Bacterial NadQ (COG4111) is a Nudix-like, ATP-responsive regulator of NAD biosynthesis

Journal of Structural Biology

BACTERIAL NADQ (COG4111) IS A NUDIX-LIKE, ATP-RESPONSIVE REGULATOR OF NAD BIOSYNTHESIS

--Manuscript Draft--

Manuscript Number:	
Article Type:	Full Length Article
Keywords:	transcription factor; NAD metabolism; X-ray crystallography; Bioinformatics; comparative genomics
Corresponding Author:	Leonardo SORCI, Ph.D. Marche Polytechnic University Ancona, ITALY
First Author:	Gabriele Minazzato
Order of Authors:	Gabriele Minazzato
	Massimiliano Gasparrini
	Annie Heroux
	Natalia V Sernova
	Dmitry A Rodionov
	Michele Cianci
	Leonardo SORCI, Ph.D.
	Nadia Raffaelli
Abstract:	Nicotinamide-adenine dinucleotide (NAD) is centrally important to metabolic reactions that involve redox chemistry. In bacteria, NAD biosynthesis is controlled by different transcription factors, depending on the species. Among the four regulators identified so far, the protein NadQ is reported to act as a repressor of the de novo NAD biosynthetic pathway in proteobacteria. Using comparative genomics, a systematic reconstruction of NadQ regulons in thousands of fully sequenced bacterial genomes has been performed, confirming that NadQ is present in α-proteobacteria and some β- and γ-proteobacteria, including pathogens like Bordetella pertussis and Neisseria meningitidis , where it likely controls de novo NAD biosynthesis. Through mobility shift assay, the DNA binding activity of NadQ from Agrobacterium tumefaciens was experimentally validated and determined to be suppressed by ATP. The crystallographic structures of NadQ in native form and in complex with ATP were determined, indicating that NadQ is a dimer, with each monomer composed of an N-terminal Nudix domain hosting the effector binding site and a C-terminal winged helix-turn-helix domain that binds DNA. Within the dimer, we found one ATP molecule bound, at saturating concentration of the ligand, in keeping with an intrinsic asymmetry of the quaternary structure. Overall, this study provided the basis for depicting a working model of NadQ regulation mechanism.
Suggested Reviewers:	SANDRA K ARMSTRONG armst018@umn.edu preliminary study on NadQ from B. pertussis
	STEVEN BRENNER brenner@compbio.berkeley.edu NUDIX PROTEIN FAMILY
	TADHG BEGLEY begley@chem.tamu.edu NAD metabolism
	YOUJUN FENG fengyj@zju.edu.cn Elucidation of functional role of NrtR, the closest NadQ homolog

	MICHAEL Y GALPERIN galperin@ncbi.nlm.nih.gov microbial genomics, particularly, reconstruction of biochemical pathways
	Jae Young Lee jylee001@dongguk.edu For the functional and structural analysis of NiaR (NAD-related transcriptional regulator)

Ancona (Italy) March 9, 2022

Dear *JMB* editor,
Please find attached our manuscript entitled “BACTERIAL NADQ (COG4111) IS A NUDIX-LIKE, ATP-RESPONSIVE REGULATOR OF NAD BIOSYNTHESIS“ by *Gabriele Minazzato, Massimiliano Gasparrini, Annie Heroux, Natalia V Sernova, Dmitry A Rodionov, Michele Cianci, Leonardo Sorci, and Nadia Raffaelli* to be considered for publication in *JMB*.

In this work, we provide a comprehensive biochemical and structural characterization of NadQ, the least studied among the known transcriptional regulators of NAD metabolism in bacteria.
Through mobility shift assays and mutagenesis, we have experimentally validated NadQ binding to a DNA region upstream of the operon involved in the first steps of the *de novo* NAD biosynthesis and demonstrated that ATP, a direct precursor of NAD, suppresses the *in vitro* binding of NadQ to its DNA target site. The resolution of the crystallographic structures of the regulator in its native form and complex with the ATP effector and DNA provided the atomic description of the mechanism of ligand and DNA recognition. Structural comparison with related Nudix proteins highlighted a novel variant of the Nudix fold, with a C-terminal four-helix bundle extension as the most noticeable feature of the NadQ Nudix domain. Finally, we systematically reconstructed NadQ regulons in thousands of fully sequenced bacterial genomes using comparative genomics, expanding our knowledge of NadQ distribution among bacteria and its controlled NAD pathways. Overall, this study presents NadQ as a master regulator of NAD biosynthesis in most proteobacteria, including some relevant pathogens.
Thanks for your consideration.
Sincerely,

The corresponding authors, Raffaelli Nadia & Sorci Leonardo, on behalf of all authors

Polytechnic University of Marche, Ancona, Italy

BACTERIAL NADQ (COG4111) IS A NUDIX-LIKE, ATP-RESPONSIVE REGULATOR OF NAD BIOSYNTHESIS

Gabriele Minazzato[1,§], Massimiliano Gasparrini[1,§], Annie Heroux[2], Natalia V Sernova[3], Dmitry A Rodionov[3,4], Michele Cianci[1], Leonardo Sorci[5,*], Nadia Raffaelli[1,*]

[1]Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy
[2]Elettra - Sincrotrone Trieste S.C.P.A.,Basovizza, Italy
[3]A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia
[4]Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, California, USA
[5]Department of Materials, Environmental Sciences and Urban Planning, Division of Bioinformatics and Biochemistry, Polytechnic University of Marche, Ancona, Italy

* To whom correspondence should be addressed. Email: n.raffaelli@staff.univpm.it; l.sorci@staff.univpm.it

[§]equally contributed

ABSTRACT

Nicotinamide-adenine dinucleotide (NAD) is centrally important to metabolic reactions that involve redox chemistry. In bacteria, NAD biosynthesis is controlled by different transcription factors, depending on the species. Among the four regulators identified so far, the protein NadQ is reported to act as a repressor of the *de novo* NAD biosynthetic pathway in proteobacteria. Using comparative genomics, a systematic reconstruction of NadQ regulons in thousands of fully sequenced bacterial genomes has been performed, confirming that NadQ is present in α-proteobacteria and some β- and γ-proteobacteria, including pathogens like *Bordetella pertussis* and *Neisseria meningitidis*, where it likely controls *de novo* NAD biosynthesis. Through mobility shift assay, the DNA binding activity of NadQ from *Agrobacterium tumefaciens* was experimentally validated and determined to be suppressed by ATP. The crystallographic structures of NadQ in native form and in complex with ATP were determined, indicating that NadQ is a dimer, with each monomer composed of an N-terminal Nudix domain hosting the effector binding site and a C-terminal winged helix-turn-helix domain that binds DNA. Within the dimer, we found one ATP molecule bound, at saturating concentration of the ligand, in keeping with an intrinsic asymmetry of the quaternary structure. Overall, this study provided the basis for depicting a working model of NadQ regulation mechanism.

***Keywords*:** Transcription factor; NAD metabolism; X-ray crystallography; bioinformatics; comparative genomics

INTRODUCTION

NAD represents both a redox coenzyme and a crucial signaling metabolite in all living organisms. Bacteria use the AMP or ADPR moiety of NAD to activate functional groups on proteins, RNA, or DNA, resulting in DNA replication, repair, and recombination, control of RNA transcription and protein translation, as well as modulation of enzymes' activity [1-4]. Homeostasis of the coenzyme is maintained by its continuous regeneration through different biosynthetic pathways, including *de novo*, recycling, and salvage routes [5, 6]. In most bacteria, the *de novo* pathway allows NAD synthesis starting from aspartate and dihydroxyacetone phosphate, whereas in a few species, including Flavobacteriales and the genus *Xanthomonas*, NAD is synthesized from tryptophan [7]. A network of pathways allows cells to recycle back to NAD the by-products of the intracellular coenzyme consumption and salvage the available exogenous precursors, i.e., nicotinic acid (niacin), nicotinamide and nicotinamide riboside, collectively known as vitamin B_3. The different metabolic routes may occur in different combinations depending on the bacterial species and the metabolic status [5]. The regulation of the biosynthetic machinery is guaranteed by four different transcription factors (TFs) that act as repressors of genes involved in both the biosynthetic pathways and the uptake of the vitamin. In most Enterobacteriaceae, NadR regulates NAD biosynthesis by sensing intracellular NAD and ATP levels [8, 9]. In particular, under physiological conditions, NAD binds NadR and the complex associates with DNA, thus repressing gene transcription. In case of low NAD and normal ATP levels, ATP binds the regulator releasing it from DNA and allowing transcription of the NAD biosynthetic genes. In bacteria from the Bacillus/Clostridium group, in Fusobacteria and Thermotogales, the regulator NiaR acts as a nicotinic acid-responsive repressor allowing cells to switch from the *de novo* biosynthesis to the salvage routes [10, 11]. In many diverse bacterial species, NAD biosynthesis is under the control of the Nudix-related transcriptional regulator (NrtR) family of TFs [10]. Some members of this family, like NrtRs from *Shewanella oneidensis* [12], *Synechocystis sp.* [10], *Pseudomonas aeruginosa* [13], and *Mycobacterium* species [14], sense the raising of ADP ribose (ADPR), a direct NAD catabolite, as a signal for the cell to replenish the NAD pool. In particular, binding of ADPR to the regulator promotes its dissociation from DNA, thus inducing NAD biosynthesis. Other members of the NrtR family, like NrtR from *Corynebacterium glutamicum*, control the *de novo* NAD biosynthetic genes using NAD as the co-repressor [15]. The least characterized among NAD-related TFs is NadQ, which earlier bioinformatics data suggested to control NAD biosynthesis in proteobacteria, and a recent report demonstrated to be a repressor of *de novo* NAD biosynthesis in *Bordetella pertussis* [16-18]. In this work, by combining bioinformatics, biochemical and structural analyses, we have characterized NadQ from *Agrobacterium tumefaciens,* providing insights into the mechanism of NadQ-mediated transcriptional regulation of NAD biosynthesis in proteobacteria.

MATERIAL AND METHODS

Cloning, expression, and purification of *At*NAdQ

The *nadQ* gene was amplified from genomic DNA of *Agrobacterium tumefaciens* str. C58 with the primers CAACGTGACCATCGGGCTTGCCCAT and TCAATTGCGGGAGAGGGGCAGTTTCGT for cloning into the pET100 vector (pET100-TOPO kit, Invitrogen). *E. coli* One Shot TOP10 Competent Cells (Invitrogen) were used for plasmid propagation. The absence of mutations in the cloned gene was confirmed by DNA sequencing at BMR Genomics (DNA Sequencing Service of Padova University, Italy, http://www.bmr-genomics.it/). For protein expression the constructs used to transform *E. coli* BL21 (DE3). Cells were grown at

37°C in Luria Bertani medium supplemented with ampicillin (100 µg/ml). After reaching an OD_{600} of 0.6, the culture was shifted at room temperature and after 20 min, protein expression was induced with 1mM IPTG at 23°C. After 15 hours of induction, cells were harvested by centrifugation at 5,000 x g for 10 min at 4°C. The pellet was resuspended in one-twentieth of the original culture volume of buffer A (50 mM HEPES/NaOH pH 7.5, 300 mM NaCl) containing 1 mM PMSF and protease inhibitor cocktail (Sigma)). Cells were disrupted by two passages in a French press cell (18,000 psi), and the cell debris was removed by centrifugation (20 min at 20,000 x g at 4°C). The supernatant was loaded onto a HisTrap FF IMAC (GE Healthcare) column (1ml), equilibrated with buffer A. After a washing step with 40 mM imidazole in buffer A, the recombinant protein was eluted with a linear gradient of imidazole, ranging from 40 mM to 300 mM, in buffer A. Imidazole was removed by a subsequent gel filtration chromatography on Sephadex G-25 resin (GE Healthcare) equilibrated and eluted with buffer A. About 40 mg of pure protein were obtained starting from 1 litre culture. The final sample was stored at -20°C until use. The Q248A/R273A NadQ mutant protein was purified following the same purification protocol of the wild-type NadQ.

The selenomethionine-substituted (SeMet) protein was produced as described [19]. Briefly, *E. coli* B834 (DE3), a methionine auxotrophic strain, was transformed with the same construct used to express the wild type protein. Transformed cells were grown at 37°C in minimal medium M9 supplemented with ampicillin (100 µg/ml) and selenomethionine (40 µg/ml). After reaching an OD_{600} of 0.5, the culture was shifted at 23°C, and protein expression was induced with 1mM IPTG. After 15 hours of induction, cells were collected, and the purification of the SeMet protein was performed as described above, except for the presence of 1 mM DTT in all buffers. About 13 mg of protein were obtained starting from 1 litre culture.

Site-directed mutagenesis

Site-directed mutagenesis was carried out using the QuickChange Lightning kit (Agilent Technologies), following the manufacturer's instructions. The plasmid pET100 harboring the wild-type *Agrobacterium tumefaciens nadQ* gene was used as a template to generate the NadQ mutant R273A (primers: AGCTTCGCCGGGGCGCCGCCGGTTTC and GAAACCGGCGGCGCCCCGGCGAAGCT), which, in turn, was used to generate the double mutant Q248A/R273A with the primers AGCCGGCGGAAATTCGCCTTGTGCAGCGTCAG and CTGACGCTGCACAAGGCGAATTTCCGCCGGCT. The mutants were sequenced to verify incorporation of the desired modification and to ensure the absence of random mutations.

Electrophoretic mobility shift assay

A 151-bp fragment containing the two predicted DNA binding sites was amplified by PCR from genomic *A. tumefaciens* DNA, by using the primers AGACCGTTCTTGATGAACG and GCGGCTGAAATCTGCTC. PCR product was purified using the High Pure PCR Product Purification kit (Roche) and spectrophotometrically quantitated.

For the electrophoretic mobility shift assay, 76 ng (800 pmol) of the DNA fragment were incubated with different amounts of wild type (or mutant) NadQ in binding buffer consisting of 10 mM Tris, pH 7.5, 50 mM KCl, 2.5% glycerol, 5 mM $MgCl_2$, 0.05% NP-40, and 0.5 mg/ml BSA, in a final volume of 20 µl. After 20 min at 25°C, samples were loaded on a 6% native polyacrylamide gel (30:0.8 acrylamide:N,N9-methylenebisacrylamide [w/w]). Electrophoresis was carried out in 0.5 x TBE (Tris–Borate–EDTA) buffer for 60 min at 120 V, at 4°C. Gels were stained for 40 min in 1 x TBE buffer, containing 2 µl Sybr Green I diluted 1:10,000.

Protein crystallization and data collection

For all crystallization trials, the purified proteins were diluted with 50 mM HEPES, pH 7.5, at room temperature to reduce NaCl concentration to about 50 mM. Samples were then concentrated by ultrafiltration, using an Amicon Ultra Centrifugal Filter (cutoff 10 kDa, Merck, Millipore), at 18°C, to a final protein concentration of 7 mg/ml. For all crystallization trials, the sitting drop vapor diffusion method was applied. One microliter of concentrated NadQ was mixed with an equal volume of reservoir solution, and it was equilibrated against 100 µl of the reservoir solution.

Native-NadQ - The best crystals were obtained in reservoir solution containing 150 mM di-Sodium DL-malate, pH 7.0, 20% (w/v) PEG 3350 (JCSG kit, JENA Bioscience). Crystals grew to their final size in four months at 18°C.

SeMet NadQ-ATP complex for phasing - Initial phases were obtained using SeMet NadQ crystals in complex with ATP at a 1:10 molar ratio. Crystals grew in 100 mM Bis-tris-propane, pH 7.0, 250 mM sodium potassium tartrate, 24 % (w/v) PEG 3400. Crystals grew in a month at 18°C to a size of 50 x 50 x 50 μm^3. Data were collected at a wavelength of λ= 0.976 Å of the Se-K edge to optimize the anomalous signal from Se atoms.

SeMet NadQ-ATP complex - Differently from the crystal batch used for phasing, the SeMet NadQ-ATP complex was co-crystallized with ATP at a 1:100 molar ratio.

NadQ-DNA complex - A 32 bp oligodeoxynucleotide duplex with sticky ends was prepared by annealing two synthetic complementary oligonucleotides (Fw, TTGACATATGCTCACAATGAGAATATGTCTTC; Rv, AAGACATATTCTCATTGTGAGCATATGTCAAG). The binding box of NadQ is underlined. The protein-dsDNA complex was formed by mixing NadQ and dsDNA at a 1:1.2 molar ratio. Crystals grew after mixing equal volumes of complex with reservoir solution containing 100 mM HEPES, pH 7.5 mM lithium chloride, 28% (w/v) PEG 6000. Crystals appeared after 15 days at 18°C.

Crystals of native NadQ and SeMet NadQ-ATP complex for phasing were transported to the synchrotron in plates, mounted in nylon loops, and flash-frozen directly at 100 K in a nitrogen gas stream. Crystals of NadQ-DNA were mounted in nylon loops and flash-frozen in liquid nitrogen for transportation. In all cases, crystals were cryopreserved in the same solution of reservoir with the addition of 20% of glycerol.

Diffraction data of native NadQ were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at beamline ID30B [20]. SeMet-NadQ-ATP data were collected at Elettra Synchrotron (Trieste, Italy), beamline XRD2 [21]. Data collection statistics are summarized in **Table 1**.

Structure determination, refinement, and analysis

Starting phases for solving the crystal structure were obtained with Se-SAD method using ATP bound SeMet NadQ crystals. Crystals belong to space group $P2_1$, with unit cell a=58.2 Å, b=134.3 Å, c=87.0 Å, and α=90.00°, β=100.04°, γ=90.00°. One of the thirty analysed crystals was of sufficient diffraction quality to collect a complete SAD dataset. The diffraction data were processed and scaled with the XDS/XSCALE program package [22]. The positions of 16 Se atoms (out of the 32 expected) were determined with SHELXD [23]; the missing Se residues were located in the N-terminus of the recombinant protein comprising 36 residues that derived from the cloning procedure and were not always completely defined in the electron density maps. SHELXE [23] was used to refine the position of the Se sites and improve the electron density maps [24]. The model so obtained was then used as starting model to refine the native NadQ structure. Automated model building was accomplished by PHENIX suite [25], which allowed the partial building of the main chains of the

four molecules (two dimers) of the asymmetric unit. The manual fitting of the side chains and solvent molecules into electron density maps were performed using COOT [26], PHENIX suite [25] and Refmac5 [27] of the CCP4 package [28], while monitoring R_{work}, R_{free}, Ramachandran plot [29] and related geometrical parameters. The NadQ-ATP structure was initially solved by direct rigid body refinement using the native form of NadQ as starting model after atom randomization to avoid any bias. The Fourier difference electron density maps at 3 sigma level clearly indicated the presence of the ATP, which was refined with full occupancies. The NadQ-DNA complex structure was solved by Molecular Replacement using PHASER [30] with one monomer of the native form as starting model. The refinement protocol for NadQ-ATP, and NadQ-DNA complexes followed the one reported for the native form. The models were checked with the PDB REDO web server [31]. Molecular graphics and root mean square deviation (RMSD) calculations were performed using Molsoft ICM-PRO version 3.8 [32]. Interface areas of macromolecular assemblies were calculated with PISA web server [33]. **Table 1** reports final refinement statistics.

Bioinformatic tools

All genomes used in this study were taken from the SEED database of bacterial genomes [34]. Orthologs of NadQ were identified as bidirectional best hit using BLASTP [35] and were confirmed by construction of phylogenetic tree using the PhyML program [36]. The tree was displayed and annotated with iTol online software [37]. Multiple sequence alignments of NadQ proteins were built in MAFFT [38]. NadQ orthologs were found mostly among three taxonomic orders of Proteobacteria, that is α-, β-, and γ-proteobacteria. Among 350 proteobacterial genomes possessing *nadQ* gene available in the SEED database in June 2020, we selected 94 genomes using the following criteria: if there are multiple genomes for the same taxonomic species, or a group of closely related species, we chose a single representative strain with preference to include the type strains (e.g. from the ATCC or DSM strain collections). The selected set of genomes included 65 α-proteobacteria, 23 β-proteobacteria, and 5 γ-proteobacteria (**Table S1**). The comparative genomic analysis of NAD biosynthesis, transport and regulatory genes was performed using the subsystems approach implemented in the SEED database and analysis tool, which combines protein similarity search, positional gene clustering, and phylogenetic gene profiling [34]. The NAD metabolism subsystem includes the functional roles listed in **Table S2**.

For identification of DNA binding sites and reconstruction of NadQ regulons, we applied the integrative comparative genomics approach as previously described [17]. Briefly, to find conserved NadQ-binding sites, we used the training sets of non-coding regulatory regions of the NAD subsystem genes. For training sets, we collected up to 300 nucleotides upstream of the translation start site and excluded the coding regions of any upstream gene if the intergenic region was less than 300 nt. 21-bp palindromic DNA motifs were identified with SignalX [39] (http://bioinf.fbb.msu.ru/SignalX/), resulting in the construction of positional weight matrix (PWM). The obtained PWM was further used to scan the studied genomes and to identify additional NadQ-binding sites using the z-scan script and GenomeExplorer software [40]. Scores of candidate sites were calculated as the sum of positional nucleotide weights. Specifically, we searched the upstream gene regions of NAD biosynthesis and salvage genes, putative niacin transporters, and *nadQ* genes and reported all candidate binding sites with scores above a PWM-specific threshold (typically between 4.5 and 6.5). The search parameters were selected allowing identification of potential sites between 300 nt upstream and 50 nt downstream of a gene start codon. Cross-species comparisons of the predicted sets of potentially regulated

genes allowed us to tentatively define regulon composition for each analysed genome. Additional BLAST searches versus non-redundant protein sequence database at NCBI identified the *nadQ* gene orthologs in two genomes from the Actinobacteria phylum. These two genomes were downloaded from NCBI RefSeq database (NZ_PDJC01000001.1 for *Propionicimonas paludicola* DSM 15597 and NZ_CP040899.1 for *Georgenia sp. Z294*) and analysed for the presence of candidate NadQ binding sites using the same PWM and Genome Explorer tool.

RESULTS

Genomic reconstruction of NadQ regulons

A bioinformatic reconstruction of the regulons controlled by NadQ in several proteobacteria predicted the regulator's DNA binding motif and showed that *nadQ* genes are often clustered with *de novo* biosynthetic genes *nadABC* [16]. Here, we have performed a systematic reconstruction of NadQ regulons in bacteria with completely sequenced genomes collected in the SEED database as of June 2020. The results captured in **Table 1** and summarized in **Figure 1A** confirm an overall occurrence of NadQ in proteobacteria, mainly in the α- and β-proteobacteria lineages. In these bacteria, NadQ mainly controls the *de novo* NAD biosynthesis operon *nadABC* and, in some Rhodobacterales and Caulobacterales, the downstream *nadD* and *NadE* genes common to both *de novo* and salvage pathways. Intriguingly, *Mesorhizobium loti* harbours a second NadQ isoform within the plasmid pMla, whereas in *Azospirillum* sp. B510 *nadQ* and its regulated operon *nadABC* are encoded in distinct plasmids, with no chromosomally coded *nadQ* gene. In *Bordetella* spp *nadQ* is in a cluster with and acts as a repressor of *nadC*, the gene of *de novo* NAD synthesis converting quinolinic acid to nicotinic acid mononucleotide [18, 41]. Notably, *Bordetella* spp. lack the first two genes of the *de novo* synthesis (*nadA* and *nadB*) and, therefore, can drive NAD synthesis from quinolinic acid. In the pathogen *Neisseria meningitidis*, *nadQ* is autoregulated, and it controls the two divergently transcribed operons *nadAB* and *nadQC* (**Figure 1A**). Remarkably, the NadQ regulon of *Caulobacter segnis* involves all the 7 NAD-related genes of this bacterium (**Figure 1A**), including *nadV* and bifunctional *nadM-nudix* genes of the amidated nicotinamide-salvage pathway to NAD [42, 43]. In γ-proteobacteria, NadQ only appears in the Moraxellaceae group, including the pathogen *Moraxella catarrhalis*. Outside proteobacteria, we detected a few occurrences of NadQ orthologs in Actinobacteria and Bacteroidetes, likely arising from horizontal gene transfer events from α- and β-proteobacteria lineages, respectively (see the phylogenetic tree in **Figure S1**). Further PWM-based searches for additional sites identified a second copy of putative NadQ site in most DNA upstream regions of the *nadABC* genes from α-proteobacteria and other target genes in γ- and β-proteobacteria (**Figure 1A** and **Table S1**). The tandem NadQ sites in Agrobacterium/Rhizobium spp. were further confirmed by phylogenetic footprinting (see multiple sequence alignment of their *nadABC* upstream regions in **Figure S2**). Our comprehensive bioinformatic analysis refined the consensus sequence of the NadQ DNA binding motif shown in **Figure 1A**.

Strikingly, in 12% of lineages with nadQ, comprising representative species of both α- and β-proteobacteria, we also detected the occurrence of NrtR. In particular, in all Comamonadaceae *nadQ* and *nrtR* co-occur, with 2 NrtR isoforms found in *Comamonas testosteroni*. This may suggest a common evolutionary origin of NadQ and NrtR in such lineages for a gene duplication event.

Overall, the systematic regulon analysis implies that NadQ is a master regulator of NAD biosynthesis.

Functional characterization of *A. tumefaciens* NadQ

To give an insight into the regulatory function of NadQ, we cloned the gene from *A. tumefaciens* and overexpressed the corresponding protein in *E. coli*. The recombinant protein was purified to homogeneity as a His_6-tagged protein through Ni-chelating chromatography, as described in the methods section. SDS-PAGE of pure NadQ revealed a molecular mass of about 39 kDa, as expected for the recombinant protein. Gel filtration experiments showed a native molecular mass of about 58 kDa, suggesting a dimeric structure of the protein in solution **(Figure S3)**. In the *A. tumefaciens* genome, two putative DNA binding motifs were found in the region upstream of the *nadABC* operon **(Figure 1B)**. An electrophoretic mobility shift assay was performed to test the binding of purified *At*NadQ to the DNA fragment containing the predicted NadQ operator site. As shown in **Figure 2A**, a substantial shift of the DNA fragment was observed upon DNA incubation with the protein. Notably, the intensity of the shifted DNA band increases in the presence of increasing amounts of NadQ. A DNA fragment of 71 bp included in the incubation mixture as a control was not affected by the presence of the protein, confirming the specificity of the NadQ–DNA interaction.

The effect of potential effectors on *At*NadQ-DNA binding was tested by incubating the protein with various NAD metabolites (**Figure 2B**). Among these, only ATP was effective in preventing binding to DNA (**Figure 2B**). The effect of ATP was dose-dependent (**Figure 2C**). The same results were obtained with both His-tagged *At*NadQ and the protein after removing the tag with enterokinase, indicating that the tag at the N-terminus of the protein does not interfere with the DNA binding ability.

Overall structure of *At*NadQ

The crystal asymmetric unit for native and ATP-bound NadQ contained four subunits (**Figure S4**). In detail, two pairs of interfaces (between subunits A and C or B and D) bury a surface area of ~586 $Å^2$. The other two pairs of interfaces (between subunits A and B and subunits D and C) bury an average surface area of 1841 $Å^2$ suggesting that homodimer AB and CD are the biological units. This further agrees with size-exclusion chromatography studies indicating that *At*NadQ exists as a dimer in solution. For each *At*NadQ form, the structures of the two dimers within the asymmetric unit were almost identical (RMSD of about 0.6 Å), with deviations mainly attributable to loop regions (**Table S3**). The dimeric structure of the native form is shown in **Figure 3A**. Each monomer comprises an N-terminal effector domain (residues 1-216) and a C-terminal DNA-binding domain (residues 218-300). The N-terminal domain shows an αβα Nudix-like fold with a landmark loop-helix-loop motif (also known as the Nudix box). The Nudix fold consists of three helices (α_1, α_6, and the short α_2) and two β sheets (made up of six β-strands (β_1-β_6) and is typical of the "Nudix Superfamily" of proteins comprising enzymes, mostly hydrolases, as well as nonenzymatic proteins [44]. In *At*NadQ, the αβα Nudix-core is flanked by a large insertion at its C-terminal region (residues 113-194) formed by a four-helix bundle (helixes α_3-α_5 and 3_{10} helixes η_2, η_3). Helixes α3 and α4 are antiparallel, while α5 goes from the bottom of α5 towards the top of α4 to depict an N-shaped footprint when observed from the front (**Figure 3B**). The C-terminal domain of *At*NadQ (residues 217-300) is a winged helix-turn-helix (wHTH) DNA-binding domain, typical of several families of bacterial TFs (40), consisting of helix α7 and α8, followed by a long β-hairpin "wing" (β7-β9) and ending with an α-helix (α9) running upward to make contact with the effector domain (see **Figure 3B** and below for the role of C-terminal end). The 18 residues-long α6 helix connects the Nudix with the wHTH domain.

The open/ATP-bound site

The Fourier difference F_o-F_c OMIT map of the *At*NadQ in complex with ATP clearly indicated a single ATP molecule bound in each *At*NadQ dimer (AB or CD**, Figure 3C,D**). We chose the CD dimer for the subsequent analysis and molecular graphics because of the overall better completeness of the crystallographic model (**Table S3**). In the attempt to saturate the effector's binding site, we used a 100-fold molar excess of the ATP ligand (at 10 mM concentration) in our crystallographic conditions. Notwithstanding, we found one ATP molecule bound per dimer. The ATP binding site is located at the interface between the two monomers (**Figure 4A**). A more significant portion binding the adenosyl moiety of ATP is harbored by one monomer (*c*), in a region comprising α4-α6, β4, and β4-β5 connecting loop. The other monomer (*d*) engages the β- and γ- phosphate groups with the N-terminal end of α1 and the C-terminal mobile loop. The adenine ring of ATP is sandwiched between the side chains of Arg201 and Tyr175 from subunit C. Together with Phe70, a continuous π-π stack interaction is generated. The two-hydroxyl groups of ribose are involved in an extensive hydrogen bond network with side chains of Arg150 and Glu176. The positively charged residues Arg73 and Arg201 form salt bridges with the oxygens of the α-phosphate of ATP. The β- and γ-phosphate are hydrogen-bonded to Gln46(d) and several surrounding water molecules with negative charges counterbalanced by potassium ions present in the binding pocket. The γ-phosphate establishes two strong H-bonding interactions with the Thr293(d) lateral group and the main-chain amide of Gly292(d). Finally, the positively charged side chain of Lys294(d) forms two salt bridges with the α- and γ-phosphates of ATP (**Figure 4A**).

The closed/unbound ATP-site

The investigation of the unbound ATP site of the NadQ-ATP complex revealed a slightly different architecture, which appears to be closed, i.e., nonconducive to ATP binding, thus explaining why only one ATP molecule is found in the NadQ dimer. In detail, in the closed binding site, Arg73(d) side chain occupies the same position of the adenine base in ATP-bound complex (i.e., is sandwiched between Tyr175(d) and Arg201(d)), thus preventing the ATP's adenine access in this site (**Figure 4A, C**). Arg73(d) position is further stabilized with two salt bridges with Asp198(d) and Glu172(d) (**Figure 4A**). The C-terminus of subunit c directly points to the unbound ATP site, further occluding the site to the ATP access (see next paragraph for the role of C-termini in NadQ asymmetry). Here, Arg201(d) is held in place by Glu172(d), but it also forms an H-bond with Leu295(c) main chain carbonyl from the C-terminus. The structural superposition of the two NadQ-ATP complex subunits clearly shows that Arg73(c) of the unbound ATP site would overlap with the adenine base (**Figure 4C**). By contrast, in the bound site, an outward shift of the X-loop, here structured as η10 helix, pulls away Arg73(d), allocating space to fit the ATP molecule (**Figure 4C**). The *At*NadQ-free form (**Figure 4B, E**) has similar binding site architectures to the *At*NadQ-ATP complex, with two ATP sites, one in a closed and the other in an open state.

The C-termini orientation favors NadQ dimer intrinsic asymmetry

As introduced in the previous paragraphs, besides the effector binding sites, the major structural difference between *At*NadQ subunits (1.9-Å average Cα RMSD) appears in the C-terminal mobile region that follows helix α9, as shown in **Figure 4D,** where the C-termini of the ATP-bound complex are superposed. Specifically, the C-terminus of monomer c (which harbors most ATP site) is stretched along the α9 helix direction, thus directly pointing and further occluding the ATP access to the other site (**Figure 4A, D**). By contrast, the d

monomer's C-terminus turns at residues 292-4 (GTK, six aminoacids from the actual terminus at N300) and extends at the dimer interface. The GTK triplet establishes strong H-bonds to the β and γ phosphates of ATP (**Figure 4A**).

Structural comparison with related Nudix proteins

Figure 5A shows a structure-guided sequence alignment of *At*NadQ with other representative members of the Nudix superfamily, including transcriptional regulators and hydrolases. Despite a poor structure-based sequence identity (<20%), *At*NadQ is most closely related to TFs of the NrtR family, namely SoNrtR (PDB: 3GZ8) *Bt*AraR (PDB: 5DEQ), and a putative nudix-like transcriptional regulator EF_2700 from *Enterococcus faecalis* (PDB ID: 2FML), with an average RMSD of 2 Å. A comparison of the topology of the *At*NadQ Nudix domain with other Nudix proteins is shown in **Figure 5B**. In general, Nudix proteins may present extensions at the N- or C-termini of the highly conserved Nudix core that vary in size and topology [44]. The most noticeable feature of *At*NadQ is a C-terminal four-helix bundle extension that precedes the DNA-binding domain. This portion is highly diversified among Nudix proteins, being unstructured as in *Bt*AraR and *Aa*Ap4A hydrolase or partially structured in α-helix as in *So*NrtR, *Syn*NadM_ADPRP, and *Ec*NudK. Similarly to *At*NadQ, the uncharacterized transcriptional regulator *Ef*_2700 hosts a large insertion in the analogous region but is organized as a β-sheet (**Figure 5B**).

Another distinctive feature of *At*NadQ pertains to the ligand binding site. Members of the Nudix superfamily recognize a wide range of nucleoside diphosphates linked to a moiety X as substrates or effector molecules. In these proteins, the ligand binding site is in the core central β-layer, with the Nudix box contacting the pyrophosphate substrate moiety, and it is delimited by additional specificity-determining loops (**Figure 5B**). In *At*NadQ, this cleft is empty, and the ligand-binding pocket is distant from the typical site, which appears to have lost its binding properties. In keeping with this observation, the Nudix box essential for the hydrolase activity of Nudix proteins has already deteriorated in NrtRs and is even more divergent in *At*NadQ, with only two conserved residues (**Figure 5A**).

Moreover, in Nudix proteins, the specificity-determining loops point toward the ligand binding cleft, while in *At*NadQ are either smaller or point outward (**Figure 5B**), underlying the loss of selective pressure for maintaining the ancestral Nudix function. Notably, the so-named "X loop" that in Nudix proteins binds the X-moiety of the ligand [44] in *At*NadQ (residues 72-82) retained the ability to interact with the ligand (**Figure 4A**). This loop is partially disordered in the open ATP site of the native NadQ but becomes ordered upon contact with ATP (**Figure 4A, B**). This conformation is implemented by the tight packing of a 3_{10} helix that imposes Arg73 in a favorable conformation to ATP binding (**Figure 4**).

The NadQ-DNA interaction

An overview of the *At*NadQ-DNA complex is presented in **Figure 6A**. The crystallographic asymmetric unit of the *At*NadQ-DNA complex contains one *At*NadQ dimer bound to a specific 32 nt DNA box. As expected, NadQ recognizes DNA through its C-terminal wHTH domain. Direct interactions between NadQ and target DNA are expected between helix α8 of the HTH motif and the major groove, while the β-hairpin wing interacts with the minor groove (**Figure 6A**). The 4.3 Å resolution electron density maps of the DNA complex for the *At*NadQ-DNA bound structure show a good fit of the HTH domain into major and minor grooves of DNA but do not allow to confidently model protein-DNA interactions or perform a meaningful structural comparison with other

structures. Thus, since the HTH domain of NadQ is quite similar to the DNA binding domain of related NrtR and AraR proteins, and crucial residues contributing to protein-DNA base interactions through the DNA recognition helix and hairpin wing, respectively, have been described [12, 45], we sought to validate functionally equivalent residues in NadQ by mutagenesis approach.

Guided by a structure-based sequence alignment (**Figure 5A**), we identified Gln248 and Arg273 as the functionally equivalent to residues of NrtR and AraR forming critical interactions with the nucleobases at the major and minor groove levels, respectively [12, 46]. We validated the roles of Gln248 and Arg273 in DNA-binding, showing that the double mutant Q248A/R273A does not bind DNA in an EMSA experiment where wild-type NadQ was used as a positive control (**Figure 6B**). Furthermore, Gln248 and Arg273 are strictly conserved in NadQ family members (see multiple sequence alignment in **Figure S5**), underlying their functional importance.

Since the roles of homologous aminoacids of NadQ Gln248 and Arg273 in NrtR and AraR were not confirmed by mutagenesis, our work provides additional cross-validation of DNA-binding determinants in the larger Nudix-related family of transcriptional regulators.

ATP modulates NadQ repressor function

ATP binding triggers conformational changes in several regions of NadQ, as shown by the superposition of NadQ-ATP and NadQ native dimers at the binding site level (**Figure 7**). First, the flexible C-terminal 8 residues become well-structured upon ATP binding with an important role played by Tyr68. While forming a pi stacking interaction with Phe70 in the free NadQ, Tyr68 undergoes a downswing upon ligand binding, adopting a different rotamer conformation stabilized by an H-bond with the main chain carbonyl of Leu295 (**Figure 7**). Notably, in the NadQ-DNA complex, up to 12 C-terminal residues (289-300) are not structured, reinforcing that this region is intrinsically flexible. In NadQ topology, this flexible C-terminal region directly links back the HTH C-terminal domain to the effector binding domain (**Figure 3B**), a connection unseen in related AraR and NrtR Nudix-like transcriptional factors. Thus, we argue that ATP-binding transmits rigidity to the two HTH motifs through α9 helix, preventing the necessary swing for a good fit into DNA grooves. Aside from the C-terminus, ATP binding-induced conformational changes appear to stabilize the X-loop with Arg73 making contact with ATP α-phosphate. Additional conformational changes occur upon ligand binding to bring Tyr175, Glu176, and Arg150 of the typical NadQ insertion module (**Figure 5B**) close to the ligand (**Figure 7**). Although the poor resolution of our NadQ-DNA complex prevents comparison with the ATP-bound complex, these observations provided initial hints to elucidate the structural mechanism underlying the decreased affinity for DNA upon ATP binding.

DISCUSSION

Prior to this work, NadQ was the least characterized among the known regulators of bacterial NAD biosynthesis. Previous studies predicted its presence in most proteobacteria and showed that *in Bordetella pertussis,* it acts as a repressor of the *nadC* gene, which is involved in the *de novo* NAD biosynthesis [16, 18]. Here, we have functionally and structurally characterized the regulator by studying the properties of recombinant NadQ from *A. tumefaciens.* Through mobility shift assays and mutagenesis, we have experimentally validated NadQ binding to a DNA region upstream of the operon involved in the first steps of

the *de novo* NAD biosynthesis, and demonstrated that ATP, a direct precursor of NAD, suppress the *in vitro* binding of NadQ to its DNA target site.

The structural characterization of the regulator revealed that among TFs with known 3D structures, NadQ is most closely related to members of the NrtR family of regulators that control NAD biosynthesis and sugar pentose utilization [12, 45]. Like NrtRs, NadQ is a dimer in which each monomer is composed of an effector-binding/dimerization Nudix domain and a winged helix-turn-helix DNA binding domain. However, while the DNA binding domain is markedly conserved, the NadQ Nudix domain is somewhat divergent. In fact, it shows a unique extension of about 80 residues that contributes to the formation of the effector binding site at the dimer interface, in a distant position from the typical substrate/ligand binding site of Nudix proteins. In NrtRs, as in several other Nudix proteins, the position of the effector binding site is dictated by a conserved Nudix motif that interacts with the pyrophosphate moiety common to all ligands and at least three loops that determine ligand specificity. In NadQ, the Nudix motif is almost entirely disrupted, and the typical specificity-determining loops are either shorter or point elsewhere. Notably, the β4-β5 connecting loop, also known as "X-loop" in other Nudix proteins [44], in NadQ plays a crucial role in contacting the ATP ligand. In addition, while NrtRs are symmetric homodimers, with two equivalent effector binding sites that the ligand can simultaneously occupy, NadQ is a homodimer with non-equivalent binding sites, i.e., they can only be occupied one at a time. Our structural analysis of the native form and the ATP- complex identified significant conformational differences of the two monomers' C-termini. In detail, when ATP is bound, the C-terminus of one monomer is conducive to ATP-binding, while the other hinders ATP-binding to the second site, which remains closed in both free and ATP-bound forms. Altogether the results of this work support a working model for NadQ as a sensor that links intracellular ATP concentration to NAD metabolism. In this model, in case of a decrease in intracellular ATP levels, NadQ binds DNA and acts as a repressor of the *de novo* NAD biosynthesis. In these conditions, we can assume that dihydroxyacetone phosphate (DHAP), a precursor of NAD in *de novo* synthesis, is better used to fuel glycolysis for ATP production than being converted to NAD, and NAD would be formed from alternative sources, like vitamin B3, through the salvage pathways. When ATP concentration builds up, ATP binds NadQ promoting its dissociation from DNA and allowing NAD synthesis through the *de novo* route.

In summary, the molecular characterization of NadQ underscored a peculiar version of the Nudix fold evolved to sense and link cellular ATP availability to NAD synthesis.

DATA AVAILABILITY STATEMENT

The atomic coordinates and structure factors for *At*NadQ have been deposited in Protein Data Bank (7Q91 for native *At*NadQ; 7Q92 for *At*NadQ-ATP complex; 7Q94 for *At*NadQ-DNA complex).

SUPPLEMENTARY DATA

Supplementary Data are available at JMB online.

ACKNOWLEDGMENTS

We thank all the facilities for the beam time and the technical support. The synchrotron MX data was collected at the PETRA III storage ring operated by EMBL Hamburg (DESY, Hamburg, Germany, beam time allocation numbers MX-610, MX-640, MX-720 to MC), at ELETTRA (Trieste, Italy, allocation number 20195606 to MC) and at ESRF (Grenoble, France, beam time allocation numbers MX-1949).

FUNDING

This research was partly supported by the Polytechnic University of Marche to NR, MC, and LS, Ministero dell'Università e della Ricerca, PRIN Project 2017CBNCYT to NR and by Fondazione Cariverona, Bando Ricerca Scientifica di Eccellenza 2018, Project NADBES 2018.0773, to NR.

CONFLICT OF INTEREST

None declared.

REFERENCES

[1] Bird JG, Zhang Y, Tian Y, Panova N, Barvik I, Greene L, et al. The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. Nature. 2016;535:444-7.

[2] Burckhardt RM, Buckner BA, Escalante-Semerena JC. Staphylococcus aureus modulates the activity of acetyl-Coenzyme A synthetase (Acs) by sirtuin-dependent reversible lysine acetylation. Mol Microbiol. 2019;112:588-604.

[3] Munir A, Banerjee A, Shuman S. NAD+-dependent synthesis of a 5'-phospho-ADP-ribosylated RNA/DNA cap by RNA 2'-phosphotransferase Tpt1. Nucleic Acids Res. 2018;46:9617-24.

[4] Sorci L, Ruggieri S, Raffaelli N. NAD homeostasis in the bacterial response to DNA/RNA damage. DNA Repair (Amst). 2014;23:17-26.

[5] Sorci L, Kurnasov O, Rodionov D, Osterman A. Genomics and Enzymology of NAD Biosynthesis. Comprehensive Natural Products II: Chemistry and Biology. 2010;7:213-57.

[6] Gazzaniga F, Stebbins R, Chang SZ, McPeek MA, Brenner C. Microbial NAD metabolism: lessons from comparative genomics. Microbiol Mol Biol Rev. 2009;73:529-41, Table of Contents.

[7] Kurnasov O, Goral V, Colabroy K, Gerdes S, Anantha S, Osterman A, et al. NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. Chem Biol. 2003;10:1195-204.

[8] Grose JH, Bergthorsson U, Roth JR. Regulation of NAD synthesis by the trifunctional NadR protein of Salmonella enterica. J Bacteriol. 2005;187:2774-82.

[9] Raffaelli N, Lorenzi T, Mariani PL, Emanuelli M, Amici A, Ruggieri S, et al. The Escherichia coli NadR regulator is endowed with nicotinamide mononucleotide adenylyltransferase activity. J Bacteriol. 1999;181:5509-11.

[10] Rodionov DA, De Ingeniis J, Mancini C, Cimadamore F, Zhang H, Osterman AL, et al. Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. Nucleic Acids Res. 2008;36:2047-59.

[11] Lee DW, Park YW, Lee MY, Jeong KH, Lee JY. Structural analysis and insight into effector binding of the niacin-responsive repressor NiaR from Bacillus halodurans. Sci Rep. 2020;10:21039.

[12] Huang N, De Ingeniis J, Galeazzi L, Mancini C, Korostelev YD, Rakhmaninova AB, et al. Structure and function of an ADP-ribose-dependent transcriptional regulator of NAD metabolism. Structure. 2009;17:939-51.

[13] Okon E, Dethlefsen S, Pelnikevich A, Barneveld AV, Munder A, Tummler B. Key role of an ADP - ribose - dependent transcriptional regulator of NAD metabolism for fitness and virulence of Pseudomonas aeruginosa. Int J Med Microbiol. 2017;307:83-94.

[14] Gao R, Wei W, Hassan BH, Li J, Deng J, Feng Y. A single regulator NrtR controls bacterial NAD+ homeostasis via its acetylation. eLife. 2019;8:e51603.

[15] Teramoto H, Inui M, Yukawa H. NdnR is an NAD-responsive transcriptional repressor of the ndnR operon involved in NAD de novo biosynthesis in Corynebacterium glutamicum. Microbiology. 2012;158:975-82.

[16] Leyn SA, Suvorova IA, Kazakov AE, Ravcheev DA, Stepanova VV, Novichkov PS, et al. Comparative genomics and evolution of transcriptional regulons in Proteobacteria. Microb Genom. 2016;2:e000061.

[17] Rodionov DA. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem Rev. 2007;107:3467-97.

[18] Brickman TJ, Suhadolc RJ, McKelvey PJ, Armstrong SK. Essential role of Bordetella NadC in a quinolinate salvage pathway for NAD biosynthesis. Mol Microbiol. 2017;103:423-38.

[19] Garavaglia S, Raffaelli N, Finaurini L, Magni G, Rizzi M. A Novel Fold Revealed by Mycobacterium tuberculosis NAD Kinase, a Key Allosteric Enzyme in NADP Biosynthesis*. Journal of Biological Chemistry. 2004;279:40980-6.

[20] McCarthy AA, Barrett R, Beteva A, Caserotto H, Dobias F, Felisaz F, et al. ID30B - a versatile beamline for macromolecular crystallography experiments at the ESRF. Journal of synchrotron radiation. 2018;25:1249-60.

[21] Lausi A, Polentarutti M, Onesti S, Plaisier JR, Busetto E, Bais G, et al. Status of the crystallography beamlines at Elettra. The European Physical Journal Plus. 2015;130:43.

[22] Kabsch W. XDS. Acta Crystallographica Section D. 2010;66:125-32.

[23] Sheldrick GM. A short history of SHELX. Acta crystallographica Section A, Foundations of crystallography. 2008;64:112-22.
[24] Sheldrick GM. Crystal structure refinement with SHELXL. Acta Crystallogr C Struct Chem. 2015;71:3-8.
[25] Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002;58:1948-54.
[26] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60:2126-32.
[27] Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67:355-67.
[28] Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994;50:760-3.
[29] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993;26:283-91.
[30] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658-74.
[31] Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. IUCrJ. 2014;1:213-20.
[32] Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. Journal of Computational Chemistry. 1994;15:488-506.
[33] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372:774-97.
[34] Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42:D206-14.
[35] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403-10.
[36] Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology. 2010;59:307-21.
[37] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 2021.
[38] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160-6.
[39] Gelfand MS, Koonin EV, Mironov AA. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. Nucleic Acids Res. 2000;28:695-705.
[40] Mironov AA, Vinokurova NP, Gelfand MS. Software for analysis of bacterial genomes. Molecular Biology. 2000;34:222-31.
[41] Sorci L, Blaby IK, Rodionova IA, De Ingeniis J, Tkachenko S, de Crecy-Lagard V, et al. Quinolinate salvage and insights for targeting NAD biosynthesis in group A streptococci. J Bacteriol. 2013;195:726-32.
[42] Huang N, Sorci L, Zhang X, Brautigam CA, Li X, Raffaelli N, et al. Bifunctional NMN adenylyltransferase/ADP-ribose pyrophosphatase: structure and function in bacterial NAD metabolism. Structure. 2008;16:196-209.
[43] Sorci L, Blaby I, De Ingeniis J, Gerdes S, Raffaelli N, de Crecy Lagard V, et al. Genomics-driven reconstruction of acinetobacter NAD metabolism: insights for antibacterial target selection. J Biol Chem. 2010;285:39490-9.
[44] Srouji JR, Xu A, Park A, Kirsch JF, Brenner SE. The evolution of function within the Nudix homology clan. Proteins. 2017;85:775-811.
[45] Chang C, Tesar C, Li X, Kim Y, Rodionov DA, Joachimiak A. A novel transcriptional regulator of L-arabinose utilization in human gut bacteria. Nucleic Acids Res. 2015;43:10546-59.

[46] Kang LW, Gabelli SB, Bianchet MA, Xu WL, Bessman MJ, Amzel LM. Structure of a coenzyme A pyrophosphatase from Deinococcus radiodurans: a member of the Nudix family. J Bacteriol. 2003;185:4110-8.

TABLE AND FIGURES LEGENDS

Table 1: Data collection and refinement statistics.

	native NadQ	SeMet-NadQ-ATP complex	NadQ-DNA complex
Wavelength (Å)	0.976	0.979	0.976
Space group	P 2_1	P 2_1	P 6_5
Cell parameters (a,b, c, Å)	58.3, 133.3, 86.8	57.92, 132.5, 86.8	104.2,104.2, 152.1
(α, β, γ, °)	90, 100.1, 90	90, 100.16, 90	90, 90, 120
Resolution range (Å)	48.51 – 2.31 (2.37-2.31)[a]	51.84 – 2.18 (2.22-2.18)[a]	77.65 – 4.30 (4.45-4.30)[a]
Unique reflections	57061	66672	6399
Multiplicity	3.8 (3.7)[a]	6.5 (6.5)[a]	7.0 (5.6)[a]
Completeness (%)	99.7 (99.6)[a]	99.5 (99.6)[a]	99.6 (99.7)[a]
Mean I/sigma(I)	9.3 (1.6)[a]	12.6 (2.1)[a]	7.14 (1.43)[a]
R_{merge}	0.089 (0.728)[a]	0.105 (0.887)[a]	0.081 (1.108)[a]
CC1/2[a]	0.997 (0.996)[a]	0.998 (0.696)[a]	0.994 (0.263)[a]
Wilson B-factor (Å^2)	38.1	37.8	-
R_{work}	0.198	0.229	0.253
R_{free} [b]	0.200	0.269	0.297
Total n. of atoms	9329	9874	5614
macromolecules	9019	9422	5614
ligands	1	89	-
waters	309	369	-
r.m.s.d.			
bond length (Å)	0.003	0.002	0.002
Angles (°)	0.55	0.44	0.45
Ramachandran			
favored (%)	96.29	94.76	93.64
allowed (%)	3.71	5.24	6.36
outliers (%)	0.0	0.0	0.0
Molprobity clashcore	3.02	5.33	5.87
Average B-factor	45.7	44.6	-
macromolecules	45.8	44.7	-
ligands	51.3	51.2	-
solvent	42.2	40.9	-

[a] Values in the highest resolution shell.

[b] calculated with 5% of total reflections

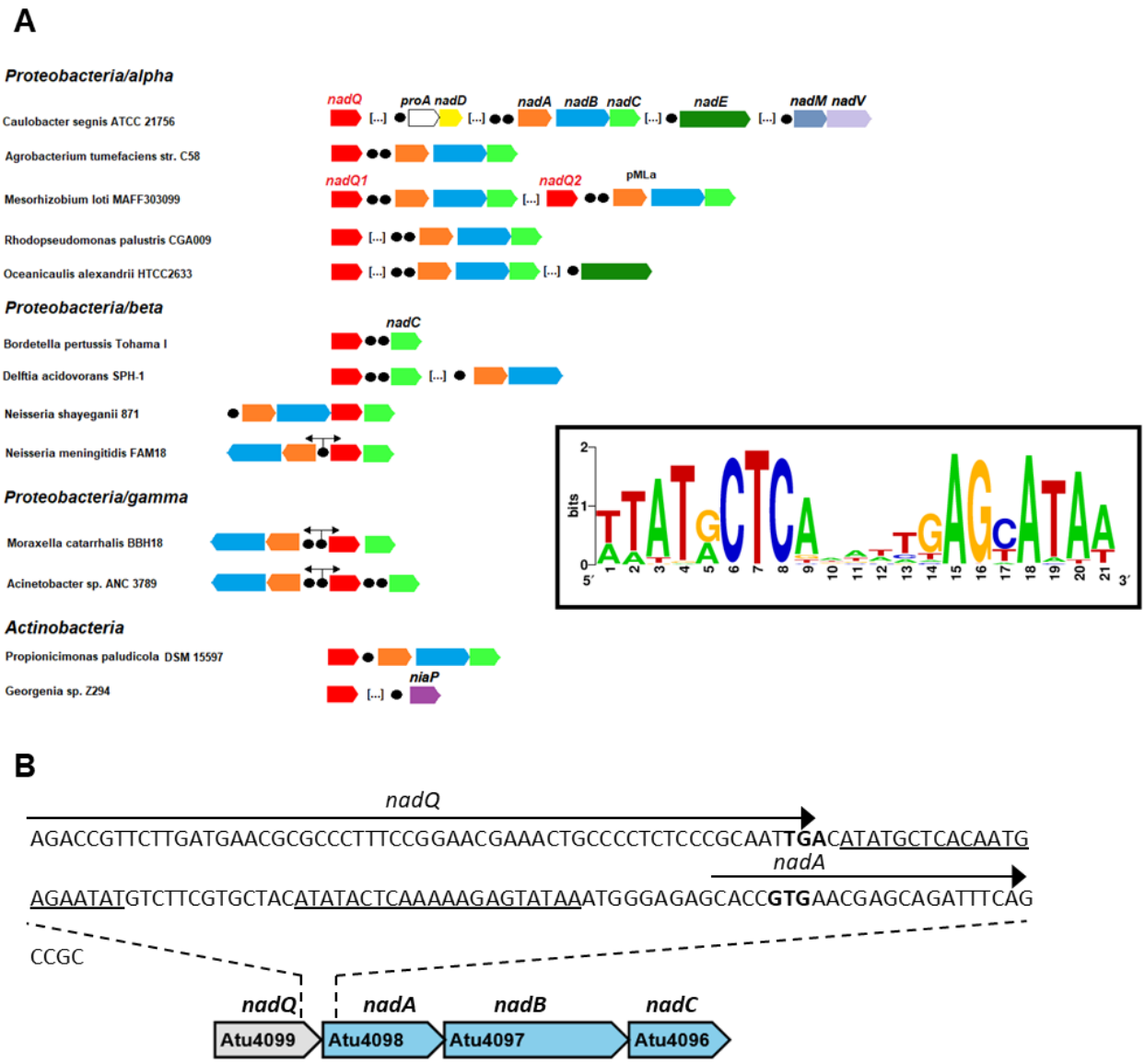


Figure 1. A) Genomic organization of the *nadQ*-containing loci and NadQ-regulated genes in representative bacterial genomes. Genes are shown as boxed arrows. Black circles mark positions of candidate NadQ-binding sites. The inset shows the sequence logo for NadQ DNA-binding site. B) *Agrobacterium tumefaciens* NadQ regulon. The genomic context of NadQ regulon is shown with boxed arrows. The sequence of the DNA fragment used for the EMSA is shown. Positions of NadQ binding sites in the *nadQ-nadA* intergenic region are underlined. The stop codon of *nadQ* and start codon of *nadA* gene are bolded.

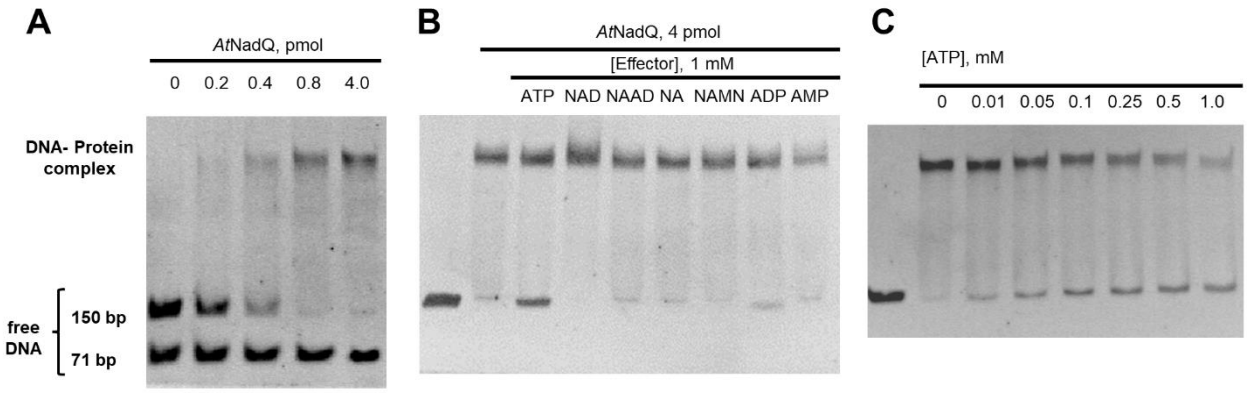


Figure 2. EMSA with *At*NadQ protein and its target DNA. A) Electrophoretic mobility of the 151 bp DNA fragment (0.8 pmol) incubated in the presence of the indicated amounts of NadQ. The specificity of interaction was confirmed by the lack of shift of the 71 bp fragment used as the control. B) Mobility of the DNA fragment incubated with 4 pmol NadQ in the absence and in the presence of the indicated compounds at 1 mM concentration (lanes 3-9). C) Mobility of DNA incubated in the absence (lane 1) and in the presence of 4.0 pmol NadQ at the indicated concentrations of ATP (lanes 2-8).

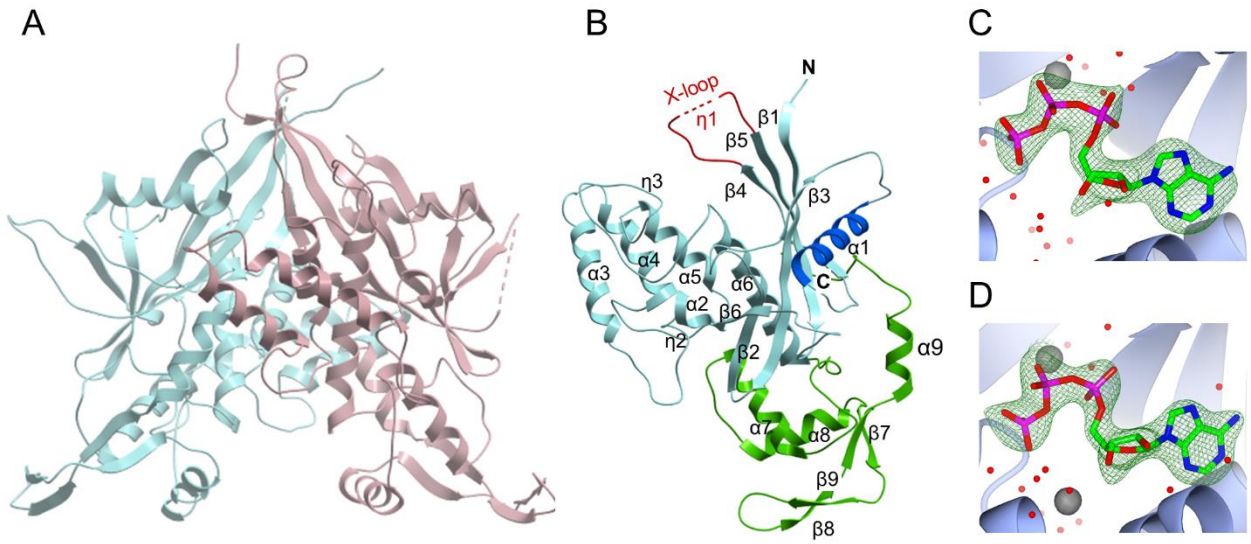


Figure 3. A) Ribbon diagram of dimeric structure of *At*NadQ in its native form (side view). B) Ribbon diagram of *At*NadQ monomer (chain C). The Nudix loop-helix-loop motif is marked in blue, and the X-loop described in the text is marked in red. The N-terminal domain is colored in cyan and the wHTH domain in green. C,D) Fourier difference (Fo-Fc) OMIT electron density maps, contoured at three σ of the ATP molecules bound to dimer AB (C) and CD (D) (resolution 2.31 Å) of the *At*NadQ-ATP complex (see also Figure 4). Water molecules are depicted as small red spheres, while potassium ions are depicted as large grey spheres.

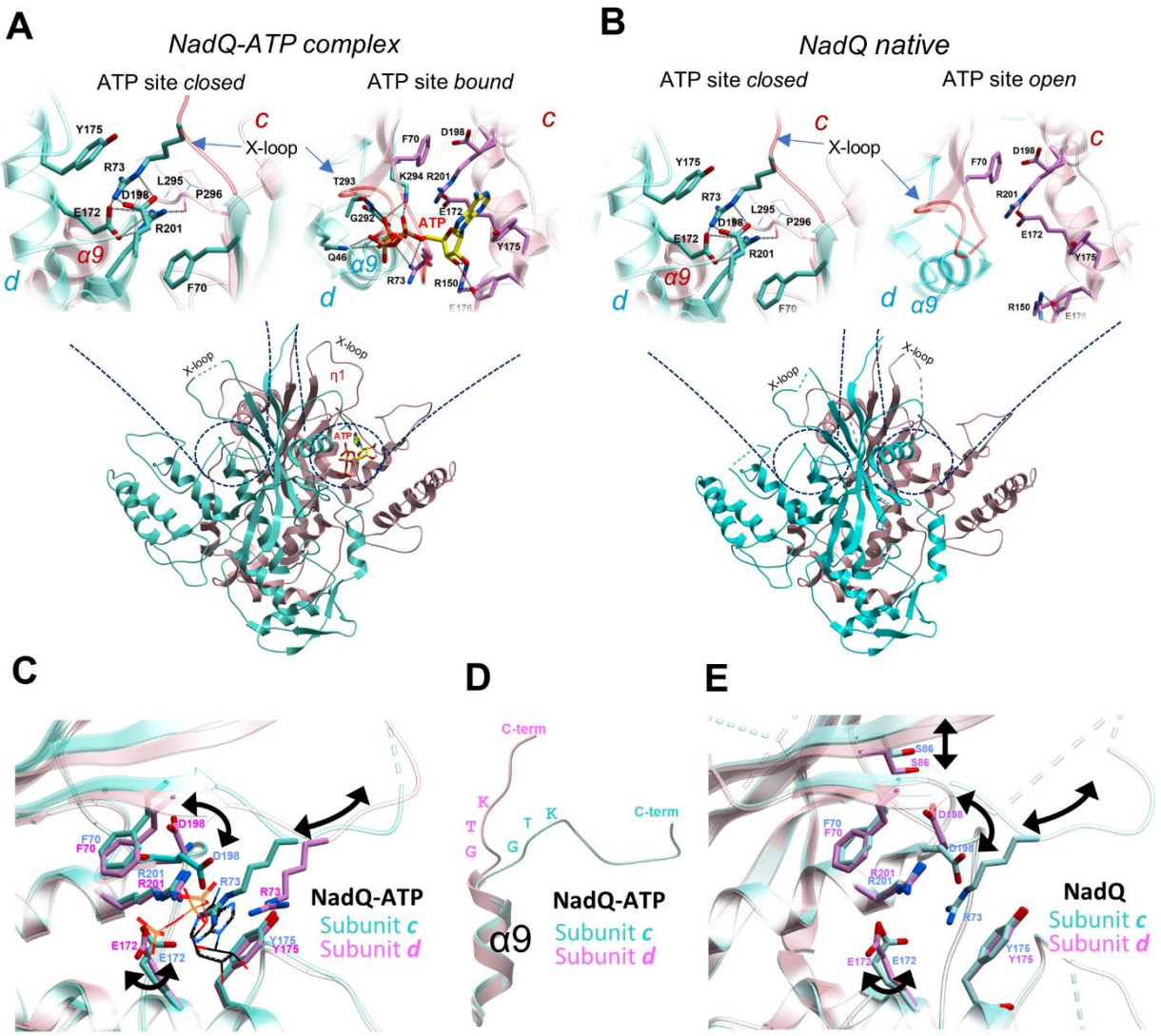


Figure 4. Ligand binding by *At*NadQ. Ribbon diagram of *At*NadQ dimeric structure with bound ATP (A) and in native form (B). The two chains are labeled and differently colored (chain d in cyan and chain c in brown). X-loops are marked. Closeups show the ATP binding pockets harbored by each monomer. The residues interacting with the ligand are labeled and represented as stick models Hydrogen bonds are indicated as black dotted lines. (C) Superposition of subunit c and d in NadQ-ATP complex dimer. (D) Different conformations of C-termini in superposed *At*NadQ chains from the dimer in complex with ATP. GTK indicates Gly292, Thr293, and Lys294. (E) Superposition of subunit c and d in NadQ native form. In panels C and D, the black arrows point to the major conformational differences between the two sites.

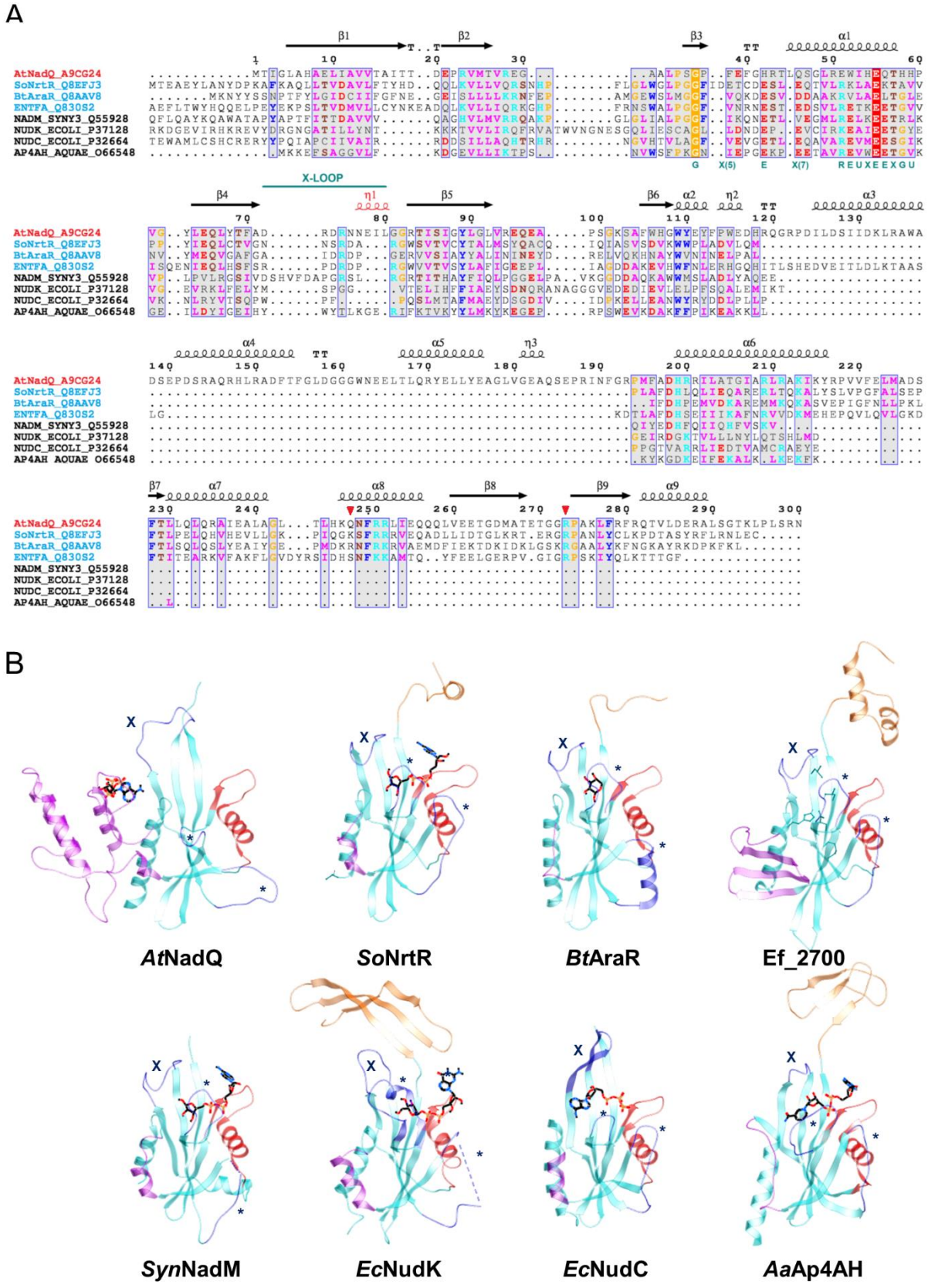


Figure 5. A) Structure-based sequence alignment of *At*NadQ with closest Nudix-like transcriptional regulators (cyan protein ID) and other representatives of the Nudix family (black protein ID). Secondary structure elements of the *At*NadQ are represented by spirals (α and 3_{10} helices) and arrows (β strands). The 3_{10} helix is

labeled η. Partially conserved residues (numbered by *At*NadQ sequence) are shown on a gray background, and universally conserved residues are highlighted by a solid background color. The Nudix signature is indicated under the sequences. Mutated residues are marked by a red triangle. Abbreviations (not described in the text): NadM_SynY3, *Synechocystis* sp. bifunctional NMN adenylyltransferase/Nudix hydrolase; NudK_Ecoli, *E. coli* GDP-mannose pyrophosphatase; NudC_Ecoli, *E. coli* NAD-capped RNA hydrolase; Ap4H_Aquae, *aquifex aeolicus* Ap4A hydrolase.

B) Structural topology of Nudix domain from *At*NadQ and related Nudix proteins. The overall topology of *At*NadQ is highly similar to known Nudix-like TFs and other enzymatic representatives of the Nudix family. For comparison, all liganded forms, if available, were used. Secondary elements of the core Nudix fold are depicted in cyan ribbons. The variable C- or N-terminal extensions are in purple or orange, respectively. The flexible X-loop that in *At*NadQ has been implicated in ATP binding is marked by "x", whereas the asterisk indicates additional loops that in other Nudix may contribute to ligand specificity. The Nudix box is in red. When present, the ligand molecule is shown in a stick representation.

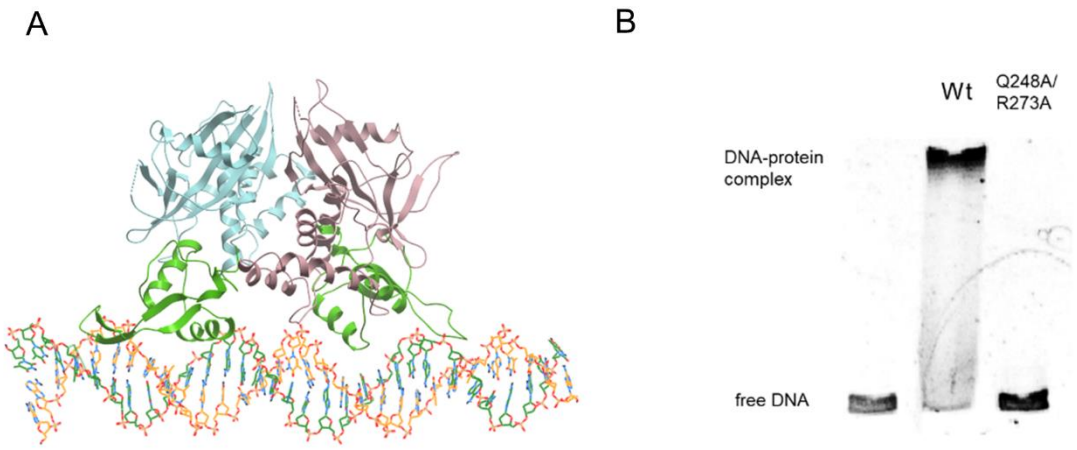


Figure 6. A) Structure of full-length *At*NadQ dimer in complex with DNA. The protein is presented as ribbons and DNA as sticks. B) EMSA of wild-type and Q248A/R273A NadQ clearly indicates that the double mutant is unable to bind target DNA.

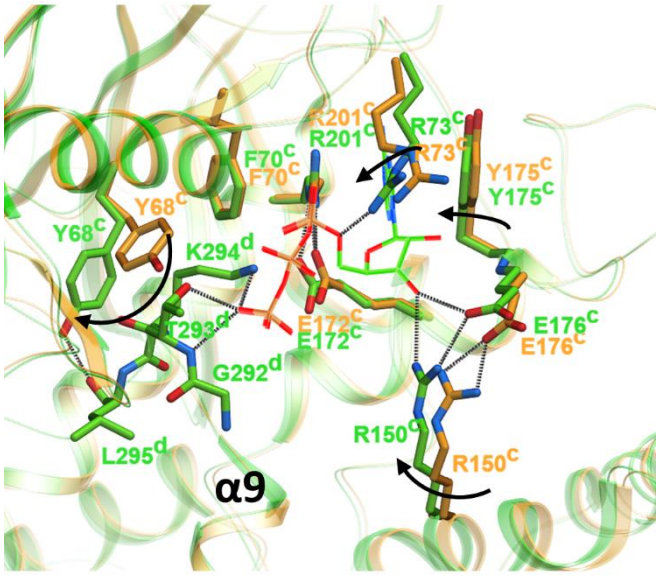
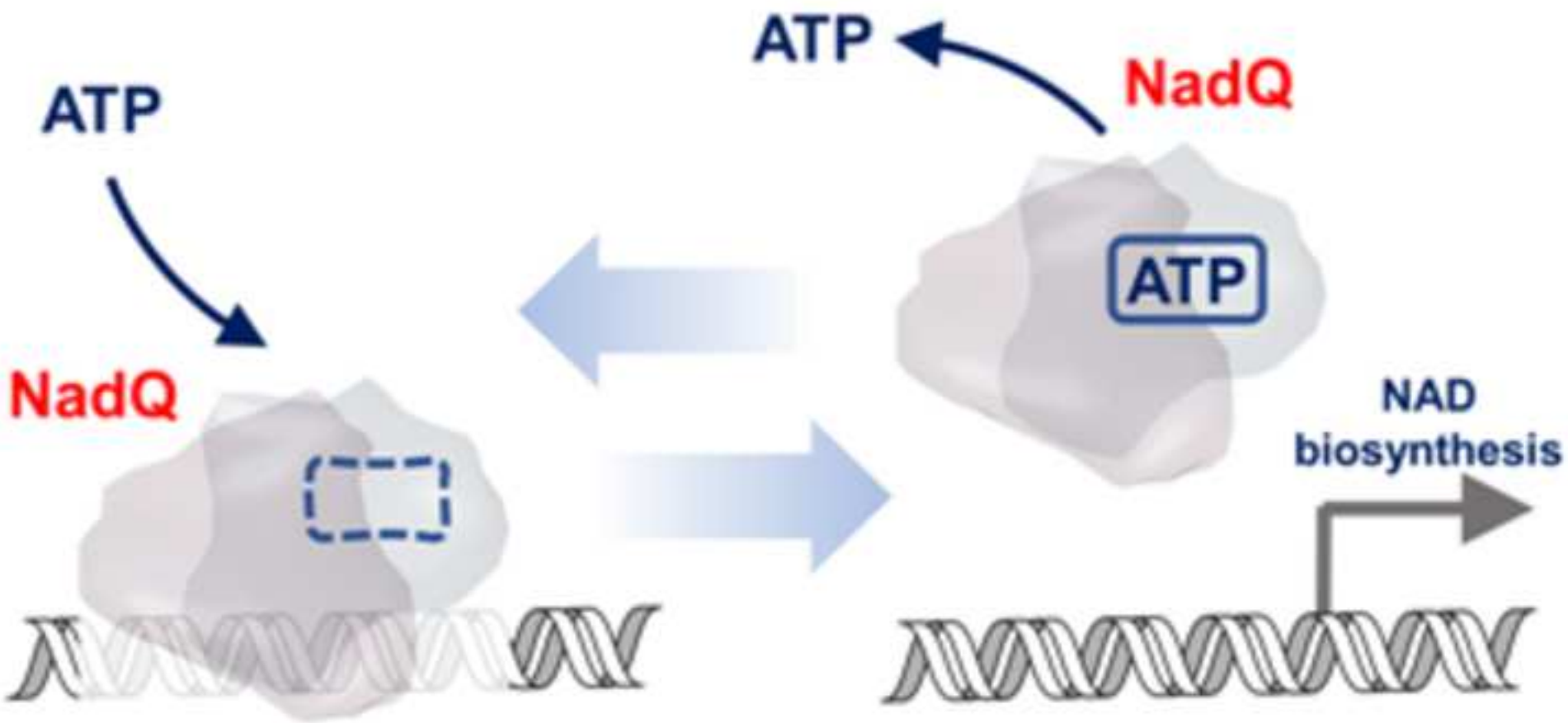


Figure 7. Superimposition of native (orange) and ATP-complexed (green) *At*NadQ dimer with a focus on the binding site. The protein is presented as ribbons; the ATP ligand and the relevant residues as sticks.

Graphical Abstract (for review)



HIGHLIGHTS

- **Regulon analysis depicts NadQ as a transcriptional regulator of NAD biosynthesis**
- **EMSA and mutagenesis demonstrates DNA binding and identifies ATP as negative effector**
- **Native, ATP- and DNA-complexed NadQ 3D structures elucidate the mechanism of action**
- **The effector domain of NadQ belongs to a novel variant of the large Nudix family**

Figure 1

Click here to access/download;Figure;Figure 1.tif

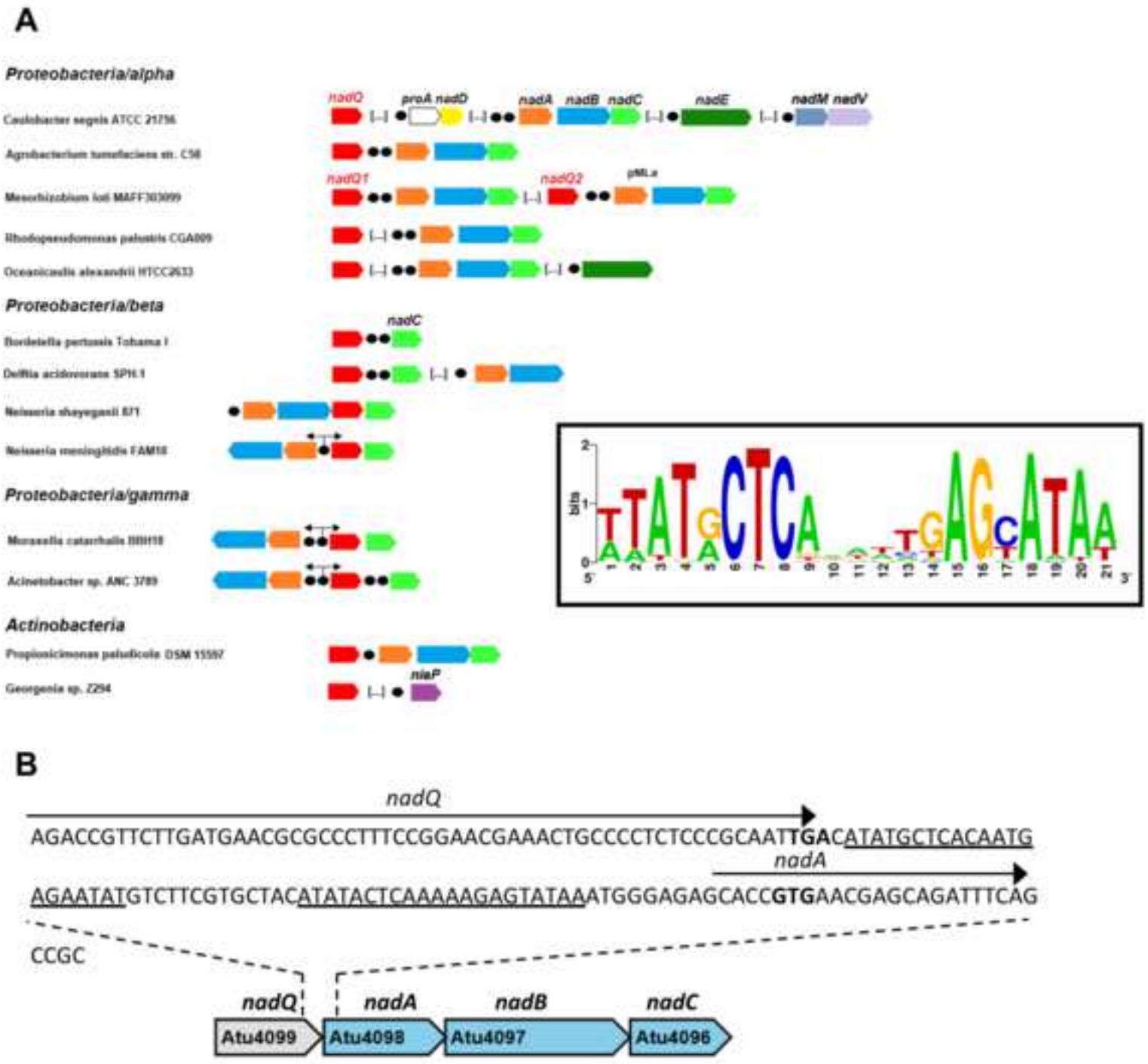


Figure 2

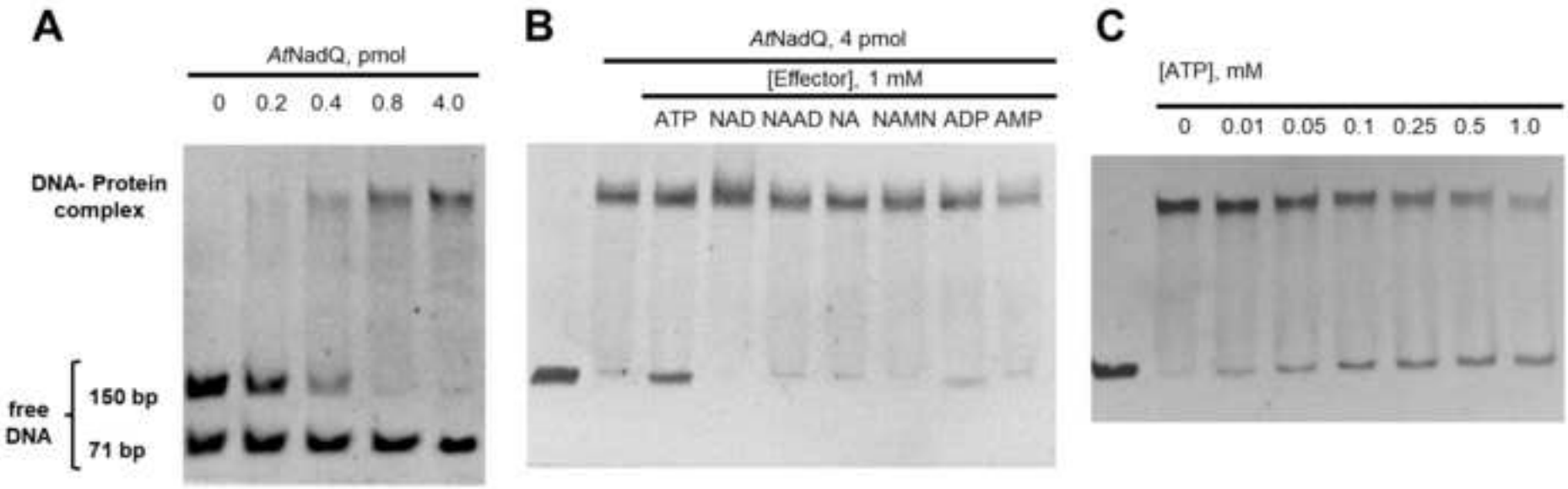


Figure 3

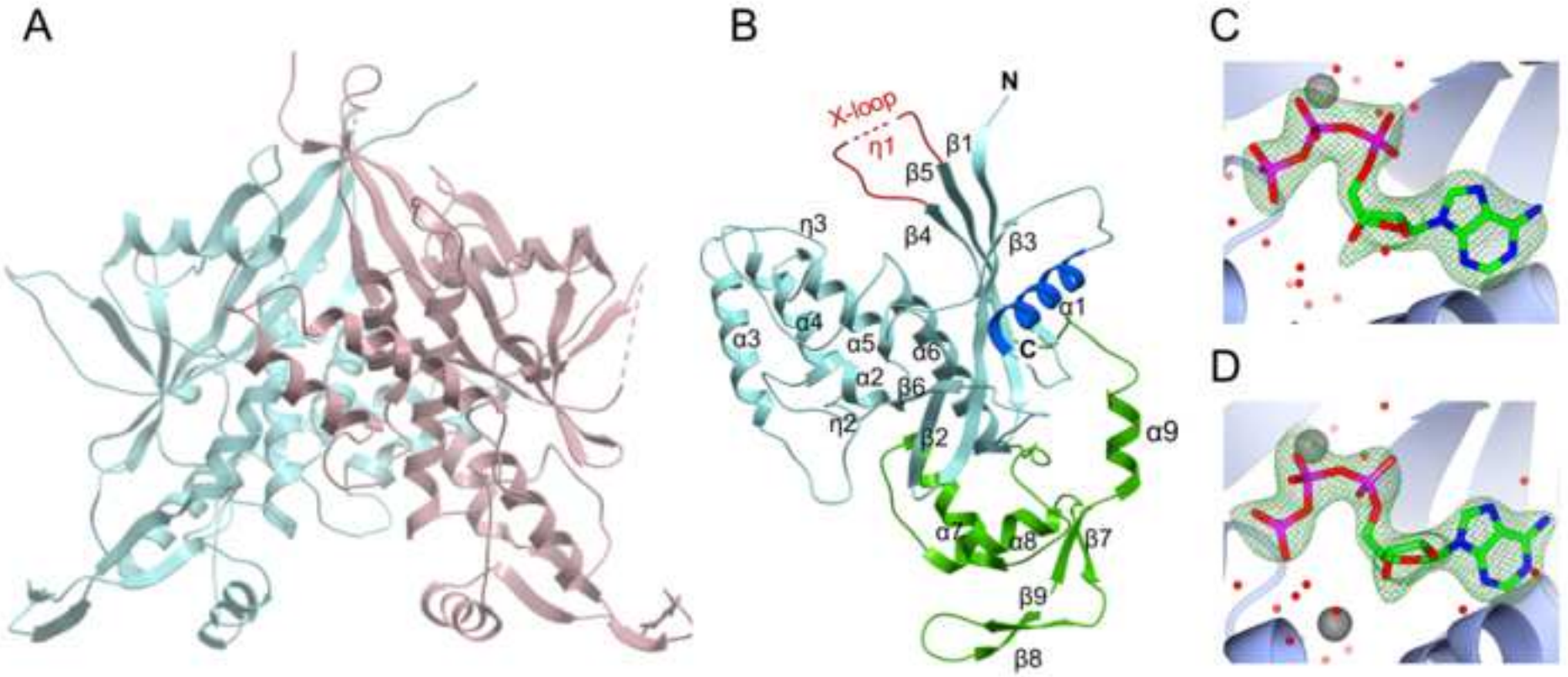


Figure 4

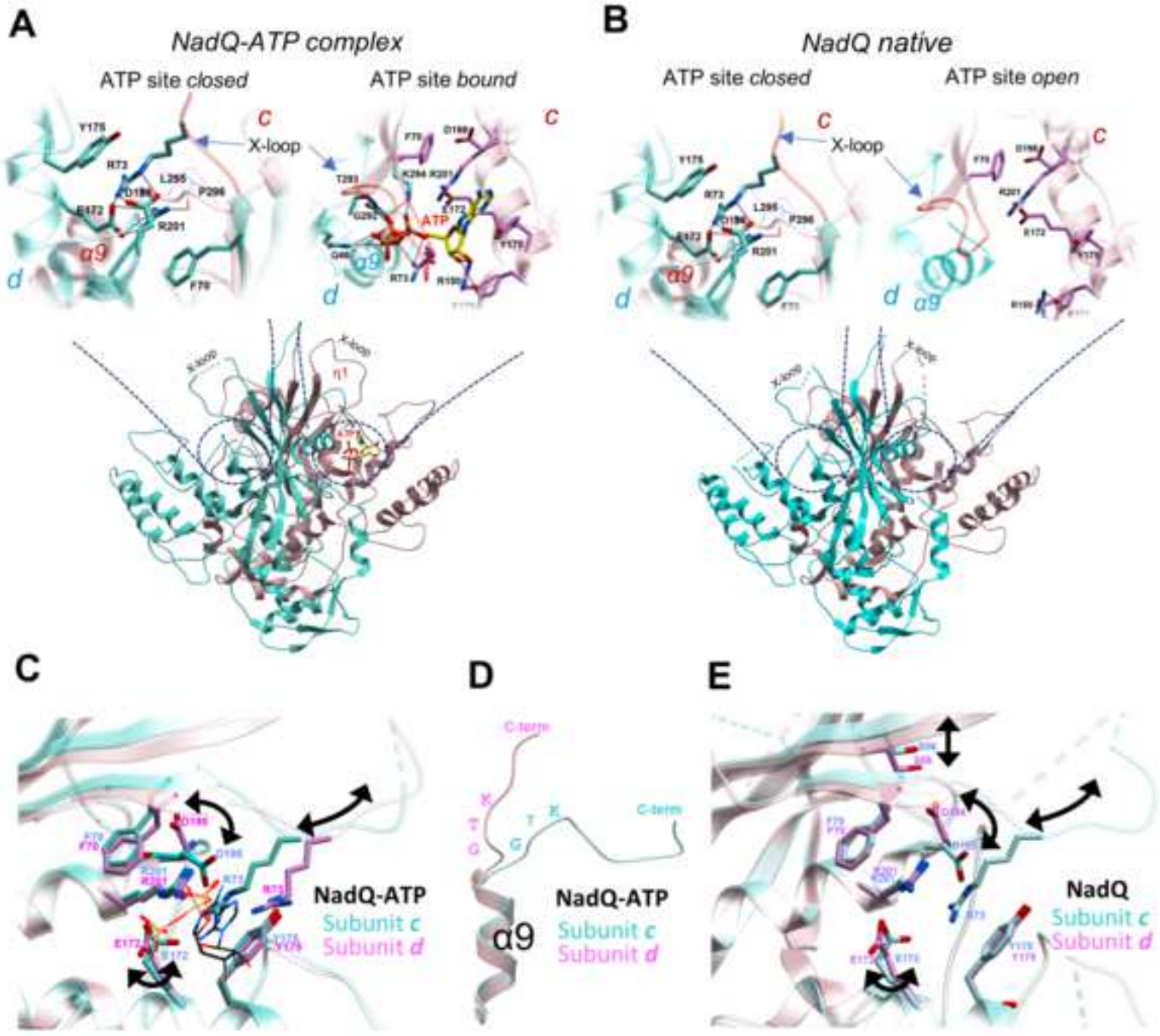


Figure 5

Click here to access/download;Figure;Figure 5.TIF

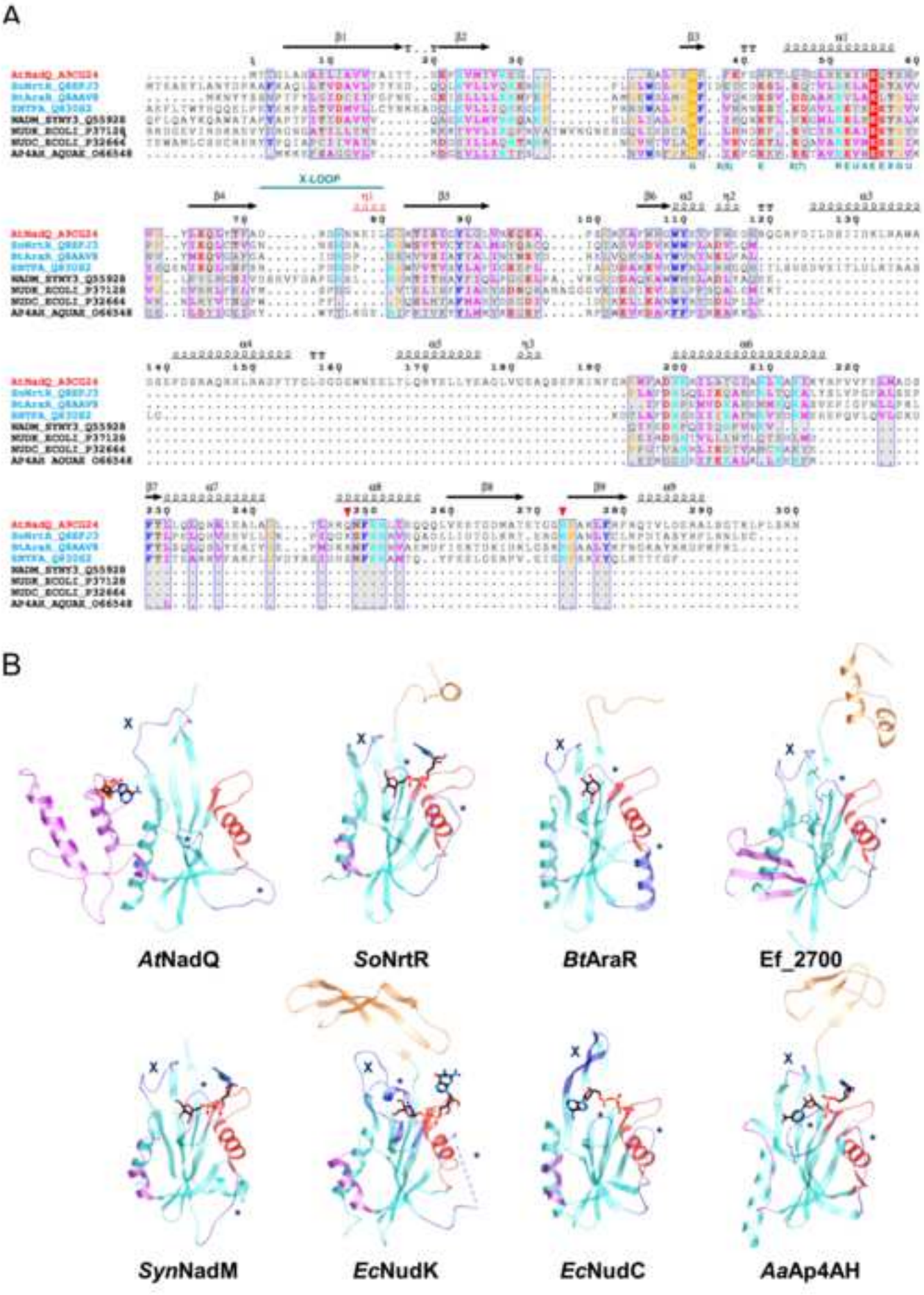


Figure 6

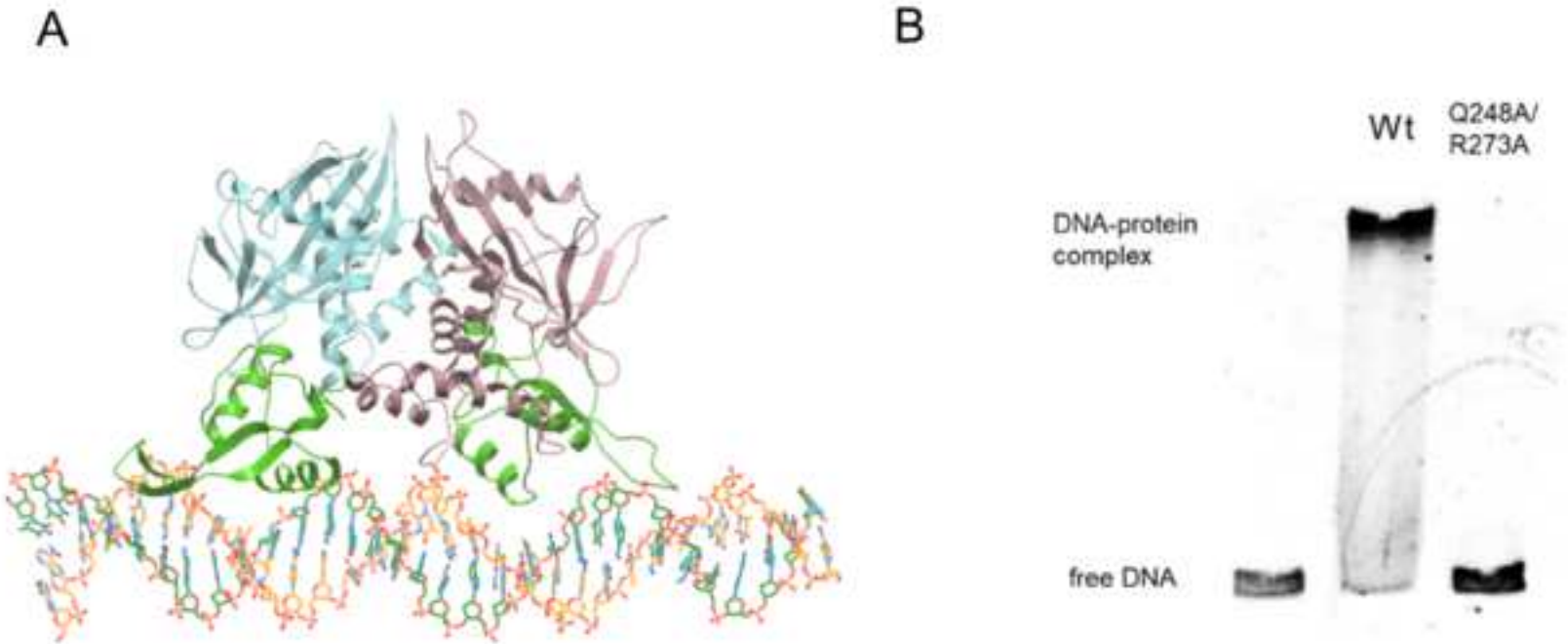
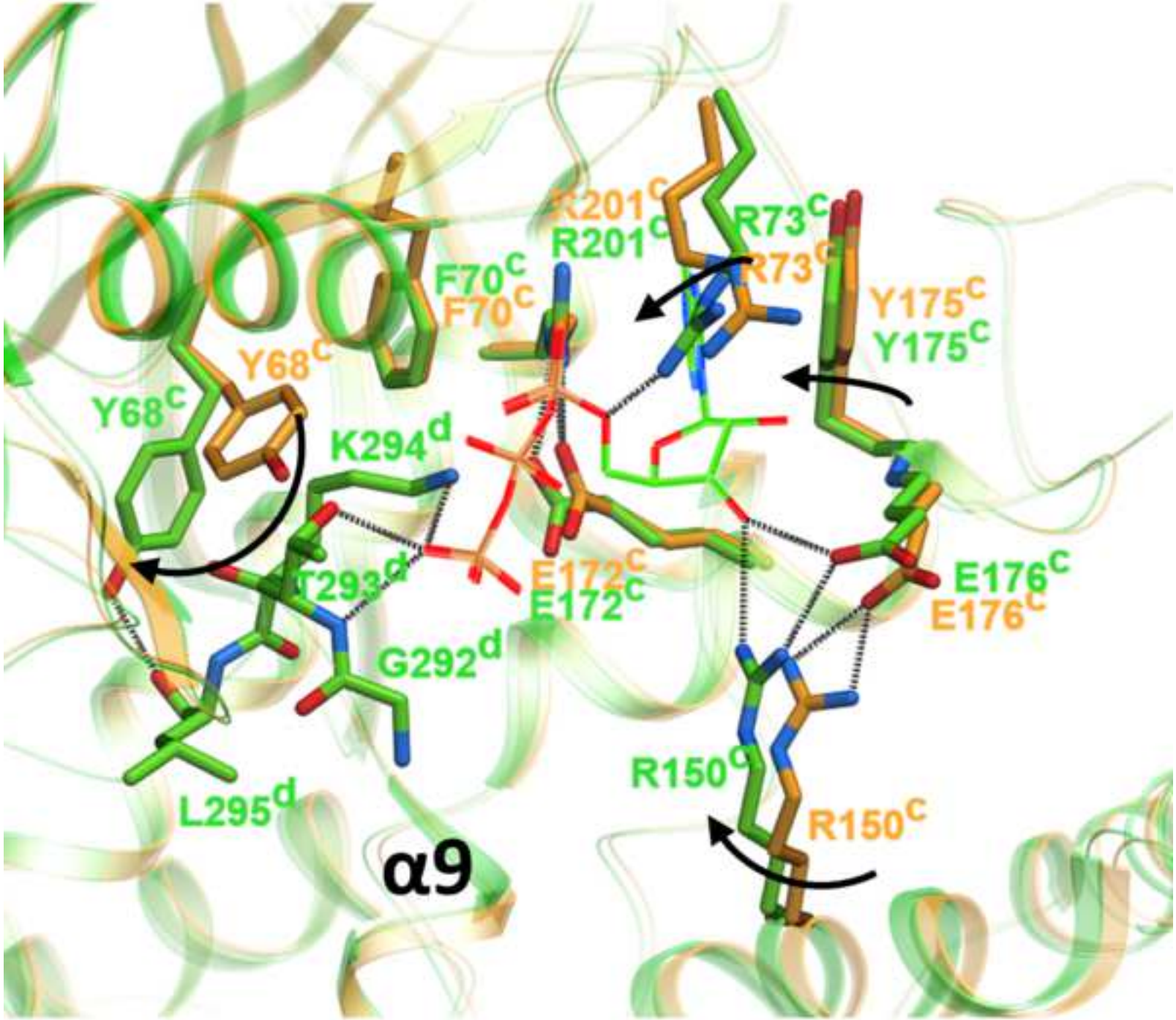


Figure 7

Click here to access/download;Figure;Figure 7_2022.tif



CRediT authorship contribution statement

Gabriele Minazzato: Formal analysis, Investigation, Data curation, Writing-review & editing. **Massimiliano Gasparrini:** Formal analysis, Investigation, Data curation, Writing-review & editing. **Annie Heroux:** Formal analysis. **Natalia V Sernova:** Formal analysis. **Dmitry A Rodionov:** Supervision, Data Curation, Methodology. **Michele Cianci:** Supervision, Methodology; Data curation; writing-review & editing. **Leonardo Sorci:** Conceptualization, Methodology, Writing-original draft, Supervision, Funding acquisition. **Nadia Raffaelli:** Conceptualization, Methodology, Writing-original draft, Supervision, Funding acquisition.