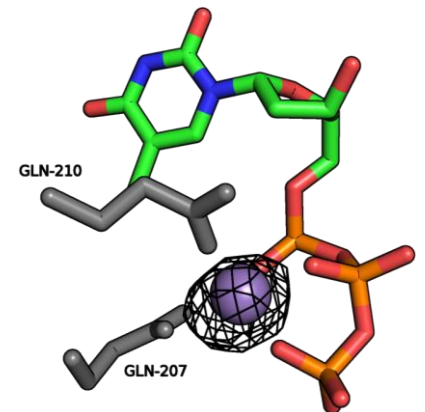
The absorption spectra for Mn gave a very clear signal for the presence of metal (Figure 19). The anomalous density map confirmed this by showing a very strong anomalous signal of 35 σ at the expected location of Mg binding, coordinating the three phosphates in the effector site. It confirms the expected Mg location in this protein (Figure 18).
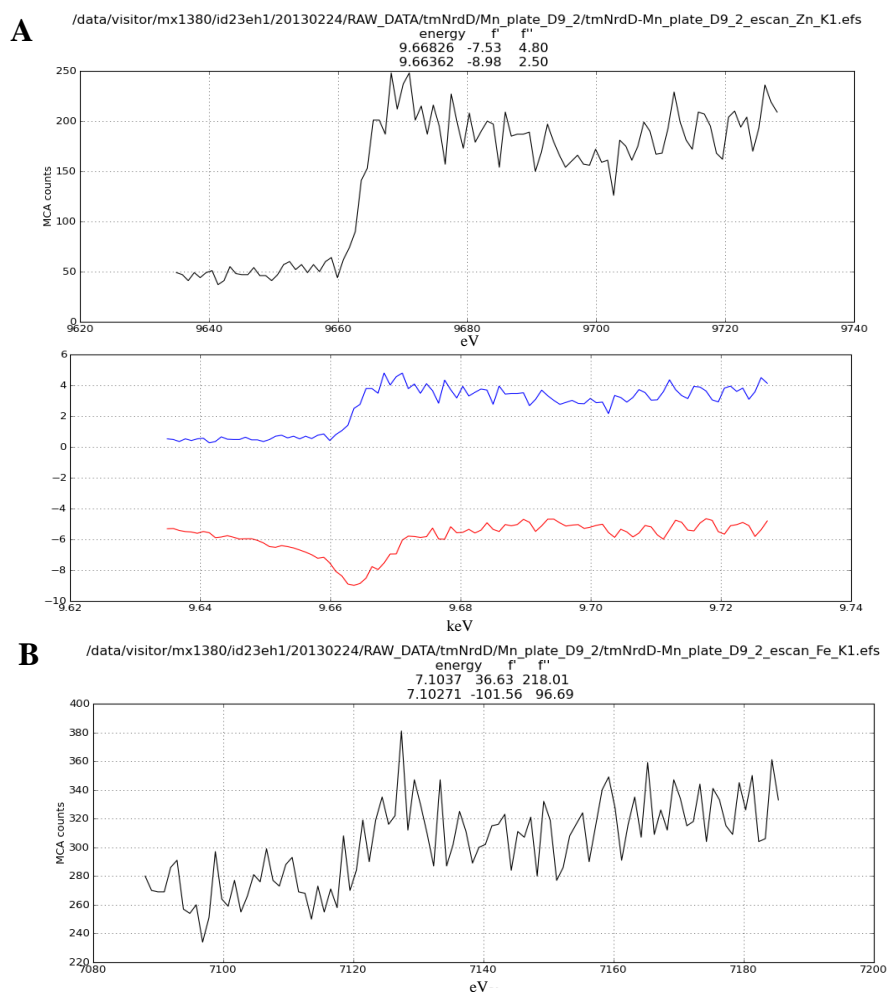

Figure 18 dTTP bound in the effector site, showing how the Mn (purple sphere) co-ordinates the dTTP phosphates with the protein residues (grey). An anomalous density map at 15 σ confirms the Mn location. (Figure from Oskar Aurelius)

Figure 19 Flourescence edge scan for Mn, at 1.89 Å. The lower spectra shows the anomalous scattering factors: f´ (red) and f´ (blue).
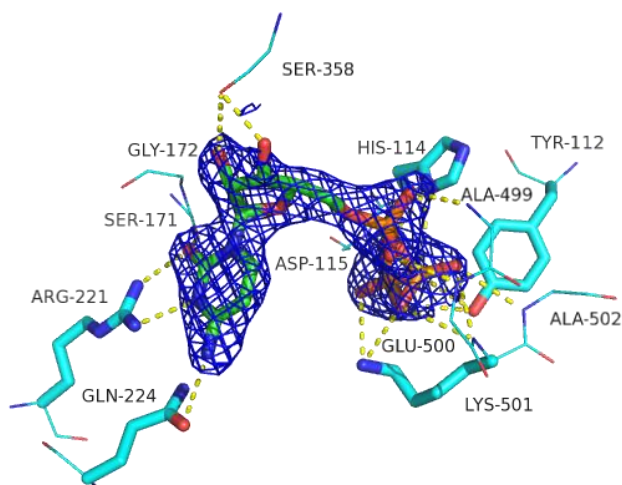
19

Figure 20 Absorption edge scans for (A) Zn, at 1.28 Å, and (B) Fe, at 1.60 Å. The lower spectra, for both A and B, show the derived anomalous scattering factors: f´ (red) and f´ (blue).

Figure 21 CTP substrate binding in the active site of tmNrdD. The protein residues are shown in cyan, with the side chains in bold. Some side chains have been omitted for clarity.

Effector-substrate complexes

Several structures of effector-substrate complexes were obtained, one of the first of these being dATP-CTP. Here the electron density clearly shows binding of both effector and substrate. dATP binds in the same way as when no substrate is present, and CTP binding is consistent with previous RNR crystal structures with bound substrate (Larsson *et al.* 2004). The CTP phosphates form many interactions with residues 499-502, as well as Tyr112, His114 and Asp115. Two residues from loop 2: Arg221 and Gln224 reach in to the active site to interact with the CTP base, while the ribose is co-ordinated by Ser358 (Figure 21).

The only other effector-substrate complexes for which clear substrate binding has been achieved are those for dTTP-GTP and dTTP-GDP. Several structures have been obtained and in all cases the effector binds as expected, but the substrate has an orientation not seen previously. In these structures the substrate phosphates are shifted one step, so that the α and β phosphates of GTP and GDP are in the β and γ positions of the CTP phosphates in the dATP-CTP structure (Figure 23). Despite this both GTP and GDP phosphates still maintain the interactions with residues 499-502, as well as Tyr112, His114 and Asp115. The ribose is flipped,

20

compared to CTP, and no longer interacts with ser358, but instead co-ordinates with Glu500, for GTP, and Tyr622, for GDP. It is instead the base, which has swung around, that interacts with Ser358 (Figure 22). This means the base is pointing in the wrong direction, as the hydroxyl groups of the ribose are now pointing away from the catalytically important residues (Figure 23). Neither GTP nor GDP interact with any of the residues from loop 2.
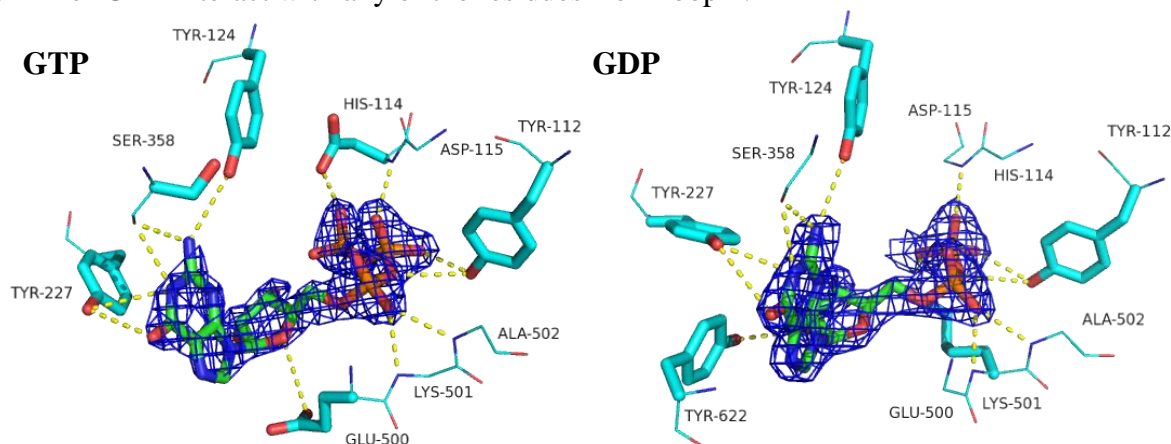

Figure 22 Substrate binding of GTP and GDP in the active site of tmNrdD. The protein residues are shown in cyan, with the side chains in bold. Some side chains have been omitted for clarity.
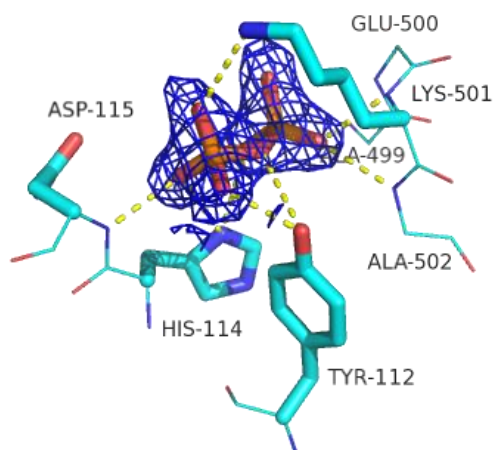
Figure 24 Representative pyrophosphate binding in the active site from the dGTP-ADP structure. The density is displayed around the pyrophosphate to 1.0 σ. The protein residues are cyan, with the side chains in bold. Some side chains have been omitted for clarity.
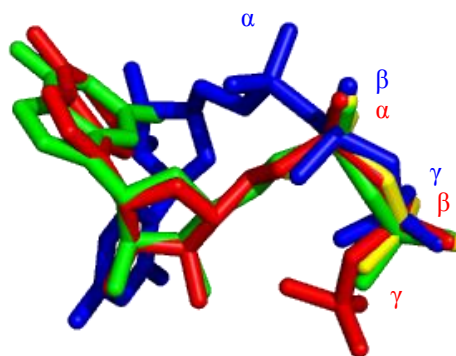
Figure 23 Superposition of the different molecules bound in the active site: CTP (blue), GTP (red), GDP (green) and pyrophosphate (yellow). The blue labels mark the different phosphates in CTP, while the red labels mark the phosphates in GTP.

The last of the expected effector-substrate combinations for which datasets were obtained were the dGTP-ATP and dGTP-ADP complexes. To investigate the possibility that, in contrast to other characterised RNRs, tmNrdD binds unexpected effector substrate combinations a dTTP-ATP dataset was also collec-ted. The effector bound in all three structures, but none of them showed clear substrate binding.

However, in all cases a couple of blobs were visible in locations in the active site that corresponded to the positions of the β and γ phosphates of CTP (Figure 23). These same blobs were also visible in the ATP structure active site. The densities were very similar in shape to pyrophosphate, which were added to the structure and this generally improved the density, with only very marginal changes in R-factor. The interactions between protein and pyrophosphate were very similar to those seen for the corresponding phosphates in the structures with bound substrate (Figure 24). For the dTTP-ATP and ATP structures a third density blob was seen in roughly the same position as the CTP α phosphate, but it was not clear enough to insert anything into it (data not shown).

The overall B-factors of the effector-substrate structures was around 30 $Å^2$, with the exception of the dTTP-GTP structure where it was 14.5 $Å^2$. For both the effector phosphates and the α and β phosphates for the substrate (equivalent to β and γ in the CTP structure) the B-factor fluctuated in the region of 20-25 $Å^2$. The exception was the dATP-CTP structure, where all bound phosphates had an average B-factor of 40 $Å^2$. The GTP γ phosphate was in a position not overlapping with any other bound phosphate and here the B-factor was around 40 $Å^2$. Appendix III contains a complete list of phosphate B-factors.

A summary of effector-substrate binding can be seen in table 6.

Table 6 Overview of the best structures available for each effector and effector-substrate combination. dNTPs bind in the effector site and NTPs in the active site, unless otherwise stated. Entries in italics represent expected effector-substrate pairs. Pyrophosphate has been shortened to pyrop.

	ATP	GTP	CTP	ADP	GDP	-
dATP	-		*Both bound*			Binds
dCTP			-			Does not bind
dGTP	*dGTP binds effector site, pyrop. in active site*	-		*dGTP bound, pyrop. in active site*		Binds
dTTP	dTTP bound, pyrop. in active site	*Both bound*			*Both bound*	Binds
-	Binds effector site, pyrop. in active site					

Discussion

Protein purification optimisation

The presence of CAT (a chloramphenicol resistance gene encoded in the RIL plasmid of this *E. coli* strain) did not seem to be reduced by the absence of chloramphenicol. However, the initial batch, grown immediately after transformation with no intervening freezing step, was grown with only kanamycin present and when this batch was purified no CAT could be detected. For all subsequent batches the cells were initially grown on agar plates containing both kanamycin and chloramphenicol, and only the growth in liquid media was without this second antibiotic. This meant that CAT would be induced during the growth on agar plates, but for one and a half days there would be no chloramphenicol present, allowing the levels to drop. It could be possible, however, that the initial induction of this protein on the agar plates was enough to keep the bacteria producing CAT throughout growth. Another possible explanation is that the cultures with and without chloramphenicol were mixed up. But this is unlikely: there were 8 litres of culture for each type, which were kept separate from the addition of the antibiotic. A cell pellet from the cultures grown with both antibiotics from the same batch as those grown with only kanamycin has yet to be purified, which could possibly give shed light on if the cultures were mixed up.

The last purified batch was homogenised by sonication, and here the yield improved by a factor of two compared to the previous batch. However, for every purified batch the yield has improved by a factor of two, due to continuous improvements to the protocol. Since only one pellet has so far been purified with sonication it is not possible to determine absolutely whether sonication significantly improves the yield. For the pellet homogenised with sonication there were two major differences compared to the previous ones: there was no presence of chloramphenicol in the liquid growth media and it was sonicated instead of run through the French press. The lack of chloramphenicol did not seem do decrease CAT levels, as discussed above, suggesting that this change in conditions did not significantly affect the cells, and as such probably would not have contributed to the increased protein yield to such an extent. Anther thing that speaks for this is that the amount of harvested cells did not increase when chloramphenicol was removed. Thus it can be assumed that sonication improves the protein yield in comparison to the French press. Further purifications should the done to verify this.

It was clear from the SDS-PAGE gels that decreasing the time of the heat precipitation step decreased the loss of tmNrdD. Since the protein originates from bacteria with an optimal growth temperature of 80°C it might be considered logical that it should not matter if the time spent at 75°C is 20 min or 5 min. However, in its native state the protein makes a complex with its activase NrdG, which might help stabilise the structure at higher temperatures. Under normal circumstances the

protein would be turned over regularly, replacing damaged ones. It is also very likely that the *T. maritima* cells have other mechanisms in place to increase protein stability. *E. coli* would lack most of these functions, which would in any case probably not be functional in the semi-purified protein mixture that tmNrdD is in during heat precipitation. All of these factors would increase heat sensitivity, meaning that even tmNrdD would be affected by the duration of this purification step.

Of the three pellets from purification the homogenisation pellet was the smallest. This is a good sign, since it means that most of the cells are broken up, and that there is little large debris. That the heat precipitation pellet is the largest is also positive: a large portion of the contaminating proteins is removed at this stage, since most proteins in *E. coli* aggregate out of solution at such high temperatures. The weight of the ammonium sulphate precipitation pellet does not correlate with the protein yield, which is not surprising as the final protein yield is only a small fraction of that pellet's total weight. This clearly demonstrates the need for further purification.

Two of the butyl Sepharose column peaks contained tmNrdD. That there are two different elution profiles is possible: if some the protein had bound nucleotide during purification its hydrophobicity might be different from the native protein, causing it to elute at a different time, happening to coincide with the elution of CAT.

The final protein yield of the last batch of purification was sufficient for crystallisation, since purification from only a few litres of culture would last for more than one month. However, it was still much lower than for Etienne, who had final protein yields of about 15 mg/l culture, using the same protocol he sent for purification of the newer batches. This was the reason the protocol was optimised, to determine what possible differences could have been introduced that caused the yield to be so much lower. The only major change done to the protocol when it was sent was the use of French press instead of sonication. Changing this back seemed to improve the yield, as discussed above, but the yield, even in the final batch was still about 3-4 times lower than expected. The increase in protein yield observed for the newer batches could have several causes, one of the major ones being more practiced handling and building up a routine for working with the protocol. For that reason it is possible that new batches produced with have further increased yield, however it is unlikely that this reason alone would bridge such a large difference. Other differences could be variations in the chemicals used for media and buffers, or different efficiencies of the columns, but again this probably does not account for this large gap. Further investigation and purification of new batches are necessary to solve this problem.

NMR and MS

The breakdown of both nucleotides and tmNrdD in NrdD2 was batch-dependent and not seen in NrdD1, strongly indicating the presence of one or more contaminating proteins. The concentration of these contaminants would be very low, since they could not be detected with either SDS-PAGE or MS. Another thing that speaks for this is the very slow rate at which breakdown occurred: for the nucleotides it could only be observed over several hours and even after weeks in the fridge there was still full length tmNrdD visible on the SDS-PAGE gel. That dTTP was not degraded could be used as a clue as to what protein or proteins was causing the degradation. However, since this degradation was batch dependent the decision was taken to produce our own protein and stop using NrdD2, rather than investigate the cause further.

Crystallisation and data collection

There was a small increase in optimal crystallisation pH for the new batches, but dialysis with MES producing so much better overall results, both in crystal appearance and data collection, might actually be a step in the other direction. Before dialysis the protein was kept in Tris buffer, pH 8, and after dialysis in MES pH 7.5. The main difference here is that Tris buffers at a higher pH than MES (Dawson *et al.* 1986). The drops were 50% reservoir with 100 mM MES and 50% protein solution with 50 mM Tris or MES, meaning this would have had some effect on the final pH of the drops. The change would not have been drastic, but the final pH for MES dialysed protein would probably have been significantly lower, resulting in an overall drop in optimal crystallisation pH.

Soaking also improved the results of data collection for the new batches, compared to co-crystallisation. If this was consistent is hard to say, however, since only one co-crystallisation screen was performed. Also, most of the crystals later screened at MAX-lab for soaking experiments were dialysed with MES, while the co-crystallisation screen was not.

Effector sites

dCTP binding has been seen in the effector site of the BT4 class III RNR, but could not be observed in this case (Larsson *et al.* 2001). It is, however, possible that lack of binding was due to suboptimal conditions or too short soaking time. That soaking generally needs a long time for binding to occur can be explained, both by the crystal packing and by being in a solution so far from *in vivo* conditions. The crystal is not a rigid structure, but its highly ordered nature can make it harder for the nucleotides to reach the effector or substrate sites. Though the solution the protein is in is favourable for maintaining the integrity of the crystal, in the case of the cryo solution where the nucleotides are introduced these conditions might not be favourable for nucleotide binding, which would also increase the time required.

The dATP, dGTP and dTTP interactions with the residues in the effector site is similar to what has been seen in previous work on tmNrdD (Aurelius 2011). There are, however, several differences to the effector binding in BT4 (Larsson *et al.* 2001). Firstly, although the phosphates are co-ordinated by a Mg ion in both cases, this metal in turn interacts differently with the effector site. In tmNrdD this role is performed by two glutamines that also interact directly with the phosphates. In the BT4 class III RNR it is instead a water molecule that co-ordinates the Mg, bridging a larger distance to Gln100 and Lys103 (BT4 class III numbering). This increase in distance allows Gln100 to co-ordinate the hydroxyl group on the deoxyribose as well, an interaction not seen in any effector complexes for tmNrdD. Here the hydroxyl group instead points in the other direction, though it still interacts with the backbone N of a glutamine (Gln147, tmNrdD numbering), the same type of inter-action as seen in BT4. The hydrogen bonding of the base is different both with regards to the residues involved and which atoms on the base that participate. In BT4 there are two key residues in effector base binding, while in tmNrdD there are two for dTTP and 4-5 for the purimidines. Though one of the key residues for loop 2 conformational changes is indeed a glutamine in tmNrdD this only interacts with dTTP, while in BT4 Gln114 interacts with everything except dGTP. The way loop 2 changes in response to various effectors is also different. Interestingly, the effectors that cause the largest and smallest conformational changes in tmNrdD (dATP and dTTP respectively) are the opposite in BT4.

That ATP binds the effector site has previously been seen in BT4 (Larsson *et al.* 2001). It is possible that also GTP may be able to bind the effector site, but no such structures have yet been solved, so this remains to be investigated. There is less interaction between ATP and the effector site, compared to dATP, indicating a weaker binding and probably lower affinity. Because ATP does not seem to change the conformation of loop 2 it is unlikely that it contributes to CTP binding in the active site, but could possibly induce binding of another substrate. One candidate would be GTP, which already binds when the loop 2 conformation is similar to the native structure. In other RNRs ATP modulates overall activity, where it binds to the ATP cone (Nordlund and Reichard 2006), but in class III RNRs that lack the activity cone there is no overall activity regulation, meaning that could not be a possible function for this binding. Overall, further studies are needed to confirm the role of ATP in the effector site, as many questions are still unanswered.

Metal characterisation

The absorption edge spectra and anomalous signal for Zn indicated that this was the metal present in the metal site on the protein. However, Fe can also give an anomalous signal at the wavelengths for Zn, meaning this result alone would not be conclusive. But since there was no presence detected on the absorption edge scan or the anomalous density map for Fe, this is a strong indicator that the metal site actually contains Zn.

The Mn co-ordinates the three phosphates of the effector with Glu207 and Glu210 in the effector site, as expected. Interestingly, though modelling in tmNrdD has always been done with Mg at this location, for the class III RNR from BT4 the Mg and β-phosphate are in swapped positions (Larsson *et al.* 2001).

Effector-substrate complexes

The dATP-CTP complex shows clear binding of both substrate and effector in accordance with previous literature (Larsson *et al.* 2004) and is the first time that such a complex has been achieved for a class III RNR. For the dTTP-GTP or dTTP-GDP structures, however, the phosphates have shifted one step, the ribose has flipped and the base is pointing in the wrong direction. For all other attempted effector-substrate complexes the only thing visible in the active site is a pyrophosphate. The one thing that was consistent throughout all substrate binding was the position of two of the phosphates (β and γ for CTP), which all interacted with the same protein residues. This site is highly electrophilic, causing binding to negatively charged molecules, such as phosphates. This has been seen in previous work where the buffer has bound to this highly electrophilic site, blocking any attempts at substrate binding (Larsson *et al.* 2001; Aurelius 2011).

GTP and GDP are in conformations not seen before, indicating two options: the protein is not active, or it is an artefact. This conformation could indicate and inactive protein because the ribose is pointing away from the catalytically important residues responsible for reduction of the 3´OH. Since there are several datasets with solved structures for this effector substrate combination where the substrate is oriented in the same way (see appendix II), the likelihood that it is an artefact decreases. If the protein would be active with the substrate in this conformation the reaction pathway would have to take a different route than what is commonly known. One such option would be that Tyr227 would swing round to the ribose, acting as a shuttle. However, this is less likely since Tyr227 is thought to contribute to maintaining the active site structure. As such, it is inconclusive what significance this new conformation of GTP and GDP has.

The only other visible active site binding was of pyrophosphates in the various structures with ATP or ADP. One possibility was that these densities belonged to phosphates from broken down nucleotides. There did not seem to be such an issue with other substrates binding, such as GTP or CTP, although it could be that only ATP and ADP was being affected. The likelihood of that decreases, however, since the ATP effector site density was very clear, yet there were still only a few blobs in the active site. It is possible that the strongly hydrophilic active site region is more sensitive to degradation products and more prone to bind them than the effector site, but overall these things speak for nucleotide degradation being possible, but unlikely.

Another possible explanation for the pyrophosphates in the active site was that the protein, being crystallised in its inactive state, would have relatively poor substrate binding. If this was the case, the substrate might still bind, but very transiently. The hydrophilic site around the phosphates would bind much more tightly than the rest of the molecule, something that is supported by all substrate structures having phosphates in those two positions, regardless of what positions the other parts of the molecules were in. This would mean that the phosphates would bind tightly enough to be visible in the electron density, but the rest of the molecule might be so mobile that it would not. It is also possible that ATP and ADP are more mobile in the active site than CTP or GTP/GDP, which would explain why it is not possible to see their densities. CTP co-ordinates with residues from loop 2, meaning it probably has more stable binding than the large purines, which do not. Similarly, the GTP and GDP conformations seen might be most stably bound in the current state of the protein, but they might adopt different ones in the active protein. Something that speaks for this is that several interactions are visible between the base and protein in the current structures, which in both cases involve the hydroxyl group missing in ATP. As such, it is possible that ATP also binds in that conformation, but since there is less to interact with it is less stable. Consequently their conformation does not have to indicate an inactive protein and there might be more substrate binding than is currently possible to visualise.

There was little variance B-factors of phosphates bound in the same position in the effector and active sites. The exception was the dATP-CTP structure, where the phosphate values were about twice as high. The dataset for this structure was to a lower resolution and of worse quality than most of the others, which explains the higher values. The overall B-factors can not be compared, since some of them are residual B-factors, while others are total B-factors and it is not possible to relate them to each other. When determining the optimal positions of the phosphates, it might have been possible to use the comparison of B-factors of phosphates in the different positions to support the hypothesis discussed above about a mobile substrate. In that case the two phosphate sites where it is possible to see a pyrophosphate should have considerably lower B-factors than the third phosphate position. However, this is only the case in the dTTP-GTP structure, while in the dATP-CTP structure it is not. Because only two structures with three phosphates in the active site are available it is not possible to determine if the differences in B-factors between phosphates support the discussion regarding a mobile substrate. To determine this more structures with bound triphosphates in the active site would be necessary.

Conclusions

Protein production of tmNrdD was generally successful with a final protein yield of about 4 mg/l culture. However, the protocol still needs optimisation and further work to increase the protein yield up to the levels achieved by Etienne. Crystallisation in MES buffer, and especially dialysis of the protein solution to remove Tris buffer, greatly improved crystallisation and data collection results. The metal site in tmNrdD has been confirmed to contain Zn, and the location of the Mg coordinating the phosphates in the effector site has also been determined.

Structures for all known specificity effectors have been obtained, as well as three substrate effector complexes. Though the dATP-CTP complex binds according to previously known data, the dTTP-GTP complex does not, with the substrate in a previously undetected conformation. In attempted structures with dGTP-ATP it was only possible to detect pyrophosphates in the active site. It is possible this is due to a larger flexibility in the substrate while the protein is inactive, than previously thought. Further work needs to be done to determine protein activity, and to attempt to obtain effectors substrate complexes with ATP, to shed light on these interesting developments.

References

Adam, R. D. (2001). "Biology of Giardia lamblia." *Clinical Microbiology Reviews* **14**(3): 447-475.

Andersson, J., S. Bodevin, M. Westman, M. Sahlin and B. M. Sjoberg (2001). "Two active site asparagines are essential for the reaction mechanism of the class III anaerobic ribonucleotide reductase from bacteriophage T4." *Journal of Biological Chemistry* **276**(44): 40457-40463.

Andersson, J., M. Westman, M. Sahlin and B. M. Sjoberg (2000). "Cysteines involved in radical generation and catalysis of class III anaerobic ribonucleotide reductase - A protein engineering study of bacteriophage T4 NrdD." *Journal of Biological Chemistry* **275**(26): 19449-19455.

Aurelius, O. (2011). *Further structural insight into anaerobic ribonucleotide reductases*. MSc Protein Science Master's thesis, Lund Univeristy.

Chabes, A., B. Georgieva, V. Domkin, X. L. Zhao, R. Rothstein and L. Thelander (2003). "Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase." *Cell* **112**(3): 391-401.

Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallographica Section D-Biological Crystallography* **66**: 12-21.

Dawson, R. M. C., D. C. Elliot, W. H. Elliot and K. M. Jones (1986). Data for Biochemical Research, Oxford Science Publ. .

Eklund, H. and M. Fontecave (1999). "Glycyl radical enzymes: a conservative structural basis for radicals." *Structure with Folding & Design* **7**(11): R257-R262.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." *Acta Crystallographica Section D-Biological Crystallography* **66**: 486-501.

Fontecave, M., R. Eliasson and P. Reichard (1989). "Oxygen-Sensitive Ribonucleoside Triphosphate Reductase Is Present in Anaerobic Escherichia-Coli." *Proceedings of the National Academy of Sciences of the United States of America* **86**(7): 2147-2151.

Gonzalez, H., J. Ottervald, K. C. Nilsson, N. Sjogren, T. Miliotis, H. Von Bahr, M. Khademi, B. Eriksson, S. Kjellstrom, A. Vegvari, R. Harris, G. Marko-Varga, K. Borg, J. Nilsson, T. Laurell, T. Olsson and B. Franzen (2009). "Identification of novel candidate protein biomarkers for the post-polio syndrome - implications for diagnosis, neurodegeneration and neuroinflammation." *Journal of proteomics* **71**(6): 670-681.

Hogbom, M. (2011). "Metal use in ribonucleotide reductase R2, di-iron, di-manganese and heterodinuclear-an intricate bioinorganic workaround to use different metals for the same reaction." *Metallomics* **3**(2): 110-120.

Hogenkamp, H. P. C., H. Follmann and R. K. Thauer (1987). "Ribonucleotide Reductase in Cell-Extracts of Methanobacterium-Thermoautotrophicum." *Febs Letters* **219**(1): 197-201.

Huber, R., T. A. Langworthy, H. Konig, M. Thomm, C. R. Woese, U. B. Sleytr and K. O. Stetter (1986). "Thermotoga-Maritima Sp-Nov Represents a New Genus of Unique Extremely Thermophilic Eubacteria Growing up to 90-Degrees-C." *Archives of Microbiology* **144**(4): 324-333.

Johansson, R., E. Torrents, D. Lundin, J. Sprenger, M. Sahlin, B. M. Sjoberg and D. T. Logan (2010). "High-resolution crystal structures of the flavoprotein NrdI in oxidized and reduced states - an unusual flavodoxin." *Febs Journal* **277**(20): 4265-4277.

Jordan, A., F. Aslund, E. Pontis, P. Reichard and A. Holmgren (1997). "Characterization of Escherichia coli NrdH - A glutaredoxin-like protein with a thioredoxin-like activity profile." *Journal of Biological Chemistry* **272**(29): 18044-18050.

Jordan, A. and P. Reichard (1998). "Ribonucleotide reductases." *Annual Review of Biochemistry* **67**: 71-98.

Kabsch, W. (2010). "XDS." *Acta Crystallographica Section D-Biological Crystallography* **66**: 125-132.

King, D. S. and P. Reichard (1995). "Mass-spectrometric determination of the radical scission site in

the anaerobic ribonucleotide reductase of *Escherichia-coli*." *Biochemical and Biophysical Research Communications* **206**(2): 731-735.

Kjellström, S. and O. N. Jensen (2004). "Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins." *Analytical Chemistry* **76**(17): 5109-5117.

Kolberg, M., K. R. Strand, P. Graff and K. K. Andersson (2004). "Structure, function, and mechanism of ribonucleotide reductases." *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1699**(1-2): 1-34.

Larsson, K. M., J. Andersson, B. M. Sjoberg, P. Nordlund and D. T. Logan (2001). "Structural basis for allosteric substrate specificity regulation in anaerobic ribonucleotide reductases." *Structure* **9**(8): 739-750.

Larsson, K. M., A. Jordan, R. Eliasson, P. Reichard, D. T. Logan and P. Nordlund (2004). "Structural mechanism of allosteric substrate specificity regulation in a ribonucleotide reductase." *Nature Structural & Molecular Biology* **11**(11): 1142-1149.

Logan, D. T. (2011). "Closing the circle on ribonucleotide reductases." *Nature Structural & Molecular Biology* **18**(3): 251-253.

Logan, D. T., J. Andersson, B. M. Sjoberg and P. Nordlund (1999). "A glycyl radical site in the crystal structure of a class III ribonucleotide reductase." *Science* **283**(5407): 1499-1504.

Logan, D. T., E. Mulliez, K. M. Larsson, S. Bodevin, M. Atta, P. E. Garnaud, B. M. Sjoberg and M. Fontecave (2003). "A metal-binding site in the catalytic subunit of anaerobic ribonucleotide reductase." *Proceedings of the National Academy of Sciences of the United States of America* **100**(7): 3826-3831.

Luttringer, F., E. Mulliez, B. Dublet, D. Lemaire and M. Fontecave (2009). "The Zn center of the anaerobic ribonucleotide reductase from E. coli." *Journal of Biological Inorganic Chemistry* **14**(6): 923-933.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." *Journal of Applied Crystallography* **40**(4): 658-674.

Mulliez, E., S. Ollagnier, M. Fontecave, R. Eliasson and P. Reichard (1995). "Formate Is the Hydrogen Donor for the Anaerobic Ribonucleotide Reductase from Escherichia-Coli." *Proceedings of the National Academy of Sciences of the United States of America* **92**(19): 8759-8762.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." *Acta Crystallographica Section D-Biological Crystallography* **67**: 355-367.

Nolan, J. (2002-03-13).   Retrieved 2013-01-23, from http://www.tulane.edu/~biochem/nolan/lectures/rna/frames/nucs.htm.

Nordlund, N. and P. Reichard (2006). "Ribonucleotide reductases." *Annual Review of Biochemistry* **75**: 681-706.

Nordlund, P. and H. Eklund (1993). "Structure and Function of the Escherichia-Coli Ribonucleotide Reductase Protein R2." *Journal of Molecular Biology* **232**(1): 123-164.

Poole, A. M., D. T. Logan and B. M. Sjoberg (2002). "The evolution of the ribonucleotide reductases: Much ado about oxygen." *Journal of Molecular Evolution* **55**(2): 180-196.

Reece, S. Y., J. M. Hodgkiss, J. Stubbe and D. G. Nocera (2006). "Proton-coupled electron transfer: the mechanistic underpinning for radical transport and catalysis in biology." *Philosophical Transactions of the Royal Society B-Biological Sciences* **361**(1472): 1351-1364.

Reichard, P., L. E. Bertani and A. Häggmark (1961). "Synthesis of Pyrimidine Deoxyribonucleoside Diphosphates with Enzymes from Escherichia coli." *Journal of Biological Chemistry* **236**(10): PC67-PC68.

Robertson, M. P. and G. F. Joyce (2012). "The Origins of the RNA World." *Cold Spring Harbor Perspectives in Biology* **4**(5).

Rofougaran, R., M. Crona, M. Vodnala, B. M. Sjoberg and A. Hofer (2008). "Oligomerization Status Directs Overall Activity Regulation of the Escherichia coli Class Ia Ribonucleotide Reductase." *Journal of Biological Chemistry* **283**(51): 35310-35318.

Rofougaran, R., M. Vodnala and A. Hofer (2006). "Enzymatically active mammalian ribonucleotide reductase exists primarily as an alpha(6)beta(2) octamer." *Journal of Biological Chemistry* **281**(38): 27705-27711.

Sintchak, M. D., G. Arjara, B. A. Kellogg, J. Stubbe and C. L. Drennan (2002). "The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer." *Nature Structural Biology* **9**(4): 293-300.

Torrents, E., P. Aloy, I. Gibert and F. Rodriguez-Trelles (2002). "Ribonucleotide reductases: Divergent evolution of an ancient enzyme." *Journal of Molecular Evolution* **55**(2): 138-152.

Torrents, E., G. Buist, A. Liu, R. Eliasson, J. Kok, I. Gibert, A. Graslund and P. Reichard (2000). "The anaerobic (Class III) ribonucleotide reductase from Lactococcus lactis - Catalytic properties and allosteric regulation of the pure enzyme system." *Journal of Biological Chemistry* **275**(4): 2463-2471.

Torrents, E., R. Eliasson, H. Wolpher, A. Graslund and P. Reichard (2001). "The anaerobic ribonucleotide reductase from Lactococcus lactis - Interactions between the two proteins NrdD and NrdG." *Journal of Biological Chemistry* **276**(36): 33488-33494.

Uhlin, U. and H. Eklund (1994). "Structure of Ribonucleotide Reductase Protein R1." *Nature* **370**(6490): 533-539.

Wang, S. C. and P. A. Frey (2007). "S-adenosylmethionine as an oxidant: the radical SAM superfamily." *Trends in Biochemical Sciences* **32**(3): 101-110.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." *Acta Crystallographica Section D-Biological Crystallography* **67**: 235-242.

Young, P., M. Ohman, M. Q. Xu, D. A. Shub and B. M. Sjoberg (1994). "Intron-Containing T4-Bacteriophage Gene Suny Encodes an Anaerobic Ribonucleotide Reductase." *Journal of Biological Chemistry* **269**(32): 20229-20232.

Appendix I: Purification protocol for production of tmNrdD (adapted from Etienne Mulliez)

Production using 1 litre cultures.

Materials
- ⚔ Sterile pipette tips and Eppendorf tubes
- ⚔ Clones:
    - ○ pRSFduet-1 with NrdD in BL21-(DE)-RIL in glycerol stock, kept at -80ºC
- ⚔ 0.5 l LB-agar:
    - ○ 5 g Tryptone
    - ○ 2.5 g yeast extract
    - ○ 5 g NaCl
    - ○ 7.5 g agar
    - ○ Before pouring into Petri dishes:
        - ▪ 500 μl 30 mg/ml kanamycin
        - ▪ 500 μl 100mg/ml chloramphenicol
- ⚔ (1 l LB-media)*x:
    - ○ 10 g NaCl
    - ○ 5 g yeast extract
    - ○ 10 g Tryptone
- ⚔ Stocks:
    - ○ 10 ml kanamycin 30 mg/ml (400 mg of 0.75 mg/mg in 10 ml water)
    - ○ 10 ml chloramphenicol 100 mg/ml (1 g in 10 ml water)
    - ○ 10 ml 250 mM IPTG
    - ○ 10 ml 150 mM Zn^{2+}
    - ○ 10 ml DTT 1M, frozen as 1 ml aliquots, made previously
    - ○ 100 ml 250 mM Tris, 1.25 M KCl pH8
        - ▪ 3.028 g Tris
        - ▪ 9.32 g KCl
        - ▪ Fill up to 80 ml ddH_2O
        - ▪ Adjust pH
        - ▪ Top up to 100 ml
    - ○ 300 ml 500 mM Tris 500 mM KCl pH8
        - ▪ 18.16 g TRIS
        - ▪ 11.184 g KCl
        - ▪ Fill up to 280 ml ddH_2O
        - ▪ Adjust pH
        - ▪ Top up to 300 ml
- ⚔ 100 ml lysis buffer:
    - ○ 50 mM Tris, 250 mM KCl (20 ml 250 mM TRIS, 1.25 M KCl pH8)
    - ○ 200 μM PMSF (3.48 mg)
    - ○ 5 mM DTT (0.5 ml 1M)
    - ○ Up to 100 ml with ddH_2O
- ⚔ 500 ml Buffer 1:
    - ○ 1 M Ammonium sulphate (63.41 g)
    - ○ 50 mM Tris, 50 mM KCl pH 8 (50 ml 500 mM pH 8)
    - ○ Fill up to 500 ml with ddH_2O
    - ○ Filter and degas the finished buffer
- ⚔ 300 ml Buffer 2:
    - ○ 50 mM TRIS, 50 mM KCl pH8 (30 ml 500 mM pH8)
    - ○ Fill up to 300 ml with ddH_2O
    - ○ Filter and degas the finished buffer

⚲ 400 ml Buffer 3:
  ◦ 50 mM Tris, 50 mM KCl pH8 (40 ml from stock)
  ◦ 5 mM DTT (2 ml from stock)
  ◦ Fill up to 400 ml with ddH$_2$O.
  ◦ Filter and degas the finished buffer.

Unless otherwise stated keep everything on ice AT ALL TIMES.

Day 1
Scrape some cells off from the glycerol stock and streak on an agar plate, containing kanamycin (30 μg/ml) and chloramphenicol (100 μg/ml).
Incubate at 37ºC over night.

Day 2
Clean-streak single colonies on to kanamycin + chloramphenicol agar plates.
Incubate at 37ºC over night.

Day 3
Innoculation
To 50 ml tubes with 10 ml LB-media per 1 litre culture intended add:
  10 μl Kanamycin stock
  10 μl Chloramphenicol stock
  10 μl Zn^{2+} stock
Innoculate a single colony of NrdD into each of the tubes.
Incubate overnight on a shaking table at 37ºC, 200 rpm.

Day 4
Growth and expression
To baffled flasks containing 1 l LB-media add:
  1 ml Kanamycin stock
  1 ml Chloramphenicol stock
  1 ml Zn^{2+} stock
  10 ml overnight culture of NrdD
Incubate at 37ºC, 200 rpm. Measure OD$_{600}$ every hour until it reaches 0.3-0.4, then more frequently until 0.5-0.7. This takes about 3 hours.
Freeze 1 ml of each culture.
Add 1 ml IPTG stock to each flask and incubate for 3 hours at 37ºC, 200 rpm.
After 1 h 30 min:
  Remove 1 ml of each culture and measure OD$_{600}$.
  Normalise the volume for each sample, including those aliquots removed previously, using 200 μl of the one with the highest OD.
  Centrifuge and resuspend the pellets in 30 μl H$_2$O.
  Use the resuspentions to run a gel, to see if there is expression. When preparing for the gel incubate at 90ºC and dilute liberally, since this becomes very viscous.
After the 3 hour incubation centrifuge the cultures at 6000 rpm, 12 minutes
Discard the supernatant.
Weigh the pellets.
Freeze and store the pellets at -80ºC.

Day 5

Homogenisation

   Dissolve the cell pellets in 20-30 ml lysis buffer. (About 4 volumes, v/w)
   Lysate 15-20 ml at a time using sonication: 80 W, 6 cycles of 20 s pulses with 60 s rest.
   Centrifuge for 20 min at 20 000 g.
   Pool the pellets, weigh and save as P1 at -20ºC, save a knife's edge separately for running a gel.
   Incubate the supernatant at 75ºC for 5 min.
   Centrifuge for 10 min at 10 000 g.
   Weigh and save the pellet as P2 at -20ºC, save a knife's edge separately for running a gel.
   Slowly saturate the supernatant with ground ammonium sulphate, up to 60% (390 g/l) during very careful stirring.
   Leave on a sample wheel on the slowest possible speed for 10 min.
   Centrifuge for 10 min at 10 000 g.
   Save the supernatant as S1 at -20ºC, after running the gel mentioned below.
   Weigh the pellet, save a knife's edge separately for running a gel and store the rest at -20ºC as P3 for use the next day.
   Run an SDS-PAGE gel. To normalise the small pieces of pellet to the supernatant S1 and ensure similar conditions dissolve the pellet in ddH$_2$O according to the following formula:

   (Assumed total V of supernatant * Weight of pellet fraction)/(Total pellet weight) = Volume of water to dissolve the pellet in
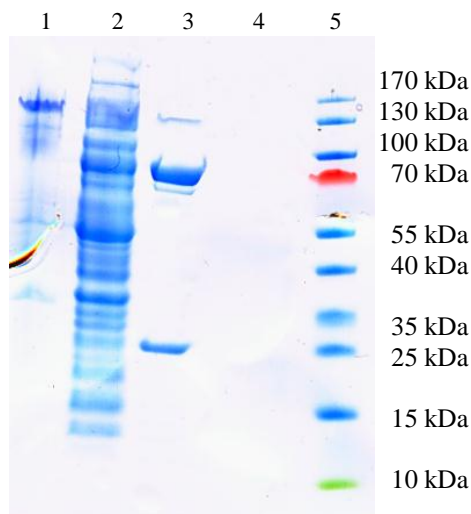
   SDS-PAGE gel of the collected pellets and supernatants. Lane 1: P1, lane 2: P2, lane 3: P3, lane 4: S1 and lane 5: ladder.

Day 6

Purification: butyl sepharose high trap column

   Equilibrate a butyl Sepharose Fast Flow HiTrap column (5 ml, GE Healthcare) with 50 ml buffer 1. If the pressure allows, keep the flow up to 5 ml/min throughout.
   Thaw P3 from day 5 and dissolve in 4 ml buffer 1.
   Measure the protein and nucleotide concentrations.
   Centrifuge for 10 min at 10 000g.
   Load the column with the total amount of protein from up to two litres of bacterial culture.
   Elute with 0-100% buffer 2, mixed with buffer 1, collecting 1 ml fractions over 100 ml.
   There are two fractions at the end of the gradient that both contain tmNrdD. The latter one also contains chloramphenicol acetyl transferase and should be discarded. Pool the fractions of the first of these peaks.
   Wash the column with 25 ml buffer 1.
   Load more protein if necessary, repeating the last three steps.

When all protein has been run through, wash with 50 ml ddH$_2$O, then with 20% EtOH.

Pool all the collected fractions and measure the concentration.

Fill two Falcon tubes with 15 ml concentration filters for 10 kDa with buffer 2 and centrifuge at 8000 rpm until all liquid is run through to wash them.

Concentrate the protein using the two tubes.

Add 100 μl DTT to 20 ml buffer 2. Add 10 ml of this to each of the concentrated protein tubes and concentrate again down to about 0.25 ml.

Measure the concentration again.

Transfer to an Eppendorf Tube, using the largest possible pipette tip, and washing the walls.

Store in the fridge over night.

Day 7

Equilibrate a Superdex 200 column (23 ml) with 50 ml ddH$_2$O, then 50 ml buffer 3 at 0.5 ml/min. This can be done over night.

Load the protein, no more than 10 mg protein or 0.3 ml sample.

Elute with 30 ml buffer 3.

TmNrdD appears as a large peak at about 10 ml. Collect 0.25 ml fractions in this area and pool them.

If necessary reload the column until all the protein has been run through.

Measure the concentration of all the pooled fractions.

Concentrate using 0.5 ml concentration tubes for 30 kDa, centrifuged at 8000g. Measure the concentration using nanodrop and, if necessary, dilute the protein samples to a final concentration of 20-25 mg/ml.

Run DLS and an SDS-PAGE gel.

Aliquot the protein in to 20-30 μl fractions.

Flash freeze in liquid nitrogen and store at -80ºC.

Appendix II: Comprehensive overview of the different crystals tested for data collection

Date	Effector	Substrate	Soak time	Data set	Space group parameters	Reso-lution	R-factor	R-free	Notes
130421	-	GTP	8h 40min	N					More than one crystal. 25% PEG400
130421	-	GTP	8h 40min	Y		2,8			Xds couldn't index. Didn't finish collecting. 25% PEG400
130421	-	GTP	9h 30min	N					25% PEG400
130421	-	ADP	10h	Y					Xds couldn't index. Didn't finish collecting. 25% PEG400
130421	-	ATP	10h 30min	Y					Xds couldn't index, didn't finish colletcing. 25% PEG400
130421	-	ATP	45min	N					ATP at 0.2 mM (1/10th normal conc)
130421	-	GTP	50min	Y	79.5 92.5 87.3 90.0 112.7 90.0	2,2	0.405	0.496	Nothing bound
130421	-	ATP	2h 20min	Y	77.6 93.2 86.5 90.0 112.2 90.0	2,6	0.362	0.437	ATP at 0.2 mM (1/10th normal conc).
130418	dGTP	ATP	3h 5min	Y	78.6 95.0 87.3 90.0 112.0 90.0	2,1			2 crystals, some ice. dGTP bound, 2 blobs in active site
130418	dGTP	ADP	3h	N					Very split
130418	dGTP	ADP	2h 55min	N					Just ice
130418	dGTP	ADP	4d	N					Lots of ice
130418	dTTP	GDP	2h 55min	Y		2,2			Slightly split. Both bind fine
130418	dGTP	ATP	4 d	N					Lots of ice
130418	dGTP	ATP	3h 15min	Y	92.6 94.6 137.2 90.0 96.7 90.0	1,9	0.250	0.306	More than one crystal in the loop. dGTP binds clearly. Two larger blobs in active site
130418	dTTP	ATP	2h 45min	N					No crystal, just some ice on the edge
130418	dTTP	ATP	2h 45min	Y	79.9 99.5 88.0 89.7 112.0 90.8	2,1	0.206	0.257	A bit cracked. Very clear density for dTTP, and for 3 blobs in active site
130418	-	GDP	2h 45min	N					Just ice
130418	-	GDP	2h 45min	N					Just ice
130418	-	GDP	4 d	N					Just ice
130418	-	ADP	2h 45min	Y					A bit split, but ok. Not solvable by xds
130418	dGTP	ADP	3h 25min	Y	93.7 95.5 138.6 90.3 96.9 89.7	1,9	0.244	0.290	A little bit split. Not very anisotropic. dGTP bound, two blobs in active site
130418	-	ATP	2h 40min	Y	78.1 98.1 87.0 90.0 111.9 90.0	1,8	0.217	0.257	No splitting. Binds effector site clearly, 3 blobs in active site
130418	dGTP	CTP	2h 45min	Y	92.5 94.8 137.1 90.0 96.7 90.1	1,9	0.491	0.541	
130418	-	ADP	2h 45min	Y					Didn't finish: the crystal wasn't great and because of no injection the intensity was very low.
130317	dATP	-	-	N					8 crystals tested
130317	-	-	-	N					5 crystals tested
130317	dATP	-	30 s	Y		3			Not bound, not long enough soaking
130317	dATP	-	1 h	Y	78.1 93.4 86.8 90.0 112.3 90.0	2,9			Bound
130317	dGTP	-	5 min	Y		3,1			Not bound, not long enough soaking
130317	dGTP	-	1h	Y	77.9 93.0 86.6 90.0 112.7 90.0	2,8	0.356	0.446	Both effectors bind, relatively clearly
130317	dCTP	-	50 min	N		3			
130317	dCTP	-	1 h	N		2,7			Very split
130317	dCTP	-	1h 5 min	Y	77.7 92.5 86.8 90.0 112.4 90.0	2,3			Lots of ice. No binding
130317	dCTP	-	3h 15 min	N					
130317	dCTP	-	3h 30 min	N					Very split
130317	dCTP	-	3h 40 min	Y	77.7 93.6 86.6 90.0 112.3 90.0	2,5			No binding
130317	dATP	-	1h 20 min	Y	77.7 94.7 87.1 90.0 111.6 90.0	2,5	0.351	0.448	Both effector sites have dATP bound, one of the clearer than the other.
130317	dTTP	GTP	2h 30 min	Y	77.9 97.8 86.7 90.0 111.9 90.0	2,3			Some ice rings. Both bind
130317	dGTP	ATP	1h 10 min	N		2,2			Very split
130317	dTTP	2xGTP	1h 25 min	Y	92.0 95.5 137.4 90.0 97.0 90.0	2	0.194	0.269	Both bind
130317	dATP	CTP	2h 30 min	N		3			
130317	dATP	CTP	2h 45 min	Y	78.1 95.5 87.3 90.0 111.8 90.0	2,5	0.214	0.256	Snow. Nice binding

130224	dTTP	-	-	Y		Zn anomalous, Mn used instead of Mg. 3 datasets
130224	dTTP	-	-	Y		Mn anomalous, Mn used instead of Mg. 5 datasets
130207	dTTP	-	-	Y		Anomalous data sets for Fe, Zn and Mn. Cocrystallised with Mn, not Mg
130207	dGTP	-	5min	N		Ice rings
130207	dGTP	-	3min	N		Some ice rings
130207	dGTP	-	1min	N		Nothing
130207	dGTP	-	1min	N		Nothing
130207	dGTP	-	3min	N		No cryo, ice
130207	dGTP	-	5min	N		dGTP soaked in drop, then transferred to cryo, then frozen.
130207	dTTP	-	-	N		Crystallised with Mn instead of Mg. 4 datasets
130125	dTTP	-	-	Y		Zn & Fe anomalous datasets, Mn used instead of Mg
121104	dTTP	-	-	N		2 crystals tested
121104	-	-	-	N		
121104	dGTP	-	-	N		3 crystals tested
121104	dATP	-	-	N/Y		2 crystals tested, 1 dataset
121104	dCTP	-	-	N		2 crystals tested
121019	-	-	-	N/Y		6 crystals tested, 1 dataset
121019	dTTP	GDP	-	Y		
121019	dATP	-	-	N		
121019	dCTP	-	-	N/Y		2 crystals tested, 1 dataset

Appendix III: B-factors of phosphates in the effector and active sites

The phosphate B-factor values are organised so that corresponding positions in the structure are in the same column. For this reason the substrate phosphate positions in brackets indicate the GTP phosphates. The mean B value is for the whole protein.

	Effectors			Substrates				Mean B value
	α	β	γ	α	β (α)	γ (β)	γ	
ATP*	19.7	24.3	24.3		26.4	17.3		26.4
	18.8	25.1	24.5		21.0	30.5		
dATP-CTP	41.4	39.0	48.7	39.7	40.5	38.1		31.0
	38.7	48.0	42.8	43.8	41.8	34.3		
dGTP-ADP*	20.6	19.3	26.3		32.7	24.9		30.0
	21.2	22.8	28.3		38.2	27.7		
	19.3	23.3	26.0		30.6	24.2		
	20.4	22.9	26.7		27.4	31.5		
dGTP-ATP*	27.4	33.6	42.0		18.4	24.0		28.5
	35.9	32.4	34.7		24.8	21.2		
	22.6	24.7	29.8		21.3	16.8		
	23.3	23.8	21.7		18.2	19.4		
dTTP-ATP*	19.8	22.3	22.0		21.2	29.4		29.4
	25.3	20.3	20.5		28.8	18.7		
dTTP-GDP	20.3	25.2	26.8		22.1	16.9		30.5
	23.3	28.3	27.5		23.4	19.0		
dTTP-GTP	17.9	19.7	22.2		20.9	20.4	45.5	14.2
	15.8	18.7	19.5		27.6	20.2	43.2	
	16.7	17.0	18.6		24.2	19.5	51.5	
	14.8	15.6	18.2		26.3	20.4	53.8	

* = the active site of this complex only contains pyrophosphate