The Purloined Letter: Bacterial Orthologs of Archaeal NMN Adenylyltransferase are Domains within Multifunctional Transcription Regulator NadR

Arcady Mushegian*

Axys Pharmaceuticals, Inc., 11099 N. Torrey Pines Road, La Jolla CA 92037, USA

Nicotinamide adenine dinucleotide (NAD$^+$) and its derivatives play a central role in cellular metabolism. Essential functions of pyridine nucleotides include: NAD$^+$ and NADP as electron-exchanging cofactors in many (de)hydrogenases; NAD$^+$ as a donor of AMP that activates the 5' end of nicked DNA in ligase reaction in bacteria; and NAD$^+$ as a precursor of poly(ADP-ribose), a compound important in cell cycle checkpoint control in eukaryotic cells. *De novo* NAD$^+$ biosynthesis and pyridine ring salvage pathways presumably include the step of linking nicotine mononucleotide (NMN) with the AMP moiety of the ATP precursor. The name, NMN adenylyltransferase (EC 2.7.7.1), has been reserved for the appropriate biochemical activity observed in various species (Natalini *et al.*, 1986; Denicola-Seoane and Anderson, 1990; Balducci *et al.*, 1995), but the search for the corresponding genes was unsuccessful for many years. Recently, NMN adenylyltransferase activity was characterized in two thermophilic archaea, *Sulfolobus solfataricus* and *Methanococcus jannaschii*, and the responsible gene, MJ0541, was identified in the completely sequenced genome of the latter species (Raffaelli *et al.*, 1997).

Sequence similarity analysis of the predicted protein product of MJ0541, using the PSI-BLAST program (Altschul *et al.*, 1997), detected closely related proteins in other archaea, *i.e.*, *Methanobacterum thermoautotrophicum*, *Archaeoglobus fulgidus* (two paralogs), and *Pyrococcus horikoshii*. Unexpectedly, this set of archaeal NMN adenylyltransferase matches was followed by a family of bacterial NadR proteins, thought to be transcriptional regulators of NAD$^+$ biosynthesis genes, with additional role in NMN transport across the membrane (Foster *et al.*, 1990). NadR proteins are found in Gram-negative, Gram-positive bacteria and in the blue-green bacteria. They are twice as long as archaeal MJ0541 orthologs, and matches to MJ0541 are in the N-terminal halves of NadR proteins. The C-terminal halves of NadR proteins in Gram-negative and Gram-positive bacteria contain the Walker-type ATP-binding motifs and are suspected to modulate transport of NMN into the cell (Penfound and Foster, 1999), whereas in Synechocystis, the C-terminal domain of NadR is related to the MutT family of phosphohydrolases (Rafaelli *et al.*, 1999).

In order to validate statistically the observed similarity between archaeal NMN adenylyltransferases and NadR proteins, the N-terminal halves of the latter were used as queries in the PSI-BLAST analysis. The sequence from *M. tuberculosis* (gi 1871585, amino acids 1-147) was used

in a database scan with the conservative E-value threshold of 10^{-3}. In this analysis, the ortholog from *P. horikoshii* passed the threshold on the first iteration with the probability of matching by chance, $p<10^{-7}$, followed, in the same iteration, by MJ0541 ($p=5 \cdot 10^{-4}$). Moreover, MJ0541 and its orthologs in archaea, and NadR proteins in Gram-positive bacteria (*M. tuberculosis*) and in blue-green bacteria (*Synechocystis sp.*) are symmetric best matches in their respective evolutionary lineages, *i.e.*, they satisfy the criteria of an orthologous group in the sense of Tatusov *et al.* (1997). It is therefore likely that the N-terminal halves of the NadR proteins in bacteria contain NMN adenylyltransferase, the essential activity in NAD$^+$ biosynthesis and salvage. It would deserve being called NAD$^+$ synthase, had that name not been already used to identify the amidotransferase that modifies the pyridine ring of the assembled dinucleotide (EC 6.3.5.1).

Iterations of the PSI-BLAST search showed that the MJ0541/NadR family is a member of the large superfamily of diverse enzymes hydrolyzing the alpha-beta pyrophosphate bond in nucleoside triphosphates which have been known to share the characteristic HxGH motif with the family I of aminoacyl-tRNA synthetases (Bork *et al.*, 1995; Venkatachalam *et al.*, 1999). The sequences retrieved by iterative scanning included: FAD synthetase, required for biosynthesis of another dinucleotide redox cofactor (this enzymes commonly occurs as a domain in a bifunctional protein that also contains riboflavin kinase domain, *e.g.*, RIBF_ECOLI, EC 2.7.7.2); PAPS synthase forming 3'phosphoadenosine 5'phosphosulfate, a universal donor in the sulfonation reaction (another occurrence in a bifunctional protein, *e.g.* MET3_YEAST; EC 2.7.7.4); cytidylyltransferases involved in lipid and lipopolysaccharide biosynthesis in bacteria and eukaryotes (phosphorylcholine transferase, CTPT_RAT, EC 2.7.7.15 and CDP-glycerol pyrophosphorylase, TAGD_BACSU, EC 2.7.7.39); pantoate-beta-alanine ligase required for biosynthesis of pantothenate coenzyme (PANC_ECOLI, EC 6.3.2.1); [citrate (pro-3S)-lyase] ligase that acetylates the prosthetic group of citrate lyase via an AMP-linked intermediate (CITC_ECOLI, EC 6.2.1.22); and a number of uncharacterized ORFs, available mostly from large-scale sequencing of microbial genomes.

The most conserved sequence motif in all these proteins is centered on the previously noted HxGH signature, but multiple sequence alignment revealed three additional motifs shared by most of these enzymes (Figure 1). The motif surrounding the HxGH signature is also well defined in distantly related class I aminoacyl-tRNA synthetases where two histidine residues are located in a loop between a beta-strand and an alpha-helix and are directly involved in interactions with the nitrogenous base of the bound ATP. Prediction of secondary structure for *E. coli* NadR protein indicates a similar arrangement for the HxGH motif in the latter (Figure 1). The counterparts of the other three motifs are not readily recognizable in aminoacyl-tRNA synthetases, and their functions in the

```
Secondary-NADR_ECOLI    eeeeeee         hhhh        eeeee        hhhhhhh       eeee
Consensus               xxUUUOxUxxxUHxGHxxUUxxA    x=xUUUxUxx    xUxxx=RuxUU    xxxUSOTxU+++UU
730107_NADR_ECOLI   65  IGVVFGKFYPLHTGHIYLIQRA 4  DELHIIMGFD 15 QPTVPDRLRWL  75 FMSISGAQIRENPF
1001812_NadR_Ssp     6  YGIYIGRFQPFHLGHLRTLNLA 3  AEQVIIILGS  9 PWRSPERMAMI  72 YPDFSSTAIRGAYF
1871585_NadR_Mtu     2  HGMVLGKFMPPHAGHVYLCEFA 3  VDELTIVVGS  4 PIPGAQRVAWM  71 VVPVTATDIRADPL
2496030_Y541_METJA   1  RGFIIGRFQPFHKGHLEVIKKI 3  VDEIIIGIGS  9 PFTAGERILMI  64 RKEYSGTEIRRRML
2648204_Af           1  RAFFVGRFQPYHLGHHEVVKNV 3  VDELIIGIGS  9 PFTAGERVLMI  70 RNEYHGTEIRRKML
2649080_Af           5  RALIFGRFQPFHLGHLKVTKWA 4  DELVLLVGMA  8 PFTAGERIWMM  67 RNLYRGSYIRKLML
2621191_Mth          4  RGLLVGRMQPFHRGHLQVIKSI 3  VDELIICIGS  9 PFTAGERVMML  67 RDRYSGTEVRRRML
3256867_Pho          2  RGLFVGRFQPVHKGHIKALEFV 3  VDEVIIGIGS  9 PFTTGERMEML  66 KDILSATEIRRRMI

125331_KDTB_ECOLI    3  RAIYPGTFDPITNGHIDIVTRA 3  FDHVILAIAA  6 MFTLEERVALA  68 WSFISSSLVKEVAR
2131252_Ect1p_yeast  8  KVWIDGCFDFTHHGHAGAILQA 5  KENGKLFCGV 13 VMNSSERYEHT  60 TYGVSTTEIIHRIL
1345855_CTPT_HUMAN  77  RVYADGIFDLFHSGHARALMQA 4  PNTYLIVGVC 12 VMNENERYDAV  60 TEGISTSDIITRIV
4626600_MUQ1_YEAST   8  KVWIDGCFDFTHHGHAGAILQA 5  KENGKLFCGV 13 VMNSSERYEHT  60 TYGVSTTEIIHRIL
135274_TAGD_BACSU    2  KVITYGTFDLLHWGHIKLLERA 3  GDYLVVAIST  7 YHSYEHRKLIL  54 TEGISTTKIKEEIA
2498241_CITC_ECOLI 146  IGCIVMNANPFTNGHRYLIQQA 3  CDWLHLFLVK  4 RFPYEDRLDLV  94 EMPISASRVRQLLA
2144246_Fom1_strwe   4  IVYVGMSADLIHPGHINILSRA 0  AELGDITIGL 11 HMTYEQRKAVV  63 TPGISSTRLHSSVK
2313990_RfaE_helpy 329  IIFTNGCFDLLHKGHASYLQKA 3  GDILIVGLNS 13 IVSEKDRAFLL  54 EEGYSTSAIIEKIK
400728_PANC_ECOLI   22  RVALVPTMGNLHDGHMKLVDEA 3  ADVVVVSIFV 14 PRTLQEDCEKL  67 EKDFQQLALIRKMV
127026_MET3_YEAST  189  RVVAFQTRNPMHRAHRELTVRA 3  ANAKVLIHPV 14 VRVYQEIIKRY  93 TLNISGTELRRRLR
1710295_RIBC_BACSU  19  SVMALGYFDGVHLGHQKVIGTA                                  DKKISSSYIRTALQ
2500977_SYE_METJA   93  MRFAPNPSGPLHIGHARAAVLN
```

Figure 1. Conserved Sequence Motifs in "HxGH" Nucleotidyltransferase Superfamily
Unique sequence identifiers in GenBank and SWISSPROT are shown. Numbers indicate distances to the N-termini of each sequence and between blocks. Bulky hydrophobic residues (I, L, F, M, V, Y, and W; U in the consensus line) are shaded, other conserved residues (including positively charged R and K, indicated by the plus sign in the consensus line; negatively charged D and E, indicated by the equal sign; and small side chain residues A, G, and S, indicated by O) are shown by bold typing. The alignment was constructed using combination of segment-pair overlap and Gibbs sampling options of the MACAW program (Schuler et al., 1991); the probabilities of occurring by chance for blocks I and IV were below 10^{-20} and below 10^{-8}, respectively. The secondary structure of NadR sequence in E. coli was predicted using the PHD program (Rost, 1996); only predictions with reliability 5 or higher are shown. In the secondary structure lane, e indicates extended structure (beta strand), and h indicates alpha helix.

nucleotidyltransferase superfamily are unknown; however, mutagenesis of phosphocholine cytidylyltransferase in Bacillus subtilis has pointed out the essential functional role of conserved arginine in motif III, as well as serine and threonine residues in motif IV (Park et al., 1997).

Interestingly, genetic evidence suggests that NadR in E. coli is a transcriptional regulator of NAD+ biosynthesis and recycling genes pncB, nadB, and nadA-pnuC (Foster et al., 1990). Moreover, NadR specifically binds to the so-called NAD box, a consensus inverted DNA repeat that is found upstream of these genes (Penfound and Foster, 1999). Notably, in Proteobacteria (Escherichia, Haemophilus, Klebsiella, Moraxella, Salmonella, and Yersinia), NadR proteins have an additional, ca. 60 amino-acid domain at the N-termini, missing in Mycobacterium and Synechocystis homologs. Database searches with this N-terminal extension detected sequence similarity to a large family of bacterial helix-turn-helix DNA-binding domains (data not shown), indicating that the observed specific binding to DNA may indeed be mediated by the N-terminal part of NadR. Thus, in Proteobacteria, NadR is a tripartite protein, comprising predicted DNA-binding module, NMN adenylyltransferase, and ATPase. NadR homologs in Mycobacterium and Synechocystis appear to be bipartite, and probably are not directly involved in the NAD box recognition.

I propose, based on sequence similarity and symmetrical orthologous relationship with archaeal proteins, that the essential and long sought-after NMN adenylyltransferase activity in bacteria resides within NadR proteins, which have been postulated to play other roles in NAD+ biosynthesis. At the time of revision of this manuscript, the NMN adenylyltransferase and ADP-ribose phosphohydrolase activities were demonstrated for the bifunctional NadR homolog in Synechocystis (Rafaelli et al., 1999). As with the coveted document in E.A. Poe's The Purloined Letter, NMN adenylyltransferase has been for years concealed from investigators by "…the comprehensive and sagacious expedient of not concealing it at all." (http://bau2.uibk.ac.at/sg/poe/works/p_letter.html).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Balducci, E., Orsomando, G., Polzonetti, V., Vita, A., Emanuelli, M., Rafaelli, N., Ruggieri, S., Magni, G., and Natalini, P. 1995. NMN adenylyltransferase from bull testis: purification and properties. Biochem. J. 310: 395-400.

Bork, P., Holm, L., Koonin, E.V., and Sander, C. 1995. The cytidylyltransferase superfamily: identification of the nucleotide-binding site and fold prediction. Proteins. 22: 259-266.

Denicola-Seoane, A., and Anderson, B.M. 1990. Studies of NAD kinase and NMN:ATP adenylyltransferase in Haemophilus influenzae. J. Gen. Microbiol. 136: 425-430.

Foster, J.W., Park, Y.K., Penfound, T., Fenger, T., and Spector, M.P. 1990. Regulation of NAD metabolism in Salmonella typhimurium: molecular sequence analysis of the bifunctional nadR regulator and the nadA-pnuC operon. J. Bacteriol. 172: 4187-4196.

Natalini, P., Ruggieri, S. Rafaelli, N., and Magni, G. 1986. Nicotinamide mononucleotide adenylyltransferase. Molecular and enzymatic properties of the homogeneous enzyme from baker's yeast. Biochem. 25: 3725-3729.

Park, Y.S., Gee, P., Sanker, S., Schurter, E.J., Zuiderweg, E.R., and Kent, C. 1997. Identification of functional conserved residues of CTP:glycerol-3-phosphate cytidylyltransferase. Role of histidines in the conserved HXGH in catalysis. J. Biol. Chem. 272: 15161-15166.

Penfound, T., and Foster, J.W. 1999. NAD-dependent DNA-binding activity of the bifunctional NadR regulator of Salmonella typhimurium. J. Bacteriol. 181: 648-655.

Raffaelli, N., Lorenzi, T., Amici, A., Emanuelli, M., Ruggieri, S., and Magni, G. 1999. Synechocystis sp. slr0787 protein is a novel bifunctional enzyme endowed with both nicotinamide mononucleotide adenylyltransferase and "Nudix" hydrolase activities. FEBS Lett. 444: 222-226.

Raffaelli, N., Pisani, F.M., Lorenzi, T., Emanuelli, M., Amici, A., Ruggieri, S., and Magni, G. 1997. Characterization of nicotinamide mononucleotide adenylyltransferase from thermophilic archaea. J. Bacteriol. 179: 7718-7723.

Rost, B. 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. Meth. Enzymol. 266: 525-539.

Schuler, G., Altschul, S.F., Lipman, D.J. 1991. A workbench for multiple alignment construction and analysis. Prot. Struct. Func. Genet. 9: 180-190.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. Science. 278: 631-637.

Venkatachalam, K.V., Fuda, H., Koonin, E.V., and Strott, C.A. 1999. Site-selected mutagenesis of a conserved nucleotide binding HXGH motif located in ATP sulfurylase domain of human bifunctional 3' phosphoadenosine 5' phosphosulfate synthase. J. Biol. Chem. 274: 2601-2604.

Note added in proof: A recently determined three-dimensional structure of E. coli KdtB protein, a phosphopantetheine adenylyltransferase related to NadR, reveals structural similarity to class I aminoacyl-tRNA synthetases and involvement of motifs I and IV in direct interactions with phosphate groups of bound ATP (Izard and Geerlof, 1999. EMBO J. 18: 2021)